Issues and Technologies Related to Privacy and Security

Libraries have a long tradition of taking extraordinary measures to ensure the privacy of those who use their facilities and access their materials. An important value surrounds the concept that individuals can access or read any material offered by the library without concern that their selections will be made available to any other person or organization. There are many scenarios involving sensitive topics where exposure of items accessed in the library or borrowed can be not just a point of controversy or embarrassment, but also a matter of personal danger.

The American Library Association addresses privacy in one of its interpretations expanding on the Library Bill of Rights, leading with this statement:

> Privacy is essential to the exercise of free speech, free thought, and free association. The courts have established a First Amendment right to receive information in a publicly funded library. Further, the courts have upheld the right to privacy based on the Bill of Rights of the U.S. Constitution. Many states provide guarantees of privacy in their constitutions and statute law. Numerous decisions in case law have defined and extended rights to privacy.[1]

The policies, processes, and procedures that libraries embrace related to print materials have been well established. Libraries regularly operate the automation systems that manage the circulation of physical materials in a manner that minimizes any possible exposure of personally identifiable information related to a patron's check-out activity. During the period of an active loan, the automation system maintains a link between the specific patron record and the item borrowed. This information, which is needed to manage the loan transaction, through linking content with an individual, would by privacy policies be treated with strict confidentiality by any library personnel with operational or technical access. Circulation systems need to be able to link content to an individual in order to ensure the return of materials and enable the sending of messages regarding items overdue, fines, or recalls. Past the point of operational need, many libraries will take measures not only to disassociate the item from the patron, but also to anonymize the transaction in a way that the link cannot be reconstructed. Such anonymization would support any historical or statistical reporting the library may need, but ensure that it is not possible to reconstruct borrowing history for any patron. These procedures protect this sensitive information from accidental exposure or from access by unauthorized individuals, as well as from requests from law enforcement or other authorities.

This issue of *Library Technology Reports* focuses on patron privacy related to how patrons interact with a library's web-based systems to access information. Since most libraries offer considerable content and services through the web, the extent to which a patron's use of these services might be vulnerable to exposure stands out as a topic of critical interest. The second statement of the ALA interpretation on privacy is especially relevant to the discussion:

> In a library (physical or virtual), the right to privacy is the right to open inquiry without having the subject of one's interest examined or scrutinized by others. Confidentiality exists when a library is in possession of personally identifiable information about users and keeps that information private on their behalf. Confidentiality extends to "information sought or received and resources consulted, borrowed, acquired or transmitted" (ALA Code of Ethics), including, but not limited to:

> database search records, reference questions and interviews, circulation records, interlibrary loan records, information about materials downloaded or placed on "hold" or "reserve," and other personally identifiable information about uses of library materials, programs, facilities, or services.[2]

One of the specific concerns surrounds how well the privacy of a patron is protected when accessing a resource provided by the library via the web. That patron may be accessing that resource from equipment inside the library itself, from a home or office computer, from the wireless network in a coffee shop, or from a distant and sensitive geographic location.

A patron's session accessing a library-provided resource can be seen as similar to a circulation record and must therefore receive the same kinds of measures to ensure its privacy. Such a session may include the sequence of data that describes a query entered into a search box, lists of items returned by the service, and items selected, as well as the text of any materials read online or downloaded. Even when a resource is accessed without an explicit sign-in, many technical clues may be available that link that session to a specific location or individual. If captured by any third party, such a bundle of information would expose even more private data than a circulation transaction.

The January 2015 issue of *Smart Libraries Newsletter* addressed library privacy and security, including a preliminary version of the vendor survey included in this report. The introduction to that study likewise provides some context to this report:

> In the consumer arena, concern for privacy may not be of central concern. Quite the contrary, details regarding any pattern of behavior that might have a direct or indirect commercial impact have become one of the major currencies of the economy of the Web. Advertising dominates as the primary business model. Very sophisticated networks have been created that gather data from both in-person interactions and online activity, primary for the purpose of targeting advertising content. In person, individuals enable tracking of their purchases through loyalty cards and retailers use many other direct or indirect mechanisms to track buying patterns. Much of the infrastructure of the internet has been infiltrated with technical mechanisms that gather and transmit information regarding the sites visited, terms searched, items purchased. From browser-based cookies to much more sophisticated techniques for tracking online behavior, considerable activity transpires behind the scenes to gather any miniscule item of data that might have some commercial value.
>
> This infrastructure that churns personal activity into targeting advertising provides the fundamental economic model for most services provided via the Web. Much of the entertainment content and productivity tools we enjoy is made possible in return for being exposed to advertisements rather than through direct payment. Facebook, for example, as the dominant social network, thrives on this ad-based economy. Since most individuals would rather not pay directly for each service they use, advertising is tolerated on the current online media just as it has been for television throughout its history.
>
> Given the pervasive gathering and transmission of personally identifiable information on the Web surrounding ad-based commerce, libraries have to be very aware of its impact on the services they deliver.
>
> As libraries offer services that allow their community members to search their collections of print, electronic, and digital resources and to access an increasing body of content online, it is critical to maintain privacy if the same ethic that has been historically held for print resources is applied to their online offerings. Given the pervasiveness of advertising networks in the deep infrastructure of the Web, libraries have to work quite hard to even know how much personally identifiable information is transmitted from the services they deliver to third parties. As libraries work to enrich their online presence with a social flavor, they may inadvertently also enable an intermingling of commercial infrastructure into the services they provide.
>
> One of the realities of the Internet lies in the ability for any third party to intercept the transmissions of information as it travels among devices and servers. Wireless networks are an especially easy target. It has to be assumed today that any information transmitted as clear text across a local network or the Internet will be intercepted and used. These purposes range from gathering personal data that might be of use for targeted advertising, to capturing data that might allow the intrusion into servers and systems to gain access to passwords, credit card numbers, sensitive documents, or other items of value.[3]

This issue of *Library Technology Reports* discusses some of the key technologies and techniques related to protecting the privacy of patrons as they interact with web-based services provided by their library. It focuses on encryption as the primary technology for protecting the privacy of online behavior, how data is stored internally, and other features that may be offered in online catalog and discovery products. This report includes two related studies. One is based on a questionnaire sent to providers to assess the capacity of the major discovery interfaces in resource management systems related to patron privacy and security. The other study examines the websites, catalogs, or discovery interfaces of large academic and public libraries, noting characteristics such as the use of secure communications and the presence of commercial tracking agents.

This report does not aim to prescribe or advocate for any specific set of privacy policies. Rather, it focuses on the technology issues surrounding privacy for those interested in not exposing personal information or search behaviors of their patrons who use library-provided web-based services. This report explores the many ways in which patron data and behavior can be easily captured in the absence of preventive measures.

Privacy for Circulation Records as a Model

Libraries treat records related to physical circulation according to practices that reflect their policies for privacy and security. These policies may include minimizing any links between content items and patron records. From an operational perspective, it is necessary to record a link between a patron record and an item that has been loaned to a patron. This link underlies the ability to track when items are due, to send reminders and notices, and to enable the patron to view lists of items currently charged and to perform renewals and other self-service actions. Once the item has been returned, many libraries will activate features of their integrated library system to disassociate the link. Data regarding a specific item borrowed by a specific patron is often anonymized, preserving only categories of items or patrons for statistical purposes. Some systems will retain a patron record identifier in an item record after it has been returned for a limited time in order to be able to trace problems. This data may then be erased or overwritten once the item is loaned to a patron.

To ensure privacy, the anonymization of library circulation transactions may be applied both to the operational databases and to any log files that reflect the transactions. Integrated library systems, like other business applications, create log or journal files that record the details of all transactions performed. These log files both provide a historic record of activity to generate statistics and also can be part of a disaster prevention and recovery procedure. In the event of a system failure, any transactions not included in backups used to restore the database can be replayed from the log files. This is a possible strategy to restore transactions that took place between the last backup and the time of the failure. A thorough anonymization of personal information must also include these transaction log files, since they could be used to reconstruct the links between content items and specific patrons.

Active database files and transaction logs will usually be backed up through routine disaster recovery procedures. Libraries interested in completely anonymizing circulation records will need to address what personal data may be retained in backup replicates. A thorough set of disaster avoidance and recovery procedures provides many layers of protection against data loss, which also makes it challenging to ensure that none of the backup copies can be used to reconstruct personalized information.

The procedures related to patron privacy are generally intended to protect specific data regarding patron reading behaviors. Destroying or anonymizing circulation records ensures that private information cannot be accidently or intentionally exposed to unauthorized parties. These procedures also protect the user of a library in cases where law enforcement authorities make a request. With thorough technical procedures in place designed to protect privacy, no personal data would exist that might be subject to such requests.

Privacy and Security for Web-Based Services

The level of attention given to circulation records to align with privacy policies may also be applicable to library websites and discovery services. The operational and technical complications involved in maintaining the privacy of circulation systems demonstrate the complex operational and technical measures involved. Similar concerns apply to patron activity conducted to access the library's web-based resources and services. Transmission of patron sessions over the Internet evokes similar issues and requires proactive measures to maintain consistency with library privacy policies. To protect privacy, organizations need to consider the protection of both "data in motion" as it traverses networks and "data at rest" as it is stored on servers. This report considers both scenarios.

The technology infrastructure of the web poses many challenges to libraries that aim to preserve patron privacy and maintain high security. Any information transmitted via the Internet, as a public network, can be easily accessed by any third party unless specific measures are taken. Web servers and associated software may expose data that may not be consistent with privacy policies. It is important for libraries to understand what information is transmitted and stored by their web-based systems in order to be able to operate these systems in ways consistent with the applicable policies.

The protocols used in the transmission of data on the Internet make it relatively easy for anyone to intercept and view that content and therefore have a direct bearing on privacy and security. The tools to eavesdrop on Internet traffic are easily acquired and do not necessarily require specialized expertise to operate. Any content transmitted over the Internet must be considered publicly viewable unless specific measures, especially encryption, are taken to protect it. But with encryption in place, such interception of data becomes almost impossible.

Basics of Secure Transmission Technologies

No network can be considered safe for the transmission of "clear text," or unencrypted data. There are just too many possible points of interception. Wireless networks provide the most convenient opportunity for gathering information via eavesdropping techniques. Any person equipped with a mobile device and easily obtained software can view all the unprotected data passing through that wireless access point. Wired networks can also be vulnerable to anyone able to physically connect. Other points of vulnerability include the organizations that provide network services. Internet service providers are able, and may be required, to capture Internet traffic and provide access to third parties, such as governmental entities.

Encryption is the primary technique used to protect data from unwanted access by third parties. It protects data transmitted across networks and stored on computers. Encryption algorithms transform data before it is transmitted into a seemingly garbled form that, if intercepted, cannot be deciphered. Most encryption technologies in use today rely on a scheme called public key infrastructure (PKI). Data is encrypted with a private key and digital signature feed into a software algorithm. The data can be decrypted with the corresponding private key. Secure communication on the web provides two important benefits: it authoritatively identifies the website, and it enables encrypted communications between the user's browser and the server providing the resource.

Without entering the deeply technical realm of encryption, there are some high-level concepts relevant to a discussion of library privacy and security concerns. The PKI infrastructure in use on the web provides secure communications by both validating the identity of a site and by transmitting data using encryption. The identity of a website transmitting securely is validated through a digital certificate. Certificates are issued through organizations that confirm the identity of the entity and are based on a hierarchy of trust. Since much of the web, especially those sites involved in e-commerce, depends on secure communications, the digital certificates used by web servers are carefully controlled. It is also possible to use self-signed certificates internally within an organization, but their use for external transactions would be apparent and flagged as not trusted. Credentials of organizations are validated before a certificate is issued by a reputable certificate authority. Compromised or otherwise problematic certificates can also be revoked. Standard validation procedures include checking certificates against revocation lists.

Digital certificates can be installed into a web server to enable encrypted transmission. When activated, pages will be transmitted using the HTTPS protocol rather than HTTP. In most cases, HTTPS traffic is associated with the TCP/IP port 443 and HTTP with port 80, although other port assignments are common. The user's browser will show an indication that the transmission is secure. Chrome, for example, presents a fully valid secure site with a green padlock and shows HTTPS in the URL, and clicking on the padlock will display the details of the certificate. Relevant details include the identity of the organization to which the certificate was issued, the certificate authority, and the technical protocols used for transmission and encryption. Before performing a sensitive transaction, a user can verify that the digital certificate indeed matches the intended organization.

Modern encryption technologies protect data even when massively powerful computers attempt to break them by brute force, such as rapidly trying all possible combinations that constitute a password or key. As computers become ever more powerful, the strength of those algorithms must likewise be improved through techniques such as longer keys. Out-of-date encryption technologies must therefore be considered insecure. A modern web browser will usually detect such vulnerabilities.

Web browsers will display the indicators of a secure transmission only when specific technical criteria have been met. Criteria include the presence of a valid certificate (including checking revocation lists) as well as current technologies for transmission and encryption. Encryption must be performed with a key of sufficient length and according to an algorithm able to defy decryption attempts. This age of massively powerful computation resources capable of decryption attempts by brute force demands the utmost caution in implementing security. The SHA-1 algorithm, which had been widely used for encryption, is now considered vulnerable, replaced by more robust protocols such as SHA-256. The secure sockets layer (SSL), in addition, is now considered obsolete and untrustworthy, with TLS 1.2 currently accepted as the trusted protocol for secure transmission on the web. Since 2014, Chrome and other browsers will flag as untrusted web servers that continue to use SSL. In 2016, it is also anticipated that sites relying on SHA-1 will likewise be flagged as untrusted. The technologies used to support communications should be considered a constantly moving target. Website operators and users who rely on secure communications must be ever vigilant and stay abreast of current standards.

The use of secure communications provides the best approach possible today for protecting the privacy of patrons as they interact with library systems. A page remains encrypted from the time it is transmitted by the web server until it is displayed on the user's browser. As a result, the information remains impervious to eavesdropping through the complete route, even if it includes unsecured wireless networks or other points of vulnerability along the way. Likewise, any information passed in the clear without encryption should be assumed to be publicly viewable.

Secure Storage

The details on secure communications apply to pages as they are transmitted from servers on the web, the concern for "data in motion." Another set of concerns relates to how data is stored. Data can be encrypted when it is stored on a network server or storage device. Such encryption would protect the data in the event of successful penetration into a server by an unauthorized entity. The most common scenario involves passwords, for which standard practice requires that they be stored in an encrypted hash and never as clear text. When a password is stored as a hash, even the site operator cannot view it. An authentication request can compare the hash of the string provided against that of the password when it was created, but the password itself cannot be reconstructed. Other sensitive elements, such as credit card numbers, would also be stored in encrypted form. Some applications designed to operate with a high level of security may also encrypt other details. For most library-related applications, routine transaction data and logs are not encrypted and depend on general system security to prevent unwanted access.

Locally Managed or Remotely Hosted

Integrated library systems, discovery services, and other library-related software may be deployed either as software that the library installs within its own technical infrastructure or as a service hosted by the vendor. The same kinds of concerns apply in either scenario. For a locally installed system, the library would bear more responsibility for its secure operation and for the procedures implemented to guard privacy and security. Hosted systems naturally place more of that burden on the vendor. Even when a system is hosted by the vendor, the library will want to understand and hopefully control the procedures in place.

The deployment methods used in hosted systems also come into play relative to these issues. One deployment model involves the hosting of individual physical or virtual servers. The same configuration options and operational procedures apply whether these server-based systems are hosted by the library or by the vendor. Each library's instance of the software can be configured individually. One library might, for example, instruct the vendor to configure the server to encrypt all traffic related to its online catalog while other libraries opt to operate without that capability. The options and features available may also depend on the version of the software implemented, which may differ across the libraries using that system. Multi-tenant platforms, where all the libraries using that system share the same instance, have the capability for uniform security configuration. It is possible for the provider of a multi-tenant application to enforce encryption for all its customers using the software. The Apollo ILS and the BiblioCore discovery service both, for example, enforce secure communications for all transactions.

Whether the servers that host the library's integrated library system, discovery service, or other systems are installed locally in the library or by an external provider impacts the route through which a patron's session is transmitted. In most cases the physical location of the service relative to the user is neutral relative to privacy concerns. Even if the user and the server were on the same local area network, the possibility remains that the transmission could be captured by others on that network or beyond. Library systems hosted externally, or content services provided over the web from the publisher's servers, traverse many intervening networks and exchange points and must be considered as vulnerable. Whether remote or local, patron sessions should be managed with encrypted transmission to ensure privacy.

In addition to the ability to capture data describing a patron session via network transmission, there are other points of vulnerability for patron privacy in a routine web session. The techniques used to support use statistics, analytics, and interactions with other services may result in exposure to external entities.

Server Logs Record Patron Activity

It is essential for any organization operating a website to be able to measure and monitor its use. In the same way that libraries often count the number of visitors to their physical facilities, the number of items loaned, reference questions received, and other services, they also track how their virtual services are accessed via their website. Use data for web-based services not only helps demonstrate the impact of the library to funding agencies and administrative authorities, it also provides essential information for designing and tuning the site to function optimally. Both commercial and nonprofit organizations that rely on their websites for critical aspects of their operations benefit from gathering extensive data regarding use patterns and performing analysis to be able to identify problems or to optimize navigation or presentation or to make other changes to improve usability and eliminate problems. The use of web analytics has become part of the essential tool kit for website administrators and user experience specialists.

Web analytics depends on data describing each interaction that takes place on the site. Web servers can be configured—and usually are—to record every page request in a log file. It is important for these server logs to be a component of the technology covered by privacy procedures. For any web server that delivers access to library services and content, these logs capture and retain details of patron interactions that may be sensitive. Server logs almost always capture every request issued, tied to a specific IP address. That IP address in itself may or may not be able to be traced to a specific individual, but it may provide clues to physical location, and data from other sources may be able to link that IP address to some level of identification.

Reviewing a few of the basics of what happens when viewing a website helps underscore the privacy concerns. In response to a request, usually evoked by clicking on a link or pressing a form button, the web server transmits a file corresponding to the URL, encoded in some flavor of HTML or XML, to the IP address associated with the web browser making the request. The entire transmission, including the URL, the page requested, any embedded scripts, and data associated with a POST directive, can, if it is not encrypted, be captured through eavesdropping hardware or software and viewed.

The web server will also record the request in its log. What the server records in its logs depends on how it was configured, but a typical log entry might resemble this one from Library Technology Guides, with the following selection of fields:

```
2016-01-03 22:56:13 64.150.189.27 GET
/libraries/search.pl ILS=Alma 80 -
107.133.80.235 Mozilla/5.0+(Windows+
NT+10.0;+WOW64)+AppleWebKit/537.36+(
KHTML,+like+Gecko)+Chrome/47.0.2526.
106+Safari/537.36 200 0 0 780 http://
librarytechnology.org
```

Which can be presented in a more readable way:

- **Date and Time Stamp:** 2016-01-01 22:56:13
- **Server IP:** 64.150.189.27
- **Method:** GET
- **URI Stem:** /libraries/search.pl
- **Query String:** ILS=Alma
- **Port:** 80
- **Client IP:** 107.133.80.235
- **User Agent:** Mozilla/5.0+(Windows+NT+10.0;+WOW64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/47.0.2526.106+Safari/537.36
- **Response Code:** 200
- **Bytes Transferred:** 780
- **Referrer:** http://librarytechnology.org

Server logs preserve a great deal of information describing a visitor session. In addition to the exact time the resource was requested, other information includes the previous page or site that the browser requested. This “referrer” data provides interesting information about what other resources funnel users to the site and internal navigation.

It should also be noted that the query string can reveal specific information about search behavior. In this case, it shows that the request involved a search for libraries using the Alma ILS. In this case, the query string was presented as part of the URL in a GET directive. If the POST directive were used instead, the same information would be transmitted via a separate data stream and placed on a temporary file on the server.

Subsequent entries from the same session would show what specific entries from the search results were displayed. This data describing information-seeking behavior is transmitted across the Internet and stored in server logs.

As we will explore in more depth below, these same mechanisms apply to online catalog, discovery service, or other library interfaces, where the behavior involved may include a search issued by a library patron, lists of items held by the library, and which specific title was selected. In the case of an e-book or other electronic resource, this data at least implies reading behavior. These sequences of data represent patron interactions that fall into the same level of concern as circulation records for physical books. These categories of data may or may not be covered by any given library’s privacy policies, but it is important to understand the technical reality that search and reading behavior is routinely exposed in the operation of web-based resources.

The data transmitted by the example above does not necessarily include personally identifying information. But it does include contextual data with the potential to be narrowed to a specific individual. The IP address identifies the device associated with the request. In some cases, the computer routinely used by an individual may have a fixed IP address, which then represents a strong link to a specific person. In other cases, the IP address recorded may be the router to the network connecting a household, organization, or larger set of devices to the Internet. The common practice of dynamically issued IP addresses further weakens the link between an IP address and a given individual.

The application generating the page transmitted may operate with additional levels of personal data. Any site with the ability for users to register and sign in with a personal account will have the potential to associate that account to specific online behaviors. In some cases, that profile can be associated with a username or handle not necessarily validated to an individual in real life. In other cases, that profile may be linked and validated to a specific individual. The automation systems used by libraries, for example, are usually validated to a specific individual with personal details such as physical addresses, phone numbers, and demographic details.

In previous times, library accounts would frequently record Social Security numbers or other official identification numbers. Fortunately this practice has largely been abandoned. Academic libraries, for example, would instead record the identification number issued by the educational institution.

From a privacy perspective, the application must securely contain personal details and behavior internally and not allow these details to be exposed externally, according to the policies and procedures in place relating to the confidentiality of patron records. Queries performed, titles selected, items currently and previously checked out, or lists of favorite items are some of the elements that may be internally stored in association with a patron's record or profile within an integrated library system or discovery environment, expanding the scope of concern beyond the records stored in databases to the log files of any web servers involved.

Tracking Tags and Web Beacons

Another mechanism that has become a routine part of web-based systems involves the use of what are commonly called web beacons or tracking tags. These tags are bundles of information sent to an external service to perform a specific function. Tracking tags may support analytics related to website usage, performance monitoring, or management of advertising content.

One of the most popular—almost ubiquitous—uses of tracking tags can be seen in sites configured for Google Analytics. This service, which Google makes available without cost, operates on the basis of collecting data transmitted with each page request. Website managers enable Google Analytics by establishing an account that is assigned an organizational identifier. Using the Google Analytics administrative tool, a snippet of code is produced that includes an institutional and site identifier, which is embedded on each page. This snippet executes JavaScript that is programmed to send specific data to Google's servers with each page request.

Google describes the categories of information it collects for any of its services.[4] The Google Analytics Developers pages provide more specific information transmitted to Google's servers for any page request.[5] At least the same level of data is sent to Google as is captured on local web server logs. But in addition to the user's HTTP request and the signatures of the web browser and the operating system of the user's computer, Google also captures contents from cookies. The first-party cookies authorized to be accessed include any from the Google family of products, which also includes the AdSense and DoubleClick advertising services.

The transmission of these data elements through web beacons to Google or other organizations does not necessarily include personally identifiable content. These data elements include details regarding the page requested, the previous page visited, time stamps, the IP address of the requestor's browser, and cookie data that provides considerable information regarding the session. It's also possible that non–personally identifiable information from a library search session might be linked with personally identifiable content captured from that individual's access to other non-library sites, with inferences of identification. If a visitor to a site that uses Google Analytics is signed into his or her Google account, there may be an increased possibility that activity carried out on that page could be directly linked to that account holder.

Mayer and Mitchell describe the privacy issues involved when a web page enables a tracking code to a third-party site:

> Web browsing history is inextricably linked to personal information. The pages a user visits can reveal her location, interests, purchases, employment status, sexual orientation, financial challenges, medical conditions, and more. Examining individual page loads is often adequate to draw many conclusions about a user; analyzing patterns of activity allows yet more inferences.
>
> When a first-party page embeds third-party content, the third-party website is ordinarily made aware of the URL of the first-party page through an HTTP referrer or equivalent. If the page embeds a script tag from a third party, the third party will also often learn the web page's title from `document.title`. Some first parties will voluntarily transmit even more information.[6]

The insertion of web beacons into library-branded pages at a minimum expands the matrix of organizations with technical data describing any given element of online behavior. The data may or may not cross any thresholds of privacy. It does seem important for libraries to be fully aware of the data transmitted to any third party relative to actions performed by their patrons through resources they provide. This report includes a survey of library websites that itemizes the web beacons detected. No further analysis was conducted to discern the specific information transmitted. This survey was conducted primarily to observe the degree to which libraries include these web beacons and which organizations receive any data regarding patron transactions carried out on library sites.

Libraries may want to conduct a thorough audit of their websites and services to gain a detailed understanding of what information is transmitted to any third parties through web beacons or similar techniques. Including these devices can be defended from a privacy perspective based on confidence of what data is transmitted and trust in the organization receiving that data. In some cases, web beacons may be enabled casually or even accidentally. It is common to include scripts and code from other sources without an exact understanding of what beacons may be embedded or what code may be executed on third-party sites.

The following two chapters include two empirical studies that relate to the treatment of privacy on library websites and discovery services. Chapter 2 provides data from a questionnaire completed by a selected set of vendors offering online catalog or discovery services that probes their capabilities and strategies regarding encryption of data transmitted and stored within their systems that may include personal information. Chapter 3 reflects data collected from two sets of large libraries regarding the secure transmission of library websites, catalogs, and discovery services and the presence of web beacons detected on the sites.

The use of cookies, another technique with privacy implications used by websites, is not covered in this report. Cookies are small data files that a web page may deposit on the computer used for session continuity, personalization features, and management of advertising content. In most cases, a cookie can be accessed only by pages associated with the organization that created it. This organization may span multiple entities with different services and activities. Google, for example, may share cookie content among its properties, including AdSense and DoubleClick. Opportunities for further study would include the use of cookies by library websites and catalogs.

Notes

1. “An Interpretation of the Library Bill of Rights, Privacy,” American Library Association, accessed January 10, 2016, www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.
2. Ibid.
3. Marshall Breeding, “Smarter Libraries through Technology: Protecting the Privacy of Library Patrons,” *Smart Libraries Newsletter* 35, no. 1 (January 2015): 1.
4. “Information We Collect,” Privacy, Google, accessed February 8, 2016, www.google.com/policies/privacy/#infocollect.
5. “Tracking Code Overview,” Google Analytics, accessed February 8, 2016, https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview?csw=1.
6. Jonathan R. Mayer and John C. Mitchell, “Third-Party Web Tracking: Policy and Technology,” in *Proceedings: 2012 IEEE Symposium on Security and Privacy, S&P 2012* (Los Alamitos, CA: IEEE Computer Society, 2012), 3, http://dx.doi.org/10.1109/SP.2012.47, available online at https://jonathanmayer.org/papers_data/trackingsurvey12.pdf.