Aiding and Abetting: Third-Party Tracking and (In)secure Connections in Public Libraries

Gabriel J. Gardner [a*]

This is an electronic version of an article published as Gabriel J. Gardner (2021) Aiding and Abetting: Third-Party Tracking and (In)secure Connections in Public Libraries, *The Serials Librarian* AHEAD OF PRINT, 1 - 19, doi: 10.1080/0361526X.2021.1943105. Serials Librarian is available online at: https://www.tandfonline.com/loi/wser20

Author Note

[a] University Library, California State University Long Beach, Long Beach, United States of America; ORCID https://orcid.org/0000-0002-9996-5587

[*] email: gabriel.gardner@csulb.edu telephone: 562.985.4976 postal: c/o University Library, California State University Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840-1901

The data that support the findings of this study are openly available in figshare at DOI: 10.6084/m9.figshare.12219863.v1

Acknowledgements

Myron Groover of McMaster University inspired this study and assisted with the research design. Janelle Brunner’s copyediting and Carolyn Caffrey Gardner’s comments greatly improved the first draft of this manuscript.

The findings, interpretations, and conclusions expressed herein are those of the author and do not necessarily reflect the policies or views of California State University Long Beach.

Aiding and Abetting: Third-Party Tracking and (In)secure Connections in Public Libraries

Abstract

Patron privacy, as articulated in the American Library Association (ALA) Code of Ethics, is a longstanding concern for librarians. In online environments, the possibility of tracking by third parties, usage of HTTPS/TLS to provide secure connections, and easy disclosure of a site's privacy policies all have implications for user privacy. This paper presents new empirical evidence about these issues and discusses their ethical implications. Data about the incidence of third-party tracking, usage of HTTPS by default, and easy discoverability of a privacy policy or terms of service (TOS) were collected for public libraries across Canada and the United States. The sample consisted of 178 public libraries; members of the Canadian Urban Libraries Council and Urban Libraries Council. Several common commercial databases (e.g. OverDrive) were also examined using the same criteria. Results show that only 12% of libraries were devoid of third-party tracking, with Google Analytics being the most common third-party tracker. While libraries may support HTTPS under certain circumstances, it was found that a majority of libraries serve neither their websites nor their online catalogs (OPACs) HTTPS by default. Regarding disclosure of possible tracking, it was found that 58% of libraries did not link to a TOS or privacy policy from their homepage. Together with previous research on the usage of privacy-enhancing tools in public libraries, these results suggest that public libraries are accessories to third-party tracking on a large scale. Implications of this fact in light of library professional ethics are discussed.

Keywords: privacy, security, tracking, HTTPS/TLS, ethics, data collection

Introduction

On June 24th, 2019, a Civil Grand July in Santa Cruz County California issued a report rebuking the Santa Cruz Public Libraries for: failing to inform patrons how their personal data was used, not adopting best practices for privacy outlined by the American Library Association, and for entering into contracts with third parties that raised liability issues related to patron privacy – among other things (York 2019). To what extent are other libraries at risk for reprimand? How are we protecting patron privacy? In the sea of libraries any particular institution is after all but one drop, and a library can preserve privacy only when it acts correctly with policies and disclosures its users are conscious of. Unless sufficient numbers of libraries take the lead on tracking disclosure, the rhetoric of privacy they employ rings hollow. Perceptive patrons will see through the facade, librarians will know they are failing to uphold their professional ethics and best practices; the whole advocacy effort on behalf of privacy could be called into question. Privacy is a complex multi-faceted issue made all the more complicated when the medium over which users are desiring privacy and which libraries offer their services is the internet. Currently, the internet more closely resembles a Benthamite panopticon more than the liberatory frontier imagined by John Perry Barlow (Assange 2014; Mozilla 2017).

This paper considers two aspects of online privacy: secure connections between users and the library content they seek, and the often invisible, web of third-parties tracking user behavior on websites. Though secure connections between users and library content is arguably more important, third-party tracking strikes at the philosophical heart of any meaningful conception and operationalization of privacy. There can be no real and effective “privacy” in a library that surreptitiously allows third-parties access to their digital environments while library patrons use these websites in ignorance of the fact that a handful of publishing and services firms are privy to their usage.

Literature Review

Professional Ethics

Librarians and their professional associations have a storied and noble history of advocacy on behalf of user privacy (Ard 2016). Indeed, Jason Griffey has suggested that in an online world of ubiquitous tracking, “the main service that public libraries might provide in the future is privacy” (Griffey 2016).

Moving from the various methods that librarians might structure the digital environments of their users for more passive privacy, there is also the question of proactively educating patrons. That efforts to educate the public about the importance of privacy are justified has been suggested by many individuals and notably since 2010 has been embodied in the American Library Association’s annual “Choose Privacy Week”. Some libraries make a stronger claim though, that not only is such education justified but necessary (Lamdan 2015; Morrone 2015). Such an attitude has been most publicly proclaimed by people associated with the Library Freedom Project which offers clear and comprehensive steps for users to take and shares their educational materials.

To take the pulse of how public libraries in Canada and the United States were approaching computing privacy and security in light of our professional ethics, this paper studies elements of those issues using library websites.

Third-Party Tracking

Third-party tracking, sometimes called "leaking", is an ubiquitous aspect of the modern internet (Englehardt and Narayanan 2016; Acar et al. 2014). Third-party tracking on the world wide web has been documented as far back as 1996 and since then has been increasing in complexity, the number of elements tracked as well as the number of methods of tracking, and prevalence unabated (Lerner et al. 2016). Common methods include: standard HTTP browser cookies (in same-site and third-party variety), Adobe Flash cookies (less of a threat with the gradual disappearance of Flash), local storage so-called "supercookies" and ETags, the Evercookie (which uses redundant multiple methods in order to make it difficult to delete), and JavaScript fingerprinting.

Some tracking mechanisms are used by law enforcement and other state entities (Albrecht and McIntyre 2014; Rankin 2014; Tate and Soltani 2014; Zetter 2014). But by and large, third-party tracking is used for commercial purposes, specifically targeted advertising (Hoofnagle et al. 2012; Cranor 2012). Of these tracking methods, JavaScript fingerprinting is perhaps the most pernicious because it allows identification of individual users with a very high degree of confidence both based on their device and also based on their behavior; the only way to avoid such fingerprinting for certain is to disable JavaScript which renders most websites non-functional (Acar et al. 2014). Furthermore, the widespread incidence of third-party tracking has been known for decades, even in the popular press (Madrigal 2012; Sullivan 2012).

In 2016, Eric Hellman looked at the websites of 123 Association of Research Libraries (ARL) member institutions. Using the Ghostery plugin for Chrome, what he found was not encouraging.[1] Google Analytics was used by 72% of ARL Libraries; 28% used Amazon to enrich their catalogs, thus transmitting query data via referrer headers. Web beacons, allowing (potential identifiable) individual tracking, were found on 13% of ARL catalog pages (Hellman 2016). Cody Hanson, in a study of 15 commercial publisher websites along the lines of that undertaken below in this study, found four publisher platforms with Facebook code, four platforms with Neustar (a marketing firm) code, six with Adobe code, four with Oracle Marketing Cloud code, 11 with AddThis (a social sharing button that cooperates with commercial data brokers), and 14 with Google code. In each case, such code allows the aforementioned parties to surreptitiously gather information about user behavior and *possibly* link library activity with pre-existing data profiles about users; he concluded: "I do not believe it is possible for use of licensed resources to be private" (Hanson 2019). Empirical work on the amount of third party tracking in library settings is largely lacking, something the present paper begins to rectify.

The possibilities for user tracking and profiling and the methods already extant and in use are astounding (Privacy International 2018). A growing and already robust market in consumer data

[1] This present study uses a very similar methodology to Hellman, 2016. I thank Myron Groover for bringing Hellman's blog post to my attention, as well as any number of the articles reviewed for this literature review.

has existed for decades. With the advent of big data as captured via web browsers and harvested from smartphones the existing streams of consumer demographic and purchase data can now be merged with online behavioral analytics data to create detailed pictures of individuals – without their knowledge – to which optimization logic is indiscriminately applied for various commercial ends (Esposti 2014). There is debate in the literature as to the richness of the probabilistic pictures that can be painted using big data; they have improved over time as machine learning algorithms have improved. Though one thing is certain, when demographic and commercial data is combined with social network information, the predictions made about people become stronger (Huey et al. 2012).

Of particular interest given the sample in this study is work done by Smith and Lyon in 2013 documenting views on surveillance and privacy among Canadian and United States samples. Both countries experienced about a 10 percent drop from 2006 to 2012 among people reporting they were knowledgeable about laws dealing with the protection of personal information from private companies. Across the same period, opinions about employers sharing employee information with third parties remained stable, with most respondents opposed; age differences were pronounced on this question with older respondents being much more opposed and those 18 to 34 only 35% opposed, suggesting that privacy advocacy has an uphill climb in the future among the young (Smith and Lyon 2013).

Library Website Security

Apart from privacy, how secure are library websites? How secure are their catalogs? The usage of Transport Layer Security which is used to verify the authenticity of websites, HTTP traffic, and keep user communications and web browsing safe from packet-sniffing and packet-injection, is the commercial and technical standard ("HTTPS" 2020). Alison Macrina, founder of the Library Freedom Project, explained the implications of using HTTPS on library websites. Libraries would never allow interlopers to stand near a circulation desk and record who checked out what, or allow for patrons to be followed as they navigate the stacks, yet by serving their websites and catalogs in unsecured HTTP, they allow for digital occurrences of a similar nature (Macrina 2015). The public image of libraries as essential democratic institutions and the trust they have earned demand the use of secure online platforms. In the context of higher education, Hellman noted only a few years ago that a paltry 20% of ARL Libraries secured their catalogs with TLS/SSL by default (Hellman, Eric 2016). Others have noted that the failure to use HTTPS with Breeding noting that only 13% of ARL library websites defaulted to HTTPS and that of 25 large public libraries he sampled, only two used HTTPS on their website and seven in their catalogs (Breeding 2016).

In addition to securing their online platforms using HTTPS, there are a variety of actions libraries can take to protect the privacy of users on library property. In communities all across the world, there are some people who primarily obtain internet access by visiting their local library and using computers provided there. Several authors have suggested that merely supplying access is not sufficient and that libraries need to configure their public access computers to provide users with a level of privacy and security above that which comes out of the box or just with programs such as *Deep Freeze*.[2] Eric Phetteplace suggested not only

[2] http://www.faronics.com/products/deep-freeze

common sense basics such as having all browsers set to never remember passwords and to load into "private" (called "incognito" in Chrome) mode, but also the installation of extensions designed to "harden" out of the box browser security. These included: HTTPS Everywhere, Web of Trust, NoScript, and the adblocker AdBlock+ (Phetteplace 2012). Such extensions would improve security for users, offer a modicum of protection against malicious sites, make potential attacks opt-in, and improve page load times by limiting some advertising and third-party tracking.

It has also been suggested that an additional step libraries might take is to default to privacy-preserving search engines, e.g. DuckDuckGo, on public computers and for libraries to promote their use (Radical Reference 2014; Gardner 2013; Phetteplace 2012). In order to provide some empirical context for library privacy and security efforts, Gardner & Groover surveyed 69 libraries in Canada and the United States, all of which fall into the sample of the present study. They found that most responding libraries failed to use the private browsing on their public terminals by default; nor did a majority have any ad-blocking software installed on those terminals (Gardner and Groover 2015). Most of the surveyed libraries defaulted to Google for search purposes; only 25% of responding libraries said they had offered public instruction in online anonymity or privacy. Those results stand in agreement to work done by Zimmer which surveyed 1,214 librarians and in which 76% said "libraries are doing all they can to prevent unauthorized access to individual's personal information". But that same survey found that only 13% responded that their libraries have offered privacy-related events for the general public (Zimmer 2014). All of this begs the question of how librarians understand "all they can" since privacy training is not ubiquitous nor is use and endorsement of tools to "harden" web browsers. These technical and empirical discussions aside, others have pointed out the ethical questions raised by systemic power imbalances between upstream commercial firms that supply so much library software and the end users who implicitly trust it (Barron and Preater 2018).

Methods

Sample

Public Libraries

This study examined the websites of 178 public libraries in North America: 45 from Canada, and 133 from the United States. These libraries were purposely selected for their membership in either the Canadian Urban Libraries Council (CULC) or the Urban Libraries Council (ULC) in the United States.[3] These organizations have some of the largest library systems in their membership and serve the urban and suburban communities in both countries. The Canadian sample libraries were located in communities that account for approximately 41% of Canada's population.[4] United States sample libraries were located in communities that account for

[3] Membership in the Urban Libraries Council changes periodically as institutions join or withdraw. For a listing of the institutions examined, see the raw data underlying this study on figshare: https://doi.org/10.6084/m9.figshare.12219863.v1

[4] Canadian population figures were obtained for every municipal library system in the CULC; libraries serving wide geographies i.e. "regional" systems were excluded. The data was obtained from the *Statistics Canada* website reporting out the results of the 2016 Census.

approximately 28% of the United States of America's population.[5] These percentage figures serve as hypothetical, upper bound estimates of the population affected by third-party tracking since not every member of these communities actually uses their local public library.

Common Content Sources

In addition to evaluating public library websites, 10 common content sources that many public libraries subscribe to were also examined to measure how many of these facilitated third-party tracking. These content sources were selected for their wide usage and the nature of the information they provide, which is commonly known to be in demand among public library users. The common content sources included the following:

- Ancestry.com, a popular source for genealogical information;
- Chilton Auto Repair Manuals, which are the leading source for automotive repair information and are widely used in do-it-yourself and professional settings;
- Consumer Reports, a well-respected and widely used source for unbiased product reviews and ratings;
- hoopla, an online streaming media service offering music, movies, television shows, and ebooks;
- Lynda, an online education source offering multi-topic video courses;[6]
- Mango Languages, an online language and culture learning suite offering videos and applications to instruct users in scores of languages;
- NoveList, the industry standard source of reading recommendations and book reviews;
- OneClickDigital, a leading provider of downloadable audiobooks and ebooks;
- OverDrive, a distributor of ebooks, audiobooks, music, and videos, to which thousands of libraries around the world subscribe; and
- zinio, an online distributor of digital copies of magazines.

Measures

Data on third-party tracking was collected using two web browser add-ons: Ghostery and Disconnect. These add-ons were chosen for this analysis due to their widespread usage among the general public and the fact that they have been used successfully in previous research on third-party tracking (Schaub et al. 2016; Mathur et al. 2018). Ghostery is used by "more than seven million" users and Disconnect by "tens of millions of people" (Ghostery n.d.; Disconnect n.d.). Typically, third-party tracking is done using JavaScript code, often called tags or scripts, to create cookies or super-cookies, which record the pages a user visits over a specified time period. This allows the third-parties to build profiles on users based on their browsing behavior; information leakage, referred to above, takes place via HTTP Referrer headers. Both Ghostery and Disconnect

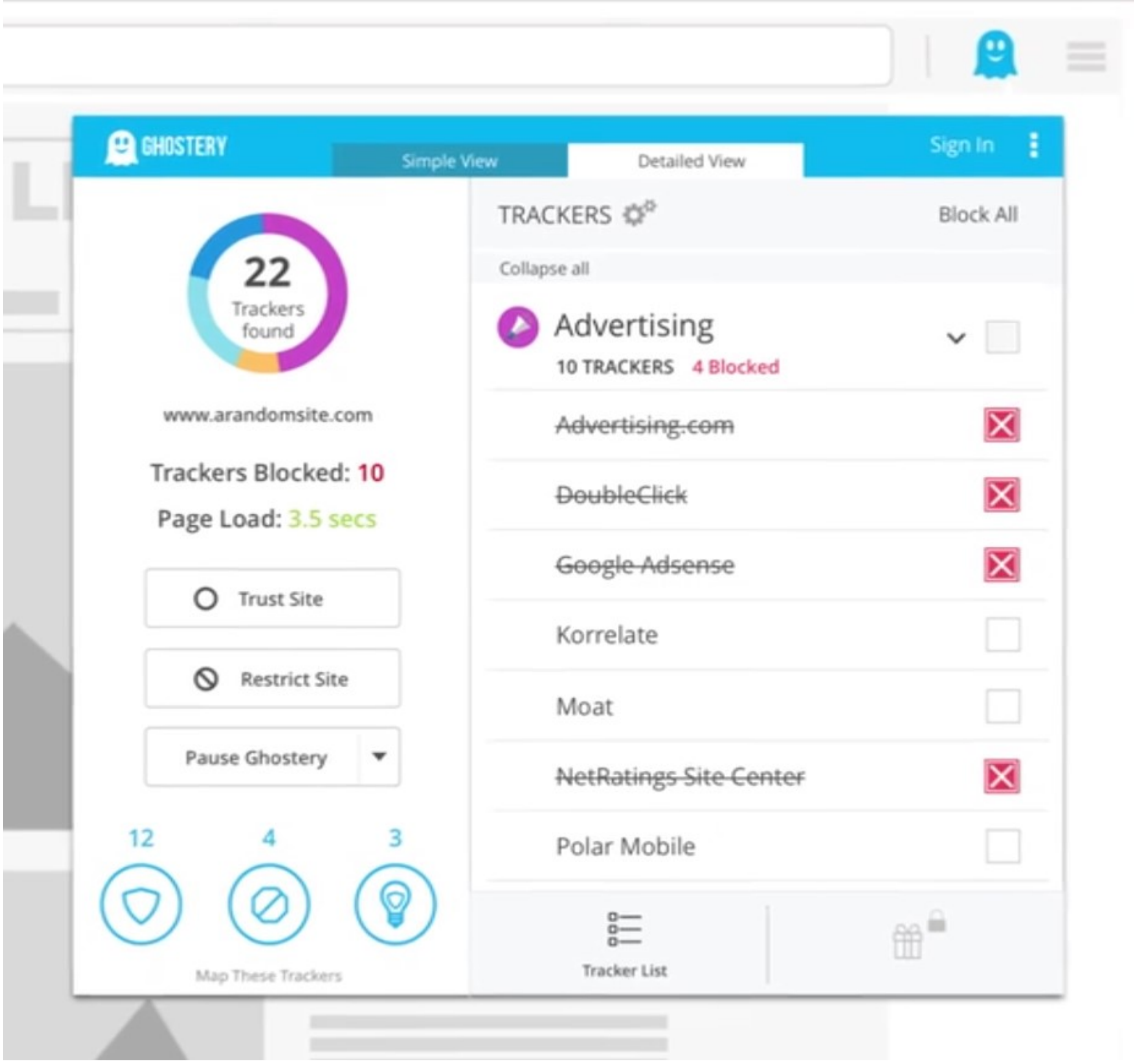
[5] United States population figures were obtained for every municipal or county library system in the ULC. The data was obtained from the *American Factfinder* website reporting out the results of the 2015 American Community Survey 5-year population estimates. When applicable, municipal library systems were excluded to avoid double counting in the USA figures if a County system was also a member of ULC; e.g. Los Angeles (Los Angeles Public Library) and Los Angeles County (County of Los Angeles Public Library).

[6] Now known as LinkedIn Learning.

work by examining the JavaScript code on web pages. Dissimilarities between the two add-ons and discrepancies between their counts for the same pages occasionally occurred; these matters are addressed in the Appendix. The counts for each are reported separately in the results below.

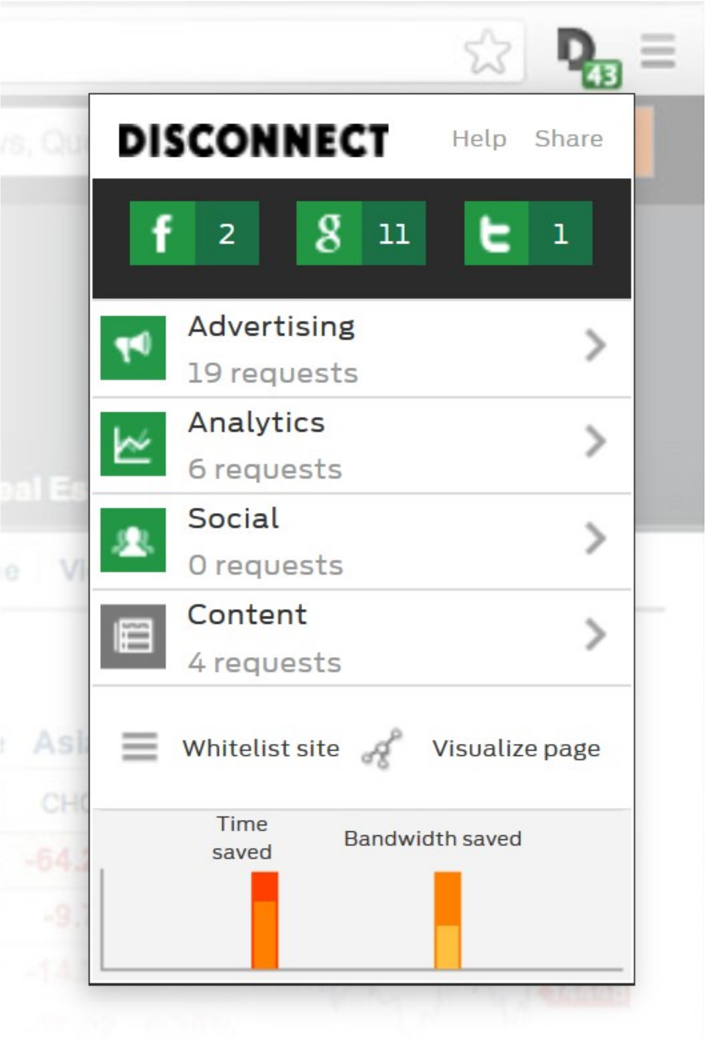
More specifically, Ghostery detects and blocks trackers and reports them in a findings window, which appears when the Ghostery icon in the browser bar is clicked (see Figure 1). This window reports the total number of trackers and also evaluates them qualitatively based on information it collects in order to classify them according to 8 mutually exclusive categories: Advertising, Comments, Customer Interaction, Essential, Pornvertising, Site Analytics, Social Media, and Audio/Video Player.

Figure 1: Ghostery



Similarly, Disconnect detects and blocks trackers and reports them in a findings window, which appears when the Disconnect icon in the browser bar is clicked (see Figure 2). The total number of trackers is visible in the Disconnect icon of the browser bar, the window displaying a qualitative analysis of each tracker, which is based on information Disconnect collects. Trackers are classified according to 7 mutually exclusive categories: Facebook, Google, Twitter, Advertising, Analytics, Social, and Content.

Figure 2: Disconnect



f 2
g 11
t 1

Advertising
19 requests

Analytics
6 requests

Social
0 requests

Content
4 requests

Additionally, in order to determine whether public libraries and common content sources informed their users of any third-party tracking, searches were performed (using Ctrl-F) to find the presence of a link on each library or source's homepage that might lead to such information. A simplistic measure of informing or not informing users was developed based on whether the homepage of a library or common content source had a linked policy statement. Some typical examples of text used to link such polices are "Privacy Policy," "Privacy Statement," "Terms of Service," "Terms of Use," and "Website Policies." Data were collected by visiting the homepage for each library and each common content source and searching for any of the proceeding words: “privacy”, “terms”, “policy”, or “policies.” (Note: for French-language libraries in Canada, the words “conditions”, “confidentialité”, and “règle” were searched.) If any of the words were found to be a link to some relevant policy, they were counted as informing their users; if the words were not found as links to such a relevant policy, the sample member was recorded as non-informing. Qualitative analysis of the linked policies, such as determining whether they *accurately* disclosed the extent of any third-party tracking, was not done. Subpages for each sample library were not examined; they were only counted as informing if a link to a statement or disclaimer was on the main homepage or landing page.[7]

Data Acquisition Procedure

All tracking data from Ghostery and Disconnect data were collected using Firefox ESR version 38.8.0., Ghostery tracking figures were collected using Ghostery (for Mozilla Firefox) version 6.2.0; Disconnect tracking figures were collected using Disconnect (for Mozilla Firefox) version 3.15.3.1.[8] Data on public libraries was obtained by visiting an online public access catalog (OPAC) search results page for each library. This was first done in Ghostery, which was then removed and replaced with Disconnect; Disconnect data was collected by visiting the same search results page again. Data on common content sources was obtained in a similar manner; however, links from the author’s respective public libraries were followed to each source.[9] This was a necessary step due to the fact that many public libraries require authentication with a proxy server using library-provided credentials in order to reach the landing homepage of the content sources that they subscribe to. No other Firefox extensions were installed during data collection. The type, i.e. brand or product, of each OPAC was also recorded for each library; this data was obtained via libraries.org, an international directory maintained by Marshall Breeding. The data collection period ran for three months, from March 2017 through May 2017. Data on potential disclosure of tracking was collected during the same period using the method previously detailed above. Whether CULC and ULC library homepages and their corresponding landing pages of the OPAC were served HTTPS by default was also recorded.

Results

Third-Party Tracking Incidence

Public Libraries

It was commonplace for libraries in the sample to facilitate third-party tracking. The results from Ghostery and Disconnect were similar as far as the total number of libraries including third-

[7] See the Limitations section for the implications of this.

[8] Firefox was used simply because it was easiest to deploy in a vanilla (out of the box) fashion on the researcher’s computer. Various web browsers do perform differently regarding their treatment of mixed HTTP/HTTPS content and some have built-in protections against third-party trackers. However the object of this study was simply to record incidence of tracking, not compare the relative merits of browsers; Firefox was sufficient for that purpose.

[9] Ghostery and Disconnect counts were measured for these sources when accessed via a Canadian public library which had access to some of them and via a public library in the United States which had access to some of them. No intra-Ghostery or intra-Disconnect discrepancies were observed between tracker counts based on geography.

party tracking in their online catalogs: Ghostery recorded 154 libraries enabling third-party tracking while Disconnect recorded 156. There were two instances where Disconnect detected third-party tracking but Ghostery did not; for the 22 libraries (12%) where Disconnect did not detect tracking, Ghostery obtained the same results. In other words, the majority of libraries in the sample allowed some type of third-party tracking of their users. Disconnect, in general, recorded more trackers than Ghostery for the same URLs. Table 1, below, shows the total counts and other measures for Ghostery and Disconnect in their examinations of library online catalogs.

Table 1

Table 1

Incidence of 3rd-Party Tracking in Public Library Catalogs

n=178

Measures	**Ghostery**	**Disconnect**
Number of Libraries Enabling 3rd-Party Tracking	154	156
Total Number of Trackers Detected	269	362
Number of Trackers per Library: Average	1.51	2.03
Number of Trackers per Library: Median	1	2
Number of Trackers per Library: Mode	1	3

Of libraries that allowed tracking, the incidence was similar across the United States and Canada. Of Canadian libraries in the sample, 89% (40) enabled third-party tracking; for libraries in the United States, 87% (116) did. (Note that the 87% figure was derived from the Disconnect total tracking count, which detected two more deployments of tracking than Ghostery.) The number of trackers detected by Ghostery and Disconnect for Canadian libraries is shown in Figure 3. Figure 4 shows the same breakdown of information for libraries in the United States.

Figure 3: Canadian Libraries – 3rd-Party Tracker Prevalence

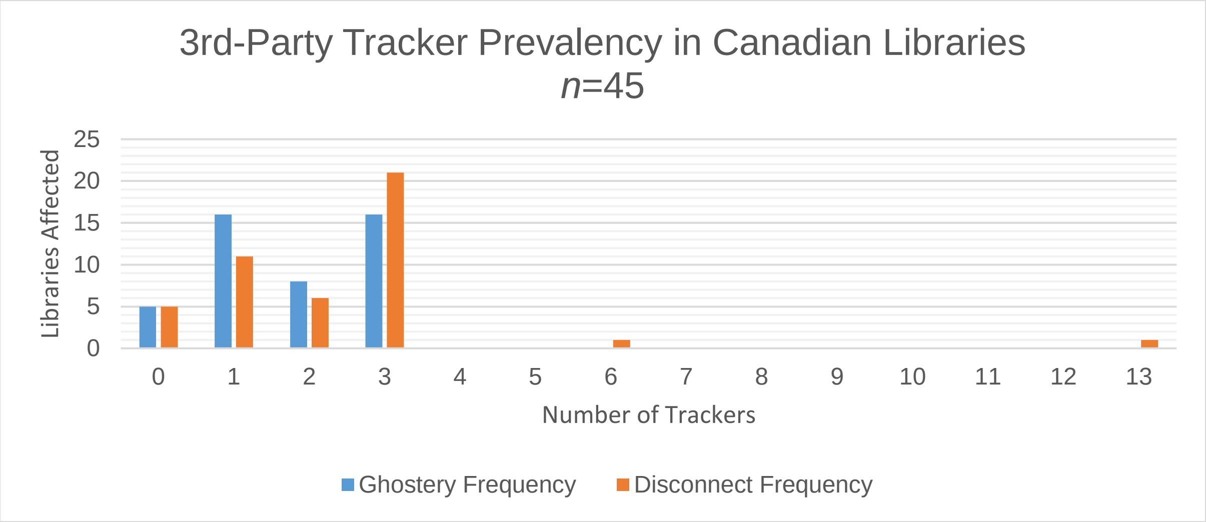
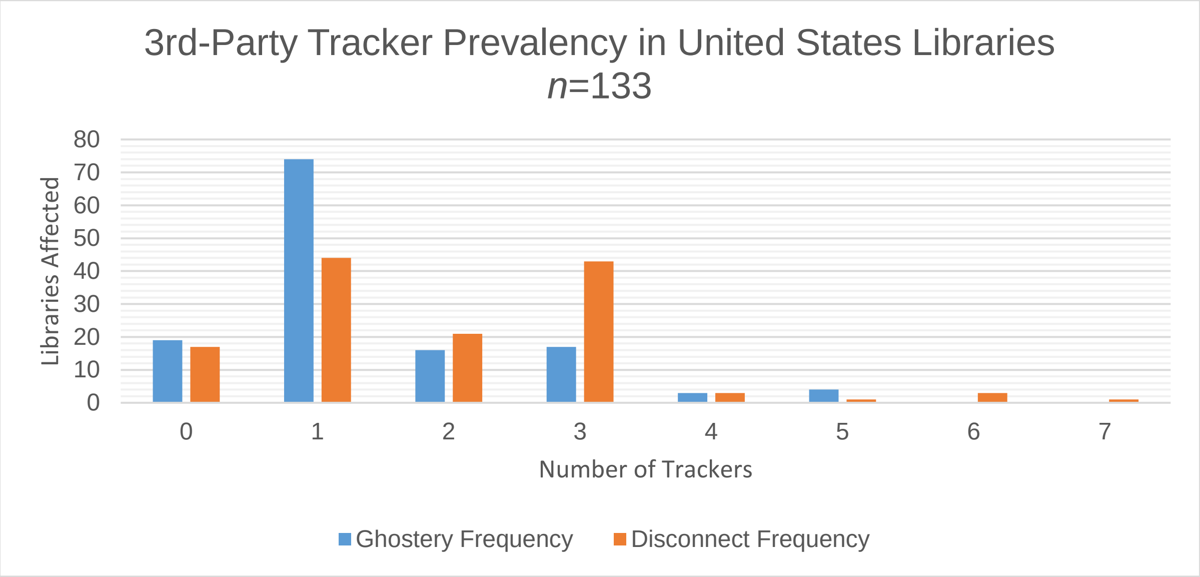


Figure 4: United States Libraries – 3rd-Party Tracker Prevalence



The various categories that Disconnect and Ghostery used to classify trackers allowed for a slightly more granular analysis. Disconnect tallied tracker counts for Google, Facebook, and Twitter and displayed them in the viewing window noted above. Facebook tracking was found by Disconnect in 40 libraries: 18 of which were in Canada, the remaining 22 in the United States. Google tracking, including Google Analytics or Google Tag Manager (or both), was found by Disconnect in 152 libraries: 40 of which were in Canada, 112 were in the United States. No libraries were found to have any third-party tracking from Twitter. Other trackers detected by Disconnect were Crazy Egg (36), New Relic (3), AddThis (3), ShareThis (1), Optimizely (1), and Zopim (1). Crazy Egg offers heatmaps, scrollmaps, click reports, and other features. New Relic offers data ingestion, storage, and visualization features along with frontend performance monitoring. AddThis is a social bookmarking service that allows users to share an item with popular social media platforms; ShareThis is a similar tracker. Optimizely is an A/B testing tool. Zopim (now called Zendesk Chat) is a live chat tool. Tracking entity as determined by Disconnect is displayed below for Canadian libraries in Figure 5. The same measures for the United States libraries are below in Figure 6.

Figure 5: Tracking in Canadian Libraries as Measured by Disconnect

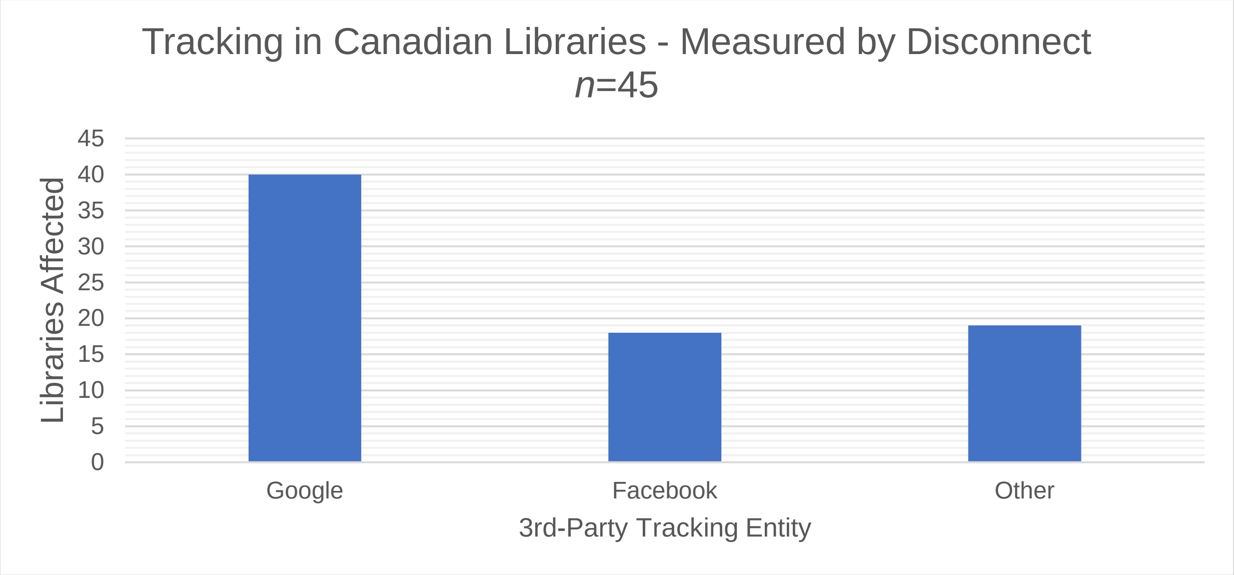
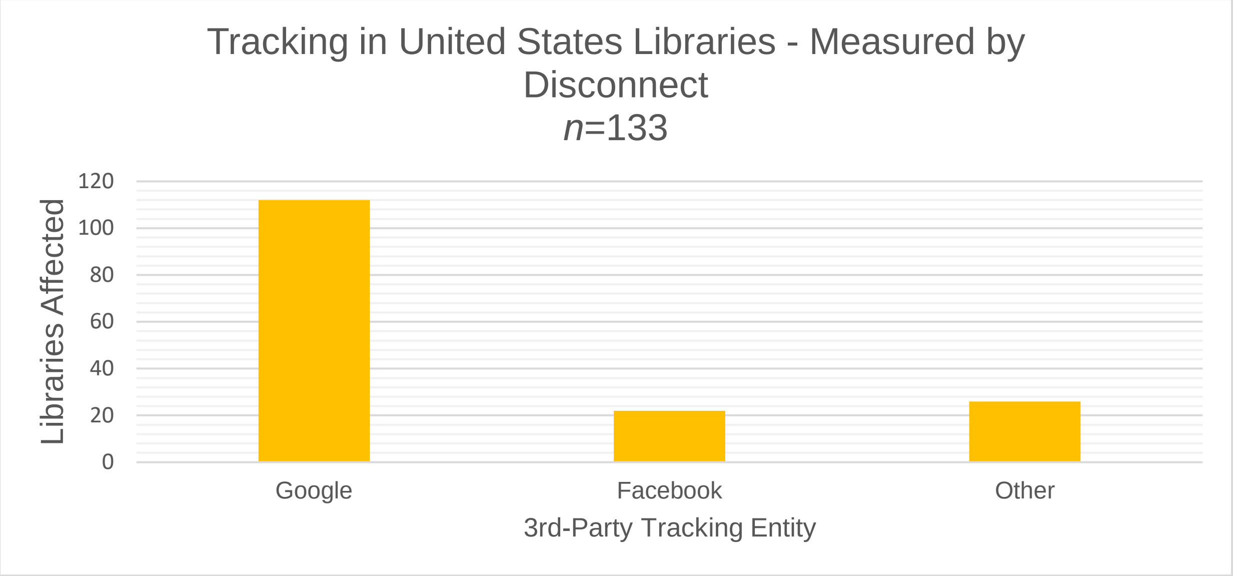
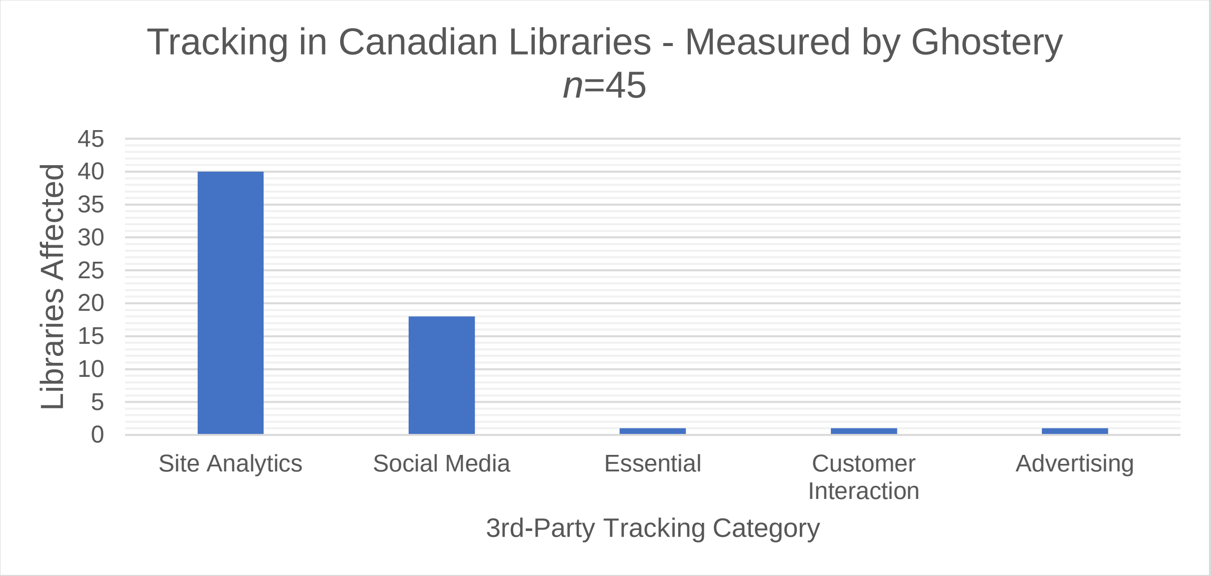


Figure 6: Tracking in United States Libraries as Measured by Disconnect



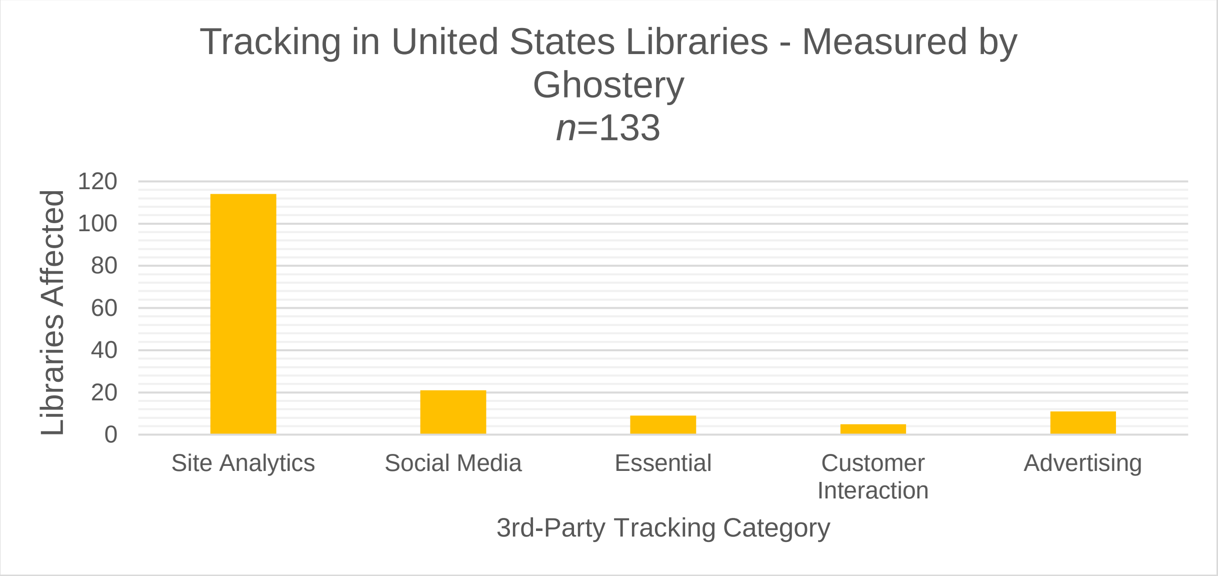
Ghostery similarly tallied tracker counts for its 8 categories of Advertising, Comments, Customer Interaction, Essential, Pornvertising, Site Analytics, Social Media, and Audio/Video Player. Facebook tracking, categorized under Social Media, was found by Ghostery in 39 libraries, 18 of which were in Canada, the other 21 in the United States. Google tracking, including Google Analytics or Google Tag Manager (or both), was found by Ghostery in 154 libraries: 40 Canadian, 114 in the United States. Other trackers detected by Ghostery included Crazy Egg (35), New Relic (9), DoubleClick (6), AddThis (4), ShareThis(1), Optimizely (1), HotJar (1), Loop11 (1), Zopim (1), SumOfMe (1), Piwik (1), and Adobe Typekit (4).[10] Tracker type as determined by Ghostery is displayed below for Canadian libraries in Figure 7. The same measures for the United States libraries are below in Figure 8.

Figure 7: Tracking in Canadian Libraries as Measured by Ghostery



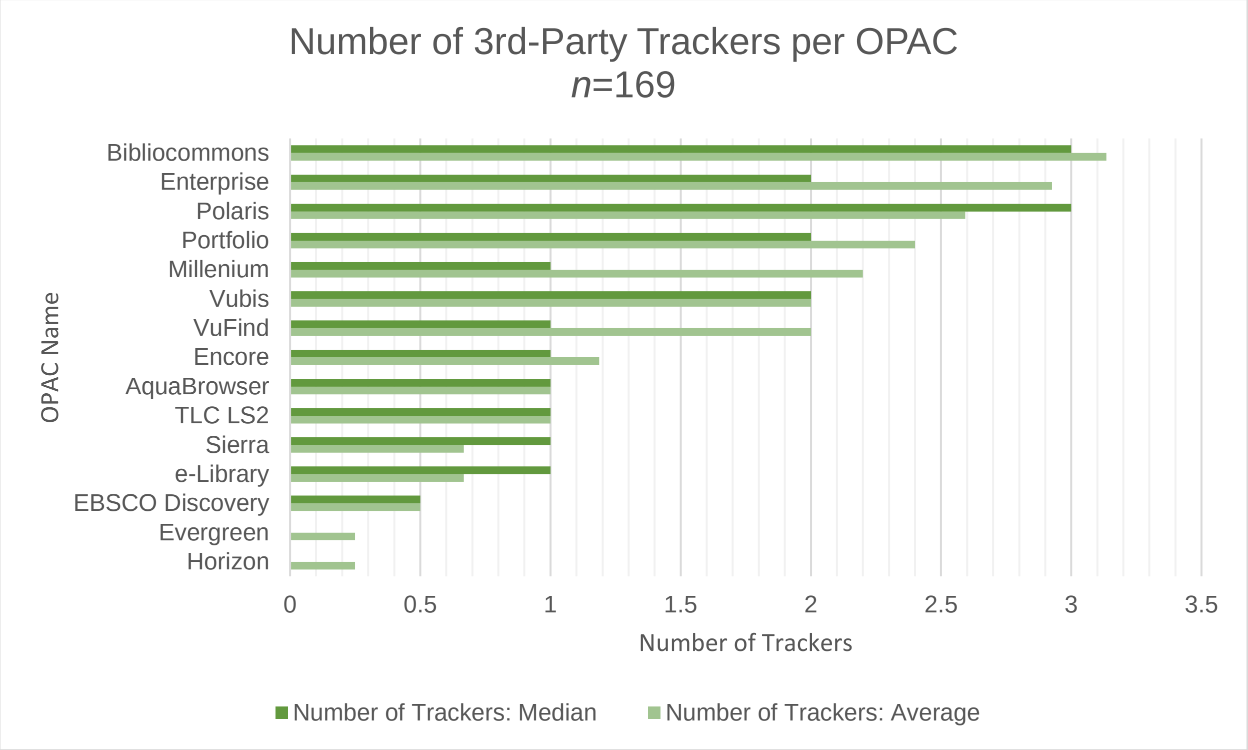
[10] There is debate about whether Adobe Typekit is actually "tracking". Adobe states that Typekit does not set any cookies in order to serve their fonts. http://blog.typekit.com/2013/06/13/clarifying-our-commitment-to-privacy/

Figure 8: Tracking in United States Libraries as Measured by Ghostery



Certain OPACS were associated with higher tracking counts. A majority of the libraries sampled (69%) used only 4 catalogs; the most used OPACs in the sample were Bibliocommons (37), Encore (32), Enterprise (27), and Polaris (27). The remaining 31% of the sample used 20 different catalogs between them. Bibliocommons, Enterprise, and Polaris usually had more trackers than other catalogs; it is unclear whether these 3 catalogs have more trackers enabled by default or whether their consistent association with higher tracker counts is a spurious relationship. Averages and medians were calculated for each catalog type that was represented more than once using the highest Ghostery or Disconnect counts. Nine catalogs appeared only once, providing insufficient data about which to generalize. Among libraries in the sample, the average number of trackers on Bibliocommons search results pages was 3.14; the average on Encore pages was 1.19; the average on Enterprise pages was 2.93, and the average on Polaris pages was 2.59. Tracker averages and medians are displayed below in Figure 9.

Figure 9: Number of 3rd-Party Trackers per Online Public Access Catalog



Common Content Sources

The presence of third-party tracking was common among the 10 common content sources examined. Only two sources, Chilton Auto Repair Manuals and NoveList did not contain third-party trackers. Among the remaining 8, there were some wide discrepancies between the total number of trackers recorded by Disconnect and the trackers recorded by Ghostery for the same URLs. Table 2 displays the total tracker counts for the examined content sources.

Table 2

Table 2

Incidence of 3rd-Party Tracking in Common Content Sources

n=10

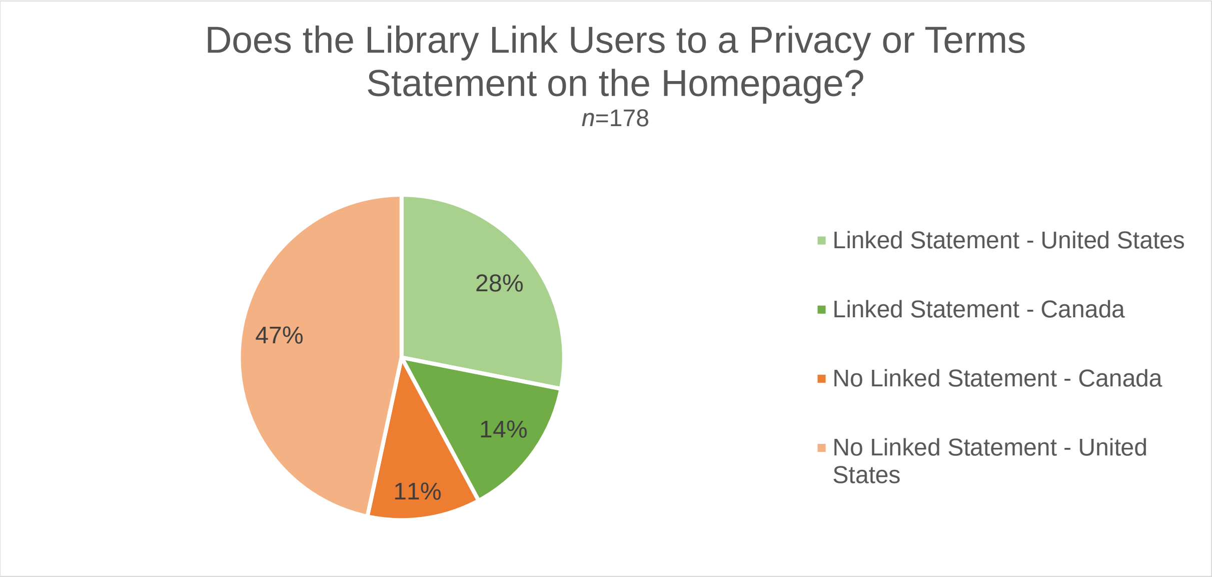
Common Content Source	**Ghostery Total Tracker Count**	**Disconnect Total Tracker Count**
Ancestry	3	11
Chilton	0	0
Consumer Reports	4	11
hoopla	4	3
Lynda	3	13
Mango Languages	3	1
NoveList	0	0
OneClickDigital	2	2
OverDrive	4	3
Zinio	3	2

Using the tracker categories delineated by Ghostery and Disconnect, the specific types of tracking enabled by these common content sources was apparent. Ghostery recorded 4 sources as allowing tracking by Facebook: Ancestry, hoopla, Lynda, and zinio. Disconnect recorded 3 sources as allowing Facebook tracking: Ancestry, hoopla, and Lynda. Both Ghostery and Disconnect recorded the same number of sources, 6, using Google Analytics: Consumer Reports, hoopla, Lynda, Mango Languages, OneClickDigital, and OverDrive. Some sites used additional analytics trackers, which included Optimizely, Inspectlet, webtrends, and New Relic.

Disclosure of Tracking

Libraries did not inform their users about third-party trackers at the same rate as tracking occurred. Fewer libraries were counted as potentially informing their users about third-party tracking than those that did potentially inform. Note that the *potential* of informing users is what was measured, not whether any linked policy statements or terms of service accurately disclosed the level of tracking. Of sample libraries in Canada, 25 presented a link on their homepage to some type of privacy statement or terms of service while 20 did not. Of sample libraries in the United States, 50 presented a link on their homepage to some type of privacy statement or terms of service while 83 did not. All 10 common content sources examined had links on their landing page leading to either a Privacy Statement or a Terms of Service statement. Figure 10 shows the percentages of CULC and ULC member libraries that link to a privacy or terms statement from their homepage.

Figure 10: Prevalence of a Link to a Privacy or Terms Statement From a Library's Homepage



For the 156 libraries in the sample that included third-party tracking scripts in their online catalogs, 69 (44%) included a link on their homepage potentially informing users of the presence of third-party tracking. (Above, it was reported that Ghostery counted only 154 libraries with third-party tracking; it is assumed for this portion of the analysis that Disconnect's detection of trackers in two additional libraries was accurate.) Thus, a majority of libraries allowing third-party tracking, 87 (56%), did not potentially inform their users about this tracking via an easily recognizable link on their homepage. Table 3 below displays the counts of third-party tracking and the counts of links to privacy or terms language for libraries in the sample. As described prior, eight of the 10 common content sources allowed third-party tracking; all 10 had a linked privacy or terms of service statement on their landing pages.

Table 3

Table 3

Prevalence of 3rd-Party Tracking and Homepage Links to Privacy or Terms Statements Among Libraries
n=178

Disclosure Potential	**3rd-Party Tracking**	**No 3rd-Party Tracking**
Link to Privacy or Terms Statement	69	6
No Link to Privacy or Terms Statement	87	16

HTTPS Usage by Default

Relatively few libraries are taking advantage of the security offered by HTTPS and continue to transmit their websites and catalogs to users over HTTP by default. In general, catalogs are more often served HTTPS, with 49% of Canadian libraries and 29% of United States libraries deploying their catalogs with the security offered by transport layer security (TLS). The number of libraries recorded using HTTPS by default for both their website and catalog was very small, just 2 Canadian libraries and 4 United States libraries offered this level of security to their patrons. One important point to stress here is that these results are only measures of the default behavior, i.e. when visiting a website with http:// in the URL in a vanilla browser, is it served up to the user in HTTP or in HTTPS? Libraries may offer HTTPS if the user adjusts the URL or is using a browser plugin such as HTTPS Everywhere; however if HTTP remains the default, many users will experience insecure connections. Results on HTTPS usage, broken down by country, are displayed in Table 4.

Table 4

Table 4

Default Usage of HTTPS When Visiting a Library Homepage and Searching Their Catalog

Location	Homepage HTTPS	OPAC HTTPS	Both Homepage & OPAC HTTPS	Neither Homepage nor OPAC HTTPS
Canada *n*=45	4 (8.9%)	22 (48.9%)	2 (4.4%)	21 (46.7%)
United States *n*=133	10 (7.5%)	38 (28.6%)	4 (3.0%)	89 (66.9%)

Note: Percentages are calculated relative to each country, not to the total sample size. Row totals do not sum to the sample size of each country because the ‘Both Homepage & OPAC HTTPS’ column double counts libraries also listed in the first two columns.

Discussion

Clearly, the results of this study demonstrate a gap between stated ethical theory and actual practice. Smith and Lyon have shown that despite their different political cultures, the similarities in how people in Canada and the United States experience surveillance are more salient than their differences (Smith and Lyon 2013). Given that, we might expect to see only minimal discrepancies between the prevalence of third-party tracking, the usage of HTTPS, and the accessibility of web privacy polices in their libraries. That is indeed what was found, though no tests of statistical significance are performed because those would not be relevant to the research question in this study, we see that Canada and the United States follow similar patterns and proportions along the various measures collected. Libraries in both countries typically have one to three third-party trackers in their web presence. The categories, defined by Disconnect, of “Google”, followed by “Facebook” in a distant second, followed by “Other” were proportionally the same in both countries. The Ghostery classification found similar results indicating that of those libraries enabling third-party tracking, most are doing so for internal site analytics reasons by way of Google Analytics and Google Tag Manager. Unfortunately, a small minority of libraries in both countries appear to enabling third-party tracking by Facebook on at least some portion of their web presence. In both countries, those lacking a link to a privacy or terms statement from a library's homepage were a minority.

An underappreciated aspect of recent development over the past decade or so is the rise of e-scores. Indeed, they've been dubbed "the new face of predictive analytics" by the American Marketing Association (Soat 2013). Cathy O'Neil has chronicled the various ways that e-scores are used; they can affect credit decisions (p.141-160), pricing by online retailers i.e. price discrimination (p. 189), auto insurance rates (p. 164-165) and of course the targeted advertising that individuals see (O'Neil 2017). By facilitating tracking, libraries are contributing information to models of the data brokers and online advertisers who compile e-scores. What is the average credit score of someone who searches the library catalog from home; how does that compare with those who search from inside the library? Are library patrons more or less likely to comparison shop online for car insurance? When libraries enable third-party tracking they are potentially allowing for the collection of information that might be incorporated into e-score models attempting to answer those very questions.

What else can libraries do to advance privacy? One compelling proposal from librarians at Cornell University is the abandonment of Google Analytics (Chandler and Wallace 2016). If Google Analytics is not abandoned entirely, libraries should configure it to maximally preserve such privacy as can be retained by: forcing a secure connection between the library website and Google's servers, implementing the Google Analytics IP Anonymization (aka IP-masking) feature, and following Google's best practices (O'Brien et al. 2018; Google n.d.). Libraries need pageview data in order to streamline and justify digital operations but there is no reason that libraries need to use the popular commercial product which by its own admission says personally identifiable information "is often inadvertently sent in these URLs" when functional alternatives exist (Google n.d.).

On the computers available for public use in-house, libraries could set an example of how to use the internet in a more secure and private way by defaulting their web search to a more private option such as DuckDuckGo (Ard 2016; Clark 2016). Along these lines, libraries should "harden" the browsers on those computers and block advertising and third-party tracking as numerous authors have suggested and provide clear documentation about why these measures have been taken (Phetteplace 2012; Gardner 2013; Gardner and Groover 2015; Clark 2016; Ayre 2017). To be specific, browsers should have the following add-ons/extensions installed: HTTPS Everywhere, one of either Privacy Badger, Disconnect, or Ghostery, and an ad-blocker such as uBlock or AdBlockPlus.[11] The presence of such software on public access computers should be supplemented with period public outreach and education, encouraging library patrons who wish to retain some degree of privacy outside of the library to install them on their own private devices (Ard 2016). One simple strategy would be to bundle such education into the existing Choose Privacy Week of the American Library Association, which already provides brief descriptions of the aforementioned tools and services in the paragraph as well as sample curricula (American Library Association n.d.; n.d.).

[11] The relative merits of each of these are beyond the scope of this paper.

Regarding the disclosure of tracking, Rebecca Miller has said it best: “We must inform people of what their daily options represent for their privacy in more meaningful language than a boilerplate click-thru agreement.” (Miller 2014) Yet at least as of 2017, 58% of libraries in the sample were failing to meet that standard. There is debate in the profession about the amount of data retention and tracking that needs to take place in order to maintain and improve virtual library services; with some arguing for a more “nuanced” approach (read: one in favor of collecting more granular data about individuals) (Varnum 2015). Yet even those individuals and organizations in favor of more tracking and data collection agree that tracking on library websites must take place with informed consent and the ability to opt-out (National Information Standards Organization 2015). The fact that so many libraries fail to make their privacy policies easily accessible should lead at least to some thoughtful self-reflection on the part of institutions and the individuals tasked with website design and maintenance; ideally, it would lead to possible public admonishment and concomitant behavior change.

As noted above, libraries which allow third-party tracking on their websites and which also fail to make their privacy policies discoverable and intelligible are in prima-facie breach of Principle III of the ALA Code of Ethics. The commercial vendors are not under the same legal or moral obligations as their customers. Ancestry, Consumer Reports, and Lynda were particularly egregious enablers of third-party tracking according to the Disconnect extension, see Table 2 above. If librarians are dissatisfied with that state of affairs, they could bring public pressure to bear on those vendors to change their practices or cease purchasing those products altogether. At minimum, libraries should be encouraging awareness of the fact that when patrons leave the confines of the library’s website, even for services the library purchases and promotes, they are subject to less stringent privacy regimes. After their ordeal with the civil grand jury, Santa Cruz Public Libraries have been a model of this behavior and developed a privacy policy page listing each of their commercial databases, the data kept by each vendor, and whether the library can see patrons’ personally identifiable information (Santa Cruz Public Libraries n.d.).

The question of HTTPS usage is a simple one: all pages under a library’s control should be served securely using HTTPS by default. Admirable work on securing websites has been underway for a few years now in the form of the Let’s Encrypt initiative of the Internet Security Research Group which is endorsed by big Silicon Valley firms. There is every reason to believe that the snapshot picture captured in time in this study (and others) has improved recently (O’Brien et al. 2018; Hellman, Eric 2016; Breeding 2016). But we as a profession should not be too quick to pat ourselves on the back considering that offering a secure connection is only one aspect of our ethical obligations in the privacy hierarchy. Furthermore, some catalogs are hosted by vendors and the libraries may not have direct control over the usage of TLS. Similarly, it is not uncommon for public libraries to have their webpage hosting taken care of by a larger governmental entity such as a city or county. In these circumstances, we should not rush to condemn them, as there are larger bureaucratic forces that may be impeding the deployment of HTTPS.

Larger institutions may also have control of the type of web analytics that a library uses on their website. In particular, Google Analytics is the de facto industry standard and commonplace on city government and university websites, including their library sub-directories, a fact that librarians use to their advantage for research and evaluative purposes (Vecchione et al. 2016;

Hess 2012). Ultimately, given the fact that HTTPS was formally specified in the year 2000 and yet in 2017 47% of Canadian libraries and 67% of United States libraries sampled still failed to default to HTTPS on their website and their OPAC (see: Table 4) suggests that much work remains to be done ("HTTPS" 2020).

Limitations

There are several limitations that apply to this study and its results. First, the sample population of public libraries in the Canadian Urban Libraries Council and Urban Libraries Council is adequate to generate conclusions about the types of tracking that goes on for users of public libraries in *urban* areas within Canada and the USA. In principle, all the issues raised regarding tracking and usage of HTTPS apply to *rural* libraries as well. It is possible that rural populations are even more dependent upon libraries for internet access. Because of lower population density, rural populations may even be more susceptible to targeted third-part tracking when using a library website or catalog from their own home or smartphone. However, because they are out-of-sample, no conclusions should be drawn about the level of tracking that rural libraries are enabling. Similarly, no one should generalize these results to the European Union context, which has more stringent privacy regulations for third parties (Linden et al. 2020). A second limitation is what constitutes "tracking" by third parties. There is ongoing debate regarding the mechanisms of tracking and the type and amount of data that is collected. Some of this debate is manifest in the way that Ghostery and Disconnect (or other similar web browser plugins such as EFF's Privacy Badger) come up with different tracker counts. Rather than wade into the definition morass, this methodology takes the different definitions and approaches given and reports both results.

Notably, the measure of tracking disclosure is crude. Some libraries may disclose tracking in depth but have the information buried several clicks into their website. These institutions would fail the measure. Other libraries may have a link to a privacy policy or terms of use on the landing page, but it may not accurately disclose the extent of third-party tracking that is happening. Since the accuracy of each library's privacy policy or terms was not verified and only the mere presence of such a document was recorded, these institutions would pass the measure.[12] Best practices are that every library should make their privacy policy and/or terms of use easily available from the landing webpage in order to allow visitors to examine that information as soon as possible should they be interested in doing so (American Library Association Office for Intellectual Freedom 2014). It is left to the reader to judge their risks and potential payoffs. Another important limitation of the disclosure data is that this study was conducted prior to the European Union's General Data Protection Regulation coming widely into effect. In compliance with those rules, it is now commonplace for users on commercial websites to be presented with web cookie or other tracking notification (Sørensen and Kosta 2019). Since users are in the process of being habituated to clicking through these notices at commercial websites, perhaps libraries may need to adjust their approach.

[12] Readers interested in pursuing this research question are urged to consult Gallagher, McMenemy, and Poulter who studied acceptable use policies with an innovative methodology (Gallagher, McMenemy, and Poulter 2015). While the privacy policies of vendors are beyond this paper, there is evidence that they fail to meet the expectations of librarians (Lambert, Parker, and Bashir 2015).

One final limitation of importance pertains to the audit of HTTPS usage on library websites and online public access catalogs. When those data were recorded in 2017 the Let's Encrypt certificate authority had launched the previous year and was gaining momentum. Let's Encrypt provides free TLS/SSL certificates in an automated user-friendly way and at time of writing has been instrumental in bringing security to over 100 million website domains (Let's Encrypt n.d.). It is very likely that some OPAC vendors, local governments, or individual public libraries themselves took advantage of Let's Encrypt's services to secure their sites since data collection. Additionally, dramatic data breaches and hacks, which have received widespread media coverage since 2017, have increased public and governmental awareness of website security and privacy issues.

Further Research

The data presented above is a static snapshot. But the internet, and the third-party tracker ecosystem, is continually evolving. One clear opportunity for further research would be to follow the incidence of third-party tracking longitudinally. A better picture of the user base affected would be captured by also studying libraries that serve rural populations. Additionally, qualitative research should be done on the accuracy of library privacy policies and terms of use. We still do not know how many libraries disclose the number of third-party trackers their users are subjected to or the accuracy of these disclosures. Do they explicitly list the firms that developed these trackers and what the firms claim to be able to legally do with the information they collect? Furthermore, what benefits do libraries derive from including third-party trackers on their websites and catalogs? Web traffic analytics are certainly useful in library administration, but it is possible to collect, store, and analyze these in-house using tools such as Matomo or Open Web Analytics. Some added benefit must be derived from relying on a commercial tool like Google Analytics, perhaps the ease of use and lack of subscription or hosting costs is why so many libraries have adopted it. Moving beyond traffic statistics, why are libraries allowing any other third-parties to track their users? Presumably, they gain operational benefits from this sacrifice of patron privacy. But we do not know their explicit motivations or whether libraries are even aware of the fact that the gains from third-party tracking explicitly force a reduction in privacy. More research into the tradeoffs of privacy for other benefits, and the motivations of the librarians who accept these tradeoffs, is required.

Conclusion

The results of this study, in concert with other recent evidence reviewed above demonstrate that many public libraries in Canada and the United States are not doing all that they can to protect the privacy of their patrons in digital environments that they ostensibly control. Librarians are marching in a compact group along a precipitous and difficult path, in which our strongest allies are each other. We are surrounded on most sides by vendors who, given the opportunity, would place our users under their almost constant surveillance. While many libraries may have bureaucratic impediments to offering their patrons a secure and private (with respect to third-parties) online experience, others are not blameless. Libraries must take active steps to secure their websites and catalogs using TLS/SSL. If libraries believe that third-party tracking provides more benefits to their operations than the cost incurred in loss of privacy, this must be disclosed to their users. If libraries opt to allow third-party tracking, their on-site computers should allow the option of viewing the website and querying the catalog using a browser that is configured to block all such tracking. Anything less than these measures is a failure to live up to the relevant clauses of the ALA Code of Ethics and a betrayal of user trust. It is particularly necessary to arouse in all who participate in the practical work of library website maintenance discontent with the status quo prevailing among librarians and an unshakable determination to rid ourselves of it. Hopefully this research will raise awareness about the level of security and privacy that Canadian and American urban library patrons are afforded and spur action.

References

Acar, Gunes, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. “The Web Never Forgets: Persistent Tracking Mechanisms in the Wild.” In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 674–689. CCS ’14. New York, NY, USA: ACM. https://doi.org/10.1145/2660267.2660347.

Albrecht, Katherine, and Liz McIntyre. 2014. “How and Why to Keep the NSA Out of Your Private Stuff - Even If You’ve ‘Got Nothing to Hide.’” *IEEE Technology & Society Magazine* 33 (4): 39–41. https://doi.org/10.1109/MTS.2014.2369571.

American Library Association. n.d. “Programs.” Choose Privacy Every Day. Accessed April 18, 2020a. https://chooseprivacyeveryday.org/programs/.

———. n.d. “Resources.” Choose Privacy Every Day. Accessed April 18, 2020b. https://chooseprivacyeveryday.org/resources/.

American Library Association Office for Intellectual Freedom. 2014. “Privacy Tool Kit.” Text. ALA Privacy Tool Kit. January 2014. http://www.ala.org/advocacy/privacy/toolkit.

Ard, BJ. 2016. “Librarians as Privacy Advocates.” *I/S: A Journal of Law and Policy for the Information Society* 13 (1): 161–74.

Assange, Julian. 2014. “Julian Assange on Living in a Surveillance Society.” *The New York Times*, December 4, 2014. http://www.nytimes.com/2014/12/04/opinion/julian-assange-on-living-in-a-surveillance-society.html.

Ayre, Lori Bowen. 2017. “Protecting Patron Privacy: Vendors, Libraries, and Patrons Each Have a Role to Play.” *Collaborative Librarianship* 9 (1): 1–5.

Barron, Simon, and Andrew J. Preater. 2018. “Critical Systems Librarianship.” In *The Politics of Theory and the Practice of Critical Librarianship*, 87–113. Sacramento, CA: Library Juice Press. https://repository.uwl.ac.uk/id/eprint/4512/.

Breeding, Marshall. 2016. “Protecting Patron Privacy.” *American Libraries Magazine*, May 31, 2016.

Chandler, Adam, and Melissa Wallace. 2016. “Using Piwik Instead of Google Analytics at the Cornell University Library.” *The Serials Librarian* 71 (3–4): 173–79. https://doi.org/10.1080/0361526X.2016.1245645.

Clark, Ian. 2016. “The Digital Divide in the Post-Snowden Era.” *Journal of Radical Librarianship* 2 (0). https://journal.radicallibrarianship.org/index.php/journal/article/view/12.

Cranor, L. F. 2012. “Can Users Control Online Behavioral Advertising Effectively?” *IEEE Security Privacy* 10 (2): 93–96. https://doi.org/10.1109/MSP.2012.32.

Disconnect. n.d. “Disconnect.” Accessed April 7, 2020. https://disconnect.me/about.

Englehardt, Steven, and Arvind Narayanan. 2016. “Online Tracking: A 1-Million-Site Measurement and Analysis.” In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS’16*, 1388–1401. Vienna, Austria: ACM Press. https://doi.org/10.1145/2976749.2978313.

Esposti, Sara Degli. 2014. “When Big Data Meets Dataveillance: The Hidden Side of Analytics.” *Surveillance & Society* 12 (2): 209–25. https://doi.org/10.24908/ss.v12i2.5113.

Gallagher, Christine, David McMenemy, and Alan Poulter. 2015. “Management of Acceptable Use of Computing Facilities in the Public Library: Avoiding a Panoptic Gaze?” *Journal of Documentation* 71 (3): 572–90. https://doi.org/10.1108/JD-04-2014-0061.

Gardner, Gabriel J. 2013. “Simple Tools to Refract PRISM in Your Library.” *Minitex Reference Notes*, August 2013.

Gardner, Gabriel J., and Myron Groover. 2015. “Web Privacy in Practice: Assessing Internet Security and Patron Privacy in North American Public Libraries.” Presented at the 2015 LITA Forum, Minneapolis, MN, November 12. https://macsphere.mcmaster.ca/handle/11375/19016.

Ghostery. n.d. “About Ghostery.” Ghostery. Accessed April 7, 2020. https://www.ghostery.com/about-ghostery/.

Google. n.d. “Best Practices to Avoid Sending Personally Identifiable Information (PII).” Google Analytics Help. Accessed April 10, 2020. https://support.google.com/analytics/answer/6366371.

Griffey, Jason. 2016. “Keep On Rockin’ In The Free World.” Keynote presented at the Lake Superior Libraries Symposium, Duluth, MN, May 20. https://speakerdeck.com/griffey/privacy-and-libraries.

Hanson, Cody. 2019. “User Tracking on Academic Publisher Platforms.” Presented at the Coalition for Networked Information Spring 2019 Member Meeting, St. Louis, Missouri, April 8. https://www.codyh.com/files/HansonCNISpring19.pdf.

Hellman, Eric. 2016. “97% of Research Library Searches Leak Privacy... and Other Disappointing Statistics.” *Go To Hellman* (blog). May 23, 2016. http://go-to-hellman.blogspot.com/2016/05/97-of-research-library-searches-leak.html.

Hess, Kirk. 2012. “Discovering Digital Library User Behavior with Google Analytics.” *Code4Lib Journal*, no. 17 (June). https://journal.code4lib.org/articles/6942.

Hoofnagle, Chris Jay, Ashkan Soltani, Nathaniel Good, Dietrich J. Wambach, and Mika D. Ayenson. 2012. “Behavioral Advertising: The Offer You Cannot Refuse.” *Harvard Law & Policy Review* 6 (2): 273–96.

“HTTPS.” 2020. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=HTTPS&oldid=949480078.

Huey, Laura, Micheal Vonn, Reg Whitaker, Paul Rosenzweig, danah boyd, Steven Margulis, Gary Marx, and Judith Rauhofer. 2012. “The Future of Privacy Online.” *Surveillance & Society* 10 (3/4). https://doi.org/10.24908/ss.v10i3/4.4551.

Lambert, April D., Michelle Parker, and Masooda Bashir. 2015. “Library Patron Privacy in Jeopardy an Analysis of the Privacy Policies of Digital Content Vendors.” *Proceedings of the Association for Information Science and Technology* 52 (1): 1–9. https://doi.org/10.1002/pra2.2015.145052010044.

Lamdan, Sarah Shik. 2015. “Social Media Privacy: A Rallying Cry to Librarians.” *Library Quarterly* 85 (3): 261–77.

Lerner, Adam, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. “Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016.” In *Proceedings of the 25th USENIX Security Symposium*, 997–1013. Austin, TX. https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_lerner.pdf.

Let’s Encrypt. n.d. “Let’s Encrypt Stats.” Let’s Encrypt - Free SSL/TLS Certificates. Accessed April 10, 2020. https://letsencrypt.org/stats/.

Linden, Thomas, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. 2020. “The Privacy Policy Landscape After the GDPR.” *Proceedings on Privacy Enhancing Technologies* 2020 (1): 47–64. https://doi.org/10.2478/popets-2020-0004.

Macrina, Alison. 2015. “Why We Need to Encrypt The Whole Web... Library Websites, Too.” *LITA Blog* (blog). January 27, 2015. http://litablog.org/2015/01/why-we-need-to-encrypt-the-whole-web-library-websites-too/.

Madrigal, Alexis C. 2012. “I’m Being Followed: How Google—and 104 Other Companies—Are Tracking Me on the Web.” *The Atlantic*, February 29, 2012. http://www.theatlantic.com/technology/archive/2012/02/im-being-followed-how-google-151-and-104-other-companies-151-are-tracking-me-on-the-web/253758/.

Mathur, Arunesh, Jessica Vitak, Arvind Narayanan, and Marshini Chetty. 2018. “Characterizing the Use of Browser-Based Blocking Extensions To Prevent Online Tracking.” In , 103–16. https://www.usenix.org/conference/soups2018/presentation/mathur.

Miller, Rebecca T. 2014. “Getting Real About Privacy: Confidentiality, Digital Literacy, and beyond | Editorial.” Library Journal. December 2, 2014. http://lj.libraryjournal.com/2014/12/opinion/editorial/getting-real-about-privacy-confidentiality-digital-literacy-and-beyond-editorial/.

Morrone, Melissa. 2015. “Privacy Matters: Anti-Surveillance Education in the Library.” Metropolitan New York Library Council. January 28, 2015. http://metro.org/articles/antisurveillance-education-in-the-library/.

Mozilla. 2017. “Hackers, Trackers and Snoops: Our Privacy Survey Results.” *Medium* (blog). March 9, 2017. https://medium.com/@mozilla/hackers-trackers-and-snoops-our-privacy-survey-results-1bfa0a728bd5.

National Information Standards Organization. 2015. “NISO Consensus Principles on Users’ Digital Privacy in Library, Publisher, and Software-Provider Systems (NISO Privacy Principles) | NISO Website.” 9781937522704. Baltimore, MD: National Information Standards Organization (NISO). https://www.niso.org/publications/privacy-principles.

O’Brien, Patrick, Scott W.H. Young, Kenning Arlitsch, and Karl Benedict. 2018. “Protecting Privacy on the Web: A Study of HTTPS and Google Analytics Implementation in Academic Library Websites.” *Online Information Review* 42 (6): 734–51. https://doi.org/10.1108/OIR-02-2018-0056.

O’Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. First Paperback Edition. New York: B/D/W/Y Broadway Books.

Phetteplace, Eric. 2012. “Hardening the Browser.” *Reference & User Services Quarterly* 51 (3): 210–14.

Privacy International. 2018. “Examples of Data Points Used In Profiling.” London, United Kingdom: Privacy International. https://privacyinternational.org/sites/default/files/2018-04/data%20points%20used%20in%20tracking_0.pdf.

Radical Reference. 2014. “We Are All Suspects: A Guide for People Navigating the Expanded Powers of Surveillance in the 21st Century.” Radical Reference. http://radicalreference.info/content/we-are-all-suspects-guide-people-navigating-expanded-powers-surveillance-21st-century.

Rankin, Kyle. 2014. “NSA: Linux Journal Is an ‘Extremist Forum’ and Its Readers Get Flagged for Extra Surveillance | Linux Journal.” *Linux Journal* (blog). July 3, 2014. http://www.linuxjournal.com/content/nsa-linux-journal-extremist-forum-and-its-readers-get-flagged-extra-surveillance.

Santa Cruz Public Libraries. n.d. “Data Privacy.” Santa Cruz Public Libraries. Accessed April 19, 2020. https://www.santacruzpl.org/data_privacy/.

Schaub, Florian, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. 2016. “Watching Them Watching Me: Browser Extensions Impact on User Privacy Awareness and Concern.” In *Proceedings 2016 Workshop on Usable Security*. San Diego, CA: Internet Society. https://doi.org/10.14722/usec.2016.23017.

Smith, Emily, and David Lyon. 2013. “Comparison of Survey Findings from Canada and the USA on Surveillance and Privacy from 2006 and 2012.” *Surveillance & Society* 11 (1/2): 190–203. https://doi.org/10.24908/ss.v11i1/2.4517.

Soat, Molly. 2013. “Incomplete Insights. (Cover Story).” Marketing News 47 (5): 32–37.

Sørensen, Jannick, and Sokol Kosta. 2019. “Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites.” In *The World Wide Web Conference*, 1590–1600. WWW ’19. San Francisco, CA, USA: Association for Computing Machinery. https://doi.org/10.1145/3308558.3313524.

Sullivan, Mark. 2012. “Data Snatchers!” *PCWorld* 30 (8): 77–85.

Tate, Barton Gellman, Julie, and Ashkan Soltani. 2014. “In NSA-Intercepted Data, Those Not Targeted Far Outnumber the Foreigners Who Are.” *The Washington Post*, July 5, 2014. http://www.washingtonpost.com/world/national-security/in-nsa-intercepted-data-those-not-targeted-far-outnumber-the-foreigners-who-are/2014/07/05/8139adf8-045a-11e4-8572-4b1b969b6322_story.html?hpid=z1.

Uzunoglu, Doruk. 2016. “Understanding Ad Blockers.” Worcester Polytechnic Institute. https://www.wpi.edu/Pubs/E-project/Available/E-project-032216-001707/unrestricted/dcuzunoglu_understanding_ad_blockers.pdf.

Varnum, Ken. 2015. “Editorial Board Thoughts: Library Analytics and Patron Privacy.” *Information Technology & Libraries*, December 2015. https://doi.org/10.6017/ital.v34i4.9151.

Vecchione, Amy, Deana Brown, Elizabeth Allen, and Amanda Baschnagel. 2016. “Tracking User Behavior with Google Analytics Events on an Academic Library Web Site.” *Journal of Web Librarianship* 10 (3): 161–75. https://doi.org/10.1080/19322909.2016.1175330.

York, Jessica A. 2019. “Grand Jury: Santa Cruz Libraries Should Better Protect Patron Privacy.” *Santa Cruz Sentinel*, June 30, 2019. https://www.santacruzsentinel.com/grand-jury-santa-cruz-libraries-should-better-protect-patron-privacy.

Zetter, Kim. 2014. “The NSA Is Targeting Users of Privacy Services, Leaked Code Shows.” WIRED. July 3, 2014. http://www.wired.com/2014/07/nsa-targets-users-of-privacy-services/.

Zimmer, Michael. 2014. “Librarians’ Attitudes Regarding Information and Internet Privacy.” *Library Quarterly* 84 (2): 123–51. https://doi.org/10.1086/675329.

Appendix

Reliability of Tracking Counts

Despite the fact that both Ghostery and Disconnect purport to use the same general methods of detecting third-party trackers, i.e. examining JavaScript elements in a loaded web page, there were occasionally discrepancies between their results. Previous research on ad-blocking software has noted differences between the two add-ons with regard to their handling of third-party tracking and advertising. Doruk Uzunoglu examined the filter lists these various ad-blockers rely on to determine if a tag or script is benign or somehow involved in tracking or advertising. Uzunoglu outlined six categories of domains listed on popular filter lists: AdTrackers, Analytics, Beacons, Other Third-Parties, Social, and Widgets; the results showed that the default Ghostery settings consistently block either more than or the same amount of domains compared to the default Disconnect settings (Uzunoglu 2016).

Our analysis found that more often than not, Ghostery and Disconnect recorded the same results when loading the same URLs. The same total number of third-party trackers results were obtained by both add-ons for 60.1% of URLs examined in this study. Disconnect recorded more trackers 32.45% of the time; Ghostery recorded more for the remaining 7.45%. When differences between the Ghostery and Disconnect counts were noted, more often than not they were small. For 75 sample members, there were discrepancies between the two tracker counts, in 63 of those instances (84%), the difference was 2 or less. There were, however, some extreme outlier cases: Figure A1 displays the frequency of each discrepancy count between the trackers. Due to the aforementioned discrepancies, counts for both trackers are reported separately in the results rather than calculating a composite tracker count.

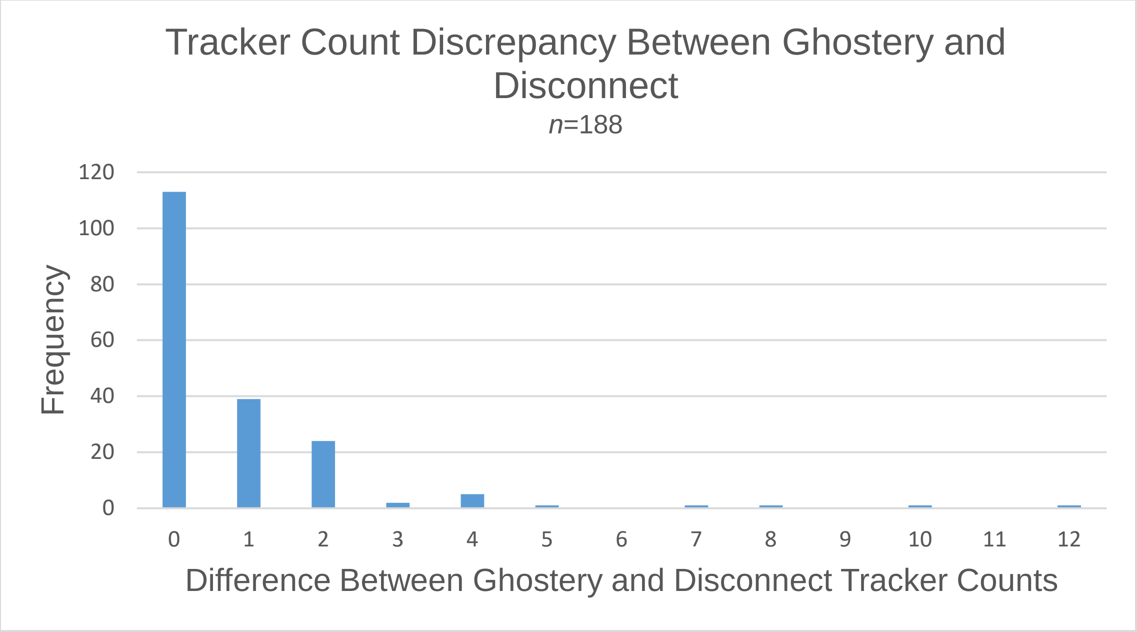


Figure A1: Tracker Count Discrepancy between Ghostery and Disconnect