Third-party services as a privacy threat on university websites

Timi Heino
University of Turku
Turku, Finland
tdhein@utu.fi

Sampsa Rauti
University of Turku
Turku, Finland
sjprau@utu.fi

Robin Carlsson
University of Turku
Turku, Finland
crcarl@utu.fi

Ville Leppänen
University of Turku
Turku, Finland
ville.leppanen@utu.fi

ABSTRACT

As the shift towards digitalization accelerates, and people increasingly rely on online services for everyday tasks, the significance of online privacy concerns keeps rising. This is also true for universities, as they continue transitioning their services and information online. In the current study, we present a technical analysis of the prevalence of third-party analytics on university websites in the EU area. We analyzed the websites of 95 universities from 19 different EU countries. The third-party analytics services employed by the studied websites were identified. The study reveals that the websites of most universities contain concerningly high numbers of analytics. These findings highlight the need to carefully choose and justify the used analytics services, and to assess the personal data websites deliver to third parties. Addressing online privacy concerns is particularly crucial for universities, as they can be considered as institutions dedicated to serving the common good.

CCS CONCEPTS

• **Security and privacy → Web application security**.

KEYWORDS

University websites, privacy, third-party analytics

ACM Reference Format:
Timi Heino, Sampsa Rauti, Robin Carlsson, and Ville Leppänen. 2023. Third-party services as a privacy threat on university websites. In *International Conference on Computer Systems and Technologies 2023 (CompSysTech '23), June 16–17, 2023, Ruse, Bulgaria.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3606305.3606335

1 INTRODUCTION

Society is becoming increasingly digitized, and people rely on online services to handle everyday tasks. Universities have also taken advantage of digitalization, as they have moved several of their services and information sources online for potential applicants, students, researchers, and the general public. To better serve their visitors and enable targeted advertising, website developers regularly use third-party analytics services on their websites to collect data on who visitors are and how they behave. Today, web analytics are also used by many universities. However, while providing many benefits, the use of third-party analytics services also raises concerns about privacy [7, 11, 13].

Many universities are publicly funded or owned by the state, and they have an important part in generating and sharing information. They are also institutions that have societal impact, grant degrees, and carry out research. While the importance of social responsibility is usually acknowledged in curricula and guidelines for researchers, universities also have the responsibility to act in a socially responsible manner as organizations. Therefore, an argument can be made that leaking personal data on website visitors to third-party analytics services is not something universities should do. Rather than leaking personal data of website visitors to third-party analytics services, universities should aspire to be at the forefront of online privacy. Given the accelerated digitalization caused by the COVID-19 pandemic, privacy issues are particularly crucial.

The presence of third party services on higher education institutional websites has not been widely studied, although Jordan et al. [6] analyze the most frequently used third-party cookies on UK higher education institutional websites. However, many university libraries appear to be interested in user privacy on their websites, regularly advising caution when embedding third-party services on library websites [1, 8]. Unfortunately, these cautionary notes seem to be overpowered by a large body of publications recommending the use of Google Analytics [2, 4, 12]. Oftentimes, privacy issues and potential harmful consequences of using third-party services are completely forgotten when touting the benefits of Google Analytics.

In order to bridge the gap in this research area, we study 95 university websites from a total of 19 EU countries. The websites are analyzed with the Website Evidence Collector tool in order to find out what kind of third-party services they employ. As a broad analysis on the prevalence of third-party services, the current study sheds light on university website visitors' privacy in different EU countries.

The rest of the paper is organized as follows. Section 2 presents the research methods of the study, including the selection process of the studied university websites and the method used to analyze the third-party services on these websites. Section 3 presents the results of the study, discussing the presence of analytics on different university websites in the EU area, as well as examining

CompSysTech '23, June 16–17, 2023, Ruse, Bulgaria

ACM ISBN 979-8-4007-0047-7/23/06.
https://doi.org/10.1145/3606305.3606335

the analytics services receiving the leaked personal data. Section 4 covers the implications and limitations of our study. Finally, Section 5 concludes the paper.

2 METHOD AND STUDY SETTING

2.1 Selection of the studied university websites

The universities in this research were chosen by reviewing the top five universities in each EU country. The selection method was as follows. Different EU countries were examined using university ranking lists and the top five were picked from the primary list. If there were less than five ranked universities in the primary list, the secondary list was used to fill the top five. Similarly, if the second list did not find enough universities, the third list was used. Before switching to another list, ranking of the previous year was also reviewed. If the three lists all failed to comprise the top five for a country, the country was dropped out from the study. The primary, secondary, and tertiary ranking lists were, respectively[1]:

(1) QS World University Ranking 2022
(2) Shanghai Academic Ranking of World Universities
(3) Timeshighereducation World University Rankings 2022

Using the method above, the top five universities in 19 EU countries were selected for a total of 95 universities. These countries included Austria, Belgium, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Netherlands, Poland, Portugal, Romania, Slovakia, Spain, and Sweden. However, 8 EU countries – Bulgaria, Croatia, Cyprus, Estonia, Latvia, Luxembourg, Malta, and Slovenia – were left out because the university ranking lists did not include 5 universities for these countries.

2.2 Analyzing third parties on the university websites

In this study, we used the Website Evidence Collector tool[2] developed and made available by The European Data Protection Supervisor (EDPS). The tool is meant to help in automating the process of inspecting websites to ensure privacy and protection of users' personal data. With the Website Evidence Collector, it is possible to collect evidence of data processing activities such as third-party web requests or the use of cookies. Before the inspection, the user sets the parameters for evidence collection, and then the inspection process is carried out automatically for the chosen websites. Another version of WEC is called Website Evidence Collector Batch (WEC-B)[3]. This tool can collect data from multiple websites in parallel and compile a joint report on the collected data from all websites. This tool was used in our study to analyze the websites in a more comprehensive manner.

In our study, we configured the Website Evidence Collector to inspect the front pages of the 95 selected university websites. The tool was configured to analyze the local and third-party connections found in the universities' front pages. From the reports of the front pages, all the local hosts, that is, any web addresses that were subdomains of the studied university websites, were collected. Next WEC-B was configured to inspect the front pages and the collected subdomains in parallel to generate a more comprehensive report of the studied websites. From the accessed pages, all third-party requests were listed and the found third parties were recorded. When visiting websites, the tool does not give consent to use cookies or enable tracking. This setup gives a clear picture of what kinds of third-party services are activated on university websites in case the user does not accept cookies.

3 RESULTS

Figure 1 shows the prevalence of the most popular third-party services on the studied websites. Google's services, largely consisting of web analytics, were most popular. The most astounding finding, however, is the prevalence of Google's services. Even though consent for data collection was never given in our experiments, Google's services were still present in 75 out of 95 cases (78.9%), which is a remarkably high portion of the studied websites. Meta/Facebook comes second with 31 university websites (32.6%).

Figure 2 shows the number of unique third parties (such as Google or Meta) for each studied country. Interestingly, the university websites in Lithuania (20) and Denmark (17) have a significantly greater number of third-party service providers than their other EU countries. In the case of Lithuania, for example, this means there are 4.0 unique third parties per one university website on average, even when consent to use cookies has not been granted. This number makes one wonder how so many essential or mandatory cookies and analytics services can be justified. Moving on, countries such as Greece, Hungary and Italy still have 10 unique third party service providers, averaging 2.0 per university website.

At the other end of the scale, Austria and Sweden only have 3 unique third parties – and only 0.6 third parties per university website. Still, even in these countries, Google's services such as Google Analytics, DoubleClick and YouTube have a relatively strong presence. Germany, on the other hand, seems to pay more attention to user privacy. Although the country has 4 unique third parties, these are not actual analytics services such as Google Analytics or Meta. Rather, they were mostly harmless third parties such as a German library portal.

Finally, Figure 3 represents the number of unique third-party domains for each studied country. Google, for example, uses a multitude of different domains for its services, such as *region1.google-analytics.com* and *doubleclick.net*. We can see that Danish universities connect to 42 different domains (8.4 per university on average). This is a very high number, although a large part of the domains are services run by Google and Microsoft. Lithuania comes second with 31 unique domains, also clearly standing out from the rest of the studied countries. In this comparison, Austria (3 domains), Germany (5) and Sweden (5) do very well again.

Explaining these results and clustering the countries based on the most obvious criteria is not straightforward. For example, the Nordic countries, Denmark, Finland and Sweden, all highly advanced technologically, are placed very differently on the scale. Therefore, in this study, the number of third parties on the studied websites does not seem to have clear correlation with the degree of

[1]The top five universities for each country were selected according to the rankings at https://www.topuniversities.com/university-rankings/world-university-rankings/2022, https://www.shanghairanking.com/rankings/arwu/2021, and https://www.timeshighereducation.com/world-university-rankings/2022/world-ranking

[2]https://joinup.ec.europa.eu/collection/free-and-open-source-software/solution/website-evidence-collector

[3]https://github.com/ovh/website-evidence-collector-batch

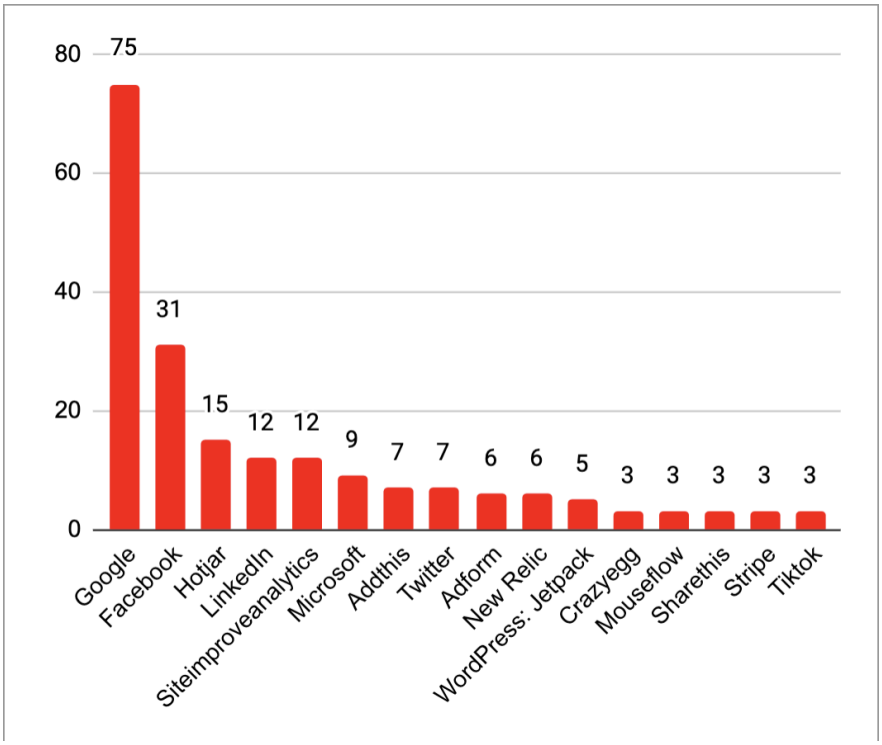

Figure 1: The most prevalent third parties found in the analysis. In the chart, each third party has only been counted once per website. For example, Google has only been counted once even though there may be several services provided by Google on one website.

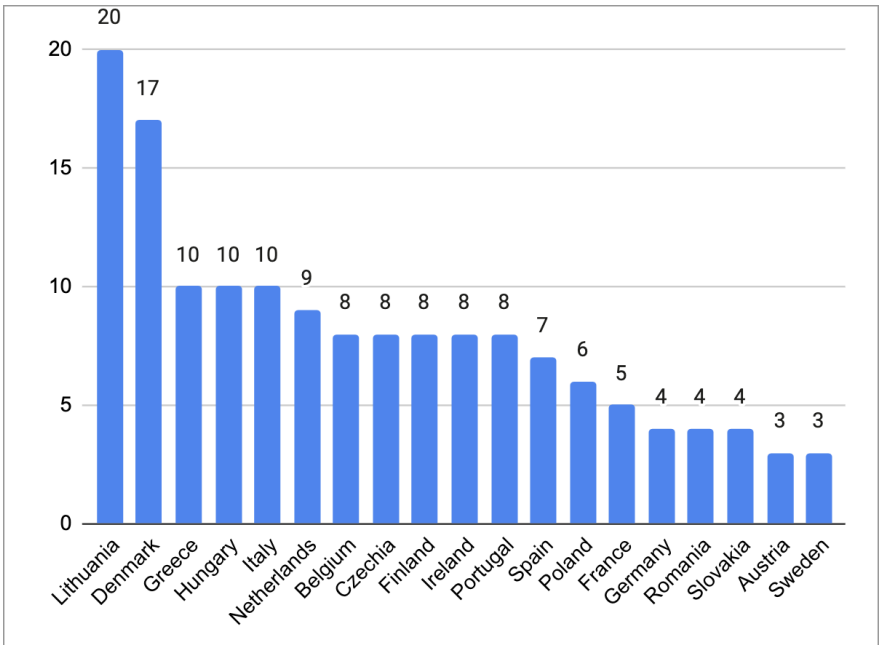

Figure 2: The number of unique third parties (e.g. Google) for each studied country.

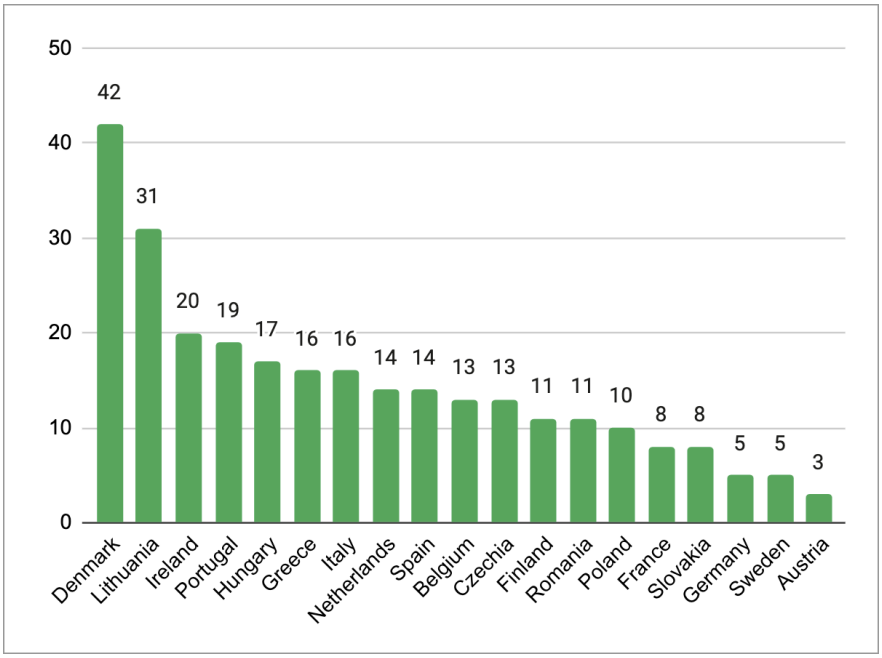

Figure 3: The number of unique third-party domains (e.g. region1.google-analytics.com) for each studied country.

digitalization. Some countries may have privacy laws and regulations that are even stricter than GDPR and there are also differences in how GDPR is interpreted and what, for instance, should be considered an "essential cookie" means and what kinds of third-party services can be present on websites even without consent. For example, Germany has the strictest data protection laws in Europe [10], which is reflected in our results. Finally, choices developers make and different platforms used to build websites can also explain differences.

Altogether, the studied EU university websites included 1.6 unique third parties and sent data to 2.9 different domains on average. In our view, these numbers can be deemed high, considering the purpose of university websites and the fact that consent to cookies was not given. Finally, it is worth noting that 20 out of 95 universities were found to also make use of a local analytics service (Matomo or Piwik), which is a positive trend in the sense that it allows the university, and not some third party, to control the collected data. However, third-party services were almost always present along with local analytics, which partly defeats the purpose.

4 DISCUSSION

In the current study, we have presented an overview of third party services found on the websites of 95 universities in the EU area. The key findings our study can be summarized as follows:

- The number of third-party services and requests to numerous related third-party domains on university websites seems to be high even when consent is not given.
- In terms of privacy, higher education websites seem to have lots of room for improvement all around the EU. It appears the level of technological development of the country does not seem to directly correlate with the number of third-party services in the current study.
- One reason for varying numbers of third-party services in various countries and web services, aside from different developers and platforms, are probably different privacy laws and regulations. There are also very diverse interpretations of what an "essential cookie" means and which third-party services are so strictly necessary that they should be used without consent.
- Over one fifth of the studied websites made use of local analytics services such as Matomo or Piwik, which is a good way to prevent personal data from falling into the hands of third parties and technology giants. However, these local services were almost always used together with third parties, and user privacy was still compromised.

When it comes to privacy issues, the websites of higher education institutes and public sector bodies in general would be a good place to start improving online privacy. Public sector bodies and institutions receiving public funding should be exemplary and strive to improve privacy of their websites [5, 9]. Universities using third-party services and tracking their website users can be considered unethical in itself [6]. An argument can be made that a university website ought to serve promoting the greater good, rather than functioning as a lucrative business benefiting third parties. Therefore, leaking visitors' browsing behavior and personal data to third parties, who use it to make profit and gain power, is not a mission of universities. The findings of the current study clearly highlight that too often, users visiting university websites have to surrender their data to third-parties. This is even true when the user does not consent to cookies and data collection, which makes data leaks even more stealthy and morally questionable. The

fact that a single company, Google, apparently receives information from a great majority of the studied websites even without consent, is also highly concerning.

The use of third-party analytics services raises many privacy concerns. When users visit websites, identifying information such as IP addresses or user and device identifiers are collected by the third-party analytics services. These services also often collect contextual information about the visited pages, which can potentially reveal sensitive personal data. Assume, for instance, that a student browsing a university website is searching for information on how to reach a psychiatrist or report a harassment case. When using the search functionality on the university website and visiting pages related to search topics, the student may inadvertently reveal sensitive information to third parties. Web developers and university personnel responsible for creating web content may not always be aware of this privacy issue and may not have the necessary technical skills or knowledge to understand how analytics services function.

Moreover, the problem is not only the use of analytics services per se, as embedding seemingly harmless instructional YouTube videos or social media share buttons on a web page can also result in data leakages. The current study did not analyze the network traffic payloads or explore what kind of personal data exactly leaks to third parties from university websites. This topic, however, as well as a closer look at what parts of the university websites handle particularly sensitive data, is a good area for further research.

In today's web development, it is common to build websites with analytics and social media buttons, and many web platforms make it very effortless to add them. However, web developers often seem to overlook privacy concerns when embedding these features on their websites. When the website contains third-party elements such as social media buttons or YouTube videos, visitors unknowingly surrender their personal data and privacy. If the use of analytics services is deemed necessary, web developers and data protection officers involved in building university websites should consider using local analytics tools like Matomo [3, 8]. This way, the data can be controlled by the university and is not sent to third parties. Moreover, when the website is being tested, examining third-party requests with tools such as Chrome DevTools should be a vital part of the process to get an understanding where the user's personal data goes. To prevent the excessive use of third-party services, each service's purpose should be clearly justified and documented.

5 CONCLUSION

In this paper, we have provided an overview of the use of third-party analytics on university websites in the EU area. The large quantity of analytics services raises concerns regarding user privacy and the way universities should operate and present themselves online. While there were some positive signals, most universities clearly use an excessive amount of analytics on their websites.

Our results highlight the need to use web analytics responsibly and ethically. Universities should be transparent about the personal data collected by the used analytics services, and only the data necessary for their intended purposes should be collected. Sensitive information should never be shared with third parties without explicit consent from the user. Web analytics can be a valuable tool on university websites when aiming to improve the user experience and achieve various organizational goals. However, analytics services have to be used responsibly and sufficient attention has to be paid to user privacy.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Adam Chandler and Melissa Wallace. 2016. Using Piwik instead of Google analytics at the Cornell university library. *The Serials Librarian* 71, 3-4 (2016), 173–179.
[2] Tabatha Farney and Nina McHale. 2013. Introducing google analytics for libraries. *Library technology reports* 49, 4 (2013), 5–8.
[3] Jonas Gamalielsson, Björn Lundell, Simon Butler, Christoffer Brax, Tomas Persson, Anders Mattsson, Tomas Gustavsson, Jonas Feist, and Erik Lönroth. 2021. Towards open government through open source software for web analytics: The case of Matomo. *JeDEM-eJournal of eDemocracy and Open Government* 13, 2 (2021), 133–153.
[4] Melanie Griffin and Tomaro I Taylor. 2018. Employing analytics to guide a data-driven review of LibGuides. *Journal of Web Librarianship* 12, 3 (2018), 147–159.
[5] Timi Heino, Robin Carlsson, Sampsa Rauti, and Ville Leppänen. 2022. 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*. 1–6.
[6] Katy Jordan. 2018. Degrees of intrusion? A survey of cookies used by UK Higher Education institutional websites and their implications. https://ssrn.com/abstract=3142312.
[7] Jonathan R Mayer and John C Mitchell. 2012. Third-party web tracking: Policy and technology. In *2012 IEEE symposium on security and privacy*. IEEE, 413–427.
[8] Denise Quintel and Robert Wilson. 2020. Analytics and privacy. *Information Technology and Libraries* 39, 3 (2020).
[9] Nik Thompson, Ravi Ravindran, and Salvatore Nicosia. 2015. Government data does not mean data governance: Lessons learned from a public sector application audit. *Government information quarterly* 32, 3 (2015), 316–322.
[10] Cody Valdez. 2016. A Glimpse at German Privacy Laws, from a Dark Past to the Strictest Data Protection Laws in Europe (but There Is Still a Long Way to Go). *Rutgers JL & Religion* 18 (2016), 430.
[11] Tim Wambach and Katharina Bräunlich. 2016. The evolution of third-party web tracking. In *International Conference on Information Systems Security and Privacy*. Springer, 130–147.
[12] Le Yang and Joy M Perrin. 2014. Tutorials on Google Analytics: How to craft a Web Analytics report for a library web site. *Journal of Web Librarianship* 8, 4 (2014), 404–417.
[13] Alexander R Zheutlin, Joshua D Niforatos, and Jeremy B Sussman. 2021. Data-tracking on government, non-profit, and commercial health-related websites. *Journal of general internal medicine* (2021), 1–3.