Third-Party Data Leaks in the Websites of Finnish Social and Healthcare Districts

Panu Puhtila[1], Esko Vuorinen[1], and Sampsa Rauti[1]

University of Turku, 20014 Turku, Finland
papuht,etvuor,sjprau@utu.fi

Abstract. With digitalization, the use of essential social and healthcare services online has become increasingly prevalent. In this paper, we conduct a survey on the websites of Finnish social and healthcare districts and determine to what extent, if any, they leak their users' personal data to third parties through the use of the collection and tracking of user data and actions with the web analytics tools. Our findings show that 82.6% of the studied websites leaked personal data to outside actors, but the extent and contents of these data leaks varied. Our study also demonstrates that in many cases, privacy policies of the studied websites do not always report personal data items transferred to third parties and fail to adequately inform users. The cookie banners of the studied websites were also found to contain several dark patterns.

Keywords: social and health websites · data leaks · data concerning health · online privacy · third-party services · SOTE

1 Introduction

The past decade has seen the proliferation of several technologies that have brought the internet to the large segments of population, ushering in an unprecedented increase in the use of online services in day-to-day activities. As technology advances, user data, especially personally identifiable information, has emerged as a valuable resource collected and monetized through targeted marketing on websites. In practice, this data collection is enabled by web analytics tools, a class of applications that monitor and survey the website user actions. In principle, these tools are designed to help the website proprietors to better manage their services. However, at the same time these tools often leak the data they gather outside the websites they are deployed in, usually to servers run and owned by the same corporations that have created these analytics tools in the first place. Laws have been enacted to prevent third-party data leaks without the user's consent and strengthen personal privacy. All too frequently, however, enforcement is lacking, and compliance with privacy regulations falls short of addressing the magnitude of the issue they intend to rectify.

This study is part of our research on third-party data leaks on Finnish websites, conducted as part of IDA (Intimacy in Data-Driven Culture) research

project. In this paper, we study the data leaks taking place in the Finnish public sector social and healthcare district websites. As an example case, we study how their online services for the alcohol- and drug-dependent people leak personal information to third parties. We will also review the privacy policies and cookie banners of these websites and determine whether their contents are in accordance with the actual data collection happening at the website. Due to the delicate nature of health related data, it is very important that no identifying information should end in wrong hands. For example, simply the knowledge that someone has sought help because of a medical condition can be extremely harmful and stigmatizing, potentially leading to severe problems in social and professional circles.

Social and healthcare services provided by the Finnish public sector have undergone a considerable change in the past few years as the result of the nationwide project called Social Welfare And Healthcare reform [17, 13, 10, 9], during which the organization, production and government of these services was transferred to a new jurisdictional body, the newly instituted wellbeing services counties. In Finnish this reform was termed "SOTE-uudistus", and due to this we have chosen to use the term "SOTE-portals" when discussing the websites inspected in this study. SOTE is a Finnish acronym of "sosiaali- ja terveys", which directly translates to "social and health". This reorganization was completed in a tight schedule, which may be a factor in explaining our results, namely, the presence of several third-party data leaks on the websites we studied. Streamlining the production of the web services at the cost of decreased protection of user privacy may have contributed to the data leaks discovered in our study. To the best of our knowledge, the current study is the first study on the privacy of the new websites of Finnish social and healthcare districts.

The rest of the paper is organized as follows. In Section II, we take a look at the previous research conducted on similar subjects. In Section III, the methodology and the setting for our research is laid bare. In Section IV, the results of our research are presented. In Section V, we discuss the implications that can be drawn from our results. Finally, in Section VI we present the conclusions of the study.

2 Previous research

The situation of the Finnish social and healthcare reform is quite unique, and research concentrating on exactly this kind of phenomenon, that is, potential data leaks in social and healthcare districts' websites after such a reform, has not been studied in previous research. However, there is an ample body of research done on the data leaks and privacy violations in medical websites in general, as well as in a more general context of social and healthcare websites. To position our study within the context of prior research, we present some of this previous work here.

As far back as 2012, Masters [14] studied the data collection on the websites of the National Medical Association members in the USA, and came to the con-

clusion that even then, 47% of the studied websites gathered user data. Huesch studied the privacy threats in medical websites in 2013 [11], and concluded that slightly over one third of the inspected websites leaked their user information to third parties. In the same year, Brown and Levy [2] published a paper which detailed a proof-of-concept design for a tool to benchmark the information collection practices of pharmaceutical websites. During the same period, Burkell and Fortier [4, 3] conducted a study which showed how medical websites did not correctly disclose their data collection activities, leading to users falsely giving consent to such practices. They also detailed how consumer health websites constantly collected personal data on their users with analytics tools to intentionally build detailed profiles of them.

More recently, Surani et al. [16] came to a conclusion that both the websites and applications used in the mental health services often leak data and exhibit also other kinds of privacy and security risks. Zheutlin et al. [20] conducted a research on how USA-based government, non-profit and commercial health-related websites collected the user data. Friedman et al. [6] show in their recent study in which ways the tracking applications in hospital websites threaten the user privacy, also putting hospitals in a legally questionable situation.

Yu et al. [19] studied a similar phenomenon, by conducting a wide-scale automated scan on tens of thousands of hospital websites all over the world, which revealed that 53.5% of them used analytics tools that collected the data of their website users. Friedman et al. researched in 2022 [5] the use of web analytics tools by abortion clinics, and came to a conclusion that the vast majority of them (99.1%) deployed at least one analytics service which leaked the user data to third-party actors. Huo et al. [12] found in their research on the patient web portal privacy that 14% of them leaked sensitive data such as names and phone numbers outside the website domain. Schnell and Roy published a paper in 2022 [15] inspecting the hospital website design, in which they concluded that the design actively inhibited the user from finding the privacy policy, thus making the users unable to consent to the data collection practices. Wesselkamp et al. [18] developed a browser extension in 2021 to detect third-party tracking cookies used on websites. Then they used it to research 385 medical websites operating in the EU area, and whether they collected the user data. The researchers discovered that 62% of the inspected websites used web analytics tools automatically, before any consent to data collection was expressed, and 15% collected the data even if the consent was not given.

Previous research appears to suggest there is a strong correlation between the use of web analytics tools in websites and data leakages to third parties, often resulting in severe breaches of personal privacy. Further investigation is essential to comprehensively understand and address this issue. Our study offers an in-depth examination of data leaks in the websites of Finnish social and healthcare districts, delving into the nature of personal data leaked to third parties.

3 Methodology and study setting

In this study, we inspected 23 Finnish social and healthcare district websites (referred as SOTE-portals from this point onwards). The websites chosen to be studied correspond to the official social and healthcare districts, which were instituted during the Social Welfare and Healthcare reform described previously. The list of these websites is also available online[1].

The data leakages in websites were studied in the following manner. First, a researcher navigated to the website. After this, Google Chrome Developer Tools (referred to as devtools from this point onwards) were turned on, and all caches were cleared and disabled. All cookies were consented to when arriving at the website. Then the website was refreshed to ensure that no cached information would distort the test results.

The testing sequence varied between different websites due to their different designs and architectures, but the general pattern was always the same: First the researcher entered a Finnish expression for drug- and alcohol-dependence "päihdeongelmat" ("problems with intoxicants") into the search console of the website in question. Then the researcher clicked the link that seemed the most relevant in finding help to alcohol- or drug-dependency from those listed by the search functionality. If clicking this link led to subsequent links, the most relevant option for attaining help was always followed, until the link trail either came to an end or led outside the studied website.

During the procedure described above, all network traffic was recorded with the devtools and saved as HAR-files[2] for later analysis. The HAR-files were further filtered to find only the instances where HTTP requests were made outside the domain of the social and healthcare provider, in other words, to third parties. We specifically concentrated on three chief factors:

1. Whether the URL address of the visited page was leaked.
2. Whether the clicking of the link to a page about getting help for addiction was leaked.
3. Whether the used search term was leaked.

The privacy policies and cookie consent banners were read and saved either as .jpg or .pdf format for later inspection. In studying the privacy policies we paid specific attention to four different factors: Whether all third parties were named in the document, and whether the three factors discussed previously (URL leaks, link click leaks, search term leaks) were mentioned at all in the privacy policy. In addition to this, we also examined whether the cookie consent banners used in the SOTE-portals exhibited design choices which can be interpreted as dark patterns. In doing this, we used the dark pattern categorizations formulated by

[1] https://soteuudistus.fi/hyvinvointialuekartta

[2] The HTTP Archive format is a file format for recording a web browser's interactions with a website.

the European Cookie Banner Taskforce[3], from which we chose four categories to examine:

1. Absence of "decline cookies" button on the first layer of the cookie banner
2. Pre-ticked consent boxes
3. Deceptive use of colors
4. Deceptive use of contrasts

It should be noted that in determining the geographical location of the servers where the leaked information was sent to we have used the `iplocation.net` service, which combines the information from eight different IP-locator services. Because these geographical locations are difficult to pinpoint in the era of cloud services, we have accepted the location to be included in our results only if all eight sources indicated the same destination.

Finally, as this paper is about leaks of personal data, we present a definition for this concept. The one given in the GDPR of the European Union and also used by the Finnish Office of the Data Protection Ombudsman is sufficient for the purposes of our study[4]. Thus, personal data is defined as "all data related to an identified or identifiable person". By this definition, technical information such as IP addresses, device identifiers, accurate location data or any data point that identifies the user of the website counts as personal data. It must also be noted that while many technical details such as device type or screen resolution alone are not sufficient to identify someone, together these data items can be used for assisting in identification of a specific person. Hence, they can be considered as personal data.

4 Results

We found that 21 out of 23 (82.6%) of the studied websites leaked personal data to third parties. A slight majority of the inspected websites, 12 out of 23 (52.17%), leaked data only to one third-party actor. Moreover, 5 out of 23 (21.7%) leaked data to two actors, 2 out of 23 (8.6%) to 3 actors and only one (4.3%) studied website leaked data to 4 different third parties. This comprises the total of 32 leaky data collection tools between 23 websites. Two of the studied websites did not leak data at all outside their own domains. Neither of these two websites had any analytics tools in use, which indicates a clear connection between the leaking of user data and the use of web analytics tools in the first place.

In total, 11 different third-party services were found to be used between the 23 websites we inspected. The most commonly encountered third-party service and analytics tool among the studied websites was Google Analytics. This result does not come as a surprise since Google Analytics is the largest "free" analytics

[3] https://edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-cookie-banner-taskforce_en

[4] https://gdpr-info.eu/

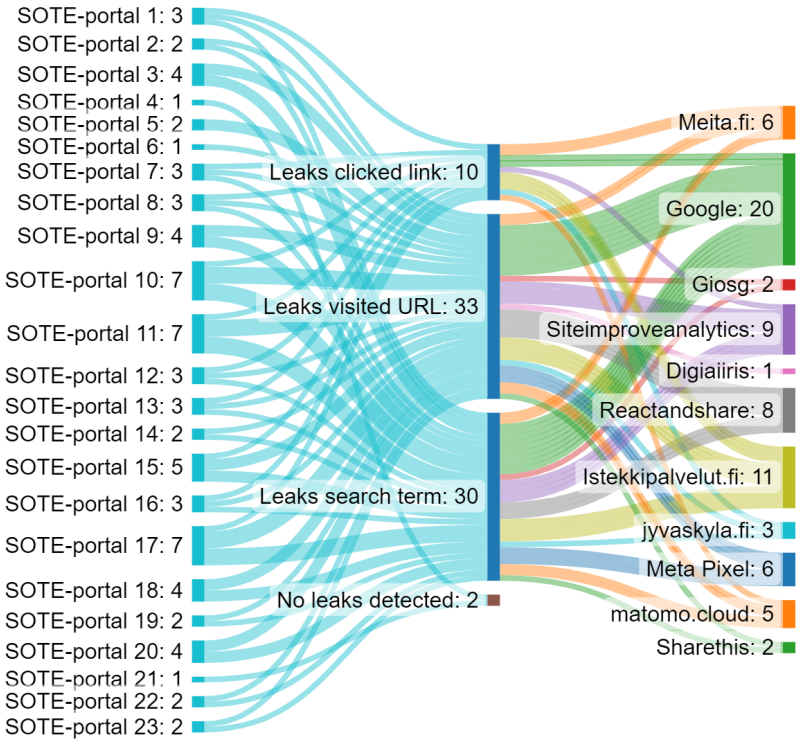

Fig. 1. Total numbers of data leaks in the SOTE-portals and the third parties the data is leaked to.

tool in circulation today, and has previously [1, 8] been shown to take the top spot as the most used analytics tool in many studies. Google Analytics was used in 9 out of 23 (39.1%) of the websites we studied in this research.

Behind Google Analytics came the services Reactandshare and Siteimproveanalytics, which were both found in 5 out of 23 websites (21.7%). Third place in popularity was taken by Istekkipalvelut.fi, which was used in 4 out of 23 websites (17.4%). Curiously, a large and often encountered website analytics tool, Meta Pixel, was quite rarely used by the websites inspected here, and was present in only 3 out of 23 (13.0%) instances. Of the remaining seven third-party tools, Meita.fi and Matomo were both encountered in two different domains (8.6%). It is important to understand, however, that Meita.fi provides a service that itself deploys Matomo as the tool to collect data. The difference here is that two of the websites had Matomo deployed locally at their own clouds, while two used the larger service package from Meita which includes Matomo analytics. Each of the remaining 5 third parties was encountered only once.

Table 1. Data leaks in the social and healthcare websites.

Website	Clicked link leaks	Visited page leaks	Search term leaks
SOTE-portal 1	X	X	X
SOTE-portal 2		X	X
SOTE-portal 3		X	X
SOTE-portal 4			
SOTE-portal 5		X	
SOTE-portal 6		X	
SOTE-portal 7	X	X	X
SOTE-portal 8	X	X	X
SOTE-portal 9		X	X
SOTE-portal 10	X	X	X
SOTE-portal 11	X	X	X
SOTE-portal 12	X	X	X
SOTE-portal 13	X	X	X
SOTE-portal 14		X	X
SOTE-portal 15	X	X	X
SOTE-portal 16	X	X	X
SOTE-portal 17	X	X	X
SOTE-portal 18		X	X
SOTE-portal 19		X	X
SOTE-portal 20		X	X
SOTE-portal 21			
SOTE-portal 22		X	X
SOTE-portal 23		X	X

It is very important to note that although Istekkipalvelut and Meita are listed as third parties here because of their own domains, they are actually in-house companies owned by public sector bodies. Therefore, data transfers to these parties are not third-party data leaks in the usual sense of the term but rather intended functionality. The same goes for jyvaskyla.fi, which is a domain of a Finnish city. Therefore, these data transfers are very different from data leaks to companies such as Google and Meta in terms of privacy risks to the user. Of course, it can still be questioned whether certain delicate page visits or link clicks, for instance, should be ever stored at all, even if the data is hosted by a trustworthy party.

The detected data transfers can be seen in Figure 1. In the figure, the numbers found besides the names of data collection tools and SOTE-portals represent the total number of all leakages in the three categories we examined. We can see that Google clearly receives the largest number of data transfers.

The most commonly leaked item of interest to our research was the URL of the visited page, which was, apart from the two websites which did not leak any data, leaked by 21 of the studied websites to at least one third-party actor. The

Table 2. Discrepancies between the studied privacy policies and actual transmitted data.

Website	Third parties mentioned	Search term	Visited page	Clicked link
SOTE-portal 1		X	X	X
SOTE-portal 2		X	X	
SOTE-portal 3		X		
SOTE-portal 4				
SOTE-portal 5	X		X	
SOTE-portal 6	X		X	
SOTE-portal 7		X	X	X
SOTE-portal 8	X	X	X	X
SOTE-portal 9		X		
SOTE-portal 10		X	X	X
SOTE-portal 11		X		X
SOTE-portal 12		X	X	X
SOTE-portal 13		X		X
SOTE-portal 14	X	X	X	
SOTE-portal 15		X		X
SOTE-portal 16		X		X
SOTE-portal 17	X	X	X	X
SOTE-portal 18		X		
SOTE-portal 19	X	X		
SOTE-portal 20		X		
SOTE-portal 21				
SOTE-portal 22		X	X	
SOTE-portal 23		X	X	

second next leaked item was the search term, which was leaked in 19 out of 23 (86.2%) websites. The information about the clicking of the link that leads to a "seeking for help" page was leaked in 10 out of 23 instances (43.5%). These findings are unacceptable, considering that we are talking about online resources meant for people who may be suffering from very stigmatizing personal problems. These results are presented in Table 1. In the table, green means that no leakages occurred, and red that they did.

Along with the data items we specifically focused on in this research, the studied websites also leaked identifying information. IP addresses and User-Agent strings, leaked with HTTP requests by default, are such data items. For user identification, Google Analytics also uses the cid number, a variable that is assigned to every unique browser-device pair. There are many other technical details that can be used as parts of the digital fingerprint for an individual user. For instance, many of the studied third parties (Google, Giosg, Siteimprovanalytics, Digiaiiris, Meta Pixel, jyvaskyla.fi and Istekkipalvelut.fi) also received the screen size of the used device, which can help in approximating the identity of the used device, and consequently, the user of the said device.

Table 3. Dark patterns found in the cookie consent banners of the SOTE-portals.

Website	Asks for consent	Reject cookies button in first layer	Pre-ticked consent boxes	Deceptive use of colors	Deceptive use of contrasts
SOTE-portal 1		X		X	X
SOTE-portal 2				X	X
SOTE-portal 3				X	X
SOTE-portal 4					
SOTE-portal 5					
SOTE-portal 6					
SOTE-portal 7				X	X
SOTE-portal 8				X	X
SOTE-portal 9				X	X
SOTE-portal 10				X	X
SOTE-portal 11				X	X
SOTE-portal 12				X	X
SOTE-portal 13				X	X
SOTE-portal 14					
SOTE-portal 15					
SOTE-portal 16				X	X
SOTE-portal 17					
SOTE-portal 18				X	X
SOTE-portal 19				X	X
SOTE-portal 20			X	X	X
SOTE-portal 21					
SOTE-portal 22					
SOTE-portal 23				X	X

Our survey of the privacy policies found similar deficiencies as were present in the actual data collection, the details of which can be seen in Table 2. The black boxes mean that the data collection of the specific type did not happen at the website and was not applicable to the privacy policy. The green color means that the privacy policy adequately described this form of data collection, and red that it did not, in the instances where data collection happened. While 15 out of 23 websites did mention all of the third parties in their privacy policies, none of the inspected websites informed the user that the information about clicking of the links or the search terms used would be collected. This result is quite frankly astounding, in a negative way. As it was so completely uniform across these different websites, it may be that the people responsible for penning down the privacy policies have not understood that these two data items can be leaked or could be related to sensitive data leaks in the first place. On the other hand, one could argue that from a legal viewpoint that mentioning the collection of just the visited URL covers at least the search term as well, since it was always leaked as a part of the URL. However, an average user can not be expected to understand that leaking an URL address may also mean that the search term leaks.

Moreover, only 9 out of 21 (39.1%) privacy policies mentioned that the visited URL would be collected, which can be considered a severe problem, as all of the sites which had any kinds of data collection did leak this piece of information. Two of the studied websites did not have any kind of privacy policy document at all. One of these also did not leak any kind of information to third parties and deployed no detectable tracking measures, which can be seen as a mitigating factor in the lack of privacy policy. The other one did leak data, however, and the lack of privacy policy or it being inaccessible to the user of the website is a direct breach of the GDPR, and it can lead to legal consequences for the parties involved. All in all, the contents of the privacy policies we encountered were inadequate, and can not be considered to give enough information to the users of the website so that they could make an informed decision about whether to consent to data collection.

Table 3 illustrates our findings in regards to dark patterns used in the design of the cookie banners. As can be seen here, the majority of the SOTE-portals asked for consent to data collection, used deceptive colors and contrasts in their cookie banners and otherwise did not use design practices that could be interpreted as dark patterns. Only one of the studied SOTE-portals lacked the "Reject cookies" button in the first layer of the cookie consent banner, and only one used pre-ticked consent boxes in their banner. Both of these attributes are considered by the Cookie Banner Taskforce to annul the consent, as the user can not be considered to be making an informed decision in either case. Five of the studied SOTE-portals did not exhibit any dark patterns in their cookie consent banner designs. Three of the studied websites – marked in black in the table – did not have any kind of cookie consent banner at all. Also, it should be noted that the singular black boxes in the column "Pre-ticked consent boxes" mean that the website in question did not present the user any option on what types of

cookies to consent to, but rather just a general "Accept/Decline cookies" option. SOTE-portal number 4 had a perfect cookie consent banner without any dark patterns, but at the same time did not use any data collection tools. It is the only portal that gets a perfect score in all of our tests.

The destinations of the leaked data, as far as we could determine, were mostly within the jurisdiction of the European Union. In 26 instances of leaky data collection tools where we could be sufficiently confident in our ability to determine the destination of the data, only once did the leaked data end up in servers outside the EU area. In this one instance, the servers were most likely located in the USA. Most common of the destinations for the data were in Finland in 10 of the inspected cases, followed by Ireland (6 instances), Sweden (5 instances) and Germany (4 instances). These results can be considered to be very good in terms of keeping the collected data within the borders of the EU.

What is noteworthy in our results was the amount of domestic data analytics providers versus global corporations we encountered during this research, as 6 out of 11 encountered data collection tools were sourced from Finnish IT companies. Two of these are well-known operators in the field of providing several kinds of information technology services for the needs of the Finnish public sector (Istekkipalvelut and Meita), and one was a service operated by a Finnish city, Jyväskylä. In the case of two websites the data collection was facilitated by Matomo which was deployed by the proprietors of these services themselves to a private cloud. The large percentage of domestic services is likely to be linked to the nature of these websites as part of the Finnish public healthcare and social service infrastructure.

In conclusion, the results we obtained can be considered concerning, but also moderately hopeful. They are alarming because it became apparent that 21 out of 23 (82.6%) of the studied websites leaked somewhat sensitive information to third parties. Every single website which had some kind of analytics tool leaked data to at least one third party. Yet, at the same time this result was hopeful, as 12 out of 23 (52.17%) of the studied SOTE-portals used no more than one tracking tool for data collection. Compared to the results of previous studies [8], which have also targeted the public sector social and healthcare related websites in Finland, this is quite promising. In the contemporary internet ecosystem where the use of web analytics tools is more the norm than an exception, and having a plethora of such third-party services deployed at every website is very common, over half of the SOTE-portals having only one such tool deployed can be considered a moderately good result in regards of user privacy. Also, the fact that the majority of the servers where the leaked data ended up were located in the EU is positive. Of course, in the cases of actors like Google and Meta, data may be transferred beyond European borders and jurisdictions, even when initially stored on EU servers.

5 Discussion

Our findings reveal that the transmission of sensitive information, including visited pages and search terms, to third parties is more common than not leaking any data. Furthermore, nearly half of the websites leaked information about users' intent to seek help, often through link clicks. These leaks are partially facilitated by unclear privacy policies that fail to adequately inform users and the presence of various dark patterns in cookie banners.

Considering the results we have obtained in this study, it is important to pose the question of whether there is any real reason for using analytics tools as parts of medical websites, whether operated by the public sector or by the private enterprise. It is extremely doubtful whether there is enough added value gained from using these tools to justify their use. The web analytics are meant, in the ideal situation, to help the website proprietors to identify how their users interact with the website. Designing web-based healthcare services in the most user-friendly ways can be achieved through means other than invasion of the privacy of the users such as usability testing. Considering the sensitive data the studied web services process, it should be obvious that such personal data should never fall into wrong hands outside the health or social service the user is interacting with. While it can be argued that the risk of abusing the data is not necessarily very large, it exists nonetheless. As this kind of exposure can lead to serious consequences for an individual's personal and professional lives, even a small risk is unacceptable.

On the bright side, it can be argued that both the relatively small amount of encountered data collection tools, and on the other hand the fact that the majority of the leaked data did not end up in servers outside the area of EU, are very good results. It is, after all, very common in leaks like this that the leaked data ends up overseas, often into servers operated in the USA. It is also quite common for the website operators to deploy many different analytics tools, which all have more or less similar data collection profiles, thus worsening the situation with leaks to numerous actors and locations.

Developers and maintainers of the SOTE-portals should ensure the user privacy by conducting a careful review of the data collection tools in use. If such surveillance tools are deemed absolutely necessary for the operation of the website in question the workings of these applications should be configured so that no leakages occur, by conducting a thorough network traffic analysis akin to what we have done in this study. Doing this demands neither special expertise, special software nor large amounts of time. In essential and critical services like SOTE-portals, there are no good justifications to not perform such testing. If a specific analytics tool cannot be configured in such a way that it would not leak the user data outside the domain it is deployed in it should be abandoned, as there are other options for web analytics that can be used without any data leaking to third parties, such as Matomo [7].

6 Conclusions

In this paper, we have studied the Finnish SOTE-portals from the perspective of user privacy, and our findings reveal clear needs of improvement. Apart from the two websites that did not use any analytics tools at all, all websites that used web analytics also transmitted sensitive personal data on their users to third parties. The SOTE-portals under examination exhibited several data leaks, including URL addresses of visited pages, search terms, and clicks on links to help-seeking pages, combined with identifying information such as IP addresses. These leaks can potentially expose sensitive personal data, such as an individual's intent to seek help for substance abuse. Since the websites examined cater to individuals dealing with health issues that can affect not only their physical well-being but also their social standing, data leaks are especially serious in this context.

Our findings indicate that severe oversight and negligence on the issue of user privacy has taken place when implementing the SOTE-portals. This situation should be remedied post-haste, as longer the current state persists, more likely it is to become entrenched as the "new normal". To fix the issues we discovered in the SOTE-portals demands that the proprietors conduct a thorough assessment of the data collection technologies they use, and discard those which can not be configured properly. Improving the privacy of social and healthcare services is not just a matter of the moral issue of exposing the users of these websites to a risk of being identified, but also legal considerations in regards to the requirements set in the GDPR of the European Union.

Acknowledgements

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

References

1. Bailey, J., Laakso, M., Nyman, L.: Look who's tracking: An analysis of the 500 websites most-visited by finnish web users. Informaatiotutkimus **38**(3–4), 20–44 (2019)
2. Brown, S.D., Levy, Y.: Towards a development of an index to measure pharmaceutical companies' online privacy practices. Online Journal of Applied Knowledge Management (OJAKM) **1**(1), 93–108 (2013)
3. Burkell, J., Fortier, A.: Consumer health websites and behavioural tracking. In: Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI (2012)
4. Burkell, J., Fortier, A.: Privacy policy disclosures of behavioural tracking on consumer health websites. In: Proceedings of the American Society for Information Science and Technology. vol. 50, pp. 1–9. Wiley Online Library (2013)
5. Friedman, A.B., Bauer, L., Gonzales, R., McCoy, M.S.: Prevalence of third-party tracking on abortion clinic web pages. JAMA Internal Medicine **182**(11), 1221–1222 (2022)

6. Friedman, A.B., Merchant, R.M., Maley, A., Farhat, K., Smith, K., Felkins, J., Gonzales, R.E., Bauer, L., McCoy, M.S.: Widespread third-party tracking on hospital websites poses privacy risks for patients and legal liability for hospitals. Health Affairs **42**(4), 508–515 (2023)
7. 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)
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. Hiilamo, H.: Why did social and healthcare services reform fail in finland? Socialmedicinsk tidskrift **97**(3), 433–441 (2020)
10. Hirvensalo, E., Asko-Seljavaara, S., Haahtela, T., Leppäniemi, A., Tukiainen, E.: Sote-uudistus ei toteuta säästöjä eikä parempaa hoitoa. Suomen lääkärilehti (2017)
11. Huesch, M.D.: Privacy threats when seeking online health information. JAMA Internal Medicine **173**(19), 1838–1840 (2013)
12. Huo, M., Bland, M., Levchenko, K.: All eyes on me: Inside third party trackers' exfiltration of phi from healthcare providers' online systems. In: Proceedings of the 21st Workshop on Privacy in the Electronic Society. p. 197–211. WPES'22, Association for Computing Machinery, New York, NY, USA (2022)
13. Jalonen, H.: Sote-uudistus: mitä, kuka, missä ja miten? (2021)
14. Masters, K.: The gathering of user data by national medical association websites. The Internet Journal of Medical Informatics **6**(2) (2012)
15. Schnell, K., Kaushik, R.: Hunting for the privacy policy – hospital website design (2022)
16. Surani, A., Bawaked, A., Wheeler, M., Kelsey, B., Roberts, N., Vincent, D., Das, S.: Security and privacy of digital mental health: An analysis of web services and mobile apps. In: Conference on Data and Applications Security and Privacy (2023)
17. Vauramo, E.: Miten sote-uudistus toteutetaan?
18. 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. p. 151–166. WPES '21, Association for Computing Machinery, New York, NY, USA (2021)
19. Yu, X., Samarasinghe, N., Mannan, M., Youssef, A.: Got sick and tracked: Privacy analysis of hospital websites. In: 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). pp. 278–286. IEEE (2022)
20. Zheutlin, A.R., Niforatos, J.D., Sussman, J.B.: Data-tracking on government, nonprofit, and commercial health-related websites. Journal of general internal medicine pp. 1–3 (2021)