Study of Third-Party Analytics Services on University Websites

Timi Heino, Sampsa Rauti, Robin Carlsson, and Ville Leppänen

University of Turku, Turku, Finland
{tdhein,sjprau,crcarl,ville.leppanen}@utu.fi

Abstract. The accelerated digitalization and the increased use of online services for everyday tasks, online privacy issues are more important than ever before. This also goes for universities which are increasingly moving information and services online. Our study provides a technical overview of prevalence of third-party analytics on university websites. Websites of 40 universities from eight different countries around the world are analyzed to reveal third-party analytics services they use. The study shows that most universities, especially in many technologically advanced western countries, have alarmingly high number of analytics services on their websites. The results emphasize the need for web developers and data protection officers to better assess what kind of data their websites deliver to third parties. This is especially important for universities, as kind of example institutions tasked with advancing the common good.

Keywords: Online privacy · Third-party analytics services · Tracking · University Websites

1 Introduction

The digitalization of society continues and online services are increasingly being used to take care of ordinary everyday tasks. Universities, too, have benefited from this development and moved many services and lots of information online for potential applicants, students, researchers and other interested parties alike. In order to serve their visitors better, website maintainers use third-party analytics services. Universities are no exception. While the analytics services reveal who the visitors are and how they behave when browsing the website, they also cause privacy concerns [6, 13, 9].

Although universities are not always in state ownership or publicly funded, they have the role of generating and sharing information. They are institutions issuing degrees, conducting research and having societal impact. While universities usually acknowledge the importance of social responsibility in their curricula and guidelines for researchers, they should also act socially responsibly as organizations. This is why, one could argue, it is not part of their role to deliver personal data about website visitors to third-party analytics services. Instead, they should aim to be exemplary in online privacy. Because of the accelerated

digitalization caused by the COVID-19 pandemic, such privacy issues are more important than ever before [11].

While the use of analytics services in higher education institutional websites has not been studied that much (see e.g. [7]), many university libraries seem to have taken the privacy of their websites more seriously, advising caution in the use of third-party analytics services [10, 2]. These critical voices seem to be overpowered by a multitude of articles recommending Google Analytics [5, 14, 3], often without sufficient consideration of privacy consequences when using a third-party analytics service.

Striving to fill the obvious gap in this research area, we analyze websites of 40 universities to find out what kind of third-party analytics services they use. The five top universities from eight countries around the world were included in the sample. The current study provides a technical analysis of prevalence of third-party analytics on university websites and also allows us to compare university websites around the world in terms of visitors' privacy.

The rest of the paper is structured as follows. Section 2 outlines the setting of the study, explaining how the websites were chosen and analyzed. Section 3 presents the results of the study, comparing the use of third-party analytics services in different countries. Section 4 discusses the impact of the findings in terms of user privacy and web development. Finally, Section 5 concludes the paper.

2 Setting of the study

In order to choose the universities for the current study, we first picked 8 countries around the world: Australia, Canada, Chile, China, Finland, Germany, India, and South Africa. For each selected country, we then chose the top five universities[1]. The websites of the selected universities were then analyzed to find third-party analytics services.

When browsing a website, we went through each link in the main menu. We also used the search option if one was available and opened one search result link. All cookies were always accepted at the beginning of the visit if a cookie banner popped up. This approach is of course not an exhaustive analysis, but it does give a clear picture of how many analytics services are invoked even during a brief visit to a given university website.

Chrome developer tools (DevTools), which can be used to analyze web requests made by a website, were used when browsing university websites, and all third-party requests (requests that go outside the university's own server/domain name) recorded during the visit were manually analyzed. Domain names of requests, payloads, and request chains were inspected to distinguish requests related to tracking and analytics from other requests. Such benign requests were, among other things, HTML and CSS element downloads and requests associated

[1] The top five universities for each country were chosen according to the listing at https://www.topuniversities.com/university-rankings/world-university-rankings/2022

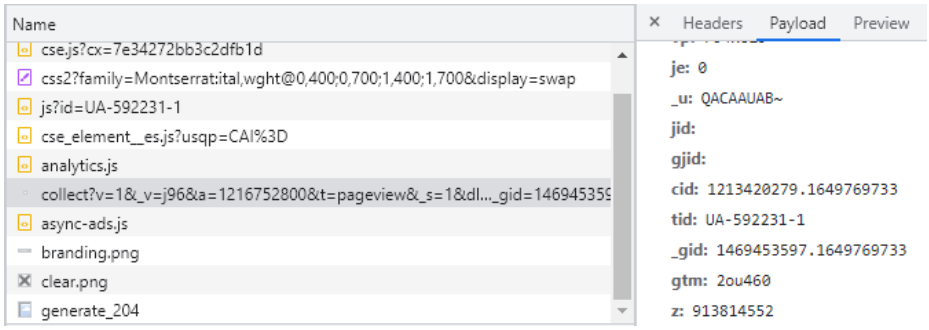

Fig. 1. A sample view of Chrome DevTools with a list of web requests and a payload of a request delivered to Google Analytics.

with the normal functionality of the website. A sample view of Chrome DevTools is shown in Figure 1.

It is worth noting that our definition of "analytics service" is quite broad here. Requests that can be used for tracking and profiling, such as third-party requests caused by embedded videos, advertisement and social media share buttons were included in our data, although their primary purpose may not always be collecting analytics.

3 Results

Table 1 shows the average number of analytics services per university website and the number of unique analytics services per country. Of the studied countries, universities in Australia had the most analytics on their websites, with a staggering average of 12.6 services per analyzed website. On the website of one Australian university, our brief test browsing revealed 16 different services potentially collecting personal data. Even the website with least analytics included 8 different analytics services. While the universities vaguely state in their privacy policy documents that analytics are used for "business and learning analytics purposes" and "improving our services", this is hardly a valid justification for using 16 different third-party services. Australian university websites also had the widest array of analytics services in the study, 27 in total. Google Analytics, Facebook and Twitter were present on all 5 studied Australian websites. There were also a couple of analytics services that seemed to be exclusive to Australia, such as Tealium.

Finland takes the second place with an average of 7.0 analytics services per website. The university with the largest number of analytics services on their website included 10 unique services, still a large number difficult to justify. Finnish universities used 13 unique analytics services in total. Not surprisingly, Google Analytics was present on all websites (5), followed by Facebook (4). Canada comes third, with 5.6 analytics services per university website on average. The highest number of analytics services on a single website was 8. The

Table 1. The average number of used analytics services per university website and the number of different services per country.

Country	Services per website (average)	Number of different analytics services
Australia	12.6	27
Finland	7.0	13
Canada	5.6	14
South Africa	5.4	12
Chile	3.0	7
India	2.2	4
China	1.4	4
Germany	1.4	4

Canadian websites included 14 unique analytics services, overtaking Finland in this respect. The scenery of different services resembles Finland, with Google Analytics rampant everywhere (5).

Thus far, it seems technologically advanced western democracies all have quite a high number of analytics services on their university websites. Although South Africa may not exactly fit in this category, it still had 12 unique services in total on the university websites. Close to Canada, South Africa had 5.4 third-party analytics services per university website on average. The assortment of analytics services also remains similar to the western countries with services such as Google Analytics (5), Facebook (3), Hotjar (2) and LinkedIn Insight (2).

Chile had 7 different analytics services on the websites, with an average of 3.0 services per website. In Chile, Google's services constituted a major part of the different analytics services on the analyzed websites, but there were also two other third party services, Hotjar and Livechatinc. Compared to Chile, India had less unique analytics services (4) planted into university websites, averaging 2.2 services per website. Furthermore, all the used services (Google Analytics, AdSense, YouTube, DoubleClick) belonged exclusively to Google. One of the Indian universities, Indian Institute of Technology Delhi, did not appear to have any analytics services on its website.

Finally, Germany and China are tied, having 4 unique analytics services on university websites, and both averaging only 1.4 services per website. It is interesting to note how well Germany does compared to other western countries. Germany is also unique in the sense that the websites do not include services primarily meant for collecting analytics (such as Google Analytics). Instead, the found services were Google AdSense, Google Ad Services and Adobe Dynamic Tag Management Assets. Germany is also exemplary in the sense that it has replaced third-party analytics services with Matomo, an open source solution that collects analytics locally [4]. Matomo was found on the websites of 4 German universities.

Considering China's usual eagerness to harvest data, it may at first be surprising there are not many analytics services present on the analyzed websites. The

reason for this is probably China's censorship and the fact that the Great Firewall blocks many western services (for instance, Google Analytics) [1]. Instead, China uses its own centralized information collection. Tracking functionality of Baidu, a large Chinese technology company, was found on 4 of the 5 studied websites. Interestingly, Baidu also seems to be one of the few private companies licensed to gather and deliver data to China's governmental analytics [8]. Only one university, Shanghai Jia Tong University, appeared to have Google's services, such as Google Analytics and DoubleClick, on its website. On the whole, China does extremely well in not leaking information outside its own borders.

Table 2. The numbers of websites using different analytics services.

Analytics service	Universities using the service
Google Analytics	28
Doubleclick	20
YouTube	17
Google Tag Manager	16
Facebook	15
Google Adsense	11
LinkedIn Insight	11
Hotjar	11
Google Ad Services	9
Twitter	7
Siteimproveanalytics	5
New Relic	4
Microsoft	4
Baidu	4
Adobe	3
AppNexus	3
The Trade Desk	3
Tealium	3
Quantcast	2
Snapchat	2
LivechatInc	2
Mpulse	1
Coveo Analytics	1
Static.srcspot	1
Qualaroo	1
Crazyegg	1
Tiktok	1
Addthis	1
Yahoo	1
ClickDimensions	1
Sharethis	1

Table 2 shows the numbers of university websites using different analytics services. For example, 28 of 40 studied university websites were found to use Google Analytics. It is easy to see that Google enjoys a position of overwhelming dominance when it comes to university website analytics. Out of the 9 most used services, 6 are provided by Google. With the exception of one German and three Chinese universities, every time a website contained any third-party analytics, Google's services were present. This gives Google a broad front-row view to general interests and behavior of academic website visitors around the globe, but also the possibility to track individuals and collect their personal data. Other major analytics providers are Facebook, LinkedIn Insight, Hotjar and Twitter.

Figure 2 gives some insight into what and how many different analytics services are used in the studied countries. The figure shows, for example, that Australia has a wide collection of different analytics services, many of which are not used in other countries. It is also easy to see how Indian websites only make use of Google's services, and how few services Germany and China employ.

4 Discussion

The results of the current study do not flatter the technologically advanced western countries like Australia and Finland, where university websites are replete with analytics. Looking at western countries, Germany is a clear winner when it comes to university website analytics and privacy. Perhaps past issues in Germany – such as the extensive surveillance by the East German secret police, Stasi, as well as the much more recent incident of alleged tapping of Chancellor Angela Merkel's phone – have had some influence on the situation. Germany has a long tradition of discussing privacy issues and healthy skepticism towards data collection. Privacy is considered a civil right, not just an option. It would be desirable that Germany's strong privacy practices would also spread to other countries, especially to those where the use of analytics services has gotten out of hand. Something can also be learned from China's protectionist approach in preventing data leaks outside of its borders.

The websites of universities and public sector bodies in general are not a bad place to start the change for better. Public sector bodies and publicly funded institutions in particular should be exemplary by improving privacy of their websites and online services [12]. It is also important to ask the question of whether it is ethical for universities to use analytics and track users [7]. A university website should not be a profit-making commodity benefiting third parties but rather a platform for advancing the common good. Institutions with societal impact and often public funding should not be giving away their users' browsing behavior and personal data to analytics providers which use it to gain profit and power. The findings of our study show that too often, users browsing university websites and looking for information have to surrender their data to third-parties, probably without fully realizing this fact. This is also especially problematic because one party, Google, seems to be receiving this information from almost all the websites studied.

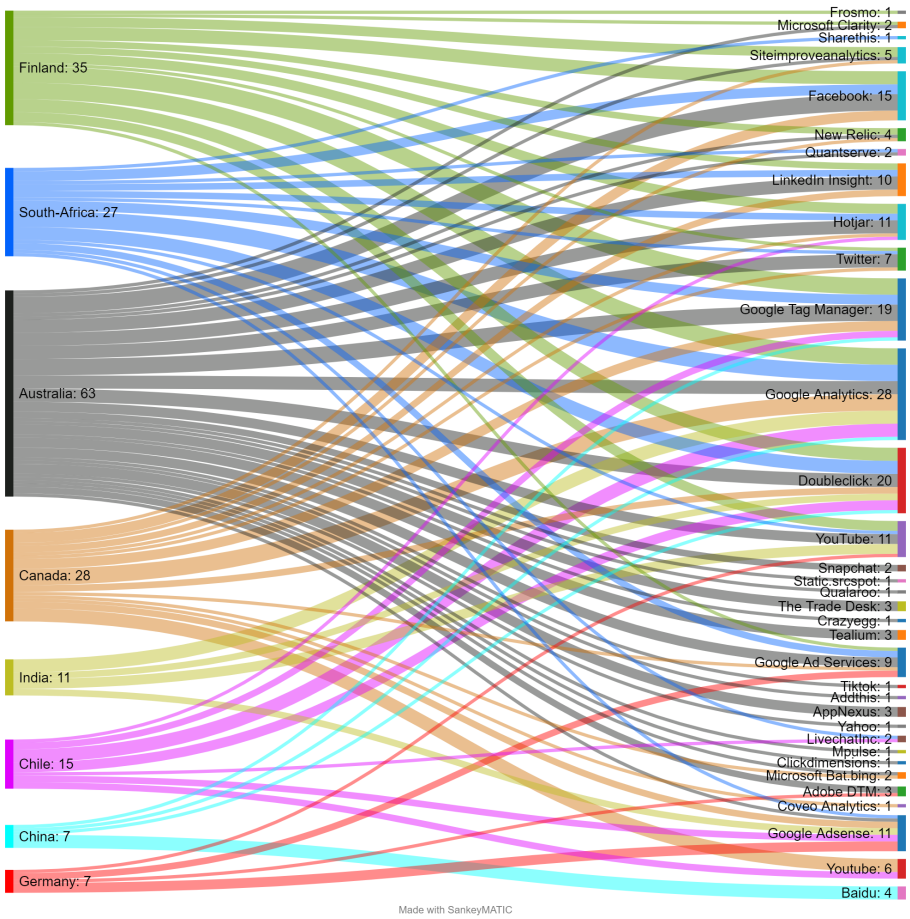

Fig. 2. An alluvial diagram showing the flow of data from the studied countries to different analytics services. The numbers indicate the found analytics service instances per country (on the the left side) and detected instances per analytics service (on the right side). The diagram clearly illustrates how Australia, for example, branches to many different third-party services, while Germany and China only use a few.

It is important to note that it is not some insignificant technical details but identifying information (such as IP addresses and user or device identifiers) that are often delivered to third parties. What is more, analytics services can also use the context information about the page the user is visiting. In some cases, this can lead to sensitive personal data leaking out. Consider, for example, a student who is searching information on how to reach an accessibility planning officer to get help with special arrangements for his or her studies. Similarly, a student could be looking to talk with psychiatrist or trying to report a harassment case. Even the fact that the student visits pages related to these themes is highly

sensitive in itself. Web developers may not always realize this dangerous connection between analytics and delicate web content. Also, as this content with the help on content management systems, university staff creating the potentially sensitive content page may not even have deeper technical skills or knowledge of the use of analytics services on the university website or how web beacons and analytics even work in general. Moreover, it may not be easy to content creators to comprehend the fact that e.g. embedding an instructional YouTube video on a web page leads to all kinds of information being leaked. Our study did not delve into what kind of personal information exactly may be leaked to third parties but this topic, along with a closer look at what kinds of potentially sensitive pages university websites include, is an interesting subject for further research.

In modern web development, analytics and social media buttons are routinely added to websites, and many web platforms and content management systems make this very easy. At the same time, developers often forget privacy – the fact that embedding e.g. social media buttons or YouTube videos on a websites costs visitors their personal data and privacy is not sufficiently taken into account. If analytics really are needed, web developers and data protection officers should consider using analytics tools that store the data locally, such as Matomo [4, 10]. Inspecting third-party requests with Chrome DevTools (like we did in the current study) or tools such as Web Evidence Collector[2] should also be an integral part of the testing phase in the web development process. Moreover, to avoid websites teeming with different analytics services, the purpose of each used service should be clearly documented and justified.

5 Conclusions

In the current study, we have provided an overview of the analytics use on university websites around the globe. The high numbers of used analytics services, especially in technically advanced western countries, raise questions about user privacy and how universities portray themselves online. While there were also some positive signs, such as adoption of local analytics in a couple of universities, the findings clearly indicate that web developers and data protection officers should pay closer attention to what data their websites send out and where. Analyzing data flows to third-party services and building websites with user privacy in mind need to become essential parts of web development.

Acknowledgements

This research has been funded by Academy of Finland project 327397, IDA – Intimacy in Data-Driven Culture.

[2] https://edps.europa.eu/edps-inspection-software_en

References

1. Chandel, S., Jingji, Z., Yunnan, Y., Jingyao, S., Zhipeng, Z.: The golden shield project of China: A decade later—an in-depth study of the great firewall. In: 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). pp. 111–119. IEEE (2019)
2. Chandler, A., Wallace, M.: Using Piwik instead of Google analytics at the Cornell university library. The Serials Librarian **71**(3-4), 173–179 (2016)
3. Farney, T., McHale, N.: Introducing google analytics for libraries. Library technology reports **49**(4), 5–8 (2013)
4. Gamalielsson, J., Lundell, B., Butler, S., Brax, C., Persson, T., Mattsson, A., Gustavsson, T., Feist, J., Lönroth, E.: Towards open government through open source software for web analytics: The case of Matomo. JeDEM-eJournal of eDemocracy and Open Government **13**(2), 133–153 (2021)
5. Griffin, M., Taylor, T.I.: Employing analytics to guide a data-driven review of LibGuides. Journal of Web Librarianship **12**(3), 147–159 (2018)
6. Heino, T., Carlsson, R., Rauti, S., Leppänen, V.: Assessing discrepancies between network traffic and privacy policies of public sector web services. In: Proceedings of the 17th International Conference on Availability, Reliability and Security. pp. 1–6 (2022)
7. Jordan, K.: Degrees of intrusion? A survey of cookies used by UK Higher Education institutional websites and their implications. https://ssrn.com/abstract=3142312 (2018)
8. Liang, F., Das, V., Kostyuk, N., Hussain, M.M.: Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. Policy & Internet **10**(4), 415–453 (2018)
9. Mayer, J.R., Mitchell, J.C.: Third-party web tracking: Policy and technology. In: 2012 IEEE symposium on security and privacy. pp. 413–427. IEEE (2012)
10. Quintel, D., Wilson, R.: Analytics and privacy. Information Technology and Libraries **39**(3) (2020)
11. Sharma, T., Bashir, M.: Use of apps in the COVID-19 response and the loss of privacy protection. Nature Medicine **26**(8), 1165–1167 (2020)
12. Thompson, N., Ravindran, R., Nicosia, S.: Government data does not mean data governance: Lessons learned from a public sector application audit. Government information quarterly **32**(3), 316–322 (2015)
13. Wambach, T., Bräunlich, K.: The evolution of third-party web tracking. In: International Conference on Information Systems Security and Privacy. pp. 130–147. Springer (2016)
14. Yang, L., Perrin, J.M.: Tutorials on Google Analytics: How to craft a Web Analytics report for a library web site. Journal of Web Librarianship **8**(4), 404–417 (2014)