Third-party data leaks on municipal websites

Sampsa Rauti[1], Robin Carlsson[1], Panu Puhtila[1], and Ville Leppänen[1]

University of Turku, 20014 Turku, Finland,
sjprau@utu.fi, crcarl@utu.fi, papuht@utu.fi, ville.leppanen@utu.fi

Abstract. This paper addresses the use of third-party services on public sector websites. We focus on Finnish municipalities and their use of third-party services, analyzing network traffic from 309 municipal websites to explore what kind of personal data is transmitted to third parties. Our findings show that 51.8% of municipal websites reveal the pages visited by the user to third parties without consent. This way, sensitive data such as search terms used by website visitors can be leaked to third parties. The paper also offers recommendations to improve privacy on public sector websites. We have also contacted the municipalities with serious data leaks and recommended privacy improvements by disabling or replacing problematic third-party services.

Keywords: Municipal websites, Web privacy, Network traffic analysis, Personal data

1 Introduction

In the continuously evolving digital landscape, the internet has become an essential platform for disseminating information and providing services. Websites act as pivotal gateways which individuals can use to access a wide variety of resources. Services offered by government institutions such as municipalities are no exception to this trend. However, even with these public sector websites, convenience and utility of virtual platforms can hide the underlying privacy-related concerns [17, 8]. In particular, third-party services often integrated into these websites present severe privacy challenges.

Municipal websites are vital conduits for citizens to access essential information on public services and community events [14, 5]. Consequently, they often handle sensitive information about users and their intentions. Therefore, leaking users' navigational paths and search terms to third parties, often without explicit consent, exposes website visitors to the potential erosion of their privacy. It is crucial to investigate to what extent municipal websites share user-generated data with third-party services.

The importance of privacy in public sector web services is also reflected by the fact that the Finnish Deputy Data Protection Ombudsman has emphasized that government bodies should "carefully consider what types of tracking technologies are necessary on their websites." Moreover, authorities should make sure that users are "able to use online services provided by authorities without data on

their website visits ending up in commercial use, for example."[1] These guidelines issued by the Deputy Data Protection Ombudsman provide a strong statement on how to correctly strike a balance between functionality and protecting user privacy.

This research paper focuses on the presence of third-party services within the websites of Finnish municipalities and the implications the use of these services has for user privacy. We analyze the network traffic of all 309 Finnish municipal websites in order to gauge what kinds of personal data items they transmit to third parties. The study finds that the majority (51.8%) of the analyzed municipal websites leak URL addresses visited by the users to at least one third party. We also attempt to alleviate the situation by contacting the municipalities with the most severe data leaks and recommending that they improve their data protection by disabling or replacing the problematic third-party services.

The remainder of the paper is organized as follows. In Section 2, an overview of related research is presented. Section 3 outlines the context of the study, including the used dataset and methodology. The results of the study, analyzing the personal data transmitted to third parties from the studied municipal websites, are discussed in Section 4. Section 5 delves into the significance of our key findings and offers recommendations to enhance the privacy measures of municipal websites and other governmental online resources in general. Lastly, Section 6 brings the paper to a close with concluding remarks.

2 Related work

While the actual data collection happening in the municipality websites has not been studied that much previously, the contents of the privacy policies displayed at such websites have received more attention from the researchers. Also, other aspects of privacy and security in these kinds of websites have been investigated to a degree, and as both of these themes are closely related to our own research, we shall here provide an overview on these topics.

Schuele [13] published already in 2005, when the whole concept of personal data collection online was still quite new, an investigation into how the municipal websites collected their users PII (Personally Identifiable Information), and whether these websites informed the users about such data collection. It should be noted that PII in this context consists of directly identifying factors such as name, social security number, credit card information, email and address, and as such it is slightly different to what kind of data collection we are studying. It should also be noted that the data collection studied in Schuele's paper refers to data willingly inputted by the users, through forms in the websites, not data mechanically harvested without any direct user involvement.

Holzer et al. [10] published in 2006 a large-scale study where the global status of digital governance of largest cities was inspected. One part of this study was the assessment of the status of privacy and security at these cities' websites, for

[1] https://tietosuoja.fi/en/-/deputy-data-protection-ombudsman-issues-reprimand-for-conveying-library-search-information-to-us-based-google

which they developed a numerical scale representing the relative level of security and user privacy. The highest ranking cities were in Oceania, with Europe coming close behind, and the worst performing cities were found to be predominantly located in South America.

Beldad et al. [2] studied Dutch municipal websites and their privacy policies, and came to a conclusion that they had severe failings in regards to the Dutch data protection laws. The privacy policies often exhibited opaque language, simply did not exist or at least were not directly available for the users. In another study, Beldad et al. [1] investigated the question of what makes the people to read the privacy policies at municipal websites, and came to the conclusion that both the older age, lower level of education and less experience in using internet correlated strongly with the user actually reading the privacy policy, whereas younger, more educated and more internet-savvy users more often neglected to read these documents.

Dias et al. [4] survey privacy policies and practices of websites of 308 Portuguese municipalities. They found that 65% of the studied municipal websites used tracking tools. Furthermore, among the municipal websites that used tracking tools, 96% either did not have a privacy policy or, if they did, the privacy policy document did not address the use of tracking technologies and third-party cookies at all. While Dias et al. consider tracking and data collection in their study, our study differs from theirs as we perform a network traffic analysis to reveal details about what kind of personal data leaks from the websites.

Kautto and Hentonen [11] studied the availability and findability of the FOI (Freedom of Information) statements and privacy policies in the Finnish municipality websites in their 2017 research, and came to the conclusion that it was almost non-existent. Even in the cases where there obviously was some kind of attempt at informing the website user about the privacy issues, such as data collection, the information was extremely hard to find due to being scattered to multiple locations across the website. From the perspective of our own research this study is very relevant, as Kautto and Henttonen studied the exact same collection of 309 current Finnish municipality websites as we did.

Gomes et al. [7] conducted a research on how much the Portuguese municipality websites used HTTPS compared to plain HTTP, to determine the extent at which the users could place their trust at these services. Their results indicate that 46.1% of the studied websites did not meet even the minimum standards for security in this regard, and only 3.1% passed their examination completely.

3 Study setting and methodology

The websites of all current 309 Finnish municipalities[2] were included in the study. Our study consists of two phases. First, we performed a network traffic analysis to assess the privacy of the studied municipal websites. Second, the municipalities that were found to have websites with inadequate user privacy were informed about the matter.

[2] https://en.wikipedia.org/wiki/List_of_Finnish_municipalities

3.1 The network traffic analysis

The network traffic of each website was analyzed in order to see what third parties the websites included and what kind of personal data was leaked to these third parties. In our experimental setting, consent was never given for cookies or data collection. The experiment included opening a municipality's front page and making a search with the website's search functionality. This experiment made it possible to explore

- what third parties were recipients of personal data
- what kind of identifying personal data items were present in this network traffic to third parties
- whether the URL addresses of the pages the user visited were sent to a third party (revealing a lot about users intentions and activities on the website)
- whether the search term the user used was leaked (also potentially revealing sensitive information about the user)

To make it possible to analyze all 309 municipalities, the network traffic analysis was mostly performed with an automatic tool we have specifically built to analyze the third parties on websites, as well as URL addresses and search terms leaking to these external actors. The technical implementation of the tool has been described in more detail in [3]. For a minority of websites on which the tool failed, the analysis was done manually with Google Chrome's Developer Tools. Failures were due to the cases where, for instance, some of the studied websites, have pop-up elements such as cookie consent banners, that were overlaid with the search functionality, preventing our tool from clicking the search button. For each website, the recorded network traffic was analyzed, and the detected third parties and the details on personal data items (did URLs and search terms leak, what kinds of identifying technical details leaked) were written down.

3.2 Informing the municipalities

As the aim of this study is also have societal impact by improving the privacy of the studied municipal websites, we contacted the municipalities with most inadequate online privacy and asked them to consider removing the third-party analytics services and possibly replacing them with better options that store data locally and do not leak users' personal data to third parties. More specifically, we contacted municipalities that were found to leak URL addresses or search terms to Meta/Facebook, Google, or both companies. Although there were several other third parties on the websites as well, these technology giants were seen as the most prominent privacy threats with the capability to identify individual users and build profiles on them for commercial use.

In the sent messages, the chosen municipalities were informed about the fact their websites have been found to leak URLs and/or search terms to third parties. We also explained how severe these leaks can be; sensitive information such as a user looking for a specific medical service may leak to Meta or Google.

Furthermore, we informed the municipalities about the Deputy Data Protection Ombudsman's statement on using third-party services in online services provided by authorities and the recommendation that users should be able to visit these websites without data on visited pages ending up in commercial use. After a period of one month, we also analyzed the responses from the municipalities to see how many of them responded and whether they promised to fix the privacy issues on their websites or not.

4 Results

4.1 Leaked personal data

The data leaked by the municipal websites can be divided into two categories: identifying data and contextual data. *Identifying data* refers to personal data that third parties can use to uniquely identify a website visitor. This can include personally identifiable information (PII) such as IP addresses, device identifiers, names, email addresses, and so on. It can also be a combination of many pieces of technical data that alone do not constitute identifying information, such as operating system, browser version, and screen size. Identifying data can be used to distinguish one individual from another. If this data were to fall into the wrong hands, third parties could use it to build profiles for users and, in certain instances, even share the data with other parties.

Contextual data contains additional details about a specific individual's behavior or situation. On a municipal website, for example, the user may use the search functionality to look for information about a sensitive topic such as drug rehabilitation within a specific region. In this case, the search term is contextual data. This data is not directly identifying on its own, but it becomes much more sensitive when it is combined with identifying personal data. Contextual data provides insights into users' behaviors and preferences, and this information can be exploited – usually commercially – when it is linked to identifying information.

Of the 309 municipal websites studied, 160 (51.8%) included third-party services. Interestingly, all 160 websites with third-party services also exposed users' visited URL addresses. In total, 296 (95.8%) out of the 309 websites featured functional search capabilities. The search term leaked in 48.6% of these websites. This indicates that during our experiments, utilizing the search functionality on a Finnish municipal website led to an almost 50% chance that the search query would be disclosed to a third party without consent.

Figure 1 shows the third parties receiving visited URL addresses. Google Analytics collected URL addresses on 132 out of 309 (42.7%) municipal websites. When we include YouTube (which on 3 occasions receives the URL through embedded videos), also owned by Google, the number increases to 135 websites (43.7%). These findings underscore the extensive reach of Google in accessing sensitive personal data on public sector websites, even in the absence of user consent. Facebook/Meta comes second with URL leaks on 21 out of 309

(6.9%) websites. Other third-party services only have few occurrences in our data. Categories of these third parties include tracking website visitor behavior (Snoobi, WP Pixel), application performance monitoring and website optimization (Siteimprove, Microsoft Application Insights, New Relic), content sharing widgets and feedback tools (ShareThis, React&Share), social media and networking platforms (Twitter, TikTok, LinkedIn), tools enhancing website search functionality (Site Search 360), AI-powered content recommendations (Rek.ai), digital advertising (Adform), and hosting platform related functionality (Wix).

Furthermore, we have also included the Matomo analytics service as a third party in the cases where it was hosted in the cloud. Generally, however, Matomo can also be hosted locally so that the website maintainer controls the collected data and it is not shared with any third party [6, 12]. The cloud-based Matomo is likely to be a safe solution as well, but in principle at least some data could still be collected by the cloud provider.

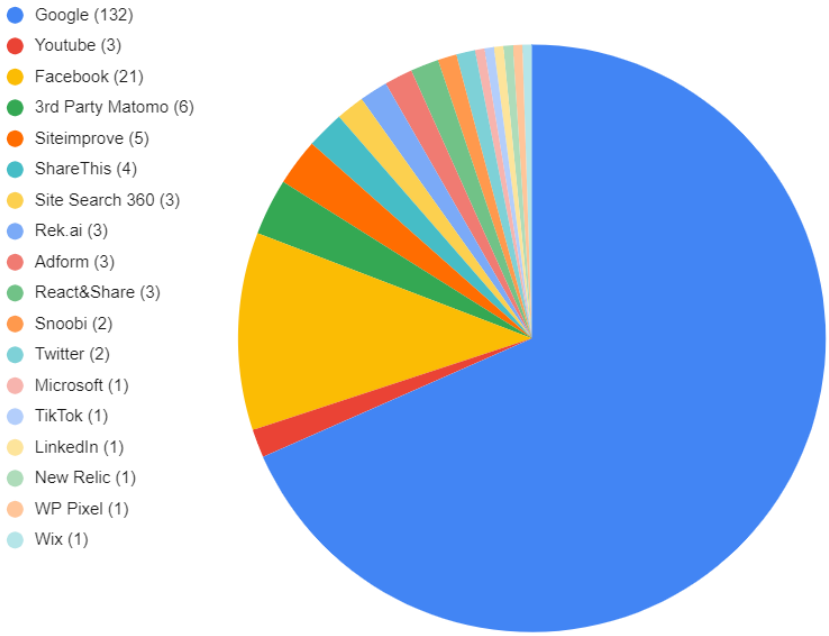

Fig. 1. The third parties receiving URL leaks from the studied municipal websites.

Figure 2 illustrates the search term leaks to third parties. The situation is quite similar to the case of URL leaks. Google Analytics receives search terms in 122 out of 309 (39.5%) websites, and Facebook/Meta is the runner-up with 18 occurrences (5.8%). The number of services receiving search terms is smaller than the number of those receiving URLs, but a significant portion of the services discussed previously still remain. The fact that nearly 40% of the municipal web-

sites studied leak search terms to Google without the user's consent is nothing short of astonishing.

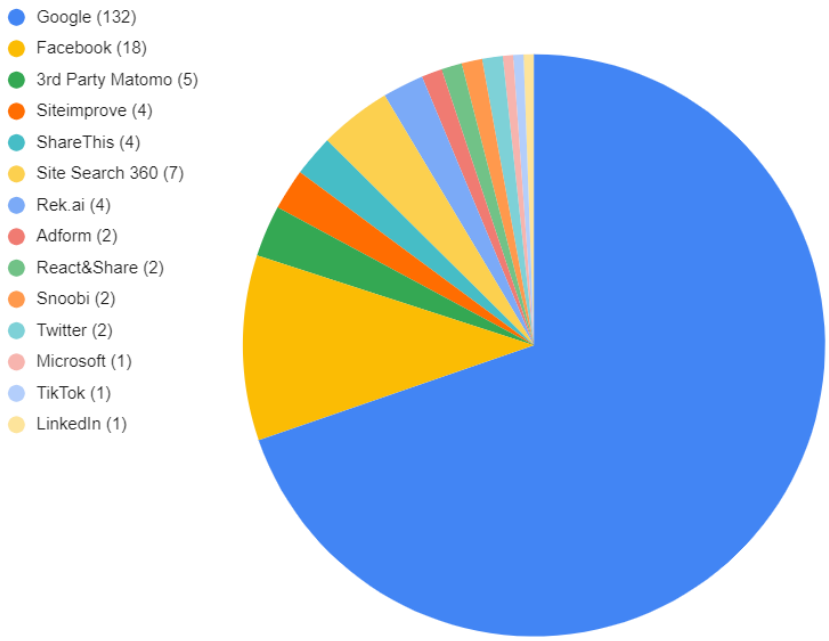

Fig. 2. The third parties receiving search terms leaks from the studied municipal websites.

Finally, we also explored whether the municipal websites asked for consent and provided a means to deny the use of cookies and data collection. Consent was requested, and there was an option to refuse cookies on 160 out of 309 websites (51.8%). Conversely, 95 websites (30.7%), a remarkably high number, did not offer the option to refuse cookies but sent the user's personal data traffic to third parties. Lastly, 54 websites (17.5%) neither asked for consent nor contained any third-party services.

4.2 Responses by municipalities

In total, 117 municipalities using Google Analytics or Meta Pixel on their websites were contacted. Additionally, 13 municipalities fixed the privacy flaws even before we contacted them – apparently, the word about our study had spread among the municipalities. We received 53 replies which is 45.3% of the contacted municipalities.

The municipalities we informed about third-party data leaks on their websites responded very positively. They expressed their gratitude for the notification and

their commitment to addressing the privacy issues. All of the municipalities that responded agreed to take the necessary measures to rectify the privacy issues on their websites.

Many municipalities indicated that they were aware of the challenges associated with using Google Analytics on public sector websites. They expressed their intentions to either remove this service or replace it with an analytics solution that offers improved privacy. Matomo was frequently mentioned as a potential alternative. While many municipalities were eager to enhance the privacy of their websites as part of an upcoming update, some also mentioned a lack of resources as a challenge.

In truth, removing or replacing an analytics service does not demand extensive resources. This was evident from the fact that several small municipalities successfully removed Google Analytics in just a few days. However, even some larger cities claimed that eliminating third-party analytics from their websites could be a substantial undertaking.

While removing an analytics service should be quite easy when done correctly, it is true that changing the used service can of course introduce some complexities in the municipality's own data collection. As one municipality noted: "The collected information contains important statistical data for planning purposes, and we need to make a plan for data extraction before deletion." Deploying the new analytics service also requires some additional planning of data collection and extraction practices.

Although municipalities were willing to improve user privacy by removing the problematic analytics services, many of them failed to do so successfully, and we had to contact them again about this issue. It appears that, despite removing the Google Analytics plugin from their content management systems, Google's tracking scripts often remained in the website's source code. This is why it is important to analyze the network traffic going to third parties to see that no tracking functionality persists.

In some cases, the lack of technical skills was also quite evident from the responses we received. For instance, a reply from one municipality mentioned that "we have deployed cookies on our website to enhance data privacy." This seems quite counter-intuitive, although it could refer to improved implementation of cookie management for better privacy.

Only one municipality was slightly reluctant to implement changes. They argued that because the use of Google Analytics is not outright illegal, and according to their interpretation, there is no strong recommendation from data protection authorities to avoid using Google Analytics, they do not see removing it as a high-priority task right now. Obviously, clearer guidelines from data protection authorities would be beneficial. As the municipality put it: "In this matter, we would have appreciated direct guidance from the relevant authorities."

Finally, 60 out of 117 (51.3%) municipalities had removed Google Analytics or Meta Pixel 3 months after we had contacted them. Some municipalities also indicated that they are going to assess the need for third-party analytics again

when the next major update of their website takes place, so more positive changes will probably happen in the future.

5 Discussion

It is evident from our results that the problem posed by the third-party analytics in regards to user privacy is not a small one. When 42.7% of these websites leak data about their users to Google Analytics without the user consent, the situation must be addressed. While there were leakages to other analytics providers too, their scale was so much smaller that it is hardly comparable to Google. On the lighter side, it is reassuring that a large number of the municipalities we contacted about these issues were very willing to make privacy improvements.

From the responses we received from the municipal authorities it is clear that there is strong variance in technological knowledge resources between different municipalities. Some of the municipal administrations we contacted were completely unaware of the issues we have brought to light, while others were already somewhat cognizant of the situation and had, at least allegedly, already discussed possible solutions to it. Then again, many municipalities also were quick to use the excuse of "it is hard to implement these changes." This is a poor excuse, as even several small municipalities which possess very limited resources were able to implement the necessary changes in a matter of days. Removing a web analytics tool from use usually means editing just a few lines of code. It may be that these kinds of objections arise out of the organizational encumbrances – the persons responding to our inquiries might not be themselves knowledgeable or responsible about the privacy matters. Also, the size and complexity of the organization – at least the case of the larger cities – may cause the information to be passed through several levels of administrative bureaucracy before reaching a person who actually can do something about the situation. Such circumstances may entice the people working in these administrations to neglect the issue in favor of "more pressing concerns", and thus come up with these excuses.

Another notable issue that came up in this investigation is the fact that almost one third of the studied municipality websites did not have proper cookie consent banners through which the user could submit consent, but nevertheless these websites leaked the user data to third parties. This kind of behavior is often a direct breach of the GDPR [16], and can potentially lead to legal consequences for these organizations.

To remedy the situation, developers working on public sector websites should conduct a thorough network traffic analysis during the testing phase of the development process, similar to what we have done in this research. This helps to assess whether there is any sensitive personal data being leaked to third parties. Conducting such a technical evaluation requires neither special expertise nor tools developers would not already have at their disposal. Carrying out this examination for a single website does not demand great investment of time either.

In choosing web analytics it would be wise to favor tools which can be deployed locally, for example Matomo, to minimize the potential of leakages. The

great irony of the situation is that it has been proven time and again that Google Analytics is a significant reason for personal data leaks [15, 9], both in terms of the amount of data being leaked and due to the nature of this data, yet it is also the most used analytics tool globally. Being the largest analytics service provider, it has also become the go-to choice for most organizations, while at the same time its use both jeopardizes the privacy of the website users and the legal standing of the website proprietors to a great extent.

Awareness about the importance of online privacy among municipal officials and the general public should be increased. This, way the importance of adhering to data protection laws and regulations would also become more apparent. It would also be beneficial for data protection authorities to offer clearer guidelines regarding the use of third-party services on public sector websites.

6 Conclusions

In this study, we have taken a survey on the user privacy aspect of the municipal websites, and found it wanting. In total, 51.8% of all municipalities in Finland leak personal data to third parties, most often to Google Analytics, even when the user has not consented to data collection. Considering that these websites are in actuality part of the Finnish governing infrastructure and as such elementary services provided for citizens, the situation can be considered bleak. This is especially the case, as the Deputy Data Protection Ombudsman has taken the stance that such essential website infrastructure should not deploy tools that collect the user data for commercial purposes.

On the brighter side, a large number of the municipalities we contacted about our findings received our findings well, and committed to improving their data collection practices. Over half (51.3%) of the municipalities we contacted also removed third-party analytics from their websites. While this is reassuring, our findings underscore the issue of a lack of technical expertise within municipal administrations, which is concerning in a technologically advanced society like Finland. As for the future research possibilities, we plan on conducting further reviews on how the privacy situation of the Finnish municipalities evolves, and whether the considerations we have brought forth are addressed. Our findings could also be compared to the other EU countries, for example, to find out whether the situation is similar elsewhere.

Acknowledgements

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

References

1. Beldad, A., de Jong, M., Steehouder, M.: Reading the least read? indicators of users' intention to consult privacy statements on municipal websites. Government Information Quarterly **27**(3), 238–244 (2010)

2. Beldad, A.D., De Jong, M., Steehouder, M.F.: When the bureaucrat promises to safeguard your online privacy: Dissecting the contents of privacy statements on Dutch municipal websites. Government Information Quarterly **26**(4), 559–566 (2009)
3. Carlsson, R., Puhtila, P., Rauti, S.: Towards an automatic tool for detecting third-party data leaks on websites. Accepted to the 10th Workshop on Software Quality Analysis, Monitoring (SQAMIA2023) (2023)
4. Dias, G.P., Gomes, H., Zúquete, A.: Privacy policies and practices in portuguese local e-government. Electronic Government, an International Journal **12**(4), 301–318 (2016)
5. Feeney, M.K., Brown, A.: Are small cities online? content, ranking, and variation of us municipal websites. Government Information Quarterly **34**(1), 62–74 (2017)
6. 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)
7. Gomes, H., Zúquete, A., Dias, G.P., Marques, F.: Usage of https by municipal websites in portugal. In: New Knowledge in Information Systems and Technologies: Volume 2. pp. 155–164. Springer (2019)
8. 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)
9. Heino, T., Rauti, S., Carlsson, R., Leppänen, V.: Study of third-party analytics services on university websites. In: International Conference on Hybrid Intelligent Systems. pp. 1284–1292. Springer (2022)
10. Holzer, M., Kim, S.T.: Digital governance in municipalities worldwide (2005): A longitudinal assessment of municipal websites throughout the world (2006)
11. Kautto, T., Henttonen, P.: Availability and findability of FOI and privacy statements on Finnish municipalities' websites. Tidskriftet Arkiv **8**(1) (2017)
12. Quintel, D., Wilson, R.: Analytics and privacy. Information Technology and Libraries **39**(3) (2020)
13. Schuele, K.: Privacy policy statements on municipal websites. The Journal of Government Financial Management **54**(2), 20 (2005)
14. Simelio-Solí, N., Ferre-Pavia, C., Herrero-Gutierrez, F.J.: Transparent information and access to citizen participation on municipal websites. Profesional de la Información **30**(2) (2021)
15. Wambach, T., Bräunlich, K.: The evolution of third-party web tracking. In: Information Systems Security and Privacy: Second International Conference, ICISSP 2016, Rome, Italy, February 19-21, 2016, Revised Selected Papers 2. pp. 130–147. Springer (2017)
16. Wesselkamp, V., Fouad, I., Santos, C., Boussad, Y., Bielova, N., Legout, A.: In-depth technical and legal analysis of tracking on health related websites with ernie extension. In: Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society. pp. 151–166 (2021)
17. Zheutlin, A.R., Niforatos, J.D., Sussman, J.B.: Data-tracking on government, nonprofit, and commercial health-related websites. Journal of general internal medicine **37**(5), 1315–1317 (2022)