ALA American Library Association

PROTECTING PRIVACY ON LIBRARY WEBSITES

CRITICAL TECHNOLOGIES AND IMPLEMENTATION TRENDS

Marshall Breeding

Library Technology Reports
Expert Guides to Library Systems and Services

OCTOBER 2019
Vol. 55 / No. 7
ISSN 0024-2586

Library Technology
REPORTS
Expert Guides to Library Systems and Services

Protecting Privacy on Library Websites: Critical Technologies and Implementation Trends

Marshall Breeding

ALA TechSource
alatechsource.org

American Library Association

Library Technology
REPORTS

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

Volume 55, Number 7
Protecting Privacy on Library Websites: Critical Technologies and Implementation Trends
ISBN: 978-0-8389-1820-3
DOI: https://doi.org/10.5860/ltr.55n7

American Library Association
50 East Huron St.
Chicago, IL 60611-2795 USA
alatechsource.org
800-545-2433, ext. 4299
312-944-6780
312-280-5275 (fax)

Advertising Representative
Samantha Imburgia
simburgia@ala.org
312-280-3244

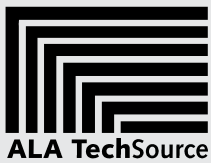
Editor
Samantha Imburgia
simburgia@ala.org
312-280-3244

Copy Editor
Judith Lauber

Production
ALA Production Services

Cover Design
Alejandra Diaz and ALA Production Services

Library Technology Reports (ISSN 0024-2586) is published eight times a year (January, March, April, June, July, September, October, and December) by American Library Association, 50 E. Huron St., Chicago, IL 60611. It is managed by ALA TechSource, a unit of the publishing department of ALA. Periodical postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to Library Technology Reports, 50 E. Huron St., Chicago, IL 60611.



alatechsource.org

About the Author

Marshall Breeding is an independent consultant, speaker, and author. He is the creator and editor of the Library Technology Guides website, editor of *Smart Libraries Newsletter*, and a columnist for *Computers in Libraries*. He has authored the annual "Automation Marketplace" feature published most recently in *American Libraries*. He has also edited or authored several books, including *Library Technology Buying Strategies*. Formerly the director for innovative technology and research for the Vanderbilt University Library, he regularly teaches workshops and gives presentations internationally at library conferences.

Abstract

This issue of *Library Technology Reports* (vol. 55, no. 7), "Protecting Privacy on Library Websites: Critical Technologies and Implementation Trends," explores the issues and technologies needed to deploy a library website with adequate protections for the privacy of those who visit. Without the implementation of standard encryption components, the online information-seeking activities of website visitors are vulnerable to exposure. Even when a site is properly encrypted, privacy can be circumvented through tracking agents placed on the site for analytics or advertising. Following discussion of the technical issues with implications for user privacy, this report includes the results of a broad study of the state of practice for these privacy-related technologies among public and academic libraries in the United States. This study reveals great progress among these libraries in the strengthening of privacy on their websites, though substantial gaps remain.

Subscriptions

alatechsource.org/subscribe

Contents

Chapter 1

Introduction

Libraries regard the protection of the confidentiality and privacy of those who make use of their content and services as a core value. Yet libraries today face many obstacles in achieving optimal privacy and security in the implementation of their websites and other aspects of their technical infrastructure. Financial resources and technical expertise to implement the latest technologies for computer and network security may not always be available. While some libraries have deployed state-of-the-art systems, others struggle to gain access to even the most basic technologies.

The challenges are not all related to availability of resources. Libraries must also deal with the tensions, if not direct contradictions, between protecting privacy and their interest in providing services that meet the expectations of their patrons. In the context of the overarching concern to protect the privacy of their patrons, libraries also desire to implement tools and technologies that provide more personalized services and that might provide opportunities for better engagement with their patrons. The boundaries are not necessarily clear between the values of protecting privacy and the tools and technologies available for personalized services.

The commercial web today works on a business model of sales and advertising based on aggressive collection and sharing of personal data. The basic fabric of commercial technology, including the infrastructure of the web and in-person retail environments, has been honed to capture as much personal data as possible. This data powers a global advertising ecosystem designed to strengthen commercial sales through ever more finely targeted placement of ads. Libraries, in contrast, embrace a model of providing services based on privacy and confidentiality. For libraries to implement websites and other technologies that reflect their values of privacy in the context of a global infrastructure optimized for commerce and advertising invariably involves difficult choices and some compromise. While libraries may not be able to entirely isolate their web-based services from commercial technologies, they can implement measures that limit exposure and that meet their expectations for protection of privacy.

This issue of *Library Technology Reports* explores these issues and the technologies needed to deploy a library website with adequate protections for the privacy of those who visit. Without the implementation of standard encryption components, the online information-seeking activities of website visitors are vulnerable to exposure. Even when a site is properly encrypted, privacy can be circumvented through tracking agents placed on the site for analytics or advertising. In some cases, tracking mechanisms may be included inadvertently, such as when they are brought in through components used for desired features. Following discussion of the technical issues with implications for user privacy, this report includes the results of a broad study of the state of practice for these privacy-related technologies among public and academic libraries in the United States. This study reveals great progress among these libraries in the strengthening of privacy on their websites, though substantial gaps remain.

Libraries Value Privacy

This report is based on the fundamental concept that the values of the profession mandate that libraries implement technology systems able to respect confidentiality and protect privacy. The American Library Association provides a clear statement of the responsibilities of libraries related to this important topic in a document that was adopted by the ALA Council initially in 2002 and subsequently updated in 2014 and 2019. The following excerpts reinforce the aspects of privacy central to this issue of *Library Technology Reports*:

> The library profession has a long-standing ethic of facilitating, not monitoring, access to information. Libraries implement this commitment through

> the adoption of and adherence to library privacy policies that are consistent with applicable federal, state, local, and where appropriate, international law. It is essential that libraries maintain an updated, publicly available privacy policy that states what data is being collected, with whom it is shared, and how long it is kept. Everyone who provides governance, administration, or service in libraries, including volunteers, has a responsibility to maintain an environment respectful and protective of the privacy of all users. It is the library's responsibility to provide ongoing privacy education and training to library workers, governing bodies, and users in order to fulfill this responsibility.
>
> . . .
>
> The American Library Association affirms that rights of privacy are necessary for intellectual freedom and are fundamental to the ethical practice of librarianship. The rapid pace of information collection and changes in technology means that users' personally identifiable information and library-use data are at increased risk of exposure. The use of new technologies in libraries that rely on the collection, use, sharing, monitoring and/or tracking of user data may come into direct conflict with the *Library Bill of Rights* and librarians' ethical responsibilities. Libraries should consider privacy in the design and delivery of all programs and services, paying careful attention to their own policies and procedures and that of any vendors with whom they work. Privacy is the foundation upon which our libraries were built and the reason libraries are such a trusted part of every community.[1]

Libraries provide access to information both through their physical facilities and through their websites. Within their physical premises, libraries take great care to ensure that information about the resources and services accessed by a patron remains private and is not shared with other individuals or organizations. The integrated library systems used to manage the lending of materials are configured to maximize patron privacy. While it is necessary to maintain a link between bibliographic records and patron records when an item is borrowed, extensive measures are taken to ensure the privacy of the transaction. While the loan is active, the connection between the item and patron data is needed to support operational tasks such as sending notices when the item is past its loan period. But once the item has been returned, libraries routinely remove all traces of the transaction from the systems involved. It is common for libraries to retain only anonymized data for concluded circulation transactions so that no records are available that reveal what items any given patron has borrowed or consulted. Log files that may otherwise hold data related to these transactions are likewise scrubbed or anonymized. Even during the interval of an active loan transaction, precautions are implemented to ensure that only specifically authorized personnel are able to view patron data, including the items on loan.

The removal of data describing completed loan transactions is only one example of the measures libraries take to ensure that no traces remain regarding the specific resources that patrons may have accessed. When asked for information regarding the resources any given patron may have accessed, even by law enforcement agencies, libraries want to be able to truthfully respond that the information is not available. Even in the event of a security breach, there should be the least possible personally identifiable information or data regarding information access. This approach toward the privacy of access to resources enables patrons to use information provided by the library without fear of judgement or reprisals.

Privacy versus Personalized Services

These measures taken to protect privacy can be seen as a constraint on the ability of the library to engage in personalized services or to enable social features. Removing data related to completed loan transactions, for example, eliminates the ability of patrons to view items they have previously borrowed, a feature most persons would expect to be available. Any e-commerce site would track all previous purchases and use data collected on items bought or viewed to present recommendations. These environments would also use data from other customers to inform recommendations: "Others who purchased this item also purchased these."

To support these kinds of personalized services, recommendations, and social sharing features, many libraries enable the collection of the associated personalized data. This collection of personalized data would usually be enabled through specific patron consent. Patrons would have the ability to opt in to retention of data on items borrowed or other types of interactions in order to receive enhanced personalized services. Whether opt-in or opt-out options are selected by default would be determined by library policy, reflected in the organization's stated privacy policies.

Protection of Online Information-Seeking Activities

Libraries use their websites to provide information regarding their facilities and services and as portals

through which their patrons can explore information resources. Libraries provide extensive collections of electronic resources and other digital content for access to the general public and to their websites. A typical interaction includes a patron typing a topic of interest into a search box, viewing results, making selections, and viewing or downloading content. These transactions include data regarding the topics of interest and items accessed by a given individual at a specific time. Even when the patron accessing the information isn't signed into a library account, technical information from the network and browser may identify, or at least imply, a specific individual.

Patron use of a library website involves data at least as sensitive as data related to physical items borrowed. For many—probably most—libraries, the quantity of information accessed by patrons through the website exceeds loans of physical materials. Achieving the same level of privacy for information access by library patrons through the website as for transactions representing physical loans requires attention to some technical details relating to the library's website and any related systems or services. Privacy for web-based transactions requires that no one can listen in on the network in a way that reveals patron information-seeking activities and that the data related to the transaction not be shared with other individuals or organizations.

This report explores issues relating to the privacy and security of data that represents the online information-seeking activities of individuals through a library website. For the purposes of this report, the term *library patron* means any person who uses library resources, including the general public. The report focuses on the technical and functional characteristics of the main library website. Online catalogs, discovery services, and the extensive portfolio of information resources that may be accessed through a library website are all subject to the same concerns but are not directly addressed in this study.

Note

1. American Library Association, *Privacy: An Interpretation of the Library Bill of Rights*, (Chicago: American Library Association, 2002, amended 2014 and 2019), http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.

Chapter 2

Key Technologies with Implications for Privacy

Encryption, Analytics, and Advertising Tracking

Need for Encryption of Websites

To be consistent with library privacy policies and values, the interactions of how persons use library-provided resources must be protected from access by any third party. When communication takes place over a network, especially the internet, it is possible for unknown parties to intercept the data. Interception or eavesdropping can take place on wireless or wired networks and can be opportunistic, targeted, or widespread. Tools for intercepting communications on local networks are readily available and inexpensive. More sophisticated surveillance equipment may be inserted into internet infrastructure to gain more widespread access. This vulnerability to capture of network communications is well known and addressed through well-established encryption techniques. Encryption of web traffic, implemented through the HTTPS protocol, ensures that the contents of the transmission cannot be viewed even if the communication stream is captured.

Given the possibilities for interception and eavesdropping, it must be assumed today that any information transmitted on the web can be captured. The contents of captured communications can be easily accessed when they are transmitted without additional protection. Only with strong encryption technologies can information transmitted across networks be considered private. Encryption does not prevent others from intercepting communications, but it ensures that no one other than the sender and receiver can view the contents and that the contents have not been altered.

The "Policy to Require Secure Connections across Federal Websites and Web" issued by the Chief Information Officer of the Office of Management and Budget of the US federal government mandates the use of HTTPS on government websites and provides a concise summary of the dangers of using HTTP: "The American people expect government websites to be secure and their interactions with those websites to be private." And later: "The unencrypted HTTP protocol does not protect data from interception or alteration, which can subject users to eavesdropping, tracking, and the modification of received data. The majority of Federal websites use HTTP as the as primary protocol to communicate over the public internet. Unencrypted HTTP connections create a privacy vulnerability and expose potentially sensitive information about users of unencrypted Federal websites and services. Data sent over HTTP is susceptible to interception, manipulation, and impersonation. This data can include browser identity, website content, search terms, and other user-submitted information."[1]

HTTPS for Identity Validation

The use of HTTPS also confirms the identity of the website. It is essential that visitors be able to confirm that any website is legitimate and is not being spoofed. The digital certificates used to encrypt the transmission from the site also include authoritative information on the organization to which the certificate was issued. Digital certificates are issued by trusted certificate authorities that validate the ownership of the certificate. To establish a secure connection, a valid certificate must be installed in the web server, and the ownership embedded in the certificate must match its domain. Any mismatch will produce an error and the page will not be secured. Visitors to the website can inspect the certificate used for an

HTTPS site to confirm that the site belongs to the expected organization.

Low Threshold of Difficulty and Expense

The means to protect communications on the web are readily available and inexpensive. Any reasonably current web server software can be configured to encrypt the content it publishes. Once the website has been configured to deliver pages with HTTPS instead of HTTP, it uses a suite of protocols for encryption technologies, including TLS or Transport Layer Security, that cannot be decrypted while the data traverses the internet.

In order to enable HTTPS on a web server, the organization must obtain a digital certificate. These certificates are issued through a "certificate authority" and come in different categories. These certificates differ in the level of validation performed for the organization and its right to use the domain:

- **Extended validation:** The certificate confirms the organization's exclusive right to use the domain and performs an extensive review of the organization details relative to official business records. Sites with this type of certificate will present the name of the organization in the URL bar of most browsers along with the indicator that the site is encrypted using HTTPS.
- **Organization validated:** The certificate authority confirms the organization's right to use the domain. If properly validated, the organization's name will be shown when the user views the details of the certificate in the browser. For sites with this type of certificate, the URL bar of the browser indicates that the site is encrypted using HTTPS.
- **Domain validated:** The certificate authority confirms the organization's right to use the domain but does not require extensive documentation regarding the organization. For sites with this type of certificate, the URL bar of the browser indicates that the site is encrypted using HTTPS.

Certificate authorities will charge higher fees for certificates requiring more extensive organizational vetting and validation. These costs currently are about $25 per year for domain validated certificates; $75 for organization validated; and $400 per year for extended validation. Wild card certificates that support multiple subdomains will also involve additional fees.[2]

The nonprofit initiative Let's Encrypt provides free digital certificates to any organization. Let's Encrypt has developed a method to automatically install, configure, and renew certificates with minimum expertise or effort. While these certificates enable encryption, they do not provide the higher level of organizational validation available through traditional certificates.

Let's Encrypt
https://letsencrypt.org

Another category of certificates are those issued by the organization itself and not through a certificate authority. These self-signed certificates can be used for basic encryption, but do not provide any assurance that the website is legitimate. These certificates are typically used for testing and will trigger a warning on most web browsers.

Sites without a digital certificate cannot encrypt pages with HTTPS and will be limited to the HTTP protocol, which delivers pages as viewable text. Again, this option does not meet the basic requirement for privacy for a library website.

Advancing to HTTPS Everywhere

The web has been in the process of transition from its initial deployment based on HTTP to universal implementation of HTTPS for more than a decade. In the earlier phases of the web, the HTTPS protocol was available, but its use was targeted to specific tasks involving sensitive information, such as the entry of credit card numbers or passwords. At that time, the process of setting up HTTPS on web servers was more complex and the additional computations needed for encryption were substantial. With current web server hardware and software, the overhead for implementing HTTPS is negligible. Today it is expected that all web traffic should be carried with HTTPS encryption. All major commercial destinations and social networks have switched entirely to HTTPS.

Google has played a major role in the transition to HTTPS. Given its dominance in search, web browsers, and general web services and infrastructure, its policies and practices have a massive impact on the broader sphere. Google Chrome, for example, currently has 63.3 percent of the market share for web browsers, with Firefox a distant second at 9.5 percent.[3]

Google has been exerting increasing pressure to entice websites to make the switch to HTTPS. This pressure comes in the form of warnings issued through its Chrome web browser and through its ranking of search results. All web browsers present some type of indicator when a site has implemented HTTPS. From the earliest phase of the web, users have been aware that they must check for this positive indicator of encryption before entering credit card information, passwords, or other sensitive information.

Pages not encrypted were given a neutral status indicator. Following a generous period of advance notice, Google changed its neutral treatment of non-HTTPS sites to a conspicuous negative indicator. Beginning in July 2018, web pages not encrypted with HTTPS via a valid digital certificate have been flagged as not secure (figure 2.1). Clicking on the information indicator presents this text: "Your connection to this site is not secure. You should not enter any sensitive information on this site (for example, passwords or credit cards) because it could be stolen by attackers."

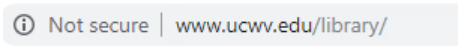


Figure 2.1. Example of Google Chrome unsecure warning

This treatment contrasts with that given pages using HTTPS (figure 2.2). Clicking on the lock icon brings up this text: "Connection is secure. Your information (for example, passwords or credit card numbers) is private when it is sent to this site."

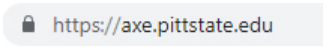


Figure 2.2. Example of secure website on Google Chrome

Meeting the HTTPS Deadline

This date of July 2018 was generally regarded as a deadline by which responsible organizations had to implement HTTPS or face the repercussions of their content being flagged as unsecure.[4] As demonstrated by the data collected in support of this report, most of the public and academic libraries in the United States have met this deadline, though a substantial portion remain not in compliance with this essential requirement.

Scott Helme, an internet security researcher, provides useful information demonstrating the progress made in the transition to HTTPS on the general web. Data describing the use of HTTPS by the top one million websites as tracked by the Alexa Internet service shows a steady climb from 6.71 percent in August 2015 to 58.44 percent in February 2019.[5] The data in this report related to the transition to HTTPS in libraries is roughly on track with the broader web trend seen in the Alexa statistics.

Alexa
https://www.alexa.com/

HTTPS and Only HTTPS

In addition to implementing HTTPS, it is also important to implement mechanisms to ensure that site visitors are not intentionally or accidentally directed to an unsecure version. Websites should be configured to always direct users to the secure version using HTTPS, even if they come to the site with a link coded with HTTP. The website should automatically redirect HTTP to HTTPS, providing protection even if the user types in http:// or comes in through an outdated version of the URL. The "HTTP Strict Transport Security (HSTS)" standard describes a protocol that can be implemented in web servers to implement comprehensive use of HTTPS.[6]

Even if a website has been configured to enable HTTPS, if it allows its pages to be accessed via HTTP, it should be considered vulnerable from the perspective of user privacy. In addition to gathering data on the number of websites for public and academic libraries implementing HTTPS, this report also assesses whether these sites implement the expected redirection behavior to ensure that HTTPS is always used.

Challenges in Implementing HTTPS

Even with the low threshold for the technical implementation, a number of challenges can hinder an organization from making the transition to HTTPS. These challenges often relate to dependencies on external resources that do not support HTTPS. In order to be validated as secure, the page, as well as any links or embedded content, including images, style sheets, and JavaScript libraries, must be delivered via HTTPS. All links to external web pages and services must also be HTTPS. If any HTTP links or content is detected, browsers will issue a conspicuous error message warning of unsecure content mixed into the page.

In the library context, avoiding these mixed content errors means that the library catalog, discovery services, and all information resources linked to from the site must be available via HTTPS links. If any of these vendors cannot conform to this requirement, the library may have to delay its own implementation of HTTPS. Since libraries' websites often exist to provide access to information resources to their patrons, ensuring comprehensive use of HTTPS throughout their portfolio of database and content products can be an extensive process. The switch to HTTPS on the library's main website may also need to be coordinated with similar changes to the online catalog, institutional repositories, blogs, or other local resources.

Some libraries may also opt to make the transition to HTTPS as part of a redesign of the library's website or a move to a new hardware or software platform. When part of a larger project, the implementation of

HTTPS may take longer than if it were an isolated task.

Libraries may also be limited by the technologies implemented by their parent institution. If the library web presence operates within the website of a university or local government, it may not have the means to make this change independently. For some libraries, working with the institutional infrastructure may mean a quicker adoption of more secure technologies.

Mandate for Libraries

Libraries have generally lagged behind the commercial sphere in the transition from HTTP to HTTPS. Despite the values-driven necessity of providing a secure and private environment for accessing library content and services, some libraries may not be well informed regarding these vulnerabilities or may lack the technical expertise or the personnel resources to implement these needed changes.

Not implementing HTTPS places libraries in an unfortunate position of their websites being flagged as not private or secure, despite their role in providing access to trusted and vetted resources. Sites implementing HTTPS will receive no such warnings, regardless of the nature of the content they publish. Although technical security and privacy configurations and the quality of content curated are entirely distinct issues, these distinctions may not be well understood by all persons. The reputation of a library can therefore be diminished if it does not attend to these critical technical details.

Analytics and Advertising Networks

Privacy concerns extend beyond configuring a server to correctly implement HTTPS encryption. Although the content of pages delivered through HTTPS cannot be viewed or altered, many other practices can compromise privacy. Even on encrypted pages, site managers can compromise the privacy of their users by including scripts or widgets that provide data to external entities. These tracking mechanisms may be positioned by the providers as innocuous but need to be well understood by organizations with heightened concerns for privacy such as libraries.

The ALA statement *Privacy: An Interpretation of the Library Bill of Rights* also addresses this topic: "Libraries should not monitor, track, or profile an individual's library use beyond operational needs. Data collected for analytical use should be limited to anonymous or aggregated data and not tied to individuals' personal data."[7]

This report studies two basic categories of tracking agents that might be added to library websites. Those related to analytics pass information regarding the use of the website to an external server, enabling website managers to observe patterns of use. The other category involves making connection to advertising networks, leaving the possibility for intermingling library sites with a presumption of privacy and commercial networks based on extraction and sharing of personal data.

Measuring Website Use through Analytics Services

Libraries, like other types of organizations, have a strong interest in measuring the use of their websites. In addition to gaining a general understanding of a site's level of use, an organization can use sophisticated analytics tools to help identify problems on the site and to inform improvements in design and functionality. Website analytics tools can take two different approaches.

- **Server log analysis:** One category is based on processing the log files produced by web servers that record each resource requested. This approach works without involvement of any external resource but may involve a higher level of difficulty. Log-based analytics require access to the internal system resources of the web server, which may be difficult in some organizations where multiple sites operate through the same server. These products also may involve the installation and configuration of the analytics software. This model of analysis was common during the earlier phase of the web but has declined due to the popularity of Google Analytics. Some organizations will use both server log analysis tools and analytics based on page tagging to get a more complete view of the use of their site. Server log tools, for example, can capture access by search indexing crawlers, which represent a substantial portion of server load, though not actual visitor activity.
- **Page tagging:** The other model relies on sending data to an external analytics service as each page is accessed. The website manager places a snippet of code on each page, usually through a standard inclusion component. The analytics tag would be included in much the same way as headers, navigation, JavaScript libraries, or style sheets to provide consistent branding and layout.

One of the topics addressed in this report relates to the use of analytics for library websites. The data collected for this library privacy study demonstrates that a large percentage of libraries use Google Analytics, a free service for measuring use and for optimizing the usability of websites. This service relies on websites

transmitting detailed usage data to Google. Libraries need to assess whether the use of this service falls within what is allowed by professional values and by the privacy policies of each library organization. From a technology perspective, we can observe that the service involves sending data describing patron information-seeking activities to a third party, which must be trusted to limit the way in which that data is used.

Google Analytics and other services from Google are designed to directly or indirectly support the company's business interests. Google earns most of its revenue through advertising. According to Statista.com, in 2018, Google reported total revenue $136 billion; of that, $120 billion came from advertising.[8] The basis of Google Analytics in the commercial advertising ecosystem warrants careful analysis to ensure that its use remains consistent with the library's privacy policies.

Google Analytics has become the dominant tool used for assessing the use of websites. As shown in the data collected for this study, it is used by all types of organizations, including libraries. Although some libraries use other tools for use statistics and analytics, this report focuses on Google Analytics given its widespread use among libraries.

Google Analytics

The implementation of Google Analytics involves two tasks, the creation of an administrative account and the inclusion of a snippet of JavaScript on each page. Each website, or "property," configured through the Google Analytics administrative console is assigned a unique identifier, which must be included in the JavaScript snippet.

Once Google Analytics is activated, each time a page is accessed on the site, information will be transmitted to Google's servers to enable detailed analysis and measurement of use patterns. The data transferred does not necessarily contain personal information about the individuals visiting the website, but it does include detailed information regarding the resources used on the site. In some cases, the data could include information regarding the topics or specific items searched for or accessed on the site. That information can be conveyed on the query string of a URL as one of the elements tracked. All resources accessed within a session are tied together through a unique identifier Google Analytics assigns and records in a browser cookie. This identifier is not associated with a specific individual through the data collected within Google Analytics.

Depending on the circumstances and interpretation, the IP address of a website visitor can be considered a personally identifiable data element. The GDPR (General Data Protection Regulation) framework of the European Union, for example, considers the IP address as personal information in some contexts.[9] Depending on the way that IP addresses are assigned, there can often be a strong correlation between an IP address and a specific device and the individual using that device.

Multiple Tracking Code Options

The code snippets that a site manager places on a web page to enable Google Analytics have changed over time. Each of these options follows the same model of page tracking associated with the site's unique identifier, though with each new version additional features have been added.

The initial Google Analytics snippet (figure 2.3), generally referred to as the Classic version, was introduced prior to HTML version 5 and supported both encrypted and unencrypted transmission of data to the Google Analytics servers. Although this version of the tracking code continues to work, Google recommends that all new sites be configured with the newer Universal analytics code.

```
script type="text/javascript">
 var gaJsHost = (("https:" == document.location.protocol) ? "https://ssl."
: "http://www.");
 document.write(unescape("%3Cscript src='" + gaJsHost + "google-
analytics.com/ga.js' type='text/javascript'%3E%3C/script%3E"));
</script>
<script type="text/javascript">
 try {
 var pageTracker = _gat._getTracker("UA-3203647-3");
 pageTracker._trackPageview();
 } catch(err) {}
</script>
```

Figure 2.3. Original Google Analytics tracking snippet

The Universal version of the Google Analytics tracking snippet uses the analytics.js JavaScript library (figure 2.4). This version always encrypts data as it is transmitted to the Google Analytics servers and includes options for anonymization of IP addresses.

In addition to directly embedding the Google Analytics code snippet into each page, the organization can also use the Google Tag Manager, another free tool from Google. This tool can enable other services that rely on tracking codes in addition to Google Analytics. While it is possible for a site to use the Google Tag Manager and not use Google Analytics, this practice is not common. The presence of the Global Site Tag tracking code for Google Tag Manager is a very strong indicator for the use of Google Analytics for pages where the other Google Analytics tracking snippets are not detected (figure 2.5). It is also possible for both the Google Tag Manager snippet and one of the Google Analytics tracking codes to be present within a web page.

If the organization has deployed Google Analytics using the Google Tag Manager, it may not be possible to detect the presence of the tracking code when inspecting the source code for the page. The Google support documentation states that only the page owner can see the tags activated through the Google Tag Manger console (see figure 2.6).[10] Browser plugins, such as Ghostery, will be able to detect the use of Google Analytics for these sites.

The default tracking code snippet currently presented through the Google Analytics console takes the form of the Global Site Tag rather than the Universal Analytics previously recommended.

Because of privacy concerns, Google Analytics includes a feature to anonymize IP addresses before they are recorded. This anonymization is essentially a truncation of the address so that it retains some useful information regarding the general location of the user. IP address anonymization can be specified in the Google Analytics JavaScript snippet, or it can also be configured in the administrative console of the Google Tag Manager.[11]

Google Analytics also includes a feature through which specific users can be tracked. This User-ID

```
<script>
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function()
{
 (i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new
Date();a=s.createElement(o),

m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(
a,m)
 })(window,document,'script','https://www.google-
analytics.com/analytics.js','ga');
 ga('create', 'UA-32191981-1', 'auto');
 ga('set', 'anonymizeIp', true);
 ga('send', 'pageview');
</script>
```

Figure 2.4. Universal version of Google Analytics tracking snippet

```
<!-- Global site tag (gtag.js) - Google Analytics -->
<script async src="https://www.googletagmanager.com/gtag/js?id=UA-
000000000-1"></script>
<script>
 window.dataLayer = window.dataLayer || [];
 function gtag(){dataLayer.push(arguments);}
 gtag('js', new Date());

 gtag('config', 'UA-32191981-1'),{'anonymize_ip': true};
</script>
```

Figure 2.5. Global Site Tag tracking code for Google Tag Manager

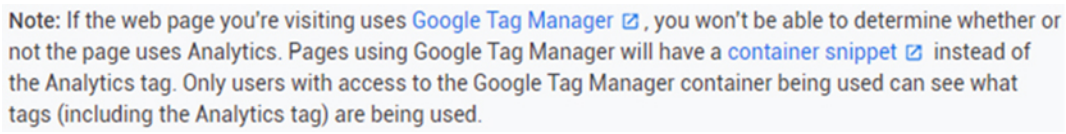


Figure 2.6. Explanation from Google Analytics support page.(Source: "Check if a Web Page Uses Analytics," Google Analytics Help, accessed July 24, 2019, https://support.google.com/analytics/answer/1032399?hl=en)

feature must be specifically configured in the Google Analytics Console, including a step agreeing to the associated privacy policy. The tracking code of the site is also updated to include the unique user identifier for the person accessing the page, which could be provided for those who have logged into the site. Activation of this feature would be inconsistent with the privacy policies of most libraries since it not only creates nonanonymized records of patron information-seeking activities, but also shares that data with Google.

When Google Analytics is used, all data relating to website use is transmitted to Google's servers. That data is used for reporting through the organization's Google Analytics account, but it may also be part of broader analytics or data mining. The Google Analytics console offers options regarding how Google employees may access the organization's data (figure 2.7). Enabling access to either Google's marketing specialists or all its sales personnel would seem inconsistent with general library practices regarding the treatment of patron use data.

Google also includes a variety of features in Google Analytics that allow an organization to enable linking with one or more Google Ads accounts. These features are useful to organizations that subscribe to Google's advertising services but would rarely be used on a library website, which usually does not offer advertising. Enabling these features allows collection of additional data and may also trigger collection of personally identifying information, such as for site visitors who are logged into a Google account. Figures 2.8 and 2.9 show the selections within the Google Analytics console that enable advertising features and extended data collection.

Advertising Networks and Social Media

The intermingling of library websites with advertising networks can introduce concerns for privacy. Analytics services involve transmission of data that may contain information-seeking activities of website visitors. Tracking codes and cookies for ad networks and social media sites represent a larger concern in regard to the privacy of patrons who access library websites. These organizations have strong interest in collecting or using information related to personal identity, interests, and past online interactions for targeting ads. In some cases, the tracking and interactions may be anonymized, and in others any current active logins, previously deposited browser cookies, or other mechanisms enable personal identification.

ProPublica has done research on the way that advertising and social networks track personal data. As far back as 2016, ProPublica reported that Google no longer separates information that it has about an individual through Gmail and other accounts and other browser data collected through DoubleClick: "The practical result of the change is that the DoubleClick ads that follow people around on the web may now be customized to them based on your name and other information Google knows about you. It also means that Google could now, if it wished to, build a complete portrait of a user by name, based on everything they write in email, every website they visit and the searches they conduct."[12]

Personal information is widely shared in the advertising ecosystem. This sharing of data across organizations can be easily observed. A search for a product on Amazon.com will cause ads for that product or similar ones to appear on Facebook and other sites. This "retargeting" mechanism is widely used by web destinations to show relevant ads based on browser

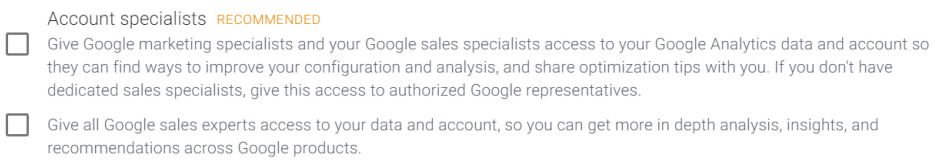


Figure 2.7. Google Analytics options for access to data by its personnel.

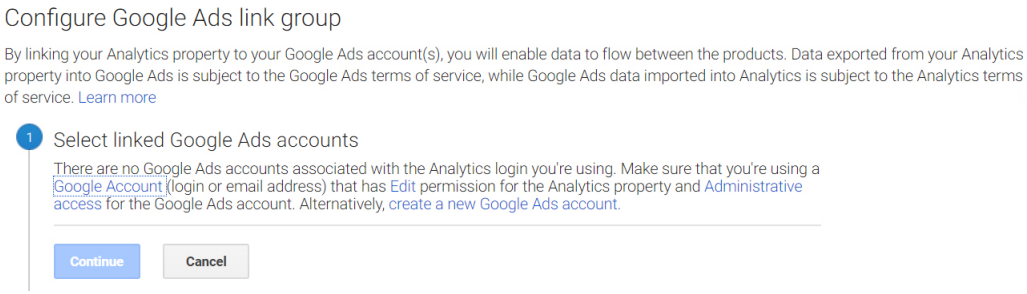


Figure 2.8. Configuring Google Ads link group

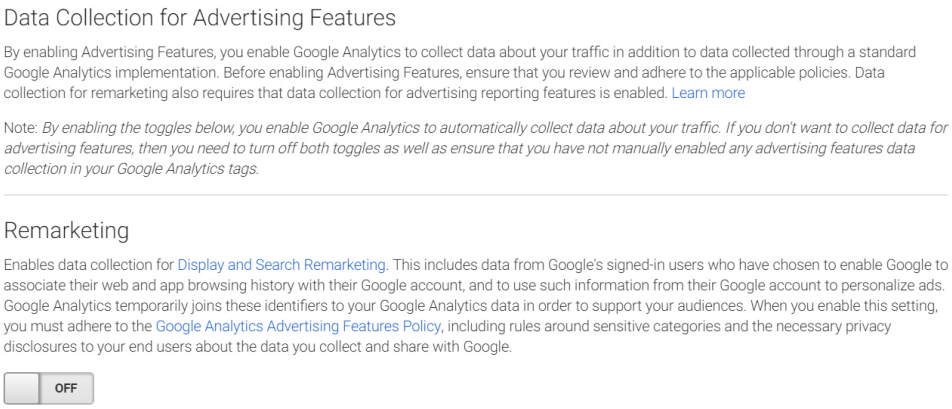


Figure 2.9. Data collection for advertising features and remarketing

history, third-party cookies, and other mechanisms.

The types of data and the mechanisms for sharing it among organizations and websites in the advertising ecosystem are complex and ever-changing. Libraries opting to enable ad-related tracking technologies will want to carefully investigate any possible external exposure of personal information or browsing history as individuals visit their websites and use their resources. Any scenario that allows content items searched for or viewed on a library website to later appear as ad suggestions on another site would not be consistent with library privacy values or most library privacy policies.

The advertising ecosystem continues to evolve toward ever more precise targeting capabilities, extending deeper into the realm of personally identifying information. One recent technique, seen with Google and Facebook, involves the concept of custom audiences. This technique involves the direct linking of known user information, such as from an organization's customer relationship management system or authentication service. In the library context, using these types of services would not be consistent with privacy protection since it involves sharing library patron data in bulk with an external organization:

> Recently, data brokers such as Facebook and Google have introduced a new feature on their advertising interfaces: custom audiences. Instead of creating audiences based on user attributes, advertisers can now upload personally identifying information (PII) about specific users; the platform then locates matching accounts and creates an audience consisting of only these users. The advertiser can then use this audience when placing ads, thereby showing their ads only to the specific users whose information they uploaded. For example, a small business may know the names and addresses of its customers; using custom audiences, the business can upload this information to Facebook, and then target these users with advertising directly. The custom audience feature has proven popular with advertisers: it allows them to directly select the users to whom their ad is shown, as opposed to only selecting the attributes of the users.[13]

In this study, a cursory screening is performed

to determine which websites may include tracking agents related to advertising or social networks. These trackers are not easily identified by the source code of the websites. A next phase of enhancements to the parsing scripts is planned that can more accurately identify these trackers.

Notes

1. Tony Scott, "Policy to Require Secure Connections across Federal Websites and Web Services," memorandum, Office of Management and Budget, June 5, 2015, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2015/m-15-13.pdf.
2. Examples from Register accessed July 26, 2019, register.com.
3. NetApplications, "Browser Market Share," June 2019, https://netmarketshare.com/browser-market-share.aspx.
4. Emily Schechter, "A Secure Web Is Here to Stay," *Google Security Blog*, February 8, 2018, https://security.googleblog.com/2018/02/a-secure-web-is-here-to-stay.html.
5. Scott Helme, "Alexa Top 1 Million Analysis—February 2019," *Scott Helme* (blog), March 11, 2019, https://scotthelme.co.uk/alexa-top-1-million-analysis-february-2019/.
6. J. Hodges, C. Jackson, and A. Barth, "HTTP Strict Transport Security (HSTS)," Internet Engineering Task Force, proposed standard, request for comments, RFC 6797, November 2012, https://tools.ietf.org/html/rfc6797.
7. American Library Association, *Privacy: An Interpretation of the Library Bill of Rights*, (Chicago: American Library Association, 2002, amended 2014 and 2019), http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.
8. See J. Clement, "Google's Revenue Worldwide from 2002 to 2018," Statista, last edited May 22, 2019, https://www.statista.com/statistics/266206/googles-annual-global-revenue/.
9. See Andrew Cormack, "IP Addresses, Privacy and the GDPR," *Jisc Community* (blog), April 4, 2018, https://community.jisc.ac.uk/blogs/regulatory-developments/article/ip-addresses-privacy-and-gdpr.
10. See Google Analytics Help, "Check if a Web Page Uses Analytics," accessed July 16, 2019, https://support.google.com/analytics/answer/1032399?hl=en.
11. See Cormack, "IP Addresses, Privacy and the GDPR."
12. Julia Angwin, "Google Has Quietly Dropped Ban on Personally Identifiable Web Tracking," ProPublica, October 21, 2016, https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking.
13. Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, and Oana Goga, "Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface" (paper, IEEE Symposium on Security and Privacy, San Francisco, CA, May 20–24, 2018), https://doi.org/10.1109/SP.2018.00014.

Chapter 3

The Current State of Practice

This section describes an ongoing study to track the progress of libraries in the implementation of technologies with implications for privacy and security. This study aims to show trends among the body of libraries considered and to help individual libraries become more aware of enhancements needed to their websites to provide better safeguards for privacy and security.

The concerns related to security and privacy issues have been widely disseminated in recent years. The level of compliance with at least nominal levels of conformance in library websites has widespread implications for library users. Increased implementation of encryption via HTTPS and the reduction of advertising trackers will provide increased protection for the private information and online behavior of library patrons as well as improve the reputation of libraries.

A longitudinal study has been underway since early 2018 to measure the implementation of security and privacy measures for public and academic libraries in the United States. This study takes advantage of data in the Libraries.org directory of libraries and automated procedures to capture the characteristics of library websites relating to privacy and security.

Methodology

This study centers on the technical characteristics of library websites in order to identify trends related to privacy and security. The methodology for the study involves automated inspection of library websites via the URLs recorded in the Libraries.org library directory. Only the main URL of each library organization is considered. Although the technical details of online catalogs, discovery services, repositories, and external information products have at least as much significance for the privacy of patron data, these were not considered in scope for this project and may be addressed in a later phase of work.

The automated scripts were developed by the author in the Perl programming language. These scripts initiate a request of the primary URL recorded for each selected library and test for a variety of technical characteristics related to privacy and security. The primary script can be run manually on demand and is also scheduled for automated execution monthly.

A reporting tool was developed to display the aggregate characteristics for each of the core selection groups. This tool includes a visualization of the portions of HTTP and HTTPS implemented across the libraries, any error codes recorded in crawling the sites, and the numbers of libraries where specific tracking agents were detected. Another reporting tool was created to display the security and privacy characteristics of each library, which can be viewed from each directory entry in Libraries.org.

Data Sources

The Libraries.org directory is a component of Library Technology Guides, a website maintained by the author that includes a variety of data repositories developed through a custom-built content management system. Data is managed through an implementation of the open source MySQL relational database. The content management system, controlling the presentation, entry, and editing of records, was written in Perl. Custom scripts developed in Perl enable the creation of specialized reports and visualizations related to any of the underlying data.

Libraries.org directory
https://librarytechnology.org/libraries

Library Technology Guides
https://librarytechnology.org

The Libraries.org directory includes a table that aggregates many different characteristics. The directory includes libraries from all countries, with over 185,000 total entries. Coverage across countries is uneven, with those in the United States having the most comprehensive and accurate data. The database includes 4,081 entries for academic libraries in the United States and 17,308 for US public libraries.

Although the Libraries.org directory includes data from all global regions, currently only the data for US public and academic libraries can be considered sufficiently complete and accurate for this type of study. Work is underway to improve data representing other counties to enable expansion of the study.

Data Structure

The table includes many different columns that describe the organizational structure, locational and demographic details, technology products implemented, statistics, and other categories. Some of the relevant columns for this study include

- **LibraryName:** the name of the library.
- **Institution:** the parent institution of the library.
- **LibraryWeb:** the URL for the library's main website.
- **LinkResponseCode:** the HTTP status code returned by the site.
- **LinkCheckDate:** the date when the site was last checked.
- **SecurityPrivacy:** a text field containing multiple name/value pairs relating to privacy and security. The multiple values structured into this field enable flexibility in what data is collected without having to add new columns to the main table.
  - **CheckDate:** the date the last automated check was performed.
  - **Protocol:** HTTP or HTTPS.
  - **Redirect:** detected behavior regarding redirection from HTTP to HTTPS.
  - **PageRetrievalStatus:** whether the automated process was able to capture the content of the web page.
  - **GoogleAnalytics:** whether Google Analytics was detected.
  - **Google Analytics Anonymize:** Is the setting enabled to anonymize Google Analytics data?
  - **Google Custom Search:** Is Google Custom Search implemented?
  - **Google Tag Manager:** Is the Google Tag Manager implemented?
  - **DoubleClick:** Tracking tag detected for Double Click?
  - **NewRelic:** Is the New Relic performance monitor enabled?
  - **CrazyEgg:** Is the Crazy Egg performance monitor enabled?
  - **Facebook Custom Audience:** Is the Pixel code for Facebook custom audience enabled?
  - **Facebook Connect:** Is Facebook Connect enabled?
  - **AddToAny:** Detection of the AddToAny sharing widget?
  - **ShareThis:** Detection of the ShareThis sharing widget?
  - **Inspectlet:** Is the Inspectlet user behavior monitoring tool implemented?
  - **TwitterAds:** Tracking tag detected for Twitter Ads?

Initial Data Collection and Cleanup

The ability to study the technical characteristics of library websites depends on maintaining accurate representations of their URLs. The links of library websites have been an element that has been maintained since the Libraries.org directory was created in 1995. When I started to prepare for the current study in 2016, the completeness and quality of these links were inconsistent. In order to assess the proportions of libraries using HTTPS, having a clean and comprehensive representation of the website URLs was essential.

A project to systematically update library website URLs for directory entries for all the public and academic libraries in the United States was accomplished in July 2017 with the assistance of J. J. Lamanna, Claire Schmieder, and other volunteers. This cleanup project involved finding valid URLs for sites where the URL was reported as broken through automated link checking and identifying working URLs for sites where they had not been previously recorded. Many libraries continue not to have websites; these libraries were also verified.

A relatively small percentage of these websites return HTTP error codes of 500. Most of these sites display through a web browser but may not respond to the testing performed through the automated script.

This work resulted in a set of records of sufficient quality to serve as the basis of the analysis of the websites of these libraries. The data set includes

- 17,308 public libraries, 16,263 of which have valid URLs recorded
- 4,081 academic libraries, 3,935 of which have valid URLs recorded

Automated Link Checking

Given the number of libraries of interest to this study, manual inspection of each site would not be feasible. Instead, automated tools were developed to probe each site and to collect specific characteristics. The Perl script used to validate links has been enhanced over time to include additional tests for redirection and for screening for tracking agents by searching the contents of the web page for specific text strings.

Manual Spot Checking

The data produced through the automated procedures was checked manually for smaller sample groups. This manual inspection was used to refine the scripts and to help identify text strings able to serve as reliable signatures of tracking agents. Manual testing included verifying whether HTTP or HTTPS was implemented through loading the page in a browser and whether expected redirection was implemented. The Google Chrome Developer Tools were used to investigate errors on websites. The Ghostery Chrome browser extension was used to verify the presence of tracking agents.

The methodology based on the inspection of the source coding used can easily underreport the tracking agents that may be employed by a site. The automated script checks only the top-level page and does not load any of the internal links that may activate tracking or advertising agents.

A browser-based utility, such as Ghostery, uses a much more sophisticated method for detecting tracking or advertising agents. Ghostery has a complete library of signatures for all known agents and processes each file linked within the page. Figure 3.1 illustrates Ghostery's ability to identify tracking agents on a website.

The less sophisticated method used for this study means that some sites that invoke tracking agents will not be counted or reported. Additional programming would be required to enhance the script used for this study to detect all cases of tracking agents.

Website Validation Script

A website validation script was developed to determine specific technical details that relate to the privacy and security issues discussed earlier in this report. The script is executed periodically to capture the current state of practice in these areas. The figures presented in this report represent data current as of July 2019 and will be continually updated and made available on Library Technology Guides.

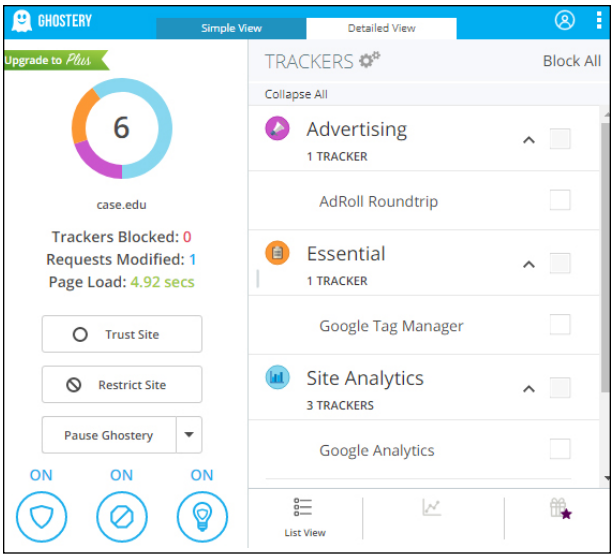


Figure 3.1. Example of Ghostery's ability to identify tracking agents

Updated figures
https://librarytechnology.org/libraries/security/report

The initial phase of the script sets the scope of the libraries to be analyzed. An SQL query is accordingly formed and run to collect the unique Record Identifiers for each directory entry in the group of interest. These interest groups include two smaller selections—members of the Association of Research Libraries and the Urban Library Council—and the two larger selections of all public libraries and all academic libraries in the United States. The script can also process individual entries. These record keys are pushed into an array used by the main control loop of the program.

Once the array has been populated, the script performs tests on each of the library records. The processing is performed in three phases.

Phase I

BASIC LINK CHECKING

Using the `LWP::UserAgent` and `HTTP::Request` Perl libraries, the script (figure 3.2) issues a request to the recorded URL held in the LibraryLink field and places the response code into a variable (`$ResponseCode`). If the page request is successful and the server also returns a redirected URL, it is recorded. This is the expected behavior if the URL has been permanently changed to a new link. The script also detects whether the redirection involves an upgrade

from an HTTP to an HTTPS link.

The detected information is then saved into the database record. If the Response Code is 200 with no redirection, the script has an option not to update the record. Any other response codes are recorded into the LinkResponseCode field and the current date is placed in LinkCheckDate. Redirected URLs are placed into LibraryWeb and the LinkResponseCode of 200.

LIMITATIONS

The basic test performed by this script for the correct deployment of HTTPS has some limitations. Though it accurately determines whether the page is transmitted with HTTPS, it does not check for important conditions that would be reported by a browser, such as whether the page has been encrypted with a valid digital certificate. It also does not check to ensure that the page does not contain any unencrypted content or links. Even though a page may be recorded as using HTTPS, it may not meet the expectations for privacy though the inclusion of mixed content, as shown in figure 3.3, where the site loads images through nonencrypted links.

Phase II

The second phase of the script (figure 3.4) assesses how each website handles redirection. If a site that has been configured to use HTTPS is accessed with a URL using the HTTP protocol, it should ideally automatically redirect to HTTPS. This redirection ensures encryption of transmission even if the user enters from an older link or types in HTTP instead of HTTPS and is classified by the script as Valid. If the site supports HTTPS but does not automatically redirect to HTTPS, it is classified as Passive. Some sites may redirect from HTTPS to HTTP, even when HTTPS is available. This behavior, possibly implemented during a testing or transition phase, is categorized by the script as Invalid. Sites that do not support HTTPS at all are classified as Unsupported. If this phase results in identifying a reliable URL not found in the first phase, it is saved into the record in the LibraryWeb field with a 200 LinkStatusCode and current LinkCheckDate.

Phase III

The final phase of the script (figure 3.5) works with the content of the page retrieved from the website. It follows a simple approach of testing for strings that can be identified as reliable signatures for specific items of interest, such as page tags for analytics or trackers for advertising networks, social networks, or e-commerce entities.

The search patterns identify selected tracking agents of interest. The text strings used to identify each tracker were initially identified through direct access to websites via the Chrome browser and the Ghostery extension. These strings are not necessarily authoritative, but are strong indicators of the tracking agent in question. Further work is needed to develop more authoritative signatures for each tracking agent. In the interim, the indicators should be considered an initial screening that needs to be reviewed manually

```
use LWP::UserAgent;
use HTTP::Request;
my $request = HTTP::Request->new(GET=>$uri);
my $ua = LWP::UserAgent->new( ssl_opts => { verify_hostname => 0 } );
$ua->agent("Mozilla/5.0 (Windows NT 10.0; Win64; x64)");
$ua->cookie_jar(HTTP::Cookies->new(file => "$tempdirectory/cookies.txt"));
$request->header('Accept' => 'text/html');
$request->accept_decodable;
my $response = $ua->request($request);
my $status_line = $response->status_line;
my $ResponseCode = $response->code;
print "Response Code: $ResponseCode\n" if ($Verbose eq "on");
my $ResponseMessage = $response->message;
print "Response Message: $ResponseMessage\n" if ($Verbose eq "on");
my $content_encoding = $response->header ('Content-Encoding');
print "Content Encoding: $content_encoding\n";
my $NewSecureURL = "off";
my $OldSecureURL = "off";
if ($response->is_success and $response->previous ) {
 $RedirectedURL = $response->request->uri;
 $LibraryWeb = $RedirectedURL;
 print "Redirected URL: $RedirectedURL\n" if ($Verbose eq "on");
 print LOG "$id: Redirected $uri to $RedirectedURL\n";
  $Redirect = "on";
 $RedirectCount++;
 $NewSecureURL = "on" if ($RedirectedURL =~ m/^https/);
 $OldSecureURL = "on" if ($uri =~ m/^https/);
 if (($OldSecureURL eq "off") && ($NewSecureURL eq "on")) {
 print "Site upgraded to Secure\n" if ($Verbose eq "on");
 $PrivacyUpgrades++;
 }
 print "Current URL: $LibraryWeb\n" if ($Verbose eq "on");
}
```

Figure 3.2
Script used for phase I

```
⚠ Mixed Content: The page at 'https://www.ackley.lib.ia.us/'  (index):313
was loaded over HTTPS, but requested an insecure image 'http://ackleypl.
follettdestiny.com/images/en/buttons/large/title.gif'. This content
should also be served over HTTPS.
⚠ Mixed Content: The page at 'https://www.ackley.lib.ia.us/'  (index):315
was loaded over HTTPS, but requested an insecure image 'http://ackleypl.
follettdestiny.com/images/en/buttons/large/author.gif'. This content
should also be served over HTTPS.
⚠ Mixed Content: The page at 'https://www.ackley.lib.ia.us/'  (index):317
was loaded over HTTPS, but requested an insecure image 'http://ackleypl.
follettdestiny.com/images/en/buttons/large/subject.gif'. This content
should also be served over HTTPS.
⚠ Mixed Content: The page at 'https://www.ackley.lib.ia.us/'  (index):319
was loaded over HTTPS, but requested an insecure image 'http://ackleypl.
follettdestiny.com/images/en/buttons/large/keyword.gif'. This content
should also be served over HTTPS.
⚠ Mixed Content: The page at 'https://www.ackley.lib.ia.us/'  (index):321
was loaded over HTTPS, but requested an insecure image 'http://ackleypl.
follettdestiny.com/images/en/buttons/large/series.gif'. This content
should also be served over HTTPS.
```

Figure 3.3. Example of mixed HTTP and HTTPS content

using Ghostery or other browser plug-ins.

One weakness of the current script is that it is based only on the HTML source of the main page of the library website. It does not check other files that may be loaded from this page, which results in an underreporting of some tracking agents. Some false positives can also take place when the string used as the signature for a given tracking agent may be used for other purposes.

```
if ($protocol eq "https") {
 my $testuri = "http:$urlstring";
 print "testing: $testuri\n" if ($Verbose eq "on");
 my $testrequest = HTTP::Request->new(GET=>$testuri);
 my $testua=LWP::UserAgent->new( ssl_opts => { verify_hostname => 0 } );
 $testua->agent("Mozilla/5.0 (Windows NT 10.0; Win64; x64)");
 $testua->cookie_jar(HTTP::Cookies->new(file =>
 "$tempdirectory/cookies.txt"));
 $testrequest->header('Accept' => 'text/html');
 $testrequest->accept_decodable;
 my $testresponse = $testua->request($testrequest);
 my $teststatus_line = $testresponse->status_line;
 my $TestRedirect = "none";
 my $TestRedirectedURL = "";
 my $TestResponseCode = $testresponse->code;
 my $TestResponseMessage = $testresponse->message;
 my $TestSecureURL = "off";
 my $TestSecureURL = "off";
 if ($testresponse->is_success and $testresponse->previous ) {
 $TestRedirectedURL = $testresponse->request->uri;
 if ($TestRedirectedURL eq $LibraryWeb) {
  ` # Passed: site uses https and automatically redirects from http
 $Redirect = "Valid";
 } else {
 #Passive: site uses https and but does not automatically redirect
 #from http\n" if ($Verbose eq "on");
 $Redirect = "Passive";
 }
} else {
 # test to see if the site has https:
 my $testuri = "https:$urlstring";
 print "testing: $testuri\n" if ($Verbose eq "on");
 my $testrequest = HTTP::Request->new(GET=>$testuri);
 my $testua = LWP::UserAgent->new(ssl_opts => {verify_hostname => 0 });
 $testua->agent("Mozilla/5.0 (Windows NT 10.0; Win64; x64)");
 $testua->cookie_jar(HTTP::Cookies->new(file =>
 "$tempdirectory/cookies.txt"));
 $testrequest->header('Accept' => 'text/html');
 $testrequest->accept_decodable;
 my $testresponse = $testua->request($testrequest);
 my $teststatus_line = $testresponse->status_line;
 my $TestRedirect = "none";
 my $TestRedirectedURL = "";
 my $TestResponseCode = $testresponse->code;
 my $TestResponseMessage = $testresponse->message;
 my $TestSecureURL = "off";
 my $TestSecureURL = "off";
 if ($testresponse->is_success and $testresponse->previous ) {
 $TestRedirectedURL = $testresponse->request->uri;
 }
 if ($TestResponseMessage eq "OK") {
 if($TestRedirectedURL eq "") {
 # failed: https is enabled but not automatically redirected
         $LibraryWeb = $testuri;
 } else {
 if ($TestRedirectedURL =~/http:/) {
 # failed: https redirects to http
 $Redirect = "Invalid";
 }

 }
 } else {
 $Redirect = "Unsupported";
 }
}
```

Figure 3.4. Script for phase II

Findings: The Current State of Practice

The study demonstrates that the library community has made rapid progress in the implementation of technologies on their websites needed to provide a reasonable degree of privacy for patron information-seeking activities. In the period from April 2018 through July 2019, there has been a dramatic improvement from less than 10 percent of academic library websites using HTTPS to 92.1 percent. Public libraries have also seen dramatic improvement, though their current implementation stands at 81.7 percent. Tables 3.1 and 3.2 show the changes in percentages for these libraries since April 2018.

Summaries by Category

Another set of reports and graphs shows additional details regarding the relevant technical characteristics across each of the interest groups (figures 3.6–3.12 and tables 3.4–3.12). A basic pie chart (figure 3.6) shows the proportions of libraries still using unencrypted HTTP transmission for their main websites. Although the percentages are dramatically better than those from the beginning of the study, it also shows

Table 3.1. Implementation of HTTPS by academic libraries in the United States

Date	Total	HTTP count	HTTP percent	HTTPS count	HTTPS percent
Apr 2018	3,960	3,569	90.1	391	9.9
Dec 2018	3,967	2,244	56.6	1,723	43.4
Mar 2019	3,954	1,370	34.6	2,584	65.4
Jul 2019	3,937	310	7.9	3,612	92.1

Table 3.2. Implementation of HTTPS by public libraries in the United States

Date	Total	HTTP count	HTTP percent	HTTPS count	HTTPS percent
Apr 2018	17,286	14,539	89.6	1,688	10.4
Dec 2018	19,728	11,717	72.1	4,539	27.9
Mar 2019	16,921	7,852	51.8	8,439	51.8
Jul 2019	16,284	2,818	18.3	12,546	81.7

```
# Get the page content and test for analytics and tracker agents
my $request = HTTP::Request->new(GET=>$LibraryWeb);
my $ua = LWP::UserAgent->new( ssl_opts => { verify_hostname => 0 } );
$ua->agent("Mozilla/5.0 (Windows NT 10.0; Win64; x64)");
$ua->cookie_jar(HTTP::Cookies->new(file => "$tempdirectory/cookies.txt"));
$request->header('Accept' => 'text/html');
$request->accept_decodable;
my $response = $ua->request($request);
my $status_line = $response->status_line;
my $ResponseCode = $response->code;
my $ResponseMessage = $response->message;
my $content_encoding = $response->header ('Content-Encoding');
my $ResponsePage = $response->decoded_content;
my $PageLength = length($ResponsePage);
my $PageRetrievalStatus = "PageRetrievalFailed";
$PageRetrievalStatus = "PageRetrievalSuccess" if ($PageLength > 0);
my $GoogleAnalytics = "NotDetected";
my $GoogleTagManager = "NotDetected";
my $GoogleDoubleClick = "NotDetected";
my $GoogleAnalyticsAnonomize = "NotDetected";
my $GoogleAnalyticsSecure = "NotDetected";
my $GoogleCustomSearch = "NotDetected";
my $NewRelic = "NotDetected";
my $CrazyEgg = "NotDetected";
my $FacebookCustomAudience = "NotDetected";
my $FacebookConnect = "NotDetected";
my $AddToAny = "NotDetected";
my $ShareThis = "NotDetected";
my $Inspectlet = "NotDetected";
my $TwitterAds = "NotDetected";

$GoogleAnalytics = "Classic" if ($ResponsePage =~ /ga.js/);
$GoogleAnalytics = "Universal" if (($ResponsePage =~ /analytics.js/) ||
 ($ResponsePage =~ /gtag.js/));
if ($GoogleAnalytics ne "NotDetected") {
 if ($ResponsePage =~ /anonymizeIp/) {
 $GoogleAnalyticsAnonomize = "Anonomized";
 } else {
 $GoogleAnalyticsAnonomize = "Non-Anonomized";
 }
 $GoogleAnalyticsSecure = "http allowed" if
 ($ResponsePage =~ /https:\/\/ssl/);
  } else {
  $GoogleAnalyticsAnonomize = "Undetermined";
 }
$GoogleDoubleClick = "Enabled" if ($ResponsePage =~ /doubleclick/);
$GoogleTagManager = "Enabled" if (($ResponsePage =~ /gtm.js/)||
 ($ResponsePage =~ /googletagmanager.com/));
$GoogleCustomSearch = "Enabled" if ($ResponsePage =~ /cse.google.com/);
$NewRelic = "Enabled" if ($ResponsePage =~ /newrelic/);
$CrazyEgg = "Enabled" if ($ResponsePage =~ /crazyegg/);
$FacebookCustomAudience = "Enabled" if ($ResponsePage =~ /FB.init/);
$FacebookCustomAudience = "Enabled" if ($ResponsePage =~ /fbevents.js/);
$FacebookConnect = "Enabled" if ($ResponsePage =~ /connect.facebook.net/);
$AddToAny = "Enabled" if ($ResponsePage =~ /addtoany/);
$ShareThis = "Enabled" if ($ResponsePage =~ /sharethis/);
$Inspectlet = "Enabled" if ($ResponsePage =~ /inspectlet/);
$TwitterAds = "Enabled" if ($ResponsePage =~ /twitter.com\/oct.js/);
 my $SecurityPrivacyValues = "";
# CheckDate: The Date of the last inspection
# Protocol: http or https
# Redirect: valid: http automatically redirects to https
# passive: https available but not redirected
# invalid: https redirects to http
# unsupprted: https not available
# PageRetrievalStatus: PageRetrievalSuccess or PageRetrievalFailed
# GoogleAnalytics: Classic or Universal
# GoogleAnalyticsAnonomize: Anonomized or Non-Anonomized
# GoogleCustomSearch : Enabled or NotDetected
# GoogleTagManager : Enabled or NotDetected
# GoogleDoubleClick : Enabled or NotDetected
# NewRelic : Enabled or NotDetected
# CrazyEgg : Enabled or NotDetected
# FacebookCustomAudience : Enabled or NotDetected
# FacebookConnect : Enabled or NotDetected
# AddToAny : Enabled or NotDetected
# Inspectlet : Enabled or NotDetected
# TwitterAds : Enabled or NotDetected

$SecurityPrivacyValues =
"$CheckDate|$Protocol|Redirect=$Redirect|$PageRetrievalStatus|GA=$GoogleAna
lytics|GAA=$GoogleAnalyticsAnonomize|GCS=$GoogleCustomSearch|GTM=$GoogleTag
Manager|DoubleClick=$GoogleDoubleClick|NewRelic=$NewRelic|CrazyEgg=$CrazyEg
g|FBCA=$FacebookCustomAudience|FBC=$FacebookConnect|ATA=$AddToAny|ST=$Share
This|Inspectlet=$Inspectlet|TwitterAds=$TwitterAds";
$SqlStatement = "UPDATE $database SET SecurityPrivacy =
\'$SecurityPrivacyValues\' WHERE RecordNumber = \'$id\'";
&executeSQL("$SqlStatement") if ($UpdateDB eq "on");
```

Figure 3.5. Script for phase III

that there are substantial numbers of libraries that are not offering basic privacy protection, long past the date in which browsers began flagging these sites as unsecure. It will be important to continue monitoring these figures to see if these remaining libraries are able make these needed improvements.

Table 3.3 describes the numbers and percentage of libraries that have implemented redirection in ways needed to ensure private communications. Although over 90 percent of academic libraries now support HTTPS, only 63 percent require it for all sessions. Almost 30 percent of these libraries do not implement redirection on their websites, so users are able to access the site with unsecured HTTP. A small number of sites redirect from HTTPS to HTTP, presumably as an interim state as encrypted configurations are implemented.

The findings regarding the proportion of libraries using some sort of tracking agent on their websites elicits more concern regarding protections in place for privacy. The implementation of Google Analytics on library websites is almost ubiquitous. A relatively small proportion use the outdated Classic tracking code, which was superseded by Universal analytics in 2012. The total number of sites using Google Analytics cannot be determined automatically from the testing script. As noted earlier, when Google Analytics has been deployed using Google Tag Manager, it is not apparent other than to the site owner what tags have been deployed. It is highly likely that those using Google Tag Manager are also using Google Analytics. We can carry this inference into our observations. Based on these assumptions, at least 3,219 out of 3,948 academic libraries, or 81 percent, use Google Analytics. Among public libraries, 10,568 out of 15,865, or 67 percent, have implemented Google Analytics. The numbers of libraries using Google Analytics that have implemented anonymization of IP addresses appears quite low, with only 335 academics and 1,386 public libraries taking advantage of this feature.

The screening for tracking agents related to advertising and social networks reveals substantial numbers of libraries enabling these connections. The most commonly implemented of this type of tracking agent is for Facebook Connect, detected in the websites of 666 academic and 2,102 public libraries. Facebook Custom Audiences, a more intrusive tracking agent, was detected in 486 academic library websites and in 690 public library sites.

A small percentage of library websites include tracking tags for advertising networks,

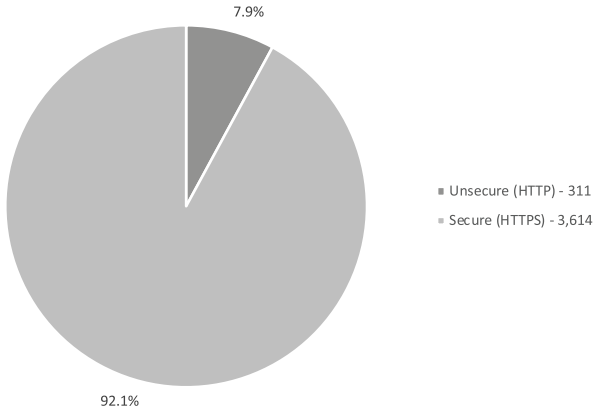


Figure 3.6. Percentage of academic library websites in the United States using HTTPS

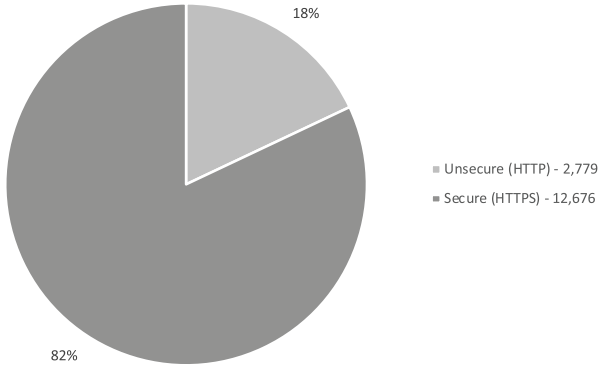


Figure 3.7. Public libraries in the United States: the percentage of 15,455 library sites using HTTPS

such as Google DoubleClick. This study includes only preliminary investigation regarding the involvement of libraries in the commercial advertising networks. Searching automatically for the signatures for the tracking tags used so far has not been reliable, with both false positives and false negatives when verified through Ghostery.

In addition to the broad groupings of public and academic libraries, this study also selected two smaller groups. The members of the Association of Research Libraries represent the top tier of academic libraries.

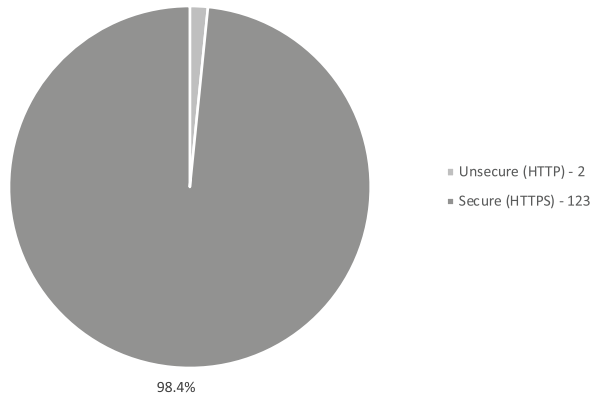


Figure 3.8. Percentage of the 125 Association of Research Libraries websites using HTTPS

The Urban Library Council is comprised of public libraries serving larger urban populations. Both of these groups are more likely to have the financial resources and the technical awareness to implement the strongest measures for patron privacy and security.

The two elite groups of libraries show much higher implementation of technologies to protect privacy than the broader populations. All but 2 ARL members

Table 3.4. Number of academic library websites in the United States and third-party tracking

Status	Count
Successful page retrieval	3,948
Failed page retrieval	2
Google Analytics Classic enabled	352
Google Analytics Universal enabled	1,630
Google Analytics Tag Manager enabled	520
Google Analytics not detected	1,448
Google Analytics total	3,219
Google Analytics anonymized	335
Google Analytics not anonymized	2,167
Google Tag Manager enabled	1,766
DoubleClick enabled	247
Facebook Custom Audiences enabled	486
Facebook Connect enabled	666
Inspectlet enabled	5

Table 3.3. Number and percent of academic libraries' websites in the United States that support HTTPS

Status	Count	Percent
Valid (supports HTTPS)	2,512	63.95
Passive (supports HTTPS, but doesn't automatically redirect to HTTPS)	1,109	28.23
Invalid (may redirect from HTTPS to HTTP, even when HTTPS is available)	40	1.02
Unsupported (does not support HTTPS)	267	6.80
Total	3,928	100.00

Table 3.5. Number and percent of public libraries' websites in the United States that support HTTPS

Status	Count	Percent
Valid (supports HTTPS)	8,460	52.58
Passive (supports HTTPS, but doesn't automatically redirect to HTTPS)	4,324	26.87
Invalid (may redirect from HTTPS to HTTP, even when HTTPS is available)	335	2.08
Unsupported (does not support HTTPS)	2,972	18.47
Total	16,091	100.00

Table 3.6. Number of public libraries' websites in the United States and third-party tracking

Status	Count
Successful page retrieval	16,270
Failed page retrieval	22
Google Analytics Classic enabled	2,039
Google Analytics Universal enabled	7,739
Google Analytics anonymized	1,305
Google Analytics not anonymized	8,473
Google Tag Manager enabled	3,053
DoubleClick enabled	742
Facebook Custom Audiences enabled	690
Facebook Connect enabled	2,070
Inspectlet enabled	3

Table 3.7. HTTPS Status of ARL members' websites

Status	Count	Percent
Valid (supports HTTPS)	101	80.80
Passive (supports HTTPS, but doesn't automatically redirect to HTTPS)	22	17.60
Invalid (may redirect from HTTPS to HTTP, even when HTTPS is available)	1	0.80
Unsupported (does not support HTTPS)	1	0.80
Total	125	100.00

Table 3.8. ARL members and third-party tracking

Status	Count
Successful page retrieval	124
Failed page retrieval	1
Google Analytics Classic enabled	11
Google Analytics Universal enabled	82
Google Analytics anonymized	25
Google Analytics not anonymized	68
Google Tag Manager enabled	42
DoubleClick enabled	1
Facebook Custom Audiences enabled	1
Facebook Connect enabled	1
Inspectlet enabled	0

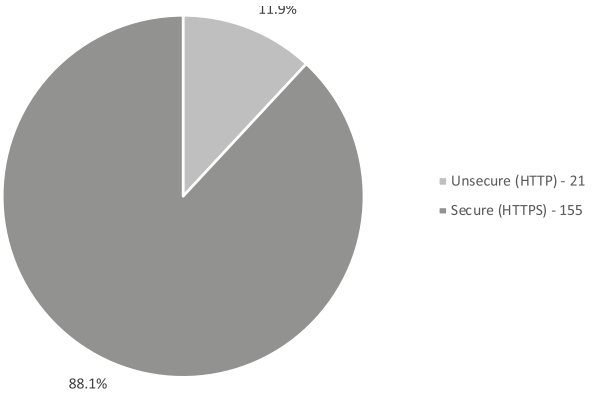


Figure 3.9. Number and percentage of Urban Libraries Council's websites using HTTPS (includes current and some former members)

implement HTTPS, though 21 out of the 178 in the ULC group, or 11.8 percent, continue to not provide HTTPS encryption. Tables 3.9 and 3.11 provide the details of each of these groups of libraries.

The data collected for each library in the study group is also presented through individual Privacy and Security Report Cards, an example of which is seen in figure 3.10. These report cards aim to provide a quick overview of how well each library has implemented technologies to protect patron privacy. Implementation of encryption and correct redirection are given green checkmark icons (shown in dark gray in figure 3.10); if the library still uses HTTP, a red X icon appears. Yellow checkmarks are provided when any of the tracking codes are detected (shown in light gray in figure 3.10). A red X is presented if Google Analytics has been implemented without the anonymization option. These report cards can be access through each library's entry in Libraries.org.

Table 3.9. This table, running multiple pages in its full form, shows findings from each ARL library's website, including whether it follows HTTPS protocol, the status of its redirect from HTTP to HTTPS, and use of third-party tracking systems, including Google Analytics, GA Anonym, Google Tag Manager, Google Custom Search, DoubleClick, and Facebook Connect. The full data set can be downloaded from the Library Technology Guides website.

ARL Members								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Arizona State University	HTTPS	Valid	Universal	X	✓	—	—	—
Auburn University	HTTPS	Valid	Universal	X	—	—	—	—
Universite Laval	HTTPS	Valid	Classic	X	—	—	—	—
Boston, MA	HTTPS	Valid	Universal	X	✓	—	—	—
Brown University	HTTPS	Valid	Classic	X	—	—	—	—
Center for Research Libraries	HTTPS	Passive	Universal	X	—	—	—	—
Columbia University	HTTPS	Valid	?	?	—	—	—	—
Cornell University	HTTPS	Valid	?	?	—	—	—	—
North Carolina State University	HTTPS	Valid	Universal	X	—	—	—	—
Dartmouth College	HTTPS	Valid	?	?	✓	—	—	—
University of Southern California	HTTPS	Valid	Universal	✓	✓	—	—	—
University of Nebraska—Lincoln	HTTPS	Valid	Universal	X	—	—	—	—
Duke University	HTTPS	Valid	Universal	X	—	—	—	—
Oklahoma State University	HTTPS	Valid	Universal	X	✓	—	—	—
New York University	HTTPS	Passive	Universal	X	✓	—	—	—
Emory University	HTTPS	Passive	Universal	X	—	—	—	—
Rice University	HTTPS	Passive	Universal	X	—	—	—	—
University of Florida	HTTPS	Valid	Classic	X	—	—	—	—
Georgia Institute of Technology	HTTPS	Valid	Universal	✓	—	—	—	—
Brigham Young University	HTTPS	Valid	?	?	—	—	—	—
Harvard University	HTTPS	Valid	?	?	✓	—	—	—
University of Notre Dame	HTTPS	Valid	?	?	✓	—	—	—
University of Connecticut	HTTPS	Valid	Universal	X	✓	—	—	—
Howard University	HTTPS	Passive	?	?	—	—	—	—
Tulane University	HTTPS	Valid	Universal	✓	—	—	—	—
Indiana University	HTTPS	Valid	Universal	✓	—	—	—	—
Iowa State University	HTTPS	Valid	Universal	✓	—	—	—	—
Case Western Reserve University	HTTPS	Valid	?	?	✓	—	—	—
Kent State University	HTTPS	Valid	Universal	X	—	—	—	—
University of Cincinnati	HTTPS	Valid	?	?	✓	—	—	—
Georgetown University	HTTPS	Valid	Universal	✓	✓	—	—	—
United States—Library of Congress	HTTPS	Valid	?	?	—	—	—	—
Louisiana State University	HTTPS	Valid	Universal	X	✓	—	—	—
University of Utah	HTTPS	Passive	Universal	X	—	—	—	—
McGill University	HTTPS	Valid	Universal	✓	—	—	—	—
University of Guelph	HTTPS	Valid	Universal	X	—	—	—	—
McMaster University	HTTPS	Valid	Universal	✓	—	—	—	—
George Washington University	HTTPS	Valid	Universal	X	—	—	—	—
Michigan State University	HTTPS	Valid	?	?	✓	—	—	—
Johns Hopkins University	HTTPS	Valid	?	?	✓	—	—	—
Massachusetts Institute of Technology	HTTPS	Valid	Universal	X	—	—	—	—

ARL Members (continued)								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Colorado State University	HTTPS	Valid	Universal	X	—	—	—	—
Southern Illinois University	HTTPS	Valid	?	?	✓	—	—	—
Boston University	HTTPS	Passive	Classic	X	—	—	—	—
United States—National Agricultural Library	HTTPS	Valid	Universal	X	—	—	—	—
National Archives and Records Administration	HTTPS	Valid	?	?	✓	—	—	—
United States—National Library of Medicine	HTTPS	Valid	?	?	✓	—	—	—
New York, NY	HTTPS	Valid	Universal	✓	—	—	—	—
New York	HTTP	Unsup-ported	Universal	X	—	—	—	—
Northwestern University	HTTPS	Valid	Universal	X	—	—	—	—
Ohio State University	HTTPS	Valid	Universal	✓	—	—	—	—
Ohio University	HTTPS	Valid	Universal	X	✓	—	—	—
University of Miami	HTTPS	Valid	Universal	X	✓	—	—	—
University of Pennsylvania	HTTPS	Passive	?	?	✓	—	—	—
Pennsylvania State University	HTTPS	Valid	Universal	✓	—	—	—	—
Princeton University	HTTPS	Passive	Universal	✓	—	—	—	—
Purdue University	HTTPS	Valid	Classic	X	—	—	—	—
Queen's University	HTTPS	Valid	Universal	✓	—	—	—	—
University of Rochester	HTTPS	Valid	Universal	✓	—	—	—	—
Rutgers University	HTTPS	Valid	Universal	✓	—	—	—	—
Simon Fraser University	HTTPS	Valid	Universal	✓	—	—	—	—
Smithsonian Institution	HTTPS	Valid	Universal	X	—	—	✓	—
Texas A&M University	HTTPS	Passive	?	?	✓	—	—	—
Florida State University	HTTPS	Valid	Universal	✓	—	—	—	✓
Syracuse University	HTTPS	Valid	Classic	X	—	—	—	—
Temple University	HTTPS	Valid	Universal	X	—	—	—	—
Texas Tech University	HTTPS	Passive	Universal	X	—	—	—	—
Boston College	HTTPS	Valid	Universal	X	—	—	—	—
University of California—Davis	HTTPS	Valid	Universal	X	—	—	—	—
University of California—Riverside	HTTPS	Valid	Universal	✓	—	—	—	—
University of California—San Diego	HTTPS	Valid	Universal	✓	—	—	—	—
University of California—Santa Barbara	HTTPS	Valid	Universal	✓	—	—	—	—
University of California—Irvine	HTTPS	Valid	Universal	X	—	—	—	—
University of California—Los Angeles (UCLA)	HTTPS	Valid	Classic	X	✓	—	—	—
University of Massachusetts—Amherst	HTTPS	Valid	Universal	X	✓	—	—	—
University of North Carolina—Chapel Hill	HTTPS	Valid	?	?	✓	—	—	—
University at Albany	HTTPS	Passive	?	?	—	—	—	—
University at Buffalo	HTTPS	Valid	?	?	✓	—	—	—
Stony Brook University	HTTPS	Valid	Universal	X	✓	—	—	—
University of Alabama	HTTPS	Valid	Universal	X	✓	—	—	—
University of Alberta	HTTPS	Valid	Universal	X	—	—	—	—
University of Arizona	HTTPS	Valid	Universal	✓	—	—	—	—

ARL Members (continued)								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
University of British Columbia	HTTPS	Passive	Universal	X	—	✓	—	—
University of Calgary	HTTPS	Valid	Universal	X	✓	—	—	—
University of California—Berkeley	HTTP	Invalid	?	?	✓	—	—	—
University of Chicago	HTTPS	Valid	?	?	—	—	—	—
University of Colorado—Boulder	HTTPS	Valid	Universal	✓	✓	—	—	—
University of Delaware	HTTPS	Valid	Universal	X	—	—	—	—
University of Georgia	HTTPS	Valid	Universal	✓	—	—	—	—
University of Hawaii—Manoa	HTTPS	Passive	Universal	X	—	—	—	—
University of Houston	HTTPS	Valid	Universal	X	—	—	—	—
University of Illinois—Chicago	HTTPS	Valid	?	?	✓	—	—	—
University of Illinois—Urbana-Champaign	HTTPS	Valid	Universal	X	—	—	—	—
University of Iowa	HTTPS	Passive	Classic	X	—	—	—	—
University of Kansas	HTTPS	Valid	Universal	X	—	—	—	—
University of Kentucky	HTTPS	Passive	Classic	X	✓	✓	—	—
University of Louisville	HTTPS	Passive	Universal	X	—	—	—	—
University of Manitoba	HTTPS	Passive	Classic	X	—	—	—	—
University of Michigan	HTTPS	Valid	Universal	✓	—	—	—	—
University of Minnesota—Twin Cities	HTTPS	Valid	Universal	✓	—	—	—	—
University of Missouri—Columbia	HTTPS	Passive	Universal	X	—	—	—	—
University of Oklahoma	HTTPS	Valid	?	?	✓	—	—	—
University of Oregon	HTTPS	Valid	Universal	X	—	—	—	—
University of Ottawa	HTTPS	Valid	Universal	X	—	—	—	—
University of Pittsburgh	HTTPS	Valid	Universal	X	—	—	—	—
University of Saskatchewan	HTTPS	Valid	Universal	X	—	✓	—	—
University of South Carolina	HTTPS	Valid	Classic	X	✓	—	—	—
University of Tennessee—Knoxville	HTTPS	Valid	Universal	X	—	—	—	—
University of Texas—Austin	HTTPS	Valid	?	?	✓	—	—	—
University of Toronto	HTTPS	Valid	?	?	✓	—	—	—
University of Virginia	HTTPS	Valid	?	?	—	—	—	—
University of Washington	HTTPS	Passive	?	?	—	—	—	—
University of Waterloo	HTTPS	Passive	Universal	X	✓	—	—	—
University of Western Ontario	HTTPS	Valid	Universal	X	✓	—	—	—
University of Wisconsin—Madison	HTTPS	Valid	Universal	X	—	—	—	—
University of Maryland	HTTPS	Valid	Universal	X	—	—	—	—
University of New Mexico	HTTPS	Valid	?	?	—	✓	—	—
Vanderbilt University	HTTPS	Valid	?	?	✓	—	—	—
Virginia Commonwealth University	HTTPS	Valid	Universal	X	—	—	—	—
Virginia Tech	HTTPS	Valid	?	?	✓	—	—	—
Washington State University	HTTPS	Passive	Universal	X	—	—	—	—
Washington University in Saint Louis	HTTPS	Valid	Universal	X	—	—	—	—
Wayne State University	HTTPS	Valid	Universal	X	—	—	—	—
Yale University	HTTPS	Valid	?	?	✓	—	—	—
York University	HTTPS	Valid	Universal	X	—	—	—	—

Table 3.10. HTTPS Status of Urban Libraries Council's websites (includes current and some former members)

Status	Count	Percent
Valid (supports HTTPS)	123	69.10
Passive (supports HTTPS, but doesn't automatically redirect to HTTPS)	34	19.10
Invalid (may redirect from HTTPS to HTTP, even when HTTPS is available)	4	2.25
Unsupported (does not support HTTPS)	17	9.55
Total	178	100.00

Table 3.11. Urban Libraries Council's websites and third-party tracking (includes current and some formal members)

Status	Count
Successful page retrieval	178
Failed page retrieval	0
Google Analytics Classic enabled	23
Google Analytics Universal enabled	127
Google Analytics anonymized	27
Google Analytics not anonymized	123
Google Tag Manager enabled	73
DoubleClick enabled	6
Facebook Custom Audiences enabled	21
Facebook Connect enabled	29
Inspectlet enabled	0

Category	Value	Explanation
Site		**Website link: https://library.nashville.org/ Nashville and Davidson County, TN; Nashville Public Library**
Protocol [https]	✓	This site uses the https protocol which ensures that the information is encrypted between the web browser and the server transmitting the page. Encryption provides a private connection in which the content cannot be viewed by any third party able to capture network traffic.
Redirection	✓	This site always uses encryption. If a link refers to an non-encrypted version of a page, it will automatically be redirected to safely encrypted version.
Google Analytics	✓	This site uses Google Analytics, service offered by Google for recording and analyzing use. This service enables Google to know each page a user might access from this site. This organization has implemented Google Analytics using the Universal Analytics method.
Google Analytics Anonymized	✓	Google Analytics has been implemented and uses the correct configuration to instruct Google to anonymize data from this site.
Google Tag Manager	✓	Google Tag Manager has been enabled on this site. This infers the use of Google Analytics as well as other applications that may track users.
Google Custom Search	?	Google Custom Search was not detected on this site.
Google DoubleClick	?	Google DoubleClick was not detected on this site.
Facebook Custom Audience	?	Facebook Custom Audiences was not detected on this site.
Facebook Connect	?	Facebook Connect was not detected on this site.
Inspectlet	?	Inspectlet was not detected on this site.
AddToAny	?	Add to Any was not detected on this site.
ShareThis	?	ShareThis was not detected on this site.
NewRelic	✓	The New Relic performance monitoring service has been enabled on this site.
Crazy Egg	✓	The CrazyEgg website optimization service has been enabled on this site.
Details:		This page was last checked on 2019-07-03.

Figure 3.10. Sample Privacy and Security Report Card: Nashville Public Library

Table 3.12. This table, running multiple pages in its full form, shows findings from each ULC library's website, including whether it follows HTTPS protocol, the status of its redirect from HTTP to HTTPS, and use of third-party tracking systems, including Google Analytics, GA Anonym, Google Tag Manager, Google Custom Search, Double Click, and Facebook Connect. The full data set can be downloaded from the Library Technology Guides website.

ULC Members								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Akron-Summit County Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Alameda County Library	HTTPS	Valid	Universal	X	✓	—	—	—
Albany Public Library	HTTPS	Valid	Universal	X	✓	—	—	✓
Albuquerque Bernalillo County Library System	HTTPS	Valid	Universal	X	—	—	—	—
Alexandria Library	HTTPS	Passive	Universal	X	—	—	—	—
Allen County Public Library	HTTP	Unsupported	Universal	X	—	—	—	—
Anchorage Public Library	HTTP	Unsupported	Universal	X	✓	—	—	—
Anne Arundel County Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Anythink Wright Farms	HTTPS	Valid	Universal	✓	—	—	—	—
Arapahoe Library District	HTTPS	Valid	Universal	X	✓	—	—	✓
Arlington County Public Library	HTTPS	Valid	?	?	—	—	—	—
Arlington Heights Memorial Library	HTTP	Invalid	Classic	X	—	—	—	—
Atlanta-Fulton Public Library	HTTP	Unsupported	Classic	X	—	—	—	—
Aurora Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Baltimore County Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Birmingham Public Library	HTTP	Unsupported	Classic	X	—	—	—	—
Boston Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Bridgeport Public Library	HTTPS	Valid	Classic	X	—	—	—	✓
Brooklyn Public Library	HTTPS	Valid	Universal			—	—	—
Broward County Library	HTTPS	Passive	?	?	—	—	—	—
Buffalo and Erie County Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Calgary Public Library	HTTPS	Valid	Universal	X	✓	—	—	✓
Camden County Library System	HTTPS	Valid	Universal	✓	—	—	—	—
Carlsbad City Library	HTTPS	Passive	Universal	X	—	—	—	—
Carmel Clay Public Library	HTTPS	Valid	Classic	X	—	—	—	—
Carnegie Library of Pittsburgh	HTTPS	Valid	?	?	✓	—	—	—
Carroll County Public Library	HTTPS	Valid	Classic	X	—	—	—	—
Central Library of Rochester and Monroe County	HTTPS	Valid	Universal	X	✓	—	—	—
Cesar Chavez Central Library	HTTP	Unsupported	Classic	X	—	—	—	—
Charlotte Mecklenburg Library	HTTPS	Valid	?	?	✓	—	—	—
Chattahoochee Valley Libraries	HTTPS	Valid	Universal	X	—	—	—	—
Chattanooga Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Chesterfield County Public Library	HTTPS	Valid	Universal	X	✓	—	—	—

ULC Members (continued)								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Chicago Public Library	HTTPS	Valid	Universal	X	✓	—	—	✓
Cleveland Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Cobb County Public Library	HTTPS	Passive	?	?	—	—	—	—
Columbus Metropolitan Library	HTTPS	Valid	?	?	✓	—	—	—
Contra Costa County Public Library	HTTPS	Passive	Classic	X	—	—	—	—
County of Los Angeles Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Cuyahoga County Public Library	HTTPS	Valid	?	?	✓	—	—	—
Dallas Public Library	HTTPS	Passive	Classic	X	—	—	—	—
Davenport Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Dayton Metro Library	HTTPS	Passive	Universal	X	—	—	—	—
DeKalb County Public Library	HTTPS	Passive	Universal	X	—	—	—	—
Denver Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Des Moines Public Library	HTTPS	Passive	Universal	X	—	—	—	✓
Detroit Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
District of Columbia Public Library	HTTPS	Valid	Universal	X	—	—	—	✓
Durham County Library	HTTPS	Valid	?	?	✓	—	—	—
East Baton Rouge Parish Library	HTTPS	Valid	Universal	X	✓	—	—	—
East Cleveland Public Library	HTTPS	Passive	Universal	✓	—	—	—	—
Eastern Oklahoma District Library System	HTTPS	Passive	?	?	—	—	—	—
Edmonton Public Library	HTTPS	Valid	Universal	X	✓	—	—	✓
El Paso Public Library	HTTPS	Passive	Universal	X	—	—	—	—
Elizabeth Free Public Library	HTTP	Unsup-ported	?	?	—	—	—	—
Enoch Pratt Free Library	HTTPS	Valid	Universal	X	—	—	—	—
Evansville Vanderburgh Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Forsyth County Public Library	HTTPS	Passive	Classic	X	—	—	—	—
Fort Vancouver Regional Library	HTTPS	Passive	Universal	✓	—	—	—	—
Fort Worth Public Library	HTTP	Unsup-ported	Universal	X	✓	—	—	—
Free Library of Philadelphia	HTTPS	Passive	Universal	X	—	—	✓	—
Fresno County Public Library	HTTP	Unsup-ported	Classic	X	—	—	—	✓
Frisco Public Library	HTTPS	Passive	Universal	✓	—	—	—	—
Gary Public Library	HTTPS	Passive	?	?	—	—	—	—
Grand Rapids Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Greensboro Library System	HTTPS	Valid	Universal	X	—	—	—	—
Gwinnett County Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Hamilton Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Harris County Public Library	HTTP	Unsup-ported	Universal	X	—	—	✓	—
Hartford Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Hayward Public Library	HTTPS	Valid	Universal	✓	—	—	—	—

ULC Members (continued)								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Hennepin County Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Houston Public Library	HTTPS	Passive	Universal	✓	—	—	—	—
Howard County Library System	HTTPS	Valid	Universal	X	—	—	—	—
Indianapolis Public Library System	HTTPS	Valid	?	?	✓	—	—	—
Jacksonville Public Library	HTTPS	Valid	Universal	✓	✓	—	—	✓
Jefferson County Public Library System	HTTPS	Valid	Universal	X	✓	—	—	✓
Joel Valdez Main Library	HTTPS	Valid	Universal	X	✓	—	—	✓
Johnson County Library	HTTPS	Valid	Universal	X	—	—	—	—
Kalamazoo Public Library	HTTPS	Valid	?	?	—	—	—	—
Kansas City Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Kent District Library	HTTPS	Valid	Universal	✓	✓	—	—	✓
Kern County Library	HTTPS	Passive	?	?	—	—	—	—
King County Library System	HTTPS	Valid	Universal	X	✓	—	✓	✓
Las Vegas-Clark County Library District	HTTPS	Valid	Universal	X	✓	—	—	—
Lee County Library System	HTTPS	Passive	Universal	X	✓	—	—	—
LeRoy Collins Leon County Public Library	HTTPS	Passive	Universal	X	—	—	—	—
Lexington Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Lincoln City Libraries	HTTPS	Valid	Universal	X	—	—	—	—
Live Oak Public Libraries	HTTPS	Passive	Universal	X	✓	—	—	—
Long Beach Public Library	HTTP	Unsup-ported	Universal	X	—	—	—	—
Los Angeles Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Loudoun County Public Library	HTTPS	Valid	Classic	X	—	—	—	—
Louisville Free Public Library—Main	HTTP	Unsup-ported	?	?	—	—	—	—
Madison Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Marin County Free Library	HTTPS	Valid	Universal	X	✓	—	—	—
Memphis Public Library and Information Center	HTTPS	Passive	Universal	X	—	—	—	—
Mesa Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Miami-Dade Public Library System	HTTPS	Valid	Universal	X	—	—	—	—
Mid-Continent Consolidated Library District	HTTPS	Valid	?	?	✓	—	—	—
Milwaukee Public Library—Central Library	HTTPS	Passive	Universal	X	—	—	✓	✓
Montgomery County Public Libraries	HTTPS	Valid	?	?	—	—	—	—
Multnomah County Library	HTTPS	Valid	Universal	X	—	—	—	—
Nashville Public Library	HTTPS	Valid	Universal	✓	✓	—	—	—
New Haven Free Public Library	HTTPS	Passive	Classic	X	—	—	—	—
New Orleans Public Library	HTTPS	Passive	Universal	X	—	—	—	—
New York Public Library	HTTPS	Valid	?	?	—	—	—	—
Newark Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Newport Beach Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Newport News Public Library	HTTPS	Passive	Universal	X	✓	—	—	—
Oakland Public Library	HTTPS	Passive	Classic	X	✓	—	—	—

ULC Members (continued)								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Ocean County Library	HTTP	Unsupported	Universal	✓	—	—	—	—
Oklahoma City Metropolitan Library System	HTTPS	Valid	Universal	X	✓	—	—	—
Omaha Public Library	HTTPS	Valid	Universal	X	✓	—	—	✓
Orange County Library System	HTTPS	Valid	Universal	X	—	—	✓	✓
Ottawa Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Palm Beach County Library System	HTTP	Unsupported	Universal	X	—	—	—	—
Palo Alto City Library	HTTPS	Valid	Universal	X	✓	—	—	—
Pasadena Public Library	HTTPS	Valid	?	?	✓	—	—	—
Phoenix Public Library	HTTPS	Passive	Universal	X	—	—	—	—
Pierce County Library System	HTTPS	Valid	Classic	X	—	—	—	—
Pikes Peak Library District	HTTPS	Valid	Universal	✓	—	—	—	✓
Pioneer Library System	HTTPS	Valid	Universal	X	✓	—	—	—
Portland Public Library	HTTPS	Valid	Classic	X	—	—	—	—
Poudre River Public Library District	HTTPS	Valid	Universal	X	—	—	—	—
Prince George's County Memorial Library System	HTTPS	Valid	Universal	X	—	—	—	—
Providence Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Public Libraries of Saginaw	HTTPS	Valid	Universal	X	✓	—	—	—
Public Library of Cincinnati and Hamilton County	HTTPS	Valid	Classic	X	—	—	—	—
Public Library of Youngstown and Mahoning County	HTTPS	Valid	Universal	X	—	—	—	—
Pueblo City-County Library District	HTTP	Unsupported	Universal	X	—	—	—	✓
Queens Borough Public Library	HTTPS	Passive	Universal	✓	✓	—	—	—
Redwood City Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Regina Central Library	HTTPS	Valid	Universal	✓	—	—	—	—
Richland Library	HTTPS	Valid	Universal	✓	—	—	—	—
Richmond Public Library	HTTPS	Valid	?	?	✓	—	—	—
Rochester Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Sacramento Public Library	HTTPS	Valid	Classic	X	✓	—	—	—
Saint Joseph County Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Saint Louis County Library	HTTPS	Valid	Universal	✓	—	—	—	—
Saint Louis Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Saint Paul Public Library	HTTPS	Passive	Universal	X	✓	—	—	✓
Salt Lake City Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Salt Lake County Library System	HTTPS	Valid	Classic	X	—	—	—	—
San Antonio Public Library	HTTPS	Valid	Classic	X	—	—	—	✓
San Diego County Library	HTTP	Invalid	?	?	—	—	—	—
San Diego Public Library	HTTPS	Valid	?	?	—	—	—	✓
San Francisco Public Library	HTTPS	Valid	Universal	X	—	—	—	—
San Jose Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
San Luis Obispo City-County Library	HTTPS	Valid	Universal	X	—	—	—	—
San Mateo County Library	HTTPS	Valid	Universal	X	✓	—	—	✓

ULC Members (continued)								
Institution	**Protocol**	**Redirect**	**Google Analytics**	**GA Anonym**	**Google Tag Manager**	**Google Custom Search**	**Double Click**	**Facebook Connect**
Santa Clara County Library District	HTTPS	Valid	Classic	✓	—	—	—	—
Santa Clara Public Library	HTTP	Invalid	Universal	X	✓	—	—	—
Santa Monica Public Library	HTTPS	Valid	Classic	X	—	—	—	—
Scottsdale Public Library System	HTTPS	Valid	?	?	✓	—	—	—
Seattle Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Skokie Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Sno-Isle Libraries	HTTPS	Valid	Universal	X	✓	—	—	—
Solano County Library	HTTP	Unsup-ported	Universal	X	✓	—	—	✓
Somerville Public Library	HTTPS	Valid	Universal	✓	—	—	—	—
Springfield City Library	HTTPS	Valid	Universal	X	—	—	—	—
Stark County District Library	HTTPS	Valid	?	?	✓	—	—	—
Sunnyvale Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Tacoma Public Library	HTTPS	Valid	Universal	X	✓	—	—	✓
Tampa-Hillsborough County Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Toledo-Lucas County Public Library	HTTP	Unsup-ported	Universal	X	✓	—	—	✓
Topeka and Shawnee County Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Toronto Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Torrance Public Library	HTTPS	Valid	Universal	X	✓	—	—	—
Tulare County Public Library	HTTPS	Valid	?	?	✓	—	✓	✓
Tulsa City-County Library	HTTPS	Valid	Universal			—	—	—
Tuscaloosa Public Library	HTTPS	Passive	Universal	X	—	—	—	✓
Virginia Beach Public Library	HTTPS	Valid	Universal	X	—	—	—	—
Waco-McLennan County Library	HTTPS	Valid	?	?	—	—	—	—
Wake County Public Libraries	HTTP	Invalid	Classic	X	✓	—	—	—
West Bloomfield Township Public Library	HTTPS	Passive	?	?	—	—	—	—
Wichita Public Library	HTTP	Unsup-ported	Universal	X	—	—	—	✓
Worcester Public Library	HTTPS	Passive	?	?	—	—	—	—

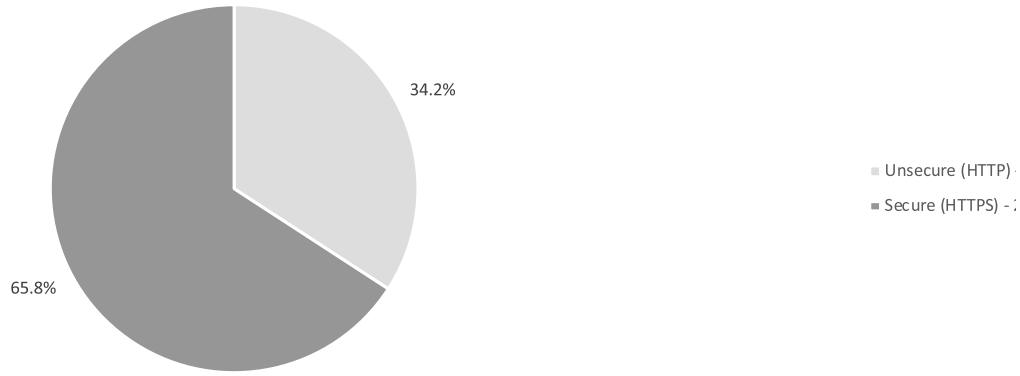


Figure 3.11. Academic libraries in the United States using HTTPS. Figure shows 3,954 of the 4,081 academic library libraries.org entries in the United States. The remaining entries either have no website link recorded or no confirmed website.

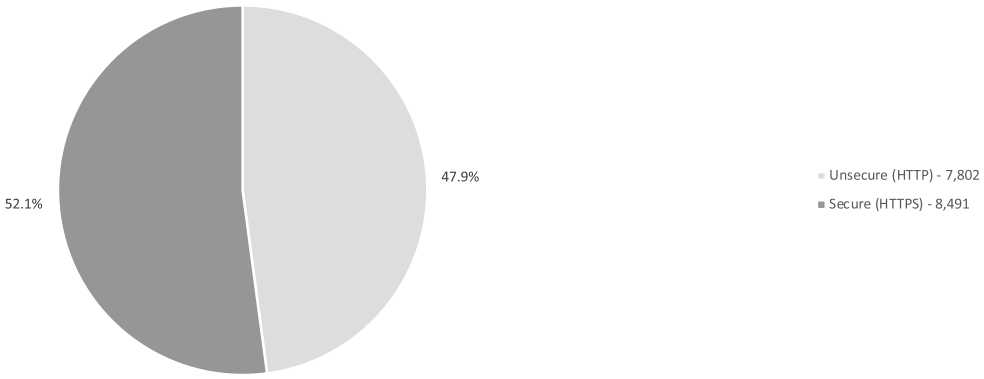


Figure 3.12. Public libraries in the United States using HTTPS. Figure shows 16,293 of the 17,310 public library libraries.org entries in the United States. The remaining entries either have no website link recorded or no confirmed website.

Looking Forward

While this study shows important progress toward library websites configured to provide stronger protections for the privacy for those that use them, much room remains for improvement. The deadline has already passed for securing web-based services, with browsers now flagging nonencrypted library websites as unsecure and not trustworthy. When considering other factors such as redirection to achieve mandatory HTTPS, the current status is not as impressive. The widespread use of tracking agents, especially when available anonymization is not implemented, means even more concern regarding privacy protections.

There will not be an easy or fast track in the deployment of properly secured HTTPS on the websites of the remaining libraries still using unencrypted communication. Libraries have generally seen slow transitions away from obsolete technologies in favor of modern alternatives. Those remaining represent a long tail of libraries with very sparse resources that also have a low level of awareness about the technical issues involved. Given the current rate of transition, I would anticipate that the number of library websites that do no use HTTPS will be less than a few percent by the end of 2020.

Privacy by Design

In the future, privacy will need to be one of the key considerations in the design of library websites if they are to be consistent with library values and meet the strategic objectives of libraries. In the same way that library websites should be responsive, work with all types of devices, and meet requirements for persons with disabilities, they should also conform to requirements for privacy protection.

Strategies for Achieving Privacy-Respecting Services

Several actions could be taken to accelerate the achievement of full compliance of privacy on library websites and related services:

- Those in leadership positions in libraries should be involved in this issue. It should not be up to the discretion of technologists. Administrators should rather hold technologists accountable to provide standard privacy protections in all systems deployed by the library.
- Professional bodies, such as the American Library Association, could further strengthen their guidance for the encryption of all web-based services used by libraries to provide access to information.
- Organizations providing or distributing funding for the implementation of library websites should require that those resources support HTTPS-only communications. I observe that many of the library websites without HTTPS encryption are funded through IMLS grants.
- Technology providers, including commercial and nonprofit, should ensure that their products are developed with the ability to operate with HTTPS-only communications and that this configuration option is enabled except in the case of unusual circumstances where such a configuration would not be possible because of local dependencies. This requirement would be especially relevant to any content management systems used to manage library websites as well as online catalogs and discovery services.
- Libraries should stipulate requirements for secure communications on all technology-related services they purchase. This requirement should apply to both browser-based interfaces and behind-the-scenes communications using standard protocols like SIP, NCIP, or Z39.50 as well as APIs.

Reducing the exposure of personal information of persons visiting websites due to the placement of tracking agents will be much more difficult to achieve. There appears to be limited awareness of privacy issues related to the tracking agents for analytics and for those related to social networks or the advertising ecosystem. Libraries are well motivated to move into the realm of big data and analytics to assess and refine their services. Libraries increasingly see personalized services and targeted marketing as ways to improve engagement with their community members and to combat the existential threats to funding and support.

Progress in mitigating the threats to privacy related to the use of tracking agents can be achieved through these measures:

- **Self-auditing of websites and related resources:** Libraries should at least be aware of the tracking agents present on their web-based services. Library personnel should use tools such as Ghostery to confirm which tracking agents have been installed. In many cases, these agents may have appeared on the library site inadvertently. Libraries often borrow scripts or widgets from other libraries or from commercial sources to achieve desired visual effects or functionality. These components may in turn invoke tracking agents. An audit of the tracking agents would inform a process to identify the specific code that invokes the agent and a review regarding which agents are viewed as tolerable within the library's privacy policies and which should be eliminated.
- **Comprehensive anonymization of tracking data:** This study shows a low rate of IP anonymization in the configuration of Google Analytics. This report provides information that the anonymization configuration of Google Analytics is more consistent with protecting the privacy of the individuals that use library-provided resources. Administrators and policy makers in the library community should make recommendations, if not mandates, that anonymization of IP addresses be implemented on any service that involves tracking agents and transmission of user activity to a third party.
- **Alternative privacy-respecting services:** Libraries have a significant interest in promoting their services to their communities. As libraries work to implement marketing strategies, they should ensure that the technologies that support these efforts do not intrude on the privacy of their users in ways that may not be intended or that are inconsistent with stated policies. While it's tempting to make use of tools and frameworks provided for free by the leading technology giants, libraries must assess any compromises that these tools require relative to user privacy and pursue or develop alternatives when needed.

Ongoing Research and Analysis

This issue of *Library Technology Reports* describes the author's ongoing project in the exploration of the trends and technologies related to the security and privacy of library websites and related systems. In this phase of the work, the study has expanded beyond a focus on the largest libraries, such as the members of the Association of Research Libraries and Urban Library Council, to the comprehensive sets of public and academic libraries in the United States. This expanded scope was made possible through the development of automated tools to identify pertinent characteristics. Identifying the proportion of libraries using HTTPS or those implementing tracking agents would not be feasible through methods based on manual inspection.

The next phase of work in this area will include refinement of the automated tools to more definitively identify tracking agents and to expand the body of libraries studied to other countries. Additional work is also needed to analyze the technical interactions between tracking agents placed on library websites and the advertising ecosystem. A clearer understanding of how traces of online information-seeking behavior performed on library websites can leak into ad networks will help inform future recommendations on what tracking agents can be allowed relative to patron privacy concerns.

A complete transition to HTTPS-only communications on library websites can be considered as basic table stakes in the struggle to protect user privacy on library websites. Enforcing encryption provides protection against hypothetical intruders that might be interested in capturing the interactions of individuals with library-provided information services. In the realm of tracking agents, the adversaries are well known. The advertising-based web ecosystem seems to continually expand its appetite for personal information. Libraries will need to be ever more vigilant in the future to ensure an impermeable firewall between their services and the surrounding ad-based commercial infrastructure.

Additional References and Resources

Brantley, Peter, Marshall Breeding, Eric Hellman, and Gary Price. "Swords, Dragons, and Spells: Libraries and User Privacy." Project briefing, CNI's December 2014 member meeting. Online video, 44:23, posted January 23, 2015. https://www.cni.org/news/video-libraries-and-user-privacy.

Breeding, Marshall. "Privacy and Security for Library Systems." *Library Technology Reports* 52, no. 4 (May/June 2016).

Breeding, Marshall. "Protecting Patron Privacy." *Smart Libraries Newsletter* 36, no. 5 (May 2017).

National Information Standards Organization. *NISO Consensus Principles on Users' Digital Privacy in Library, Publisher, and Software-Provider Systems (NISO Privacy Principles)*. White paper. Baltimore, MD: NISO, December 10, 2015. https://www.niso.org/publications/privacy-principles.

O'Brien, Patrick, Scott W. H. Young, Kenning Arlitsch, and Karl Benedict. "Protecting Privacy on the Web: A Study of HTTPS and Google Analytics Implementation in Academic Library Websites." *Online Information Review* 42, no. 6 (2018): 734–51, https://doi.org/10.1108/OIR-02-2018-0056.

Santa Cruz County Civil Grand Jury. "Patron Privacy at Santa Cruz Public Libraries: Trust and Transparency in the Age of Data Analytics." June 24, 2019. http://www.co.santa-cruz.ca.us/Portals/0/County/GrandJury/GJ2019_final/SantaCruzPublicLibrariesReport.pdf.

Notes

Statement of Ownership, Management, and Circulation

Library Technology Reports, Publication No. 024-897, is published eight times a year by the American Library Association, 50 East Huron St., Chicago (Cook), Illinois 60611-2795. The editor is Samantha Imburgia, American Library Association, 50 East Huron Street, Chicago, IL 60611-2795. Annual subscription price, $340.00. Printed in U.S.A. with periodicals class postage paid at Chicago, Illinois, and at additional mailing offices. As a nonprofit organization authorized to mail at special rates (DMM Section 424.12 only), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.

(Average figures denote the average number of copies printed each issue during the preceding twelve months; actual figures denote actual number of copies of single issue published neared to filing date: August/September 2019 issue.) Total number of copies printed: average, 585; actual, 577. Paid distribution outside the mails including sales through dealers and carriers, street vendors, counter sales, and other paid distribution outside the USPS: average 73; actual, 53. Total paid distribution: average, 357; actual, 329. Free or nominal rate copies mailed at other classes through the USPS (e.g., First-Class mail): average, 0; actual, 0. Free or nominal rate distribution outside the mail (carriers or other means): average, 16; actual, 15. Total free or nominal rate distribution: average, 16; actual, 15. Office use, leftover, unaccounted, spoiled after printing: average, 213; actual, 233. Total: average, 585; actual, 577. Percentage paid: average, 97.29%; actual, 96.89%.

Statement of Ownership, Management and Circulation (PS Form 3526, July 2014) filed with the United States Post Office Postmaster in Chicago, September 9, 2019.

Library Technology REPORTS

Upcoming Issues	
November/ December 55:8	**Blockchain in Libraries** by Michael Meth
January 56:1	**Digital Rights Management: Barriers, Problems, and Solutions** by Mirela Roncevic
February/ March 56:2	**Digital Disruption** by Bohyun Kim

Subscribe
alatechsource.org/subscribe

Purchase single copies in the ALA Store
alastore.ala.org

alatechsource.org

ALA TechSource, a unit of the publishing department of the American Library Association