FEATURE

Who's Reading Who?

Exploring the State of Third-Party Tracking Technology in Open Access Journal Content

Author _ Bill Marino (wmarino1@emich.edu), Associate Professor, Online Learning Librarian, Eastern Michigan University

The shift in scholarly communication from print to electronic format—and from ownership to access-based delivery models—has changed the dynamics of control related to library collections (Breeding 2019; Singley 2020). External providers now deliver most scholarly content to users, reducing library control to primarily print collections. User data gained through tracking technology, which often collects information without a user's knowledge or agency, is a product of this shift that has been identified by previous studies (Hinchliffe, Zimmerman, and Altman 2018; Hanson 2019). This tracking data, aggregated by third parties who use it to form user profiles that can then be commodified, "challenges libraries' historical assumptions about privacy and anonymity" (Hanson 2019, under "Aggregated Identities").

Particularly troubling is the presence of tracking technology in open access literature (Hinchliffe, Zimmerman, and Altman 2018), which the Budapest Open Access Initiative hailed as a public good that promises "completely free and unrestricted" peer-reviewed research (Chan et al. 2002). Like Barbrook and Cameron's (1996) Californian Ideology, which describes tension in the emerging internet between, on the one hand, a virtual community that exhibits a free exchange of information and ideals, and on the other, a neoliberal inspired electronic marketplace, open access literature is pulled between two opposing sides, with one firmly anchored in the ideals that publicly funded research should be available to all people regardless of economic or social status (Chan et al. 2002; Stebbins 2013; Science Europe n.d.), and the other operating within the internet's market ecosystem. While the Budapest Open Access Declaration acknowledged the need for novel funding models to support costs incurred by publication, the presence of tracking technology raises the question of whether we have merely shifted the product, selling the reader rather than the content. If this is the case, it would run counter to both the American Library Association's (ALA) and International Federation of Library Associations and Institutions' (IFLA) stance on user privacy rights (American Library Association 2017; 2019; International Federation of Library Associations and Institutions 2012; 2014).

This study further explores the prevalence of third-party tracking in open access literature, focusing on

English-language publications in the Directory of Open Access Journals (DOAJ). It builds on the work of previous studies to consider what parties are doing the tracking and for what purposes, but also introduces simple mathematical models from privacy studies in computer science (Englehardt and Narayanan 2016; Yu et al. 2016; Karaj et al. 2019) to determine the invasiveness of scripts based on their behavior in the real world. This is achieved through considering how often they are encountered and how often they are likely to gather data that can identify individual users. In doing so, it identifies a tracker profile that appears unique to open access literature and considers how much user data is forfeited.

Literature Review

Corporate Influence and the Internet

Concern over corporate control of the internet has a long history. Borgman (2015) points to policy changes in the early 1990s that led to what she calls "the commodity internet," replete with "new business models, shifts in the balance of stakeholders, and unforeseen challenges of security and privacy [that] are contributing to redesign of the infrastructure" (46). The rise of the corporation led to the individual being labeled as "samples, data, markets, or 'banks'" (Deleuze 1992, 5), and established roles of producer and consumer became clouded as consumers began contributing to social processes surrounding products (Lazzarato 1996). Schiller (1996) warned of a world in which corporate interests usurp the traditional role of social and civic institutions, threatening "the public good" and transforming communications channels into marketing tools whose ultimate purpose is sustained economic growth, a concern that is also reflected in Barbrook and Cameron's (1996) Californian Ideology.

These concerns are still alive today. In higher education, they exist in a large body of literature that explores corporate influence in the academy through the adoption of various educational technology products and the privacy ramifications associated with their use. Common themes that emerge are outdated institutional interpretations of privacy, the ever-changing nature of data, and the ability of corporations to gain increased access to student information and operate without transparency or liability (Brown and Klein 2020; Jones et al. 2020; Paris, Reynolds, and McGowan 2021). Parks (2017) described an environment where "participation in the system of higher education in the United States now implicitly requires that students consent to sharing their personal information with third parties with little transparency or control over their own information" (27).

Library literature also shows concern with corporate influence, as several studies have focused on the implications of the use of third-party resources, particularly the use of products within Google's ecosystem. O'Brien et al. (2018) investigated the use of Google Analytics products on academic library sites, finding that most implemented Google Analytics or Google Tag Manager, yet few used connection or security features to protect user data (741–42). Breeding (2019) warned libraries that the "basis of Google Analytics in the commercial advertising ecosystem" may not be compatible with privacy policies (12), and other studies echo this concern, with Chandler and Wallace (2016) and Quintel and Wilson (2020) exploring the use of Piwik and Matomo, open-source alternatives to Google Analytics.

A theme found in library privacy literature is the need to balance innovation of services with professional ethics. This is true of library efforts to participate in campus learning analytics initiatives, where the possibility to measure library impact on student success and secure funding must be weighed against preserving patron privacy (Hartman-Caverly 2019; Hwang and Hanson 2021; Jones 2019; Jones et al. 2020; Oakleaf 2016; 2018; Oliphant and Brundin 2019; Selwyn 2019; Slade and Tait 2019; Travis and Ramirez 2020). However, there is evidence that, like the academy, many libraries are ill-prepared for this task.

A pair of studies (Zimmer 2014; Tummon and McKinnon 2018) explored perspectives on privacy among US and Canadian librarians and found that most librarians are concerned with the amount of personal data that is collected by companies and the government. However, Tummon and McKinnon (2018) also noted that "librarians are alarmingly unaware of the practices at their own libraries regarding online and patron privacy" (92), and Hanson (2019) warns that libraries are operating on a dated definition of personally identifiable information (PII). Considering recent legislative efforts at the state level that modify the definition of PII (e.g., the California Consumer Privacy Act of 2018 and Illinois Biometric Information Privacy Act), it becomes increasingly important that librarians remain educated and engaged on this topic.

Libraries: Tracking the Trackers

Other studies in the library literature have begun to explore third-party tracking on library sites and resources. Breeding (2016) explored websites from ARL libraries as well as the 25 largest public libraries in the United States for instances of tracking scripts, and Marino (2021) examined ARL library homepages for third-party tracking cookies. Both noted a high prevalence of tracking software on library websites.

Two studies explored the presence of tracking technology in scholarly literature. Hinchliffe, Zimmerman and Altman

(2018) analyzed tracking technology on publisher sites, performing a comparison of EBSCO, EBSCO Open, Elsevier, ProQuest, ProQuest Open Dissertation, Springer LINK, Springer Open, Taylor & Francis, and Wiley, in terms of encryption (HTTPS), ad placement, external sources, cookies, reader apps, and fingerprinting. Hanson (2019) analyzed the source code of the most frequently accessed article from 15 publisher platforms at the University of Minnesota for instances of third-party assets loaded on the page. Widespread tracking was noted in both open access and subscription-based resources, and Hinchliffe, Zimmerman, and Altman (2018) concluded that laws and licenses were ineffective in protecting user privacy. Universally, these studies call for additional work that considers the implications of third-party influence on the privacy of library users.

The Growth of Open Access Literature

Open access publications have grown in both volume and impact over the past two decades. Piwowar et al (2018) found that 44.7% of all articles published in 2015 were open access, and multiple studies have shown that open access articles exhibit a citation advantage over subscription-based alternatives (Piwowar et al. 2018; Breugelmans et al. 2018; Arendt, Peacemaker, and Miller 2019). However, Hinchliffe, Zimmerman, and Altman (2018) noted the less favorable side of open access, identifying third-party tracking technology on open access material available from large publisher platforms.

The Behavior of Trackers

Computer science literature has provided a tool set that can be used for further exploration of tracking scripts' behavior. This work focuses on analysis of the trackers themselves (Englehardt and Narayanan 2016; Yu et al. 2016; Karaj et al. 2019). Englehardt and Narayanan (2016) noted a long tail, where relatively few trackers perform most of the user tracking. Yu et al. (2016) identified problems with the traditional domain-level approach to tracker blocking—where all scripts originating from a particular domain are blocked—noting that tracking scripts exhibit "mixed behavior." They do not pass information that can identify a particular user every time they are loaded. Karaj et al. (2019) identify the need to analyze trackers in relation to user's real-world behavior, opting to measure tracker reach across the browsing history of more than 5 million users representing "multiple countries, ISPs, and browser configurations" (1). Their work has led to the WhoTracks.Me website, which provides detailed monthly data on a tracker's reach and tracking frequency that can be used in targeted analyses.

Research Questions

This study builds on previous attempts in the library literature (Hinchliffe, Zimmerman, and Altman 2018; Hanson 2019) to catalog third-party tracking, expanding to include mathematical concepts from other privacy-related disciplines and data from the Ghostery/WhoTracksMe database. It concentrates specifically on English-language open access material from the DOAJ, considering the following questions:

- How prevalent is third-party tracking in open access journals?
- What parties are doing the tracking? What purposes do they serve?
- How invasive are the tracking scripts?
- Are user autonomy options available? Do they affect third-party tracking?

Methods

Hanson's (2019) methods for discovering and logging tracker resources served as a basis for the collection methods used in this study. Data was collected between May 24 and June 18, 2021 using Chrome versions 90.0.4430.212 and 91.0.4472.114 for Mac (Chrome 91 was released on May 25 with no substantial changes that could affect this study—see Bommana 2021; LePage 2021). Chrome was selected because it is the most popular browser worldwide, holding a 64.75% market share as of May 2021 (StatCounter n.d.), and it has no built-in cookie or script blocking features at the time of the study. Unlike Hanson, who used the Ghostery plug-in, Privacy Badger was selected to help with identification of tracking code that was loaded. It does allow a user to opt-into a "heuristics" learning mode that analyzes sites for tracking activity and adapts to meet an individual's browsing habits (Arrieta et al. 2020), but the default (non-heuristics) mode was used for this study.

The researcher downloaded journal metadata dated May 21, 2021 from the DOAJ. A set consisting of English language content from journals bearing the DOAJ seal (see https://doaj.org/apply/seal/) that had published more than one article in the last five years was identified. This resulted in 1323 journals published by 140 publishers. The decision was made to analyze content at the publisher level, as it was noted that most publishers use a single platform across their journal content. For each publisher, the researcher accessed the first journal listed in the DOAJ metadata record using the URL supplied in its entry. The researcher then selected one article from each journal to test, which was visited using a clean Chrome browser (all browsing data and cookies

cleared) with the Privacy Badger extension (version 2021.2.2) installed—no other extensions were present.

The researcher logged the following information in a spreadsheet:

- Article information (i.e., Publisher name, Journal title, URL, etc.),
- Whether users had any control over their privacy preferences (if the user was provided with a choice, the collection process was repeated twice for those articles—one page with tracking rejected and a second with it accepted),
- if user autonomy functions existed, whether they were opt-in or opt-out in nature,
- the number of trackers that Privacy Badger blocked for a given page,
- the blocked script's domain/subdomain address.

To identify the entity behind each blocked script, the study turned to Karaj et al (2019), running its address through Ghostery's WhoTracksMe site (https://whotracks.me). This step allows domains/subdomains that serve the same tracking script to be aggregated, ensuring that if multiple instances of a tracker are loaded on a single page, they are only counted once. For example, a page could load the same Twitter tracker script from several Twitter-owned subdomains (e.g., twitter.com, t.co, twimg.com, pbs.twimg.com, cdn.syndication.twimg.com, ton.twimg.com, etc.), but all would be counted one single time. Tracker purpose was determined by the WhoTracksMe's purpose categories (see https://whotracks.me/blog/tracker_categories.html for category definitions).

Next, the site reach value was determined for each tracker, third-party that set tracking scripts, and purpose category across all pages analyzed using Karaj et al.'s (2019) formula:

Site reach = [unique pages where variable was seen]/[unique pages]

This value between 0 and 1 (with a value of 1 meaning that it was present on all pages) establishes the frequency with which a particular variable occurred across the sample set. It also provides a value that can be compared against Ghostery's cumulative site reach value for the top 10,000 websites in June 2021 (*Cumulative site reach = [Tracker site reach top 10k (Ghostery)]/10,000*) to indicate whether our sample of open access articles are indicative of web content in general.

To gauge invasiveness of a tracker, this study introduces the invasiveness product, which allows the researcher to estimate how many times a tracker is engaged in tracking specific individuals based on the sample set size. It is calculated by multiplying a tracker's measured site reach value by its utilized tracking content value for June 2021. The utilized tracking content value is available via the Ghostery/Who TracksMe database (https://whotracks.me/explorer.html) and measures the proportion of pages on which a potential tracker transmits an unique identifier that is able to track a specific user across the web, either via cookies or fingerprinting. It is adjusted monthly and accounts for Yu et al.'s (2016) "mixed behavior."

*Invasiveness product = [Site reach value] * [Utilized tracking content value (Ghostery)]*

For example, the Google Static tracker has an invasiveness product of 0.079. This means that in a sample consisting of 100 webpages, it is estimated to track specific users on roughly 8 of them. The invasiveness product was determined for each known tracker in the sample set.

Finally, the effect of user autonomy options was considered by comparing the number of tracking scripts logged when a user rejected tracking versus the number logged when a user accepted it. Whether these features were opt-in or opt-out in nature was considered. In cases where a user had the enhanced ability to enable or disable specific categories of trackers, an all or nothing approach was used—either all categories were disabled or all categories were enabled.

Results

Prevalence of Third-Party Tracking

Of the 140 articles selected for testing, 2 exhibited network errors that prevented data collection. Sixteen had statements that provided user autonomy and were tested twice. This resulted in a total of 154 pages tested. 138 unique journals were represented (see appendix 1 for list of journals, publishers, and platforms tested).

132 of the 154 (85.7%) pages had tracking scripts, with an average of 4.2 scripts loaded per page. A total of 645 tracking scripts were loaded across the sample set, with 96.4% (n = 622) being linked to 47 known trackers in the WhoTracksMe database (see appendix 2).

Like Englehardt and Narayanan (2016), this study noted a long tail with regards to both the trackers and the third parties responsible for them. A small number of trackers had high site reach values; the remaining values dropped off quickly. Likewise, a limited number of third parties were responsible for setting most of the tracking scripts (see figure 2).

On average, the sample set had a unique tracker profile that was not representative of the web's top sites. Most

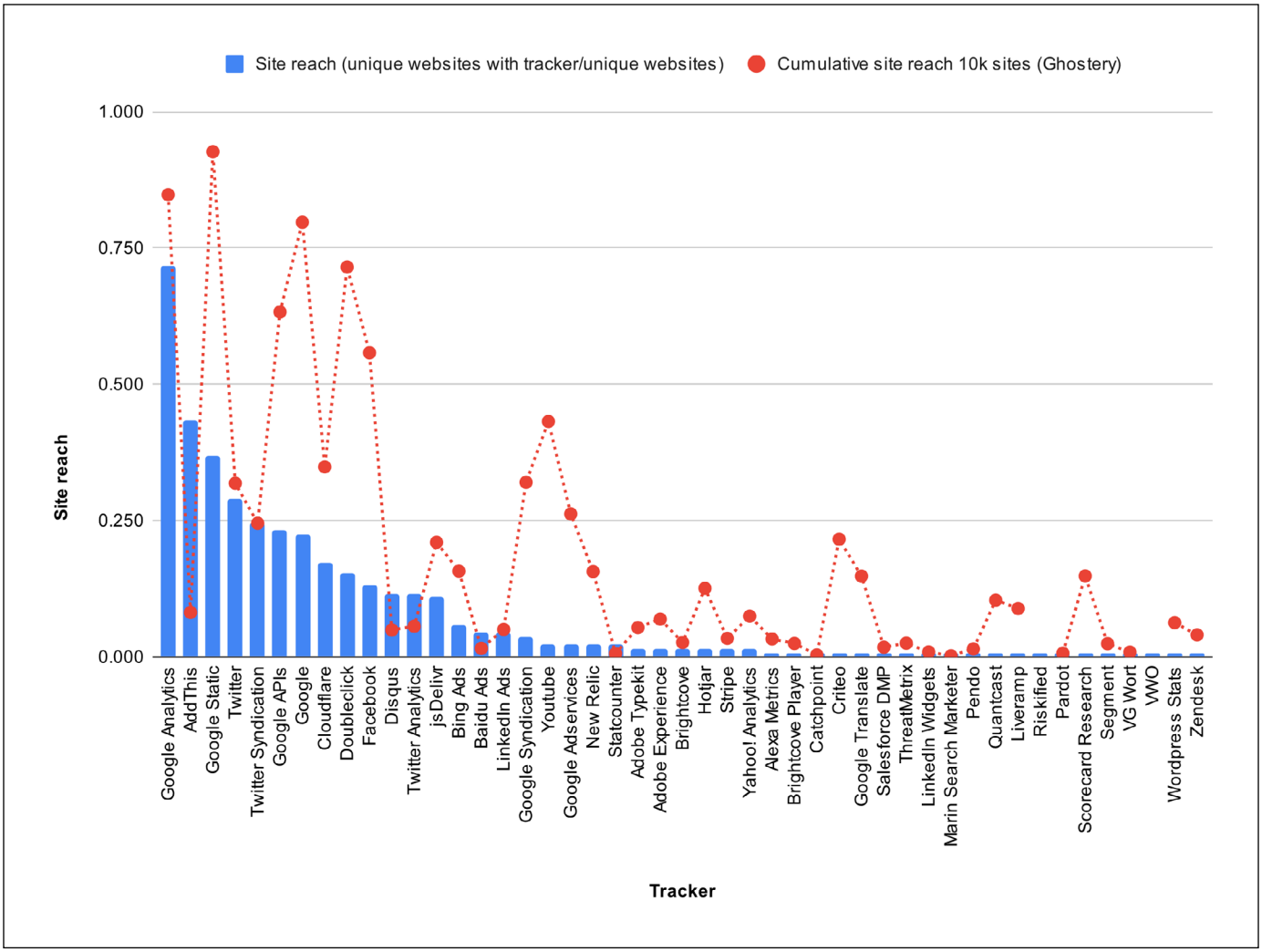

Figure 1. Trackers by site reach value (unique pages = 154)

trackers had comparable or lower site reach values than they did across Ghostery's top 10,000 sites (see figure 1), perhaps suggesting that tracking is less prevalent in open access articles than other areas of the web. Many trackers that typically had high site reach values (e.g., Doubleclick, Facebook, etc.) appeared further down the long tail in the study's sample. Conversely, some trackers that had low site reach values across the top 10,000 sites appeared frequently in our set. Oracle's AddThis tracker was more than 5 times as likely to appear on a page in the sample set, with a site reach value of 0.435 versus 0.082 in Ghostery's top 10,000 sites. Disqus and Twitter Analytics were more than twice as likely to appear in the sample set.

Thirty-three third parties loaded trackers on the pages. Google, Twitter, Microsoft, Adobe, and Brightcove were responsible for multiple trackers. Google set the most at 9, Twitter and Microsoft set 3, and Adobe and Brightcove set 2. However, the number of scripts that a party set did not always correlate with site reach. When tracking scripts were aggregated by the third parties responsible for them, Adobe and Brightcove both appeared further down the long tail (see figure 2).

Tracker Purpose

Eight categories of trackers—advertising, audio video player, CDN, comments, customer interaction, essential, site

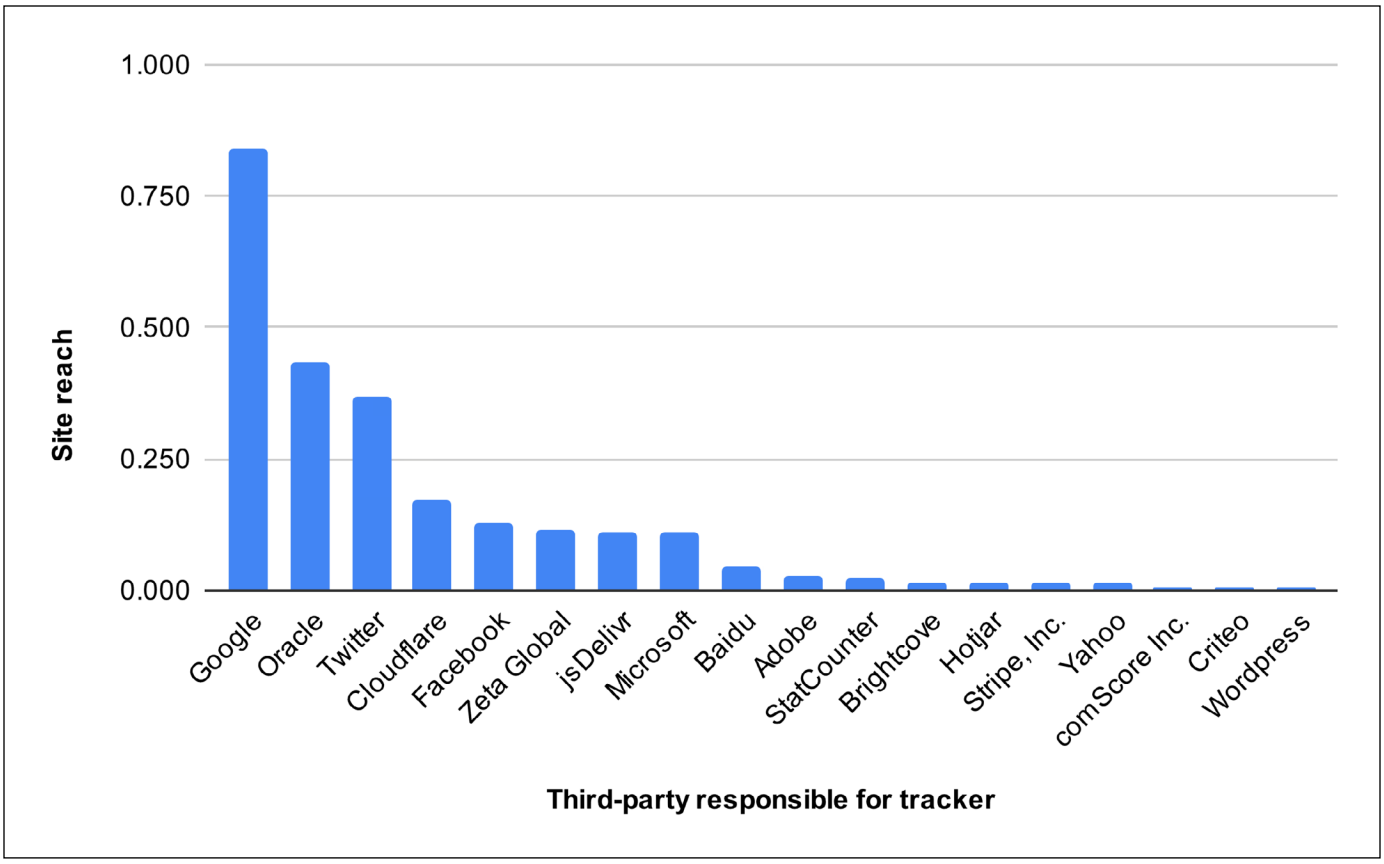

Figure 2. Third parties responsible for trackers by site reach value (unique pages = 154)

analytics and social media—were noted. Of the 47 known trackers, the majority placed in the site analytics (36.2%, n = 17) or advertising (29.8%, n = 14) categories (see figure 3). Most third parties set trackers that fell in a single category. However, Google, Twitter, Microsoft, and Adobe each set trackers serving multiple purposes (see figure 4).

Again, the number of unique trackers in a category did not always correlate with the category's site reach (see figure 5). As expected, the site analytics and advertising categories had high site reach values. However, the CDN category, despite making up only 8.5% of the tracker set, had a site reach value of 0.616, meaning that it appeared on nearly 62% of the pages analyzed. The social media category, which made up 6.4% of the total tracker set, had a site reach value of 0.290, appearing on 29% of the pages, and the comments category, which comprised only 2.1% of the trackers, boasted a site reach value of 0.116, appearing on a nearly 12% of the pages in the sample.

Invasiveness

Only one tracker, Google (set by the google.com domain), had an invasiveness product greater than 0.100, indicating that it can identify individuals across at least 10% of the sample set. With its invasiveness product value of 0.176, it is estimated to be setting unique identifiers that track users on 27 of the 154 (17.5%) pages analyzed (see figure 6). Four trackers had an invasiveness product greater than 0.050—Twitter Analytics (0.096), Twitter Syndication (0.083), Discus (0.080), and Twitter (0.078).

High site reach ranking did not translate to high invasiveness. Except for Twitter's trackers Twitter Syndication and Twitter, no other trackers posting a site reach value in the top five had an invasiveness product greater than 0.030. Google Analytics, which had the highest site reach value in the sample set, posted an invasiveness product of 0.005, indicating that it is estimated to use unique identifiers to track users on only 1 of the 154 pages (see appendix 2 for full results).

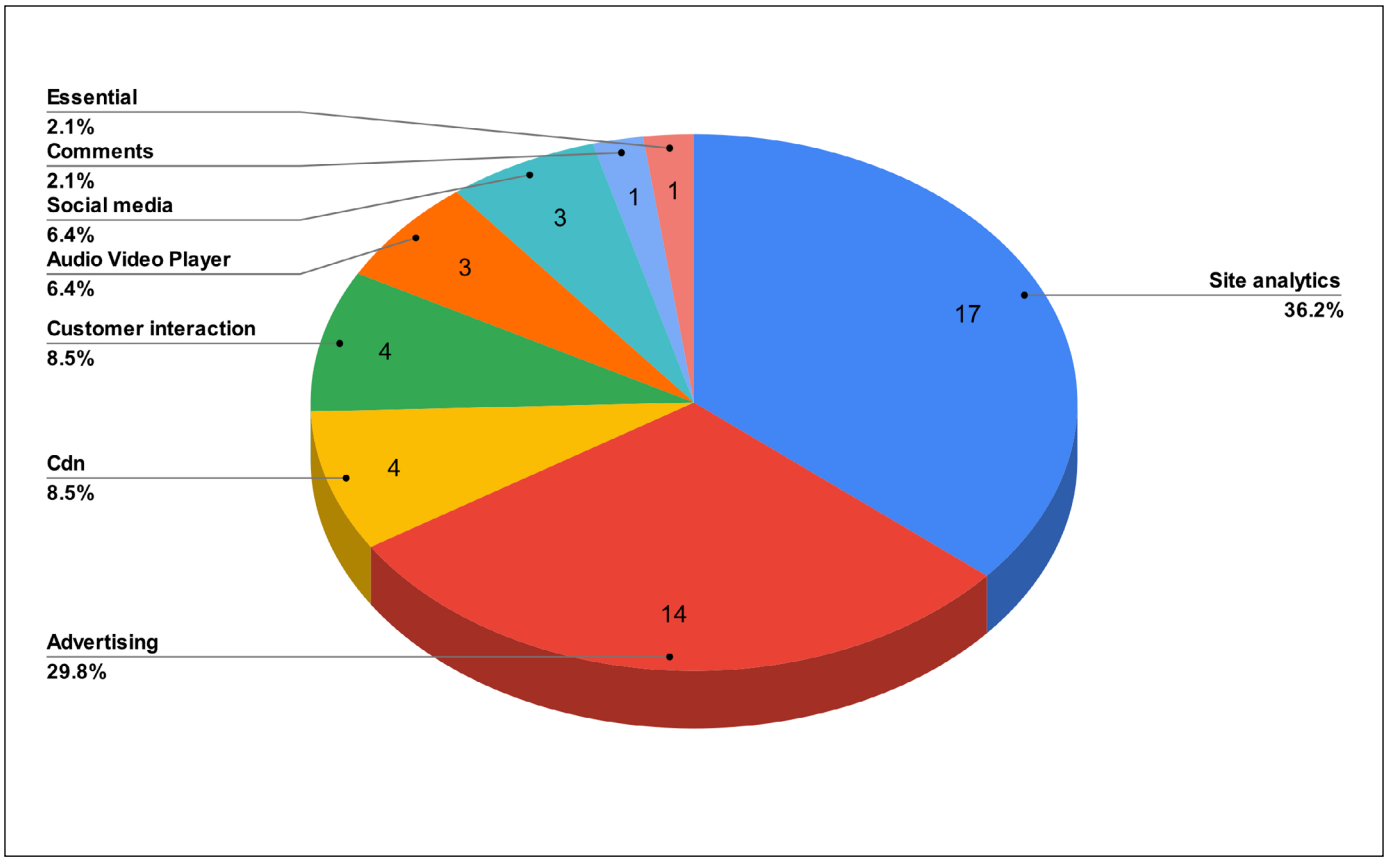

Figure 3. Trackers by purpose category (n=47)

Perhaps most surprising is the number of trackers that had an invasiveness product so low that they are estimated not to track users at all. Twenty-three (48.9%) trackers had an invasiveness product value less than 0.002. A few different scenarios may be responsible for this low value: both a low site reach value and a low utilized tracking content value, meaning that the tracker is not logged on many pages and does not often track users when it is logged; a high site reach value and a low utilized tracking content value, where the tracker appears frequently but does not track often when logged; or a low site reach value and a high utilized tracking content value, where the tracker does utilize tracking frequently when logged but doesn't get logged frequently.

User Autonomy Options

Ten of the 16 articles (62.5%) that offered user autonomy options logged fewer tracking scripts when tracking was rejected, with 2 of the 10 logging zero tracking scripts after the researcher disallowed tracking. For the remaining six, user selections had no effect on the number of tracking scripts loaded or cookies logged.

Of the 10 articles that logged fewer trackers based on user autonomy options, 7 were opt-in, meaning that tracking was disabled by default, but users could enable it at their discretion. Three of the 10 were opt-out. The researcher noted that two of these opt-out sites still logged cookies from trackers after all available options were exercised to block tracking. The exact reason for this is unknown and beyond the scope of this study, but it may be due to an error in the set-up of the cookie management software for the site, or the site may be loading content that is a part of a tracker network with the ability to load and track for other entities.

Discussion

Like previous studies (Hinchliffe, Zimmerman, and Altman 2018; Hanson 2019), third party trackers were found on most of the pages analyzed, confirming that open access literature is by no means immune to Borgman's commodity internet. Corporate influence has expanded to open access

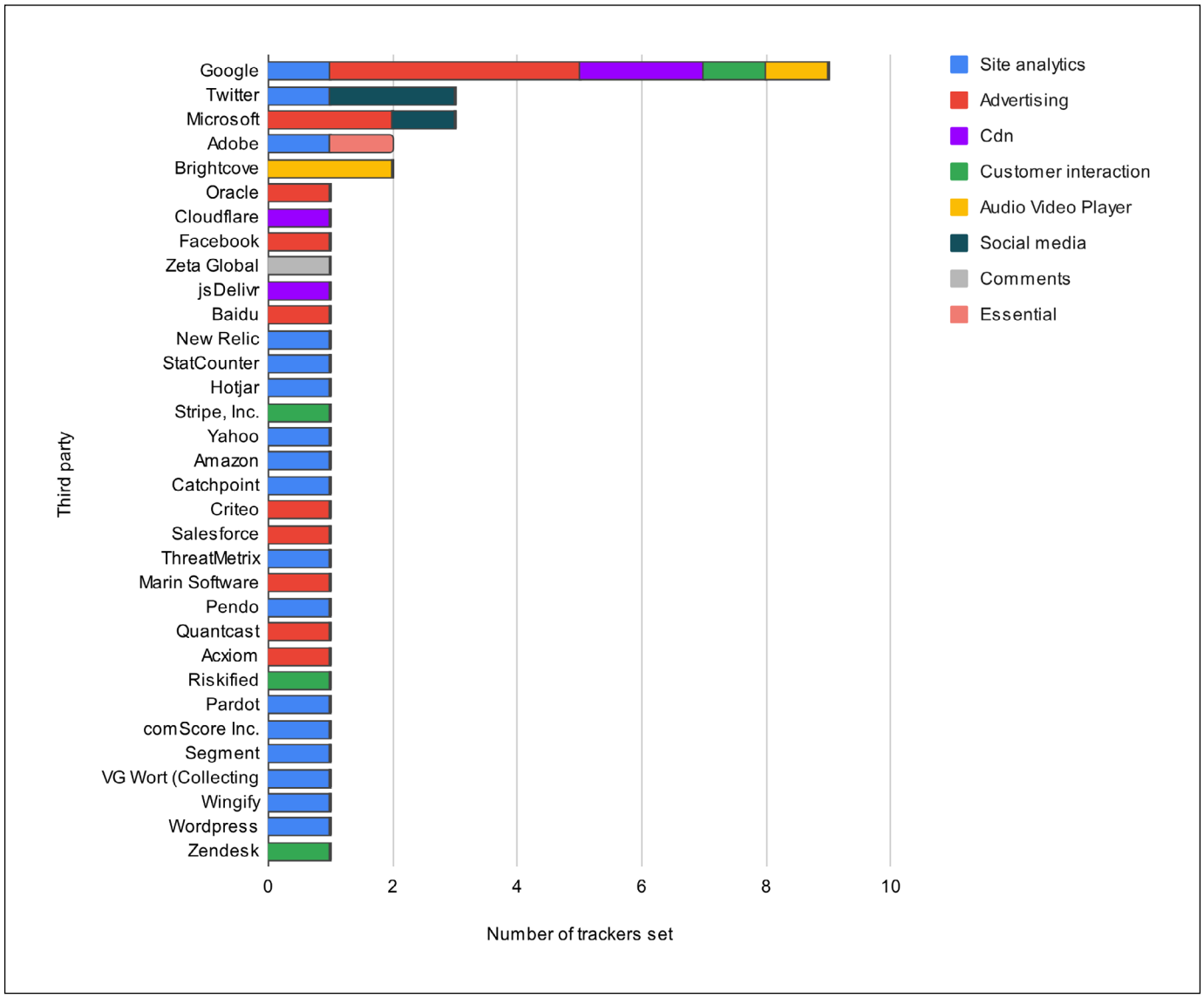

Figure 4 . Tracker purpose by third party (n=47)

literature through the various hosting platforms' websites, and users are given agency over this tracking only a small percentage of the time. However, in using site reach values to analyze how often a user is likely to encounter a specific tracker cumulatively across the set, the results revealed that relatively few trackers had a broad reach. It's tempting to theorize that, because of this low site reach, trackers are encountered less frequently, gain less information, and as a result, are less invasive. However, the reality is more complex, and as posited by Karaj et al. (2019), the sample's relationship to the overall web must be considered.

Our set of open access journals had a unique tracker profile when compared to the 10,000 most popular sites on the web. A low observed site reach value in the sample set does not necessarily mean that users will encounter the tracker any less frequently in their overall web activity. In fact, cumulative site reach values across Ghostery's top 10,000 sites seem to indicate that most of the identified trackers appear more frequently than observed in the sample, and still have the potential to gather personal data and build aggregate profiles.

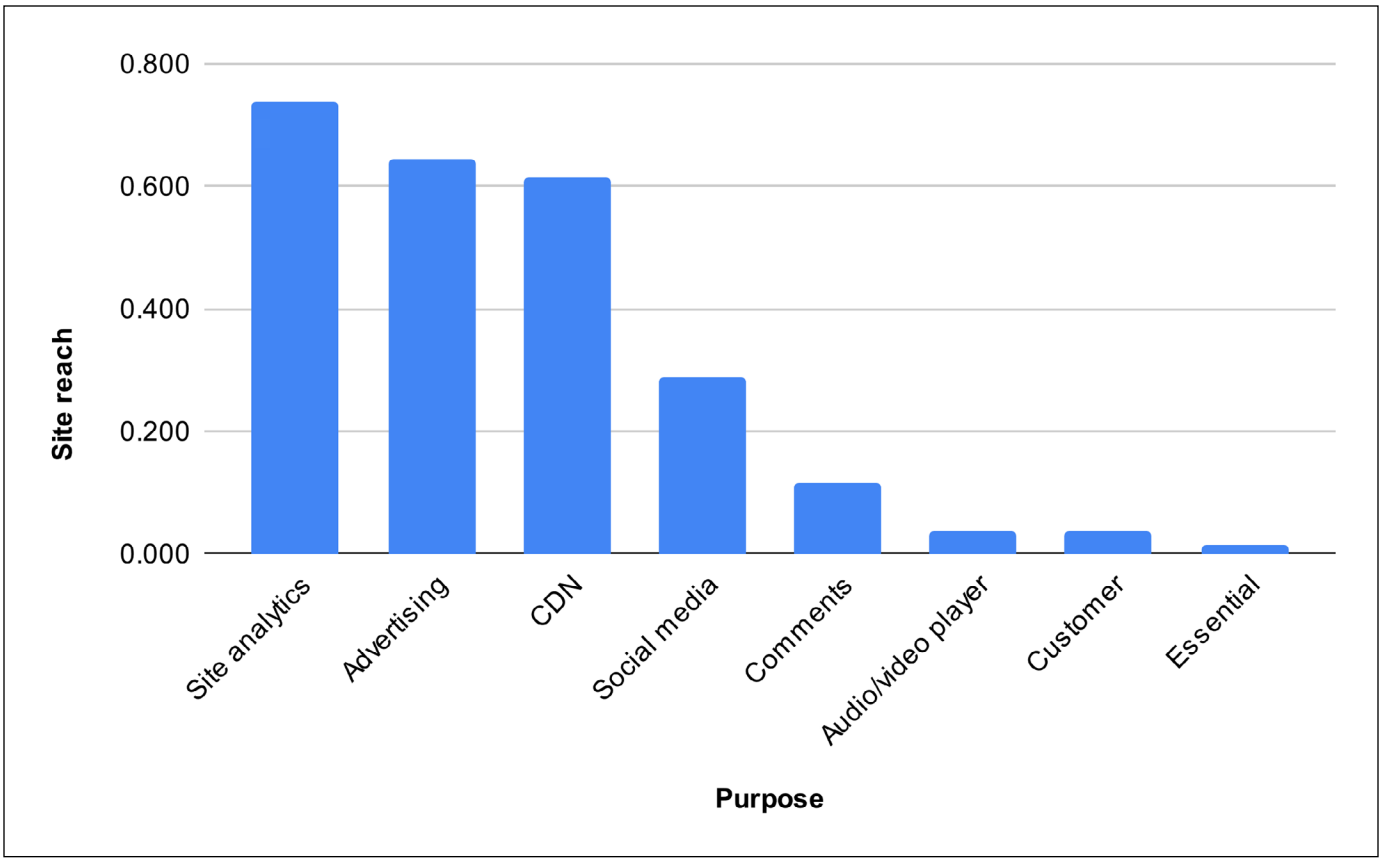

Figure 5 . Tracker categories by site reach value

To consider the likelihood of a tracker gathering information that could identify a specific user, the study also adopted Yu et al.'s (2016) observation of mixed behavior with regards to trackers—that tracking scripts are not always actively tracking. The proposal of an invasiveness product using site reach values and Ghostery's utilized tracking content value is unique and allows us to estimate how often user data is vulnerable to a given tracker in a set. Estimates made from this value show that very few of the trackers encountered in DOAJ content were highly invasive; most shocking was the estimate that nearly half of the group would likely not track at all. This shows that, taken in isolation, DOAJ content remains relatively safe with regards to user privacy. However, very few users use the web solely to visit open access articles and this finding may not hold up when the set is expanded to reflect users' real world browsing habits. Further studies should take this into account, using the invasiveness product value to analyze scholarly communication in terms of its relationship to a user's overall browsing history.

Given Breeding's (2019) warning to libraries regarding Google, the presence of its trackers, along with those social media platforms inhabiting the high end of the site reach and invasiveness long-tails, should give pause. Their use must be further evaluated and, perhaps, reconsidered. Not only did Google dominate over other entities in terms of the overall number of trackers logged and site reach value of its trackers, but it also posted several of the most invasive scripts logged in the study. Twitter and Facebook both logged higher than average invasiveness product values, as did AddThis, which allows users to share content with their social networks. While nearly impossible to entirely step away from the Google/social media ecosystem—Google set scripts in five of the identified purpose categories—studies that present alternatives (Chandler and Wallace 2016; Quintel and Wilson 2020) should be revisited and additional research should further investigate the value added by these services and explore viable, privacy-respecting alternatives to the most problematic.

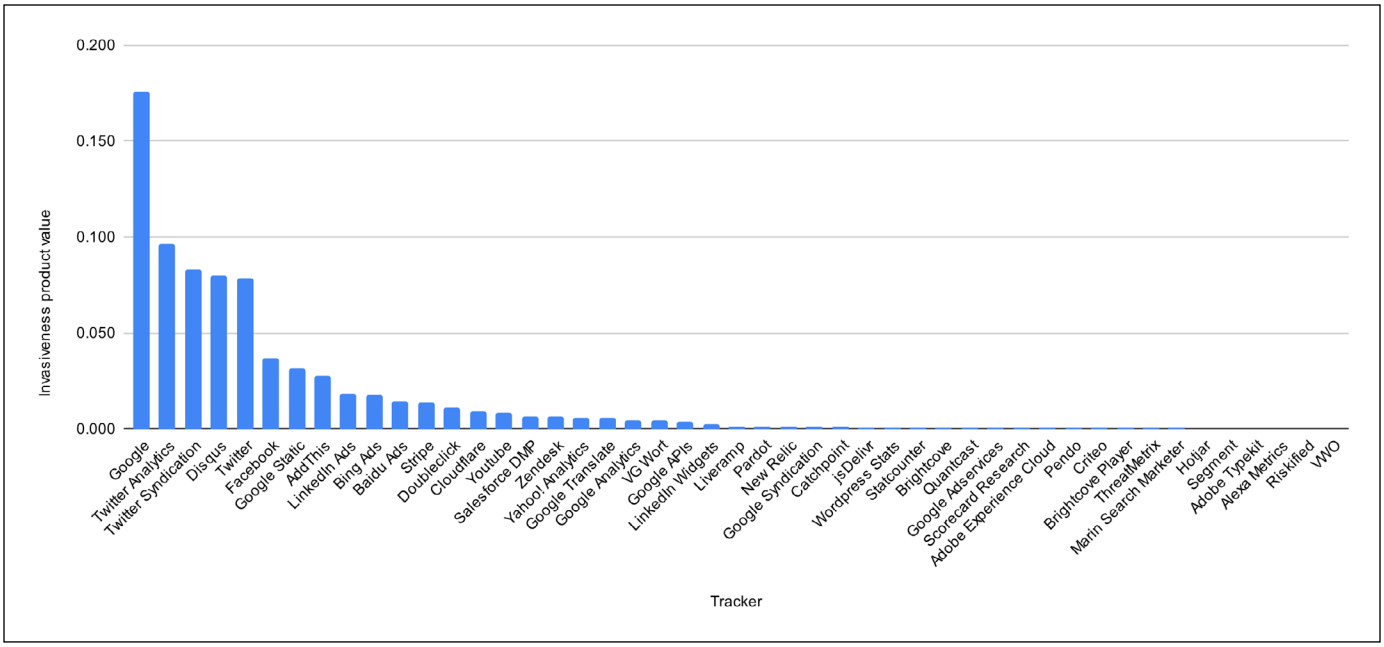

Figure 6 . Tracker by Invasiveness Product

Finally, the results indicate that user autonomy options are neither widespread nor fully developed. Only 16 articles (11.6%, n = 138) offered the user any control over tracking content, and some of these had options that proved ineffective at disabling the tracking they were meant to prevent. Further studies should focus exclusively on the efficacy of these autonomy features.

Limitations

This study is not without shortcomings. It represents a snapshot of a small sample of scholarly publishing at a set point in time. It does not consider users' real world browsing habits, which are unique and can influence their susceptibility to tracking. The internet is not static. Given the nature of tracking technology, the entities that set trackers as well as the scripts loaded on sites will change often. Likewise, the frequency with which they track users is mercurial—Ghostery's data is released monthly to account for these changes.

A small number of tracking scripts (n = 23, 3.6%) could not be identified or tied to any entity in the WhoTracksMe database. This lack of information made it impossible to determine the purpose of these scripts or their invasiveness.

Finally, this study did not explore the interconnected nature of tracking scripts. Scripts can use their access to a site's DOM, or document object model, to pass data to affiliates. Not only might a given tracker be forfeiting user information to many additional entities, but end users may also be subject to an ever-changing number of privacy policies based on algorithmic actions out of their control.

Conclusions

This study confirms that tracking technology is widespread in DOAJ content, but considers various caveats—site reach and invasiveness product values—to conclude that:

- DOAJ content has a unique tracker profile that deviates from the web's most popular 10,000 sites.
- Only about one-quarter (13 of the 47) of the identified trackers appeared on greater than 10% of the sample.
- Users are most likely to encounter trackers in the analytics, advertising, CDN, and social media categories when visiting DOAJ content.
- Most of the trackers were not highly invasive, with only 1 tracker (Google) estimated to track identifying information across more than 10% of the sample. 48.9% of the trackers were estimated not to track identifying information on individual users at all.
- User autonomy options are still not prevalent, appearing on only 16 of the articles tested, and only moderately effective when encountered, with only 10 of the 16 exhibiting fewer tracking scripts when users disallowed tracking.
- Due to the variability of tracker behavior and uniqueness of the sample's tracker profile, there is a need for further

studies that examine publisher content out of isolation, in the context of a user's overall web use.

Finally, there are actions that librarians can take to combat third-party tracking in scholarly communications that center on education and advocacy. First, librarians can continue to educate both themselves and users on privacy matters that affect library resources (Singley 2020; Jones et al. 2020; Brown and Klein 2020; Paris, Reynolds, and McGowan 2021). Librarians must monitor and adapt to evolving definitions of PII (Hanson 2019) and be aware of third-party tracking on the resources that they provide. Those who teach should include discussions of data privacy and user rights in their curriculum. Those who deal with library collections should pay particular attention to data handling and sharing portions of vendor contracts and pressure publishers to ensure that their platforms respect user privacy (Hinchliffe, Zimmerman, and Altman 2018). The Licensing Privacy project at the University of Illinois at Urbana-Champaign (https://publish.illinois.edu/licensingprivacy/) provides a good start. Those who serve in a support capacity for an academic journal should research the platform options, plug-ins, and privacy features that can be implemented. Where possible, the time has come for the library to rethink its relationship with Google. This calls for additional research that explores viable, privacy-respecting alternatives to Google services (see Chandler and Wallace 2016; Quintel and Wilson 2020). Finally, librarians must advocate for common sense privacy policies that provide transparency and autonomy—transparency on what information is being collected, who has access to that information, and how it is being used; and user autonomy that gives users real decision rights over what information can be collected.

References

American Library Association. 2017. "Privacy." Advocacy, Legislation & Issues. http://www.ala.org/advocacy/privacy.

———. 2019. "Privacy: An Interpretation of the Library Bill of Rights." Issues & Advocacy. June 24. https://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.

Arendt, Julie, Bettina Peacemaker, and Hillary Miller. 2019. "Same Question, Different World: Replicating an Open Access Research Impact Study." *College & Research Libraries* 80 (3): 303–18.

Arrieta, Andrés, Bennett Cyphers, Alexei Miagkov, and Daly Barnett. 2020. "Privacy Badger Is Changing to Protect You Better." Electronic Frontier Foundation. October 7. https://www.eff.org/deeplinks/2020/10/privacy-badger-changing-protect-you-better.

Barbrook, Richard, and Andy Cameron. 1996. "The Californian Ideology." *Science as Culture* 6 (1): 44–72. https://doi.org/10.1080/09505439609526455.

Bommana, Prudhvikumar. 2021. "Stable Channel Update for Desktop." *Chrome Releases* (blog). May 25, 2021. https://chromereleases.googleblog.com/2021/05/stable-channel-update-for-desktop_25.html.

Borgman, Christine. 2015. *Big Data, Little Data, No Data*. Cambridge, MA: The MIT Press.

Breeding, Marshall. 2016. "Chapter 3. Data from Library Implementations." *Library Technology Reports* 52 (4): 29–35.

———. 2019. "Chapter 2. Key Technologies with Implications for Privacy: Encryption, Analytics, and Advertising Tracking." *Library Technology Reports* 55 (7): 8–16.

Breugelmans, J. Gabrielle, Guillaume Roberge, Chantale Tippett, Matt Durning, David Brooke Struck, and Michael M. Makanga. 2018. "Scientific Impact Increases When Researchers Publish in Open Access and International Collaboration: A Bibliometric Analysis on Poverty-Related Disease Papers." *PLoS One* 13 (9): e0203156. https://doi.org/10.1371/journal.pone.0203156.

Brown, Michael, and Carrie Klein. 2020. "Whose Data? Which Rights? Whose Power? A Policy Discourse Analysis of Student Privacy Policy Documents." *The Journal of Higher Education* 91 (7): 1149–78. https://doi.org/10.1080/00221546.2020.1770045.

Chan, Leslie, Darius Cuplinskas, Michael Eisen, Fred Friend, Yana Genova, Jean-Claude Guédon, Melissa Hagemann, et al. 2002. "Budapest Open Access Initiative." February 14. https://www.budapestopenaccessinitiative.org/read.

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.

Deleuze, Gilles. 1992. "Postscript on the Societies of Control." *October* 59: 3–7.

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*, 1388–1401. Vienna Austria: ACM. https://doi.org/10.1145/2976749.2978313.

Hanson, Cody. 2019. "User Tracking on Academic Publisher Platforms." April 8, 2019. https://www.codyh.com/writing/tracking.html.

Hartman-Caverly, Sarah. 2019. "Human Nature Is Not a Machine: On Liberty, Attention Engineering, and Learning Analytics." *Library Trends* 68 (1): 24–53. https://doi.org/10.1353/lib.2019.0029.

Hinchliffe, Lisa Janicke, Katie Zimmerman, and Micah Altman. 2018. "Evaluating and Closing Privacy Gaps for Online

Library Services." Presented at the CNI Fall 2018 Project Briefing, Washington, DC, December 7. https://www.cni.org/topics/user-services/evaluating-and-closing-privacy-gaps-for-online-library-services.

Hwang, Soo-Yeon, and Michael Hanson. 2021. "Learning Analytics and Privacy: A Library Perspective." *Internet Reference Services Quarterly* 25 (3): 69–72. https://doi.org/10.1080/10875301.2021.1946456.

International Federation of Library Associations and Institutions. 2012. "IFLA Code of Ethics for Librarians and Other Information Workers (Full Version)." August. https://www.ifla.org/publications/node/11092.

———. 2014. "Internet Manifesto 2014." August 2014. https://www.ifla.org/publications/node/224.

Jones, Kyle M. L. 2019. "'Just Because You Can Doesn't Mean You Should': Practitioner Perceptions of Learning Analytics Ethics." *Portal: Libraries and the Academy* 19 (3): 407–28. https://doi.org/10.1353/pla.2019.0025.

Jones, Kyle M. L., Kristin A. Briney, Abigail Goben, Dorothea Salo, Andrew Asher, and Michael R. Perry. 2020. "A Comprehensive Primer to Library Learning Analytics Practices, Initiatives, and Privacy Issues." *College & Research Libraries* 81 (3): 570–91. https://doi.org/10.5860/crl.81.3.570.

Karaj, Arjaldo, Sam Macbeth, Rémi Berson, and Josep M. Pujol. 2018. "WhoTracks .Me: Shedding Light on the Opaque World of Online Tracking." ArXiv:1804.08959 [cs.CY], April 24. http://arxiv.org/abs/1804.08959.

Lazzarato, Maurizio. 1996. "Immaterial Labor." In *Radical Thought in Italy: A Potential Politics*, edited by Paolo Virno and Michael Hardt, new edition, 133–48. University of Minnesota Press.

LePage, Pete. 2021. "New in Chrome 91." Chrome Developers. May 26, 2021. https://developer.chrome.com/blog/new-in-chrome-91/.

Marino, Bill. 2021. "Privacy Concerns and the Prevalence of Third-Party Tracking Cookies on ARL Library Homepages." *Reference Services Review* 49 (2): 115–31. https://doi.org/10.1108/RSR-03-2021-0009.

Oakleaf, Megan. 2016. "Getting Ready & Getting Started: Academic Librarian Involvement in Institutional Learning Analytics Initiatives." *The Journal of Academic Librarianship* 42 (4): 472–75. https://doi.org/10.1016/j.acalib.2016.05.013.

———. 2018. "The Problems and Promise of Learning Analytics for Increasing and Demonstrating Library Value and Impact." *Information and Learning Science* 119 (1/2): 16–24. https://doi.org/10.1108/ILS-08-2017-0080.

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.

Oliphant, Tami, and Michael R. Brundin. 2019. "Conflicting Values: An Exploration of the Tensions between Learning Analytics and Academic Librarianship." *Library Trends* 68 (1): 5–23. https://doi.org/10.1353/lib.2019.0028.

Paris, Britt, Rebecca Reynolds, and Catherine McGowan. 2021. "Sins of Omission: Critical Informatics Perspectives on Privacy in E-learning Systems in Higher Education." *Journal of the Association for Information Science and Technology* 73 (5): 708-725. https://doi.org/10.1002/asi.24575.

Parks, Cecelia. 2017. "Beyond Compliance: Students and FERPA in the Age of Big Data." *Journal of Intellectual Freedom & Privacy* 2 (2): 23–33. https://doi.org/10.5860/jifp.v2i2.6253.

Piwowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. 2018. "The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles." *PeerJ* 6: e4375. https://doi.org/10.7717/peerj.4375.

Quintel, Denise, and Robert Wilson. 2020. "Analytics and Privacy: Using Matomo in EBSCO's Discovery Service." *Information Technology and Libraries* 39 (3). https://doi.org/10.6017/ital.v39i3.12219.

Schiller, Herbert. 1996. *Information Inequality: The Deepening Social Crisis in America.* New York: Routledge.

Science Europe. n.d. "'Plan S' and 'COAlition S'—Accelerating the Transition to Full and Immediate Open Access to Scientific Publications." Accessed January 24, 2020. https://www.coalition-s.org/.

Selwyn, Neil. 2019. "What's the Problem with Learning Analytics?" *Journal of Learning Analytics* 6 (3): 11.

Singley, Emily. 2020. "A Holistic Approach to User Privacy in Academic Libraries." *The Journal of Academic Librarianship* 46 (3): 102151. https://doi.org/10.1016/j.acalib.2020.102151.

Slade, Sharon, and Alan Tait. 2019. "Global Guidelines: Ethics in Learning Analytics." International Council for Open and Distance Education.

StatCounter. n.d. "Browser Market Share Worldwide." StatCounter Global Stats. Accessed April 13, 2022. https://gs.statcounter.com/browser-market-share/all/worldwide/2021.

Stebbins, Michael. 2013. "Expanding Public Access to the Results of Federally Funded Research." Whitehouse.Gov. February 22, 2013. https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research.

Travis, Tiffini A., and Christian Ramirez. 2020. "Big Data and Academic Libraries: The Quest for Informed Decision-Making." *portal : Libraries and the Academy* 20 (1): 33–47. https://doi.org/10.1353/pla.2020.0003.

Tummon, Nikki, and Dawn McKinnon. 2018. "Attitudes and Practices of Canadian Academic Librarians Regarding Library and Online Privacy: A National Study." *Library & Information*

Science Research 40 (2): 86–97. https://doi.org/10.1016/j.lisr.2018.05.002.

Yu, Zhonghao, Sam Macbeth, Konark Modi, and Josep M. Pujol. 2016. "Tracking the Trackers." In *Proceedings of the 25th International Conference on World Wide Web*, 121–32. Montréal, Canada: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872427.2883028.

Zimmer, Michael. 2014. "Librarians' Attitudes Regarding Information and Internet Privacy." *The Library Quarterly: Information, Community, Policy* 84 (2): 123–51. https://doi.org/10.1086/675329.

Appendix 1. Publishers, Journals, and Articles Tested

Publisher name	Journal tested	Journal website	Article tested
Academy of Science of South Africa	South African Journal of Science	http://www.sajs.co.za	https://www.sajs.co.za/article/view/3577
Advancements in Agricultural Development Inc	Advancements in Agricultural Development	https://agdevresearch.org/index.php/aad	http://agdevresearch.org/index.php/aad/article/view/87
AIP Publishing LLC	AIP Advances	http://aipadvances.aip.org/	http://dx.doi.org/10.1063/1.4968177
AIP Publishing LLC and ACA	Structural Dynamics	http://sd.aip.org	http://dx.doi.org/10.1063/1.4883975
Alanya Hamdullah Emin Paşa Üniversitesi	Journal of Contemporary Urban Affairs	https://www.ijcua.com/index.php/ijcua	https://ijcua.com/index.php/ijcua/article/view/7
American Physical Society	Physical Review X	http://prx.aps.org/	http://doi.org/10.1103/PhysRevX.5.041034
American Psychological Association	Archives of Scientific Psychology	http://www.apa.org/pubs/journals/arc/	https://psycnet.apa.org/fulltext/2019-71045-001.html
American Society for Microbiology	mBio	http://mbio.asm.org/	http://mbio.asm.org/cgi/content/full/7/1/e01931-15
AOSIS	Onderstepoort Journal of Veterinary Research	http://www.ojvr.org	https://ojvr.org/index.php/ojvr/article/view/1595
Aperio	Journal of Modern Philosophy	https://jmphil.org	https://jmphil.org/articles/66
Arkat USA, Inc.	ARKIVOC	http://www.arkat-usa.org	https://www.arkat-usa.org/arkivoc-journal/browse-arkivoc/ark.5550190.p010.407
Association for Learning Technology	Research in Learning Technology	https://journal.alt.ac.uk/index.php/rlt/index	https://journal.alt.ac.uk/index.php/rlt/article/view/2446/2815
Association for Medical Education in Europe (AMEE)	MedEdPublish	https://www.mededpublish.org/home	https://www.mededpublish.org/Manuscripts/1009
Association International de Management Stratégique (AIMS)	M@n@gement	https://management-aims.com/index.php/mgmt	https://management-aims.com/index.php/mgmt/article/view/4501/12129
Australasian Association for Information Systems	Australasian Journal of Information Systems	http://journal.acs.org.au/index.php/ajis/index	http://journal.acs.org.au/index.php/ajis/article/view/1098

Publisher name	Journal tested	Journal website	Article tested
Australian International Academic Centre PTY. LTD.	Advances in Bioscience and Clinical Medicine	http://www.journals.aiac.org.au/index.php/ABCMED/index	http://journals.aiac.org.au/index.php/ABCMED/article/view/3598
Bangladesh Pharmacological Society	Bangladesh Journal of Pharmacology	http://www.banglajol.info/index.php/BJP/index	https://www.banglajol.info/index.php/BJP/article/view/45800
Beilstein-Institut	Beilstein Journal of Organic Chemistry	http://www.beilstein-journals.org/bjoc	https://doi.org/10.3762/bjoc.10.332
Betasciencepress Publishers	Journal of Applied Bioanalysis	https://betasciencepress-publishing.com/journals/journal-of-applied-bioanalysis/	https://doi.org/10.17145/jab.18.019
BMC	Journal of Cardiovascular Magnetic Resonance	http://www.jcmr-online.com/	https://doi.org/10.1186/1532-429X-15-S1-W4
Canadian Science Publishing	FACETS	http://www.facetsjournal.com/	https://www.facetsjournal.com/doi/10.1139/facets-2021-0023
Cappadocia University	Ecocene: Cappadocia Journal of Environmental Humanities	http://ecocene.kapadokya.edu.tr	https://ecocene.kapadokya.edu.tr/index.php/ecocene/article/view/30
Cardiff University Press	Welsh Economic Review	https://wer.cardiffuniversitypress.org/	https://wer.cardiffuniversitypress.org/articles/254
Centre for Security Governance	Stability : International Journal of Security and Development	http://www.stabilityjournal.org/	https://www.stabilityjournal.org/articles/600
Cogitatio	Urban Planning	http://www.cogitatiopress.com/urbanplanning	https://www.cogitatiopress.com/urbanplanning/article/view/2619
Copernicus Publications	Archives Animal Breeding	http://www.archives-animal-breeding.net/	https://www.arch-anim-breed.net/61/481/2018/aab-61-481-2018.pdf
CSRC Publishing	Journal of Accounting and Finance in Emerging Economies	http://publishing.globalcsrc.org/jafee/	http://publishing.globalcsrc.org/ojs/index.php/jafee/article/view/100
D. G. Pylarinos	Engineering, Technology & Applied Science Research	http://www.etasr.com/index.php/ETASR	https://etasr.com/index.php/ETASR/article/view/4008
De Gruyter	Science and Engineering of Composite Materials	https://www.degruyter.com/view/j/secm	http://www.degruyter.com/view/j/secm.2019.26.issue-1/secm-2019-0032/secm-2019-0032.xml?format=INT
Department of Art History, University of Birmingham	Journal of Art Historiography	https://arthistoriography.wordpress.com	https://arthistoriography.files.wordpress.com/2017/11/mateo-rev.pdf
Diponegoro University	International Journal of Renewable Energy Development	http://ejournal.undip.ac.id/index.php/ijred	http://ejournal.undip.ac.id/index.php/ijred/article/view/8872
EDP Sciences	Parasite	http://www.parasite-journal.org/	http://dx.doi.org/10.1051/parasite/2014070

Publisher name	Journal tested	Journal website	Article tested
EL-Med-Pub	Journal of Neonatal Surgery	http://www.jneonatalsurg.com/	https://jneonatalsurg.com/ojs/index.php/jns/article/view/335
Emerald Publishing	International Journal of Climate Change Strategies and Management	http://www.emeraldgrouppublishing.com/products/journals/journals.htm?id=ijccsm	https://www.emeraldinsight.com/doi/pdfplus/10.1108/IJCCSM-05-2016-0074
European Federation of Psychology Students' Associations	Journal of European Psychology Students	http://jeps.efpsa.org/	https://jeps.efpsa.org/articles/333
European Publishing	Tobacco Induced Diseases	http://www.tobaccoinduceddiseases.org	http://www.tobaccoinduceddiseases.org/Perceptions-of-lung-cancer-screening-and-smoking-behavior-nchange-among-Chinese-immigrants,133579,0,2.html
EXARC	EXARC Journal	https://exarc.net/journal	https://exarc.net/ark:/88735/10163
F1000 Research Ltd	F1000Research	https://f1000research.com	https://f1000research.com/articles/9-1498/v1
Felix-Verlag	ILIRIA International Review	http://iliriapublications.org/index.php/iir/index	http://iliriapublications.org/index.php/iir/article/view/45
Firenze University Press	Phytopathologia Mediterranea	http://www.fupress.net/index.php/pm/index	https://oajournals.fupress.net/index.php/pm/article/view/11840
Forum Kunst und Markt	Journal for Art Market Studies	https://fokum-jams.org	https://fokum-jams.org/index.php/jams/article/view/7
Frontiers Media S.A.	International Journal of Public Health	https://www.ssph-journal.org/journals/international-journal-of-public-health	https://www.ssph-journal.org/articles/10.3389/ijph.2021.1604045/full
Genetics Society of America	G3: Genes, Genomes, Genetics	http://www.g3journal.org	http://g3journal.org/lookup/doi/10.1534/g3.117.300232
Geological Survey of Denmark and Greenland	Geological Survey of Denmark and Greenland Bulletin	https://geusbulletin.org/index.php/geusb/index	https://doi.org/10.34194/GEUSB-201943-03-03
German Medical Science GMS Publishing House	GMS Ophthalmology Cases	https://www.egms.de/en/journals/oc/	http://www.egms.de/static/en/journals/oc/2019-9/oc000097.shtml
Gonzaga Library Publishing	Journal of Hate Studies	https://jhs.press.gonzaga.edu/	https://jhs.press.gonzaga.edu/articles/147
H.S. Skovoroda Kharkiv National Pedagogical University	Pedagogy of Physical Culture and Sports	https://sportpedagogy.org.ua/index.php/ppcs	https://sportpedagogy.org.ua/index.php/ppcs/article/view/1618
Helsinki University Press	Redescriptions	https://journal-redescriptions.org	https://journal-redescriptions.org/articles/19
Hindawi - SAGE Publishing	Adsorption Science & Technology	https://www.hindawi.com/journals/ast/	https://doi.org/10.1177/0263617416659490

Publisher name	Journal tested	Journal website	Article tested
Hindawi Limited	Behavioural Neurology	https://www.hindawi.com/journals/bn/	http://dx.doi.org/10.1155/2020/9370891
Hindawi-Wiley	Journal of Food Quality	https://www.hindawi.com/journals/jfq/	http://dx.doi.org/10.1155/2021/6654211
IJHCR Publication	International Journal of Health and Clinical Research	http://ijhcr.com/index.php/ijhcr/about	http://ijhcr.com/index.php/ijhcr/article/view/8/8
IJPHY	International Journal of Physiotherapy	https://www.ijphy.org	https://www.ijphy.org/index.php/journal/article/view/703
IMR (Innovative Medical Research) Press Limited	Reviews in Cardiovascular Medicine	https://rcm.imrpress.com	https://rcm.imrpress.com/fileup/2153-8174/PDF/1609227825483-1867379003.pdf
Institute of Paleobiology PAS	Acta Palaeontologica Polonica	http://www.app.pan.pl/	http://www.app.pan.pl/archive/published/app63/app005332018.pdf
International Association for Court Administration	International Journal for Court Administration	https://www.iacajournal.org	https://www.iacajournal.org/articles/237
International Biogeography Society	Frontiers of Biogeography	http://escholarship.org/uc/fb	http://escholarship.org/uc/item/6nt6b38b
International Centre For Transactional Analysis Qualifications	International Journal of Transactional Analysis Research	http://www.ijtarp.org	https://www.ijtarp.org/article/view/20783
International Medical Society	International Archives of Medicine	http://imedicalsociety.org/ojs/index.php/iam/index	http://imedicalpublisher.com/ojs/index.php/iam/article/view/2931
International Society for Engineering Education (IGIP), Kassel University Press	International Journal of Engineering Pedagogy (iJEP)	http://www.i-jep.org	http://online-journals.org/index.php/i-jep/article/view/8099
International Union of Crystallography	IUCrJ	http://www.iucrj.org	http://scripts.iucr.org/cgi-bin/paper?S2052252517014324
Ivano-Frankivsk National Medical University	Galician Medical Journal	https://ifnmujournal.com/gmj/	https://ifnmujournal.com/gmj/article/view/1196
Japan Epidemiological Association	Journal of Epidemiology	http://jeaweb.jp/english/journa/index.html	https://www.jstage.jst.go.jp/article/jea/29/3/29_JE20180196/_pdf
JMIR Publications	Journal of Medical Internet Research	https://www.jmir.org	http://www.jmir.org/2020/6/e17930/
Joint Implant Surgery & Research Foundation	Reconstructive Review	http://reconstructivereview.org	https://reconstructivereview.org/ojs/index.php/rr/article/view/201
Kamje Press/xmlink	Annals of Occupational and Environmental Medicine	http://aoemj.biomedcentral.com	http://link.springer.com/article/10.1186/s40557-017-0184-x
KenzPub	Nuclear Receptor Research	http://www.kenzpub.com/journals/nurr/	https://www.kenzpub.com/journals/nurr/2019/101435/

Publisher name	Journal tested	Journal website	Article tested
Kiel Institute for the World Economy	Economics : the Open-Access, Open-Assessment e-Journal	http://www.economics-ejournal.org/	http://www.economics-ejournal.org/economics/discussionpapers/2017-86
Knowledge E	Sudan Journal of Medical Sciences	https://knepublishing.com/index.php/SJMS	https://doi.org/10.18502/sjms.v16i1.8938
Korea Information Processing Society-Computer Software Research Group	Human-Centric Computing and Information Sciences	http://www.hcis-journal.com/	http://link.springer.com/article/10.1186/s13673-020-00220-2
Levy Library Press	Annals of Global Health	https://www.annalsofglobalhealth.org/	https://annalsofglobalhealth.org/articles/3291
Librelloph	Journal of Human Security	http://www.librelloph.com/journalofhumansecurity	http://www.librelloph.com/journalofhumansecurity/article/view/503
LLC "CPC "Business Perspectives"	Problems and Perspectives in Management	https://businessperspectives.org/journals/problems-and-perspectives-in-management?category_id=30	https://businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/14829/PPM_2021_01_Rybina.pdf
MDPI AG	Medicina	http://www.mdpi.com/journal/medicina	https://www.mdpi.com/1010-660X/56/5/220
Nandan Nawn	Ecology, Economy and Society ‚Äì The INSEE Journal	https://ecoinsee.org/journal/ojs/index.php/ees/index	https://ecoinsee.org/journal/ojs/index.php/ees/article/view/344
National Iranian Oil Company (NIOC) Health Organization	The International Journal of Occupational and Environmental Medicine	http://www.theijoem.com/	http://www.theijoem.com/ijoem/index.php/ijoem/article/view/1164
National Numeracy Network	Numeracy	http://scholarcommons.usf.edu/numeracy/	http://scholarcommons.usf.edu/numeracy/vol9/iss2/art7/
National Research Nuclear University (MEPhI)	Nuclear Energy and Technology	https://nucet.pensoft.net/	http://www.sciencedirect.com/science/article/pii/S2452303817300663
Nature Publishing Group	Scientific Reports	http://www.nature.com/srep/index.html	https://doi.org/10.1038/s41598-021-90872-6
New Bulgarian University	English Studies at NBU	https://esnbu.org	http://esnbu.org/data/files/2018/2018-2-3-schroeder-p117-130.pdf
Norwegian Polar Institute	Polar Research	https://polarresearch.net/index.php/polar/index	https://polarresearch.net/index.php/polar/article/view/4458/11076
Open Library of Humanities	Digital Medievalist	https://journal.digitalmedievalist.org/	https://journal.digitalmedievalist.org/articles/55
Österreichische Apotheker-Verlagsgesellschaft m. b. H.	Scientia Pharmaceutica	http://www.mdpi.com/journal/scipharm	https://www.mdpi.com/2218-0532/89/1/5

Publisher name	Journal tested	Journal website	Article tested
PAGEPress Publications	European Journal of Histochemistry	http://www.ejh.it/	https://www.ejh.it/index.php/ejh/article/view/3213
Palacky University Olomouc	Acta Gymnica	http://gymnica.upol.cz	http://gymnica.upol.cz/artkey/gym-201604-0004_Physical_fitness_of_primary_school_children_in_the_reflection_of_different_levels_of_gross_motor_coordination.php
Papers in Physics	Papers in Physics	https://www.papersinphysics.org/index.php/papersinphysics/index	https://www.papersinphysics.org/papersinphysics/article/view/638
PeerJ Inc.	PeerJ	https://peerj.com/	https://peerj.com/articles/9570/
Pensoft Publishers	Folia Medica	https://foliamedica.bg	https://foliamedica.bg/article/54171/
Performance Philosophy	Performance Philosophy	https://www.performancephilosophy.org/journal	http://www.performancephilosophy.org/journal/article/view/201
Polish Botanical Society	Acta Mycologica	https://pbsociety.org.pl/journals/index.php/am/index	https://pbsociety.org.pl/journals/index.php/am/article/view/8557
PsychOpen	Europe's Journal of Psychology	http://ejop.psychopen.eu/index.php/ejop	http://ejop.psychopen.eu/article/view/1007
Public Library of Science (PLoS)	PLoS Biology	http://journals.plos.org/plosbiology/	http://europepmc.org/articles/PMC5369665?pdf=render
Queensland University of Technology	International Journal for Crime, Justice and Social Democracy	https://www.crimejusticejournal.com/index	https://www.crimejusticejournal.com/article/view/1659
RCVS Knowledge	Veterinary Evidence	https://veterinaryevidence.org/index.php/ve	https://veterinaryevidence.org/index.php/ve/article/view/72
SAGE Publishing	International Journal of Engineering Business Management	https://journals.sagepub.com/home/enb	https://doi.org/10.1177/1847979016670526
Scandinavian Military Studies	Scandinavian Journal of Military Studies	https://sjms.nu/	https://sjms.nu/articles/67
School of English, Aristotle University of Thessaloniki, Greece	Ex-centric Narratives: Journal of Anglophone Literature, Culture and Media	http://ejournals.lib.auth.gr/ExCentric/index	http://ejournals.lib.auth.gr/ExCentric/article/view/5997
SciELO	Journal of Venomous Animals and Toxins including Tropical Diseases	http://www.scielo.br/jvatitd	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1678-91992016000100325&lng=en&tlng=en
Sciendo	IZA Journal of Labor Economics	http://www.izajole.com/	http://www.degruyter.com/view/j/izajole.2020.9.issue-1/izajole-2020-0002/izajole-2020-0002.xml?format=INT

Publisher name	Journal tested	Journal website	Article tested
Scientific Medical Association of Moldova	The Moldovan Medical Journal	http://www.moldmedjournal.md	http://moldmedjournal.md/wp-content/uploads/2021/05/moldovan-med-j-2021-64-2-surev-et-al-full-text.pdf
SciPost	SciPost Physics	https://scipost.org/SciPostPhys	https://scipost.org/SciPostPhys.1.2.016
SEEd	Clinical Management Issues	https://journals.seedmedicalpublishers.com/index.php/CMI	https://journals.seedmedicalpublishers.com/index.php/cmi/article/view/1298
SEEd Medical Publishers	Farmeconomia: Health Economics and Therapeutic Pathways	https://journals.seedmedicalpublishers.com/index.php/FE/index	https://journals.seedmedicalpublishers.com/index.php/FE/article/view/1237
Septentrio Academic Publishing	Rangifer	https://septentrio.uit.no/index.php/rangifer	https://septentrio.uit.no/index.php/rangifer/article/view/4630
Shared Science Publishers OG	Microbial Cell	http://microbialcell.com/	http://microbialcell.com/researcharticles/2021a-ranalli-microbial-cell/
Slovenian Chemical Society	Acta Chimica Slovenica	http://acta.chem-soc.si/	https://journals.matheo.si/index.php/ACSi/article/view/2920
Society for Sociological Science	Sociological Science	https://www.sociologicalscience.com/	https://sociologicalscience.com/articles-v8-4-73/
South African National Biodiversity Institut	Bothalia: African Biodiversity & Conservation	http://www.abcjournal.org	https://abcjournal.org/index.php/abc/article/view/2099
Springer	Intereconomics	https://www.springer.com/journal/10272	https://doi.org/10.1007/s10272-021-0961-1
SpringerOpen	Journal of Inequalities and Applications	http://www.journalofinequalitiesandapplications.com/	https://doi.org/10.1186/s13660-020-02535-1
Stockholm University Press	Designs for Learning	http://www.designsforlearning.nu/	https://www.designsforlearning.nu/articles/97
Swedish Nutrition Foundation	Food & Nutrition Research	http://foodandnutritionresearch.net/index.php/fnr	https://foodandnutritionresearch.net/index.php/fnr/article/view/5453/13390
Taylor & Francis Group	European Journal of Psychotraumatology	https://www.tandfonline.com/toc/zept20/current	http://dx.doi.org/10.1080/20008198.2019.1706297
The Company of Biologists	Disease Models & Mechanisms	https://journals.biologists.com/dmm	http://dmm.biologists.org/content/9/3/271
The Ohio State University Libraries	Empirical Musicology Review	http://emusicology.org/	http://emusicology.org/article/view/6113
The Royal Society	Open Biology	https://royalsocietypublishing.org/journal/rsob	https://royalsocietypublishing.org/doi/pdf/10.1098/rsob.170121
Ubiquity Press	Psychologica Belgica	http://www.psychologicabelgica.com/	https://www.psychologicabelgica.com/articles/475
UCL Press	Archaeology International	https://www.uclpress.co.uk/pages/archaeology-international	https://www.ai-journal.com/articles/384

Publisher name	Journal tested	Journal website	Article tested
Universitas Ahmad Dahlan	IJAIN (International Journal of Advances in Intelligent Informatics)	http://ijain.org/index.php/IJAIN/index	http://ijain.org/index.php/IJAIN/article/view/426
University Library System, University of Pittsburgh	Journal of World-Systems Research	https://jwsr.pitt.edu/ojs/index.php/jwsr	http://jwsr.pitt.edu/ojs/index.php/jwsr/article/view/652
University of Alberta	Evidence Based Library and Information Practice	https://journals.library.ualberta.ca/eblip/index.php/EBLIP	https://journals.library.ualberta.ca/eblip/index.php/EBLIP/article/view/29634
University of Bologna	Journal of Formalized Reasoning	http://jfr.unibo.it/	https://jfr.unibo.it/article/view/8751
University of California Press	Collabra: Psychology	http://www.collabra.org	https://www.collabra.org/articles/218
University of Jyvaskyla	Human Technology	https://humantechnology.jyu.fi/	https://humantechnology.jyu.fi/archive/vol-13/issue-2-1/puolakanaho_latvala
University of Kansas	Journal of Montessori Research	http://journals.ku.edu/jmr	https://journals.ku.edu/jmr/article/view/15122
University of South Florida	Undergraduate Journal of Mathematical Modeling: One + Two	http://scholarcommons.usf.edu/ujmm/	http://scholarcommons.usf.edu/ujmm/vol6/iss2/4/
University of Victoria Libraries	KULA	https://kula.uvic.ca/	https://kula.uvic.ca/articles/63
University of York	Internet Archaeology	http://intarch.ac.uk/	http://intarch.ac.uk/journal/issue50/17/index.html
Upsala Medical Society	Upsala Journal of Medical Sciences	https://ujms.net/index.php/ujms/index	https://ujms.net/index.php/ujms/article/view/6118/13543
Utrecht University Library Open Access Journals (Publishing Services)	Liber Quarterly: The Journal of European Research Libraries	https://www.liberquarterly.eu/	http://www.liberquarterly.eu/articles/10.18352/lq.10185/
Utrecht University School of Law	Utrecht Law Review	https://www.utrechtlawreview.org	http://www.utrechtlawreview.org/articles/10.18352/ulr.350/
UTS ePRESS	PORTAL: Journal of Multidisciplinary International Studies	https://epress.lib.uts.edu.au/journals/index.php/portal	https://epress.lib.uts.edu.au/journals/index.php/portal/article/view/7400
Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften	Quantum	http://quantum-journal.org/	https://quantum-journal.org/papers/q-2019-01-06-115/pdf/
Veterinary World	Veterinary World	http://www.veterinaryworld.org	http://www.veterinaryworld.org/Vol.13/March-2020/27.pdf
Volcanica	Volcanica	https://www.jvolcanica.org/ojs/index.php/volcanica/index	http://www.jvolcanica.org/ojs/index.php/volcanica/article/view/38

Publisher name	Journal tested	Journal website	Article tested
Wellcome	Wellcome Open Research	https://wellcomeopenresearch.org/	https://wellcomeopenresearch.org/articles/6-107/v1
White Rose University Press	Undergraduate Journal of Politics and International Relations	https://www.ujpir-journal.com/	https://www.ujpir-journal.com/articles/76
Wiley	Molecular Oncology	https://febs.onlinelibrary.wiley.com/journal/18780261	https://doi.org/10.1002/1878-0261.12692
World Century Publishing Corporation	China Quarterly of International Strategic Studies	https://www.worldscientific.com/cqiss	http://www.worldscientific.com/doi/pdf/10.1142/S2377740018500252
Yale University	British Art Studies	http://www.britishartstudies.ac.uk	http://britishartstudies.ac.uk/issues/issue-index/issue-4/thomas-rowlandson

Appendix 2. Known Trackers Identified in Sample Set

Tracker	Third-party responsible for script	Purpose Category	Site reach (sample set)	Cumulative site reach - top 10k website (Ghostery)	Utilized tracking content value (Ghostery)	Invasiveness product	Estimated tracking reach in sample set (n = 154)
Google Analytics	Google	Site analytics	0.717	0.848	0.006	0.005	1
AddThis	Oracle	Advertising	0.435	0.082	0.063	0.028	4
Google Static	Google	CDN	0.370	0.927	0.086	0.032	5
Twitter	Twitter	Social media	0.290	0.319	0.270	0.078	12
Twitter Syndication	Twitter	Social media	0.246	0.245	0.336	0.083	13
Google APIs	Google	CDN	0.232	0.633	0.018	0.004	1
Google	Google	Advertising	0.225	0.797	0.783	0.176	27
Cloudflare	Cloudflare	CDN	0.174	0.349	0.051	0.009	1
Doubleclick	Google	Advertising	0.152	0.715	0.074	0.011	2
Facebook	Facebook	Advertising	0.130	0.558	0.280	0.036	6
Disqus	Zeta Global	Comments	0.116	0.049	0.687	0.080	12
Twitter Analytics	Twitter	Site analytics	0.116	0.056	0.830	0.096	15
jsDelivr	jsDelivr	CDN	0.109	0.210	0.007	0.001	0
Bing Ads	Microsoft	Advertising	0.058	0.157	0.308	0.018	3
Baidu Ads	Baidu	Advertising	0.043	0.016	0.334	0.015	2
LinkedIn Ads	Microsoft	Advertising	0.043	0.050	0.425	0.018	3

Tracker	Third-party responsible for script	Purpose Category	Site reach (sample set)	Cumulative site reach - top 10k website (Ghostery)	Utilized tracking content value (Ghostery)	Invasiveness product	Estimated tracking reach in sample set (n = 154)
Google Syndication	Google	Advertising	0.036	0.320	0.024	0.001	0
Google Adservices	Google	Advertising	0.022	0.262	0.023	0.000	0
New Relic	New Relic	Site analytics	0.022	0.157	0.042	0.001	0
Statcounter	StatCounter	Site analytics	0.022	0.006	0.024	0.001	0
Youtube	Google	Audio Video player	0.022	0.432	0.371	0.008	1
Adobe Experience Cloud	Adobe	Site analytics	0.014	0.069	0.029	0.000	0
Adobe Typekit	Adobe	Essential	0.014	0.054	0.000	0.000	0
Brightcove	Brightcove	Audio Video Player	0.014	0.026	0.036	0.001	0
Hotjar	Hotjar	Site analytics	0.014	0.126	0.007	0.000	0
Stripe	Stripe, Inc.	Customer interaction	0.014	0.034	0.943	0.014	2
Yahoo! Analytics	Yahoo	Site analytics	0.014	0.075	0.402	0.006	1
Alexa Metrics	Amazon	Site analytics	0.007	0.033	0.001	0.000	0
Brightcove Player	Brightcove	Audio Video Player	0.007	0.025	0.039	0.000	0
Catchpoint	Catchpoint Systems	Site analytics	0.007	0.004	0.110	0.001	0
Criteo	Criteo	Advertising	0.007	0.216	0.054	0.000	0
Google Translate	Google	Customer interaction	0.007	0.148	0.775	0.006	1
LinkedIn Widgets	Microsoft	Social media	0.007	0.009	0.319	0.002	0
Liveramp	Acxiom	Advertising	0.007	0.089	0.153	0.001	0
Marin Search Marketer	Marin Software	Advertising	0.007	0.002	0.020	0.000	0
Pardot	Pardot	Site analytics	0.007	0.007	0.145	0.001	0
Pendo	Pendo	Site analytics	0.007	0.015	0.055	0.000	0
Quantcast	Quantcast International Ltd.	Advertising	0.007	0.104	0.070	0.001	0
Riskified	Riskified	Customer interaction	0.007	Data not available in Ghostery			
Salesforce DMP	Salesforce	Advertising	0.007	0.018	0.878	0.006	1

Tracker	Third-party responsible for script	Purpose Category	Site reach (sample set)	Cumulative site reach - top 10k website (Ghostery)	Utilized tracking content value (Ghostery)	Invasiveness product	Estimated tracking reach in sample set (n = 154)
Scorecard Research Beacon	comScore Inc.	Site analytics	0.007	0.148	0.065	0.000	0
Segment	Segment	Site analytics	0.007	0.024	0.002	0.000	0
ThreatMetrix	ThreatMetrix (LexisNexis Risk Solutions FL)	Site analytics	0.007	0.025	0.021	0.000	0
VG Wort	VG Wort (Collecting Society)	Site analytics	0.007	0.009	0.574	0.004	1
VWO	Wingify	Site analytics	0.007	Data not available in Ghostery			
Wordpress Stats	Wordpress	Site analytics	0.007	0.063	0.073	0.001	0
Zendesk	Zendesk	Customer interaction	0.007	0.040	0.876	0.006	1