Health and Technology (2024) 14:375–392
https://doi.org/10.1007/s12553-024-00819-w

ORIGINAL PAPER

Analyzing third-party data leaks on online pharmacy websites

Sampsa Rauti[1] · Robin Carlsson[1] · Sini Mickelsson[2] · Tuomas Mäkilä[1] · Timi Heino[1] · Elina Pirjatanniemi[3] · Ville Leppänen[1]

Received: 11 July 2023 / Accepted: 18 January 2024 / Published online: 3 February 2024

Abstract

Purpose With digitalization, using essential digital services such as online services has become increasingly common. These services process sensitive health related data, such as customers' prescription medicine orders, which makes ensuring stringent data privacy crucial. The current study examines third parties such as analytics services on Finnish pharmacy websites and investigates the nature and contents of data leaks on these websites.

Methods We perform an extensive network traffic analysis to reveal data leaks among 163 Finnish online pharmacies. We also study a set of privacy policies of these online pharmacies, and provide a legal analysis regarding the interpretation of the concept of data concerning health in the context of online pharmacies.

Results Our findings reveal serious data leaks among Finnish online pharmacies. We found 145 pharmacies had third-party services on their websites and only 18 did not. Out of all 163 online pharmacies, 57 (35.0 %) leaked a specific prescription medicine name connected with identifying personal data on the customer. We argue that the information concerning purchases on the prescription medicines should be interpreted as data concerning health to ensure efficient protection of customers' right to data protection and privacy.

Conclusions We hope that these concerning results will serve as a wake-up call for the developers and maintainers of online pharmacies and other web services processing sensitive data. Any third-party services incorporated into websites processing sensitive personal data should be closely inspected in terms of data leaks, or preferably not used at all.

Keywords Online pharmacies · Data leaks · Web privacy · Data concerning health · Sensitive data

1 Introduction

Today, information and communication technology is an important component in delivering essential services, providing a viable alternative to the onsite service. In particular, digital services can offer great benefits to vulnerable groups such as the people with severe health conditions, and those living in remote areas [1, 2]. The COVID-19 pandemic has further increased the use of online services and accelerated the digital shift [3]. As digital services have become more prevalent and accessible to a greater share of the population, the legislators have also strived to ensure the quality, security and equal access of these services.

As more people turn to digital services, protecting customers' online privacy becomes increasingly crucial. Online pharmacies, which we will be discussing in this article, are one very important category of services with such privacy concerns. These services handle sensitive health related information, such as prescription medicine orders, making it essential to ensure strict data privacy. In general, pharmacies are only supposed to use the customer's sensitive health information for providing good patient care and fulfilling their legal obligations. In traditional brick-and-mortar pharmacies, the need to protect the privacy and confidentiality of clients and patients has been taken into account when designing premises and customer service procedures [4]. For example, consultations with a pharmacist should be conducted in an area where sensitive information such as medication and medical conditions are not likely to be

✉ Sampsa Rauti
sjprau@utu.fi

[1] Department of Computing, University of Turku, Turku, Finland

[2] Faculty of Law, University of Turku, Turku, Finland

[3] Institute for Human Rights, Åbo Akademi University, Turku, Finland

overheard by bystanders. It has also been shown that the level of privacy provided by pharmacy facilities affects customers' willingness to utilize the service [5].

Despite the importance of confidentiality and privacy, these principles are not always adequately upheld in online versions of essential services. Third-party analytics and tracking tools are one notable threat to privacy [6, 7]. Common across the internet, this third-party functionality is widely present even in online health services [8, 9] and on websites provided by public sector bodies [10, 11]. When proper attention is not paid to privacy during the design process, these web services can inadvertently share sensitive personal information with third parties. With this in mind, the current paper studies Finnish online pharmacies and focuses on the use case in which a customer first searches for a prescription medicine on the pharmacy website and then orders the medicine in question. In particular, our experiment involves testing whether the customer's intent to order a specific prescription medicine leaks to third party services.

The contributions of the current paper are as follows. We present an extensive analysis of privacy among Finnish online pharmacies. We provide a comprehensive network traffic analysis of third-party services and data leaks present in a representative dataset of 48 Finnish online pharmacies. With automatic analysis making use of the Web Evidence Collector tool, we extend the earlier in-depth analysis to study data leaks in all 163 Finnish pharmacies that sell prescription medicines. Additionally, we also study the privacy policies of 20 online pharmacies to get an understanding how well the user is informed about data processing activities. The current paper also contains a legal analysis of why information concerning purchases on prescription medicines should be considered as data concerning health under the EU General Data Protection Regulation (GDPR).[1] Notably, our findings reveal data leaks in dozens of studied online pharmacies. We have informed the Finnish data protection authorities about this issue, and the large majority of data leaks have already been fixed. Consequently, the study has had a significant societal impact and also includes an analysis of the change in the numbers of data leaks after the issue was addressed by the authorities. Finally, to the best of our knowledge, excluding our earlier paper on third-party data leaks in online pharmacies which used a much smaller dataset of 20 pharmacies [12], this is the first study to provide an extensive analysis of data leaks in online pharmacies. The current study provides an overview on the online privacy of a broadly digitalized essential service sector encompassing the entire country of Finland.

The rest of the paper is structured as follows. Section 2 reviews the related work. Section 3 presents the study setting, data set and methodology. Section 4 discusses the findings of the study, providing an analysis on personal data the studied online pharmacies leak to third parties. Along with the technical discussion, we also study the transparency of privacy policy documents found on pharmacy websites and provide a legal analysis on the notion of data concerning health in this context. Section 5 discusses the implications of the data leaks and offers recommendations on how these serious privacy issues can be mitigated. Section 6 presents our conclusions.

2 Literature review

The literature on online pharmacies seems to mainly concentrate on the challenges of illegal pharmacies and potentially unsafe medications sold by them [13–15]. When privacy is discussed, it is usually in the context of illegal online pharmacies, where the user's personal data has a high risk of getting into the hands of criminals [16]. Common web application vulnerabilities on online pharmacy websites have been studied [17] without paying much attention to third-party data leaks. However, the issue of third-party services and web analytics in legal online pharmacies seems to be almost entirely absent from the scientific literature.

In Sophos's security blog, Vaas [18] discusses the case of the GoodRx mobile application, which is in many ways similar to our findings in the current paper. A couple of years ago, it was revealed that a mobile app helping customers to save money on prescription medicines, GoodRx, was found to leak users' sensitive data to 20 other companies such as Google and Facebook. The application shared the names of medications that users were researching, including medicines related to sensitive conditions such as HIV, mental health, fertility, and sexual dysfunction.

Zheutlin et al. conducted a study [19] focusing on the prevalence of third-party data collection within the online pharmacy sector. The findings revealed that more than two-thirds of accredited online pharmacies in the US utilized multiple third-party analytics services that collect visitors' personal data. Not surprisingly, Google Analytics and Facebook were found to be the most commonly employed data-tracking tools. While the authors expressed concerns about third-party data tracking and the possibility of health related data such as prescription medicine names leaking, they only studied presence of analytics services and did not examine what kind of data is sent to these third parties. Our current research can be viewed as an extension of their work, as we

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EU (General Data Protection Regulation). Official Journal of the European Union, L 119, 4 May 2016.

delve into greater detail regarding the nature of sensitive personal data sent to third parties.

Kuzma et al. study privacy policies of online pharmacies around the globe [20]. The authors observe that overall, only approximately 70% of the studied pharmacies has privacy policies in place on their websites. The quality and comprehensiveness of these documents were generally found to be poor. The research findings show that there is a substantial amount of work to be done in order to adequately address the privacy needs of pharmacy website visitors. The authors particularly note that there is a need for additional details and greater transparency in the studied privacy statements. Brown and Levy [21] have presented an index to measure the privacy violations of pharmaceutical companies. Their analysis however, seems to focus more on traditional personal data as names and contact information of users, largely leaving out the technical data items we discuss in the current paper.

Huo et al. [8] examined health data leaks and third-party analytics on 459 online patient portals. They found that Google Analytics was present in 14% of patient portals. Additionally, sensitive health data leaked to third parties on 9 websites, including details such as prescribed medicines and laboratory results. It seems many health care operators and maintainers of medical websites do not have sufficient technological skills to notice and solve these data privacy issues. In various studies on designing web-based medical platforms (see e.g. [22, 23]), third-party analytics are employed as a part of the suggested solutions, probably without fully appreciating the privacy risks. The findings of Huo et al., like the results of our current study, emphasize the absence of a privacy-by-design approach and underscore the significance of educating website developers and data protection officers regarding the privacy issues third parties may cause.

In general, the discussion of the use of these services on medical websites is plentiful [24–28]. The message in all of these publications is the same: third-party analytics collect sensitive personal data without properly informing the users or obtaining consent. Recently, third-party analytics have also been found to severely jeopardize the user privacy on hospital websites [29, 30]. Surprisingly, there also exists research, such as a paper by Kes et al. [31], which aims to highlight the positive impacts of gathering personal data on medical websites. The authors contend that the collection of personal data can be beneficial, even at the expense of users' privacy. They argue that this practice helps the online service to get to know the user, which ultimately leads to a more tailored user experience, also benefiting the user.

Zheutlin et al. [32] conducted an investigation of government, non-profit, and commercial health-related websites regarding their data-tracking practices. The authors conclude that it is quite common for these websites to share user information with third parties. An average of 2.11 third-party analytics services were found per analyzed website. Nowadays, seeking and accessing health information online is often not a private endeavor. This privacy issue extends beyond health-related websites and affects various critical services. Even public sector entities can inadvertently expose sensitive personal data to third parties through their websites [10].

In order to effectively inspect network traffic and study compliance concerns on websites, appropriate tools and methodologies should be utilized. Users should be able to tell whether websites really adhere to the data privacy preferences they have selected. Martinez et al. [33] propose novel algorithms and a metric to evaluate user tracking compliance and the confidence in analytics services. When it comes to technical analysis, it is essential to be able to accurately identify personal data when examining the information shared with third-party services. To tackle this challenge, Liu et al. [34] have devised a novel approach for detecting different types of personal data present in network traffic.

The legal concepts of "personal data" and "non personal data" have been studied, for instance, by Finck and Pallas from an interdisciplinary perspective, combining computer science with legal analysis. They present that perfect anonymization is impossible and the residual risk of re-identification should be accepted within the legal definition of personal data [35]. Purtova has addressed the issues relating to the broad interpretation of personal data, stating that the development of information technology will make the GDPR 'the law of everything' that is in practice impossible to comply with [36]. Accordingly, this study argues in favor a broad interpretation of the personal data notion in the context of online pharmacy data leaks.

Finally, several legal analyses have been conducted concerning the limits of "data concerning health" notion under the GDPR, proposing different methods for assessing the limits of the concept, such as Schäfke-Zell's seven step threshold analysis [37], two variables (intrinsic sensitiveness and computational distance) approach of Malgieri and Comandé [38] and Taka's four step health data assessment for well-being and lifestyle data [39]. In this study, we refer to the criteria provided by these authors as part of reasoning for the applicability of the health data notion in the context of online pharmacies.

Compared to existing literature, our study extends the research on third-party data leaks to cover online pharmacies, also presenting a technical in-depth analysis of the studied data leaks and their potential consequences. Our study can also be seen as a continuation to previous studies analyzing data leaks on medical websites. Furthermore, we outline the dangers of using the analytics services of Big Tech companies that can easily connect the action of

Table 1 Web platforms the pharmacy websites used and their popularity (out of the 163 studied pharmacies)

Platform	Number of pharmacies	Description
Platform 1	125	A generic open-source e-commerce platform for online stores
Platform 2	20	A Finnish platform specifically for online pharmacies
Platform 3	15	A Finnish platform specifically for online pharmacies
Platform 4	2	A generic cloud-based e-commerce platform for online stores
Platform 5	1	Other

users to users' real identities. Lastly, we present a novel legal analysis of data concerning health in the context of online pharmacies.

3 Study setting and methodology

3.1 Selecting the online pharmacies

A list of legal online pharmacies in Finland was retrieved from the website of Finnish Medicines Agency, Fimea. The list contained 259 entries at the time of viewing. In addition to websites, the list contained mobile applications and phone services. We pruned the list manually to only include entries with links, resulting in a new list of 186 entries.

The entries containing web links were subsequently grouped into clusters based on which e-commerce platform the online store was built on. This was done with a Python script by identifying platform-specific elements in the pharmacy websites' source code. The automatic sorting was verified by hand as the websites were each visited manually. The sorted lists were amended when necessary, for example when a pharmacy used separate platforms for the home page and store.

Finally, inactive online pharmacies and pharmacies not selling prescription medicines online were filtered out from the list. The remaining 163 pharmacies, divided into clusters by the used platform, are shown in Table 1. As we can see, there are five platforms that are given pseudonyms P1 through P5, and the corresponding clusters (sets of pharmacies with the same platform) will be referred to as C1 through C5. Plaforms 1 and 4 were generic e-commerce platforms for online stores while Platforms 2 and 3 were designed for online pharmacies specifically. One website used an unknown platform or technology, referred to as Platform 5 here.

3.2 Manual network traffic analysis

In the current study, the network traffic of 163 pharmacies was recorded. We chose to test pharmacies in clusters C2–C5 manually. Additionally, a random sample of 10 pharmacies from cluster C1 (that contained 125 online pharmacies altogether) were tested by hand to get a more comprehensive understanding of this cluster. Therefore, 48 pharmacies in total were analyzed manually.

At the beginning of each test, the cookies set by the website were deleted, after which the page was reloaded. Full consent to the use of cookies was given when asked, and the process for ordering a specific psychiatric prescription medicine (Risperdal, an antipsychotic drug primarily used to treat conditions such as schizophrenia) was initiated.

The testing sequence was started on the pharmacy front page, from which we first carried out a search. From the results page, we continued to the medicine's product page, if one was available. Some pharmacies required a sign-in at this point, after which the test was continued by navigating to the first pages of the actual ordering process. The test was finished before actually ordering the medicine. The exact testing path from page to page slightly varied between different online pharmacies and e-commerce platforms. For example, the P2-based online pharmacies provided no product pages at all for prescription medicines.

The network traffic generated during experiments was saved in HAR log files – a standardized format used to record and store detailed information about network requests and responses. The logs were then analyzed carefully and personal data was extracted from the network traffic. First, this includes identifying information such as different user and device specific identifiers, IP addresses, as well as several technical details such as operating system and its version, browser information, screen size etc. The second category is contextual information, that in this case deals with the user visiting the individual pages and viewing and ordering medicines. More specifically, in our test sequence, there are four kinds of contextual information that can be leaked to third parties:

- *L1*: The fact that user visits the pharmacy website
- *L2*: The fact that user views a specific medicine
- *L3*: The fact that user intends to order some prescription medicine (but the medicine name is not transferred)
- *L4*: The fact that user intends to order a specific prescription medicine (the medicine name is transferred)

While none of these details should leak to a third party, the fourth one, linking a specific prescription medicine to a specific user, constitutes the most severe data leak.

3.3 Automatic network traffic analysis

As mentioned previously, 10 pharmacies from cluster C1 were tested manually. However, we also analyzed the product pages of the remaining P1-based pharmacies using the Website Evidence Collector tool[2] published by European Data Protection Supervisor. The tool can list the third-party services on a website and shows the payloads of the sent HTTP requests. However, it does have some limitations, such as not being able to simulate all browsing sequences such as searches, inability to distinguish the traffic of individual pages from each other, and only collecting third-party services found on premade lists. This is why we chose to primarily analyze the chosen online pharmacies by hand.

The uniform structure of the P1-based online pharmacies made it possible to automatically generate a list of URLs to be visited for the tool, containing the chosen medicine's product page for each pharmacy. Like in the previous manual tests, full consent to cookies was also given. This automated test setup allowed us to get a broader overview of what kind of third-party services product pages had in online pharmacies built with P1, and whether data on the prescription medicine the user is interested in is sent to third parties.

3.4 Privacy policies

In addition to network traffic analysis described above, the privacy policy documents of online pharmacies were also studied. For this purpose, we randomly chose the privacy policies of 20 online pharmacies included in the manual network traffic analysis. Our aim was get an understanding whether they were written transparently. In other words, we wanted to see how well the privacy policies correspond to actual recorded network traffic. The privacy policies were read to find out whether anything about sending data on intended prescription medicine orders was mentioned. The documents were read by two researchers and conflicting interpretations were discussed until a consensus was reached.

3.5 Legal analysis and the definitions: Personal data, data leaks, and third parties

This article takes an interdisciplinary perspective to assess the results of online pharmacy network traffic analysis, combining software engineering with legal analysis. Legal analysis is particularly used in defining the limits of "data concerning health" notion in the GDPR in the context of online pharmacies. By analyzing the legislative text of the GDPR and its interpretations by the Court of Justice of the European Union (CJEU), the European Data Protection Board (EDPB) and its predecessor, the Article 29 Working Party (WP29), as well as in reference to the assessment criteria proposed in academic literature, we assess whether the transferred information, mainly the information about the purchases on the prescription medicines (*L4* of Section 3.2), can be interpret as data concerning health under the GDPR. Doctrinal study of law approach is also utilized in mapping the regulatory underpinnings of the referred definitions in the study.

Personal data is a crucial concept in our research. In the current study, this term has the same meaning as in the GDPR. According to Article 4(1) of the GDPR "personal data" encompasses any information relating to an identified or identifiable person. Put differently, it is data that can be used to identify an individual directly or indirectly. The GDPR states that an individual can be identifiable specifically based on a reference to an identifier like name, device identifier, accurate location data, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the individual. Traditionally the scope of personal data notion has been interpreted broadly (see e.g. [36, p. 41–42, 40, p. 4, 41, para. 49, 42, 43, p. 113]). In this study we adopt this broad interpretation as a starting point.

In this study, we refer to the data transfers to third parties as *data leaks* to highlight the unnecessary and harmful impact they potentially have on the privacy of online pharmacy customers, instead of referring to actual data breaches in the sense of unlawful access to personal data. By "third parties", we mean the external companies or service providers involved in processing personal data. For simplicity, in this study we use the term "third party" in a broader meaning than what has been given to it in Article 4(10) of the GDPR, including data processors, joint controllers and other recipients of data. By contrast, Article 4(10) of the GDPR defines "third parties" as entities or individuals other than the data subject, controller, processor and persons who are authorized by controller or processor to process personal data.

For clarity, we note that in our analysis we do not assess the legal roles of the parties receiving the data leaks as that would require having access to the contractual details within the agreements between the parties. Thus, the recipients of data that we call *third parties* may include entities acting as a data processor, joint controller or separate controller. While Article 4(7) of the GDPR defines a controller as the individual or entity who alone or jointly with others determines the purposes and means of the processing of personal data, under Articles 4(8) and 28 of the GDPR data processors

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

are processing personal data only on behalf of the controller based on an agreement. However, we argue that even though data processors should be obliged to confidentiality and to follow controller's instructions by an agreement, there always remains a risk that the data processor uses the data for its own purposes and, thus, unnecessary data transfers of sensitive data to the data processors should be avoided. This is particularly the case where data processors are so called Big Tech companies who have a high bargaining power over the terms and agreements concerning data processing and low transparency over their data processing practices. It goes without saying that the risk of data being used for unwanted purposes is even higher with the recipients acting as controllers as they may use the data for whatever (lawful) purposes they prefer.

3.6 Ethical considerations

Our research follows the guidelines of the Finnish National Board on Research Integrity. Our experiments did not cause any harm to the tested pharmacies and websites. For example, the medicine ordering process was always stopped before engaging with pharmacists on the chat, and no fake accounts were created during the registration process on the pharmacy websites.

Moreover, in the current study, we chose not to name the studied pharmacies or disclose their web addresses. This is done to protect the reputation of the affected pharmacies and to avoid singling them out negatively. In our view, this decision aligns best with ethical research practices. Also, the focus of our study is not on individual pharmacies but rather on a broader phenomenon, the identified serious privacy issues and the potential solutions. Finally, keeping the pharmacy names anonymous serves the dual purpose of allowing Finnish data protection authorities to conduct investigations without unnecessary interference and providing pharmacies with the space to address and rectify privacy issues found on their websites.

4 Results

4.1 Manual network traffic analysis

The third-party services found on the 48 online pharmacy websites in our manual network analysis are shown in Table 2. Websites of 4 pharmacies did not have third-party services. These pharmacies have been omitted from the table. Pharmacy 26 had 6 unique third-party services, while several other pharmacy websites included 3 services. These numbers can be considered large as we only followed a test sequence which consisted of a few different pages processing sensitive data. Google was the most frequently used third-party service (35 occurrences), followed by Pingdom (20), a Swedish website monitoring tool, and Facebook (13). A noteworthy detail is that we have excluded Giosg, which

Table 2 Pharmacies and detected third-party analytics services

	Third-party analytics services
Pharmacy 1	Facebook, Google, New Relic
Pharmacy 2	Crazy Egg, Facebook, Google
Pharmacy 3	Google
Pharmacy 4	Facebook, Google
Pharmacy 5	Google
Pharmacy 7	Google
Pharmacy 8	Google
Pharmacy 9	Google
Pharmacy 11	Google
Pharmacy 12	Google
Pharmacy 13	Google, Hotjar
Pharmacy 15	Facebook, Google
Pharmacy 16	Google, Bing
Pharmacy 17	Google, Yandex
Pharmacy 18	Facebook, Google, Bing
Pharmacy 20	Google
Pharmacy 21	Facebook, Google
Pharmacy 22	Google
Pharmacy 23	Google
Pharmacy 24	Facebook, Google
Pharmacy 25	Facebook, Google
Pharmacy 26	Facebook, Google, Bing Clarity, Omnisend, Twitter
Pharmacy 27	Facebook, Google, Bing
Pharmacy 28	Facebook, Google
Pharmacy 29	Google, Pingdom
Pharmacy 30	Facebook, Google, Pingdom
Pharmacy 31	Pingdom
Pharmacy 32	Pingdom
Pharmacy 33	Google, Pingdom
Pharmacy 34	Google, Pingdom
Pharmacy 35	Google, Pingdom
Pharmacy 36	Pingdom
Pharmacy 37	Pingdom
Pharmacy 38	Pingdom
Pharmacy 39	Google, Pingdom
Pharmacy 40	Google, Pingdom
Pharmacy 41	Pingdom
Pharmacy 42	Google, Pingdom
Pharmacy 43	Pingdom
Pharmacy 44	Google, Pingdom
Pharmacy 45	Google, Pingdom
Pharmacy 46	Pingdom
Pharmacy 47	Facebook, Google, Pingdom
Pharmacy 48	Pingdom

appeared 9 times, from this table. This is not an analytics service like other third parties but a live chat service based in Finland, enabling the customer to chat with a pharmacist, which is a service pharmacies are legally obligated to provide in Finland.

The personal data shared with third-party services includes various pieces of identifying technical data. One of the most important of these is the IP address of the device accompanying each network packet, a crucial data point when trying to identify an individual. Along with IP addresses, unique identifiers associated with devices and users, often stored in cookies on the user's computer, can be used to single out specific users. The User-Agent headers provide details about the operating system and browser in every HTTP request. There are also several other data items, such as screen resolution, window size, language and country, that can be combined to identify a specific user.

Recital 26 of the preamble to the GDPR stipulates that when assessing whether an individual can be identified, it is essential to consider all means reasonably likely to be used to a person's direct or indirect identification (see also [44, para. 302–307, 45, p. 11–14]). This encompasses all objective facts like the associated expenses and time involved in the identification process, along with the existing technological capabilities and advancements. In accordance with the ruling in the Breyer case by the CJEU, IP addresses can be considered to constitute personal data, even if one must obtain additional information from a third party to identify a specific individual [41, para. 49]. Even though it is not required that all the information enabling the person to be identified must be in the hands of a single entity [41, para. 42–43, 46, para. 45–46, 47, para. 90], we argue that Big Tech companies that as such have access to extensive amounts of data and are operating in the technology field are likely to have the means to identify a person effectively.

The CJEU has also assessed in its recent case Meta Platforms and Others the combining and using of personal data, including sensitive personal data, for behavioral advertising purposes within the context of Meta group [48]. In general the case reflects the aspect that Big Tech companies, such as Meta, receive and link data from various sources [48, para. 26–27]. Furthermore, the court maintained that where a social network user visits websites or apps to which sensitive data relate and enters information into them when registering or when placing online orders, collection of data from the visits and of the information entered by the user, the linking of all those data with the user's social network account and the use of those data by the operator, must be regarded as processing of sensitive data, where that data processing allows sensitive data to be revealed. Data collection may happen by means of integrated interfaces, cookies or similar storage technologies.

Recital 26 of the GDPR states that data protection provisions do not apply to data that has been anonymized. Google Analytics has an IP anonymization feature (IP masking) that partially or fully omits the collected IP address. The effectiveness of this anonymization method can be questioned, however. Although anonymization of the IP address is carried out, several other technical data items are delivered along with it. This makes identifying the user possible, for large analytics companies such as Google. Therefore, it is questionable whether the data can be considered anonymous "in such a manner that the data subject is not or no longer identifiable", as the recital 26 of the preamble to the GDPR puts it. Also, as the anonymization process is carried out on Google's own servers, Google is in practice processing the IP address data also in its identifiable form.

Recital 30 of the preamble to the GDPR maintains that persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as IP addresses, cookie identifiers or other identifiers, leaving traces which can be used to identify them. While the IP address is a very important piece of data when third parties seek to identify a specific user [49], it is also worth noting that by using cookies, large analytics providers such as Google and Facebook/Meta can link the actions of users on different websites to their Google or Facebook accounts. Very often, this also means that the real name of the user can easily be linked to an action such as purchasing prescription medicines. For example, Google Analytics uses a cookie that contains a unique identifier, client ID (cid), for each individual browser-device pair. The main cookie of Google Analytics (called _ga), lasts for 2 years and enables Google to distinguish users from one another.[3]

Moreover, Google Signals, which was released in 2018, enables cross-device tracking of users in Google Analytics.[4] Google Signals implements this by solution by leveraging its own data from users logged into Google accounts. Therefore, an individual using their Google account on a computer, tablet, or phone can be easily identified as the same user across different browsers and devices. In this case, too, the actions of the user can be linked to their real name. This is of course even more serious than leaking just an IP address combined with sensitive information. Meta/Facebook has also introduced a similar cross-device tracking based on Facebook accounts. Due to the factors discussed above, we argue that when it comes to the selected online pharmacies, there is a substantial risk that the leaked sensitive data can be linked to the individual using the pharmacy website, especially when it comes to technology giants.

[3] https://policies.google.com/technologies/cookies?hl=en-US.

[4] https://support.google.com/analytics/answer/9445345?hl=en.

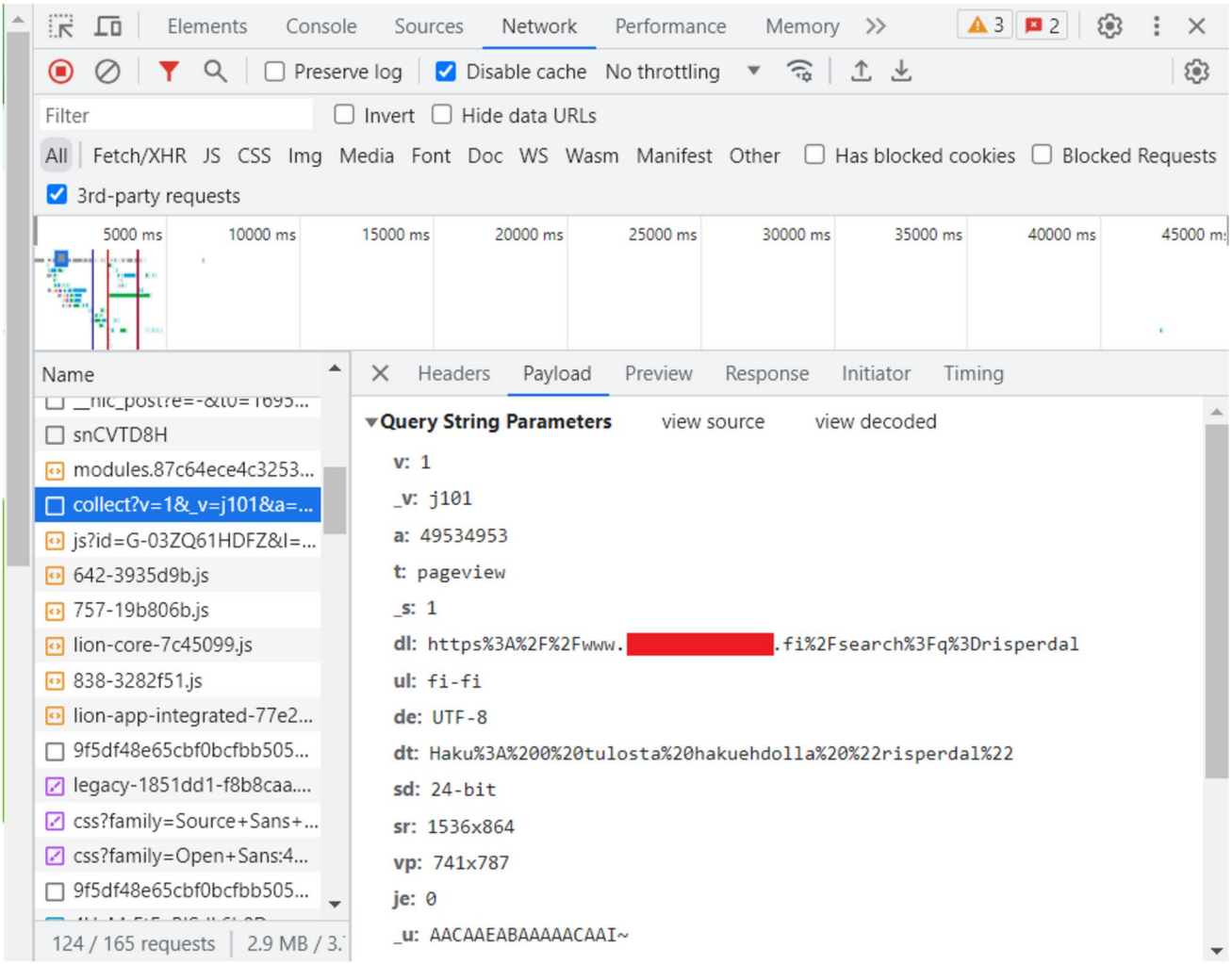

Fig. 1 A screenshot showing a medicine name (Risperdal) being leaked to Google as a part of an URL address. The website name has been omitted from the figure

While identifying the user is crucial, the most sensitive piece of information potentially leaked to third parties in online pharmacies concerns URL addresses of the pages visited by the customer (who is identified by a device identifier or IP address, for instance). There are usually two pages of interest: the product page of a particular prescription medicine, and the page for ordering prescription medicines. A visit on the product page indicates interest in the medicine in question. Respectively, visiting the order page implies that the customer has the intention of ordering medicines. These details are definitely personal data items that should never leak to third parties. When these two page visits (viewing a product and intending to place an order) can be connected to each other by a third party, meaning that the customer is intending to order a specific prescription medicine, a compelling argument can be made that sensitive data concerning health is leaking. Figure 1 shows an example screenshot of a medicine name leaking in an URL address.

We found that there are two main ways for a third party service to deduce the connection between a specific medicine and a purchase. First, the connection is quite obvious when the page visited previously (referer) is transmitted to the third party when the customer arrives at the medicine order page. In this scenario, the third party can promptly discern that the order originates from a product page for a specific medicine, which signifies a clear intention to purchase that specific prescription medicine. In most cases, the medicine name was even a part of the product page's URL address (although a working and unique product page URL without a medicine name is enough, as the third party can simply check the contents of the page and learn about the medicine). The second type of connection between a medicine and order is created when a third party does not get the information about the previous page directly, but the third party is present on both the product page and the order page. By analyzing the sequence of timestamps associated with

Table 3 The data items related to viewing and ordering prescription medicines delivered to third parties

	Product page URL sent	Intent to order sent	Order and medicine can be connected
Pharmacy 1	X	X	X
Pharmacy 2	X	X	X
Pharmacy 3	X	X	X
Pharmacy 4	X	X	X
Pharmacy 9	X	X	X
Pharmacy 11		X	
Pharmacy 12	X	X	X
Pharmacy 13	X	X	X
Pharmacy 15	X	X	X
Pharmacy 16	X	X	X
Pharmacy 17	X	X	X
Pharmacy 18	X	X	X
Pharmacy 20	X	X	X
Pharmacy 21	X	X	X
Pharmacy 22	X	X	X
Pharmacy 23	X	X	X
Pharmacy 24	X	X	X
Pharmacy 25	X	X	X
Pharmacy 26	X	X	X
Pharmacy 27		X	
Pharmacy 28		X	
Pharmacy 29		X	
Pharmacy 30		X	
Pharmacy 31		X	
Pharmacy 32		X	
Pharmacy 33		X	
Pharmacy 34		X	
Pharmacy 35		X	
Pharmacy 36		X	
Pharmacy 37		X	
Pharmacy 38		X	
Pharmacy 39		X	
Pharmacy 40		X	
Pharmacy 41		X	
Pharmacy 42		X	
Pharmacy 43		X	
Pharmacy 44		X	
Pharmacy 45		X	
Pharmacy 46		X	
Pharmacy 47		X	
Pharmacy 48		X	

page visits, the third party can infer that the two pages have been visited in succession. This leads to the conclusion that with a high probability, an order has been placed for the medicine that was viewed previously. Although this latter case requires more analysis, in both of these cases the third party can be quite confident that the customer has a definite intent to purchase a specific prescription medicine.

Table 3 shows what kinds of sensitive data is delivered to third parties as the user proceeds to order a prescription medicine on the studied pharmacy websites. Out of 48 websites, the ones that transferred any kind of sensitive data to third-party analytics – 41 websites in total – are presented in the table. The columns in this table correspond to the data leak types L2–L4 presented in Section 3.2. The first column of the table displays whether the URL address of the medicine product page the user is viewing is sent to third parties. The second column designates whether the intent to make a medicine order was leaked (note that this does not mean a specific medicine name leaked). Last but not least, the third column indicates whether the intention to make an order can be connected to a specific prescription medicine.

We can see that the URL of a medicine product page was leaked in 18 cases out of 48 (37.5%). In 41 cases (85.4%), the intention to make an order was leaked to at least one third party – 23 of these data leaks did not have information on the specific medicine which was ordered. The specific prescription medicine name could be connected to an order made by a particular user in 18 cases (37.5%). These findings give reason for great concern and also warrant further study with a larger set of pharmacies (carried out in Section 4.4).

The different platforms used to build the studied pharmacy websites are indicated in different colors in Table 3. The table shows that when pharmacies from cluster C1 (shown in red) contain third-party services, they leak several types of sensitive data. As P1 is a generic e-commerce platform, it is not really planned for handling sensitive data. When building webstores with this platform, the integration of Google Analytics to the website is made effortless. However, it is worth noting that some pharmacies in C1 (e.g. Pharmacies 6 and 10) have chosen not to use any analytics. Cluster C3 (shown in yellow) exhibits a pattern similar to many pharmacies in C1, where all studied data items are usually leaked to third-party analytics services. This is somewhat surprising, as it is a platform specifically designed for online pharmacies. P3 highlights both employing analytics services and the importance of security in its advertising. Unfortunately, these goals appear to be in conflict when it comes to practice. C2 (shown in green), also an online pharmacy platform, fares better. P2-based pharmacies do not have prescription medicine product pages, so their address cannot be leaked. This design choice also protects data on a specific prescription medicine from leaking in the order phase. Nevertheless, third-party services are used on every single website built on this platform. Finally, clusters C4 and C5 (in blue and white) are each represented by only one pharmacy that has analytics. These pharmacies (like the ones in C2) send data on intent to order to third parties, without leaking a medicine name.

A notable finding in our study is also that pharmacy customer's sensitive health related information was likely transferred outside the EU or EEA at least in some cases. For example, the name of prescription medicine was leaked to Bing/Microsoft (3 cases) and Yandex (1 case) which were found to likely be located in North America

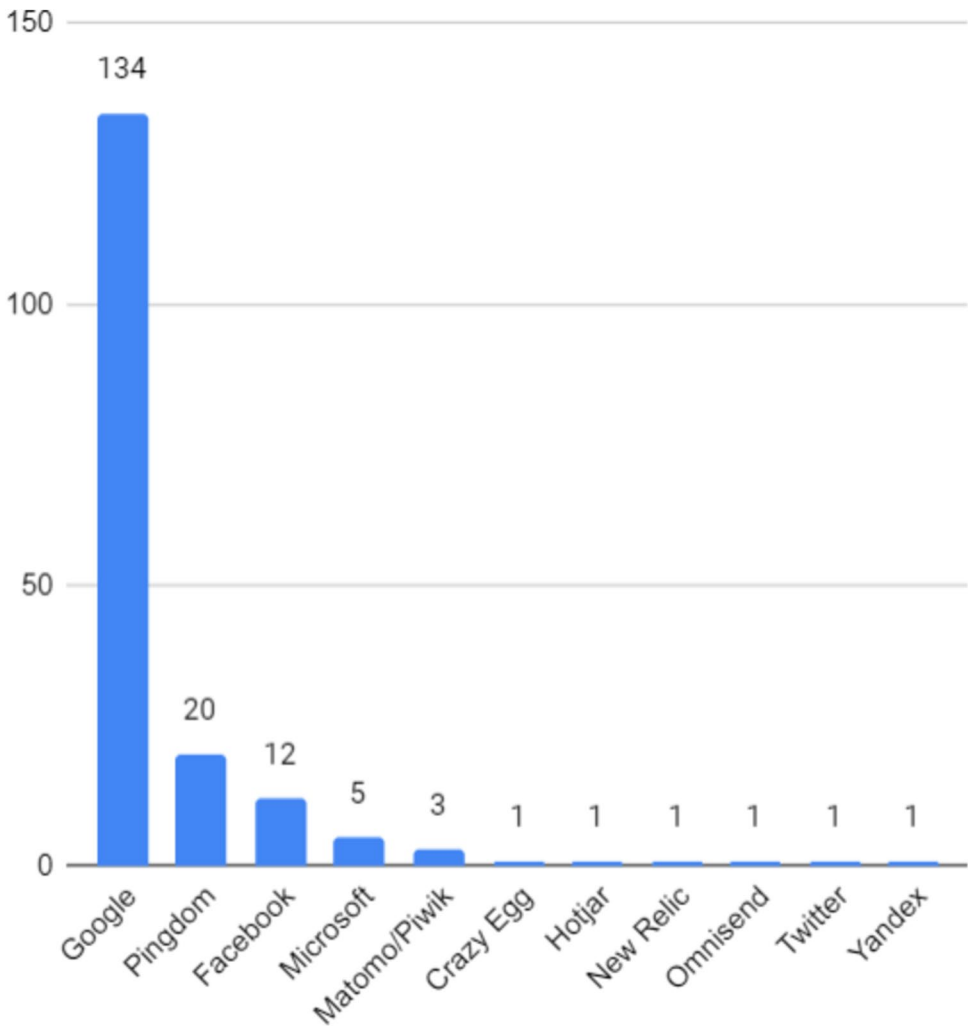

Fig. 2 The presence of third-party services on 163 Finnish online pharmacy websites

and Russian Federation, respectively.[5] However, it is also worth noting that third-party service providers can have access to the data located in Europe from outside the EU or EEA which is regarded as a data transfer. Our analysis only covers the client-side functionality of websites. Personal data transfers out of EEA are subject to special safeguards under the GDPR as they may involve noteworthy risks to the data subjects e.g. due to access requests by public authorities (see e.g. [50]). This has lead to data protection supervisory authorities to reject, for example, the use of Google Analytics in many cases (see e.g. [51–53]) and to cast a record fine of 1.2 billion euros to Meta Platforms Ireland Ltd [54]. Transferring sensitive health related data out of EEA definitely raises concerns about whether an equivalent level of protection has been granted to the data as required by the GDPR.

4.2 Automatic analysis with a larger dataset

To get a better understanding on the complete set of Finnish online pharmacies, we ran an automatic analysis including all 163 pharmacies. As a result of this more extensive analysis, 145 pharmacies had third-party services on their web pages and only 18 did not. Out of all 163 online pharmacies, 57 (35.0%) leaked the medicine name either on the product page or during ordering. It goes without saying that this number is remarkably high – over one third of online pharmacies leaked the customer's sensitive health data to third parties!

Figure 2 shows the third-party services found in the Finnish online pharmacies. While Google's top place in the diagram is not surprising, 134 pharmacy websites is a very high number. Google's services are present on 81.6% of all studied online pharmacies. Pingdom, which seems to be

[5] We used the iplocation.net service to locate the recipients of personal information, which in turn used 8 different geolocation services to determine the locations of servers. In the era of cloud services, however, it can often be challenging to accurately and certainly locate the recipient of data.

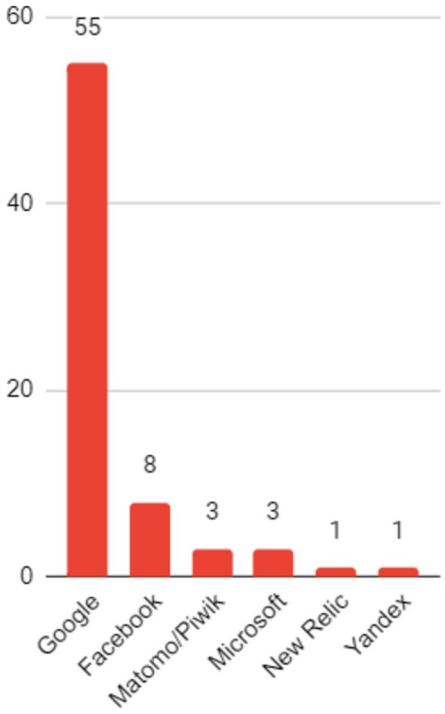

Fig. 3 The cases in which third parties received information on a specific medicine viewed or ordered by the user (counted once per pharmacy) on 163 Finnish online pharmacy websites

an analytics service of choice on Platform 2, has 20 occurrences, while Facebook/Meta has 12 and Microsoft has 4. Finally, while Matomo/Piwik is an analytics service which is often used locally and retains the pharmacy's control of data (without surrendering it to a third party), in our dataset this was not always the case. We have counted the instances of Matomo/Piwik analytics service in the cases where it was not used locally, but instead, data was sent to an external domain not owned by the pharmacy.

Figure 3 shows the cases where the data on a specific prescription medicine is leaked to a third party, either on the product page or when ordering the medicine. The fact that the medicine viewed or ordered by the user is leaked to Google on 55 pharmacy websites is astounding. This is over third (33.7%) of all Finnish online pharmacies! It also means that in the set of all 57 pharmacies that leak sensitive health data, Google's services receive this data in 96.4% of the cases. Compared to Google, Facebook's presence as a receiver of sensitive data is not that high (8 cases, 4.9% of all pharmacies). It is also worth noting that in this second diagram Pingdom has disappeared – despite the relatively high presence, it never collects information on the exact prescription medicine.

Figure 4 shows the connections between the used platforms and medicine name data leaks. Only two platforms, P1 ans P3, ever leak the medicine name. In total, 35.2% (44/125) of the P1-based pharmacies leak the medicine name, while 86.7% (13/15) of the P3-based pharmacies leak the medicine name. In both clusters C1 and C3, the use of Google's services is the largest culprit for data leaks – causing 42 leaks in C1 and 13 leaks in C3. Clusters C2, C4 and C5 were not found to have any medicine name leaks in our experiments.

4.3 Change in the number of third parties

In addition to scientific pursuits, one of the objectives of the current study was to improve the privacy of Finnish online pharmacies. To this end, our findings were reported to the Finnish data protection authorities. Dozens of online pharmacies are still being investigated as of this article being written. Initially, the privacy issues were discovered and reported to authorities in April 2022. The data protection authority started an investigation on a larger scale in November 2022, and at this point, we revisited the situation. Finally, we carried out the last analysis in March 2023 when all pharmacies had been given a fair possibility to improve their websites.

Figure 5 shows the number of third-party services on all studied Finnish online pharmacy websites ($N = 48$) as a function of time. Looking at the initial situation compared to March 2023, the number of third parties has dropped from 79 to 18, and is now less than fourth (22.8%) of the initial number of third-party services. Figure 6, on the other hand, shows the number of instances where information on a specific medicine leaks to third parties. Again, comparing the initial situation with the current one we can see that the drop is even more drastic here: from 33 leaks to 4. The number in March 2023 is only 12.1% of the initial number. This shows the significant impact our findings have already had on the privacy of the analyzed pharmacies.

Despite the good outcome, the charts also show that there are individual cases where new privacy risks have been introduced even after the authorities have been involved. For example, in Fig. 6 we can see that Snapchat has suddenly popped up as a new destination for data on prescription medicine names. Our goal in future is to keep monitoring the situation, and if necessary, work with the pharmacies to reduce the number of prescription medicine leaks to zero.

In the figures, it is worth noting that one pharmacy can contain numerous third parties and several leaks, so the numbers reflect data leaks rather than unique pharmacies. Also, three (3) online pharmacies have been on a long maintenance break after the data leaks were discovered, which may have minor effects on the reported numbers.

4.4 Analysis of privacy policies

We analyzed privacy policies of 20 pharmacies included in the manual network traffic analysis. Among these online

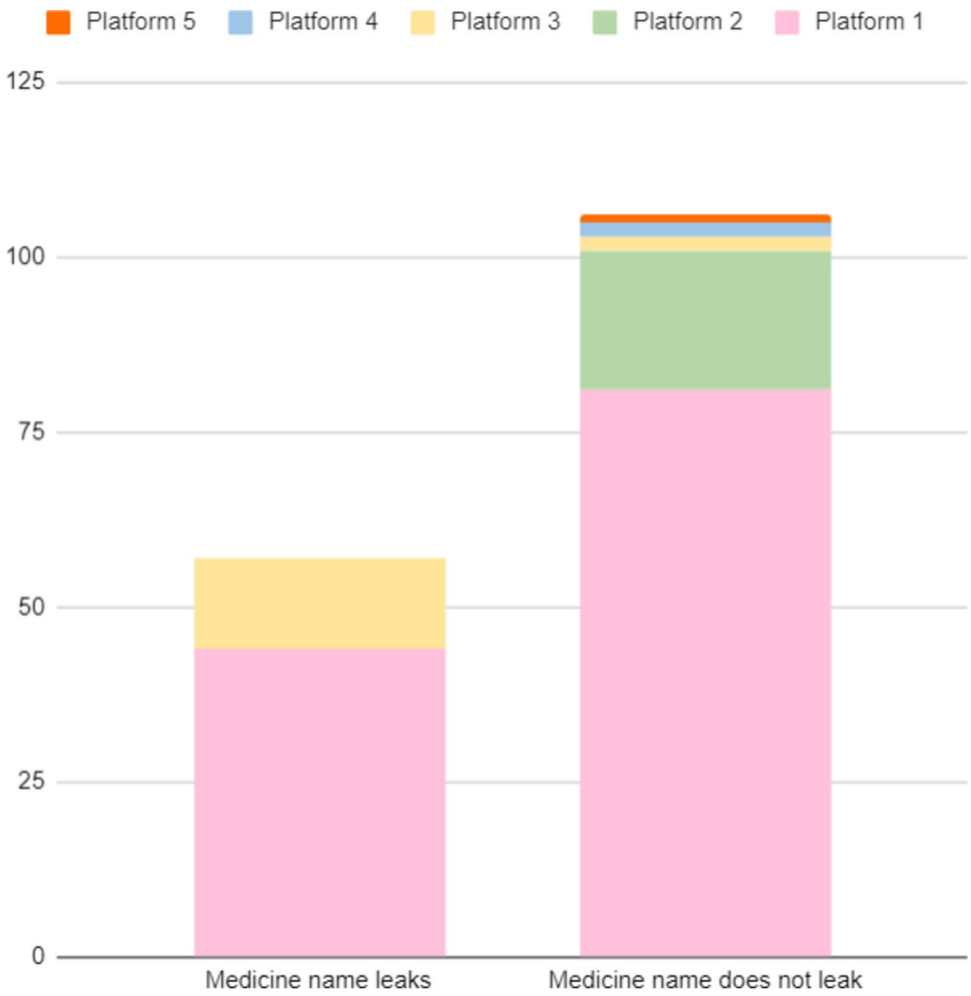

Fig. 4 A Comparison of the used platforms: the pharmacies leaking medicine name and pharmacies without medicine name leaks

pharmacies, 16 were found to send personal data to third-party services. In the privacy policies on these websites, 10 out of 16 pharmacies denied sending any data on products (medicines) users have viewed or ordered. This, of course, completely contradicts our network traffic analysis which proves that many of these websites share data about prescription medicines with third parties.

Three of the analyzed pharmacies admitted in their privacy policies that these kinds of transfers to third parties can happen. However, none of them explicitly stated that information about intended prescription medicine orders is handed over to third parties. The used expressions were more subtle, stating that the personal data that is collected on the website can, among several other personal data items, include data on visited URLs or ordered products. It is left to the user to realize that the visited URL can also mean a product page of a specific prescription medicine. The user may not also immediately realize that data on an ordered product could mean sensitive data on a specific prescription medicine.

In another section of these privacy policies, it was then explained that the collected personal data can also be shared with third parties. One of the analyzed privacy policy documents (Pharmacy 11) did not state clearly whether personal data can be sent to third parties. Although the cookie banners are outside the scope of the current study, it is worth noting that Pharmacy 1 very clearly explained in its cookie consent banner that data on prescriptions or medication is not collected. This statement and our findings blatantly contradict each other.

Based on this analysis, it is obvious that the majority of the analyzed privacy policies did not adequately inform the customer about the fact that their sensitive health data is turned over to third parties. Curiously, in several cases privacy policy documents had clearly been directly copied from other online pharmacies. On many occasions, large sections of the documents were identical and only a few details were changed in the text, without sufficient attention to the applicability to the specific pharmacy and contents of the document in general. Indeed, numerous

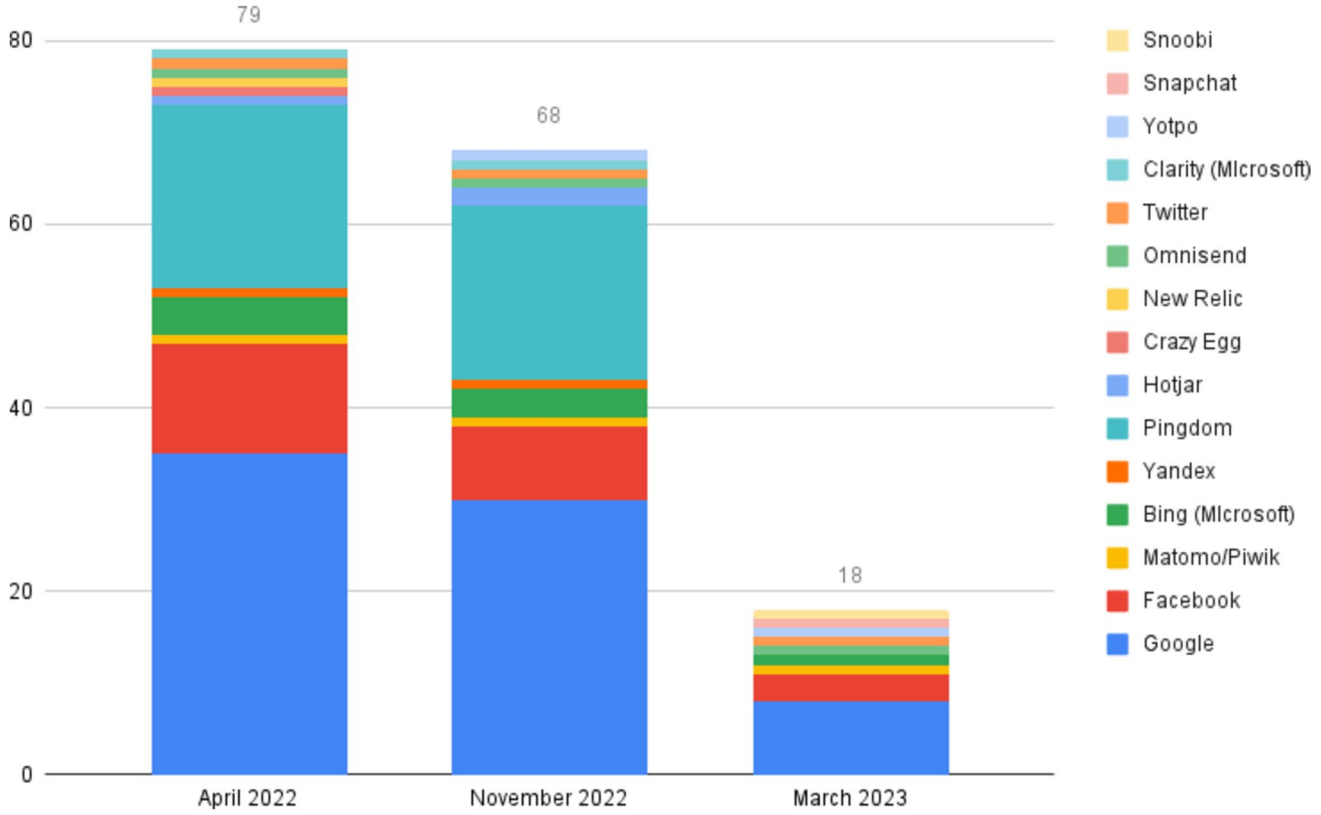

Fig. 5 A change in the number of third-party services on the Finnish online pharmacy websites

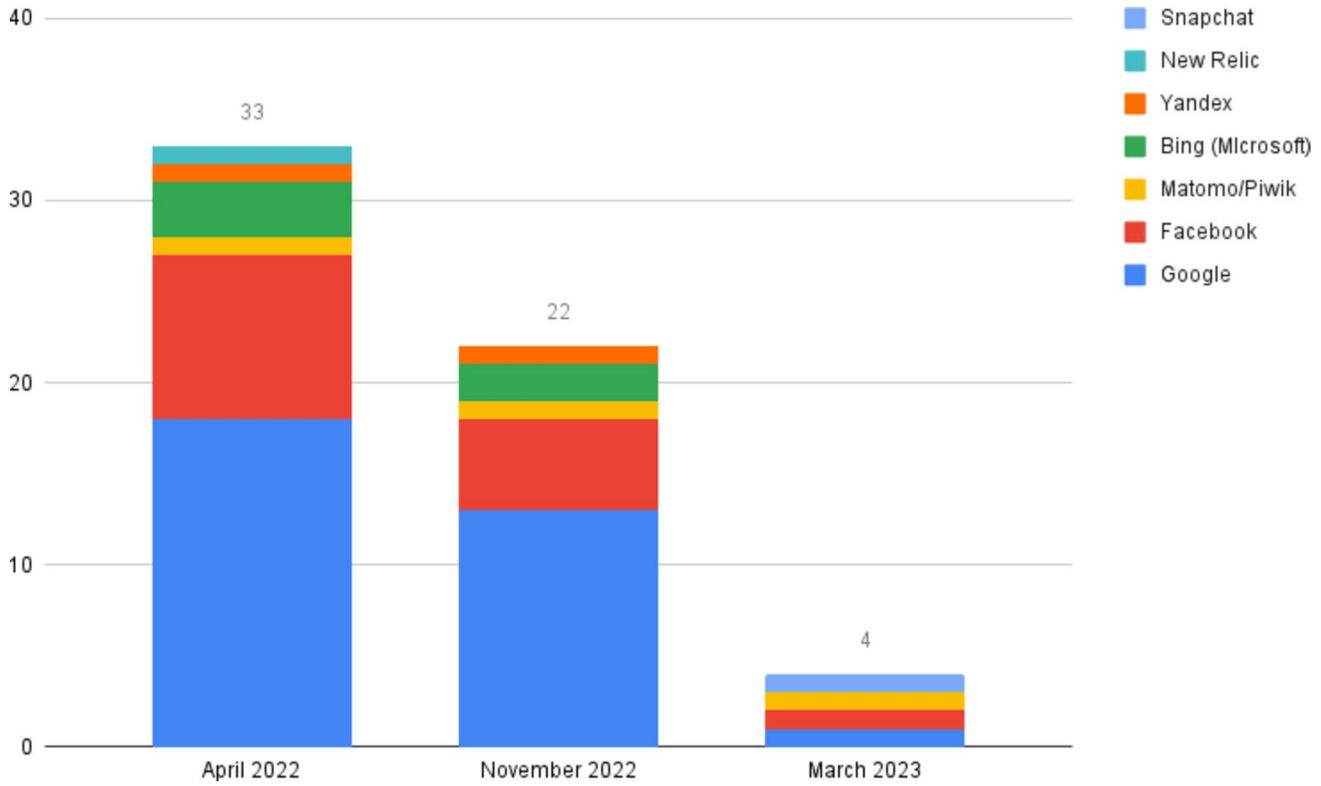

Fig. 6 A change in the number of instances where a specific medicine name leaks to third parties

privacy policies even had the same misspellings at several points in the text.

The lack of transparency in the studied privacy policies in terms of found data leaks is not that surprising, however, when we take into account that the web developers and data protection officers have most likely been totally unaware of these leaks before they were informed about the situation.

4.5 Legal analysis: Data concerning health

This section discusses why information concerning purchases on prescription medicines should be considered as data concerning health under the GDPR. In the presented study, it was discovered that the URL addresses of the websites on which prescription medicines could be ordered were sent to analytic service providers among with other identifiable data such as IP address and device identifiers. In some cases, the said URL address directly contained the name and the details of the prescription medicine. The combination of collected data also revealed the person's intention to buy the medicine in question. Information concerning a person's medication can lead into interpretations about his or her health status and, thus, can also have significant impacts on the individual in case misused. Thus, we argue that this kind of information should be considered as "data concerning health" under the GDPR to ensure that the individual's right to data protection and privacy are efficiently protected.

Health data as sensitive data is subject to special protection under data protection and privacy laws [55, para. 126]. Under Article 9(1) of the GDPR the processing of sensitive data is by default prohibited. This applies regardless of whether the information in question is correct and of whether the controller is acting with the aim of obtaining that sensitive data [48, para. 69]. Sensitive data can only be processed lawfully in case a special legal ground, such as explicit consent, is applied and additional safeguards, such as conducting a Data Protection Impact Assessment in high risk cases, implemented. Processing of sensitive data should particularly be considered when implementing data protection by design approach into services and products as well as in defining technical and organizational safeguards for processing activities. According to Bygrave and Tosoni, health data needs special protection as it can reveal the essential vulnerabilities of a person, exposing the person in question to negative consequences such as stigma and discrimination [56, p. 218]. Respecting the confidentiality of health data is also important from the perspective of general interest as it is essential for ensuring trust in health care services [56, p. 218, 57].

Article 4(15) of the GDPR defines "data concerning health" as personal data relating to the physical or mental health of a person. This includes the provision of health care services, which reveal information about his or her health status. According to the recital 35 of the GDPR, health data includes all data pertaining to health status of a person which reveal information relating to his or her past, current or future physical or mental health status. This includes, for example, any information on a disease, disability, disease risk, medical history, clinical treatment or biomedical state of the person. Furthermore, the GDPR acknowledges that data concerning health needs to be granted special protection, as the use of such sensitive data may potentially have significant adverse impacts on individuals.[6]

Several factors support the interpretation that information concerning purchase of a prescription medicine should constitute data concerning health under the GDPR in the context of the studied online pharmacies. Firstly, in light of the to preamble to the GDPR and the case law of the CJEU, the term "data concerning health" should be given a wide interpretation (see e.g. [55, 58, p. 5, 59]). In case Vyriausioji tarnybinės etikos komisija the CJEU maintains that personal data that indirectly reveals sensitive information concerning the individual is to be interpret as sensitive data [59, para. 128]. Secondly, WP29 has explicitly stated that health data, which is a broader term than pure "medical data", also includes data about the purchase of medical products, devices and services, when health status can be inferred from the data [60, p. 2]. WP29's interpretation concerned the term "health data" under the Directive 95/46/EU, the predecessor of the GDPR. Nevertheless, it can be assumed that this interpretation would apply also under the GDPR [37, p. 36]. Thirdly, as acknowledged in the GDPR, processing of information concerning the health status of a person can result in a high risk to the rights and freedom of individuals. Disclosing information concerning a person's prescription medicines can be assumed to have similar risks to the individual.

Nevertheless, the scope of the term potentially has certain limitations when addressing indirect information that should be taken into consideration. Schäfke-Zell presents that to clarify the gray areas of the term "data concerning health", the scope of the term has sometimes been defined in legal commentaries by the purpose of the processing activity rather than by the categories of data in question [37, p. 34–36]. Similar kind of approach has been presented also, for example, in the guideline of the EDPB concerning processing of personal data through video devices, presenting that data is to be regarded as sensitive data in case the material is processed to actually deduce sensitive information from it [61]. Furthermore, it should also be noted that e.g. in case Dionyssopoulou, the CJEU determined that a mere reference without any disclosure of

[6] Recital 51 of the preamble to the GDPR.

data concerning a person's health did not constitute health data [56, p. 221, 62, 63].

However, in the context of this study, as the information concerning person's medication purchases can be used to determine sensitive information concerning the person's health status (cf. e.g. [37, p. 40–41, 38, p. 238–239, 60, p. 4–5]), the information is collected together with several other identifiers (cf. e.g. [37, p. 39–40, 60, p. 4–5]), and the Big Tech related analytics services have the resources to analyze data by advanced technologies (cf. [38, p. 239–241]), a broad interpretation of the term "data concerning health" should be applied. This is particularly because the use of this data for unwanted purposes is likely to have significant impact on the customers in question (cf. e.g. [39, p. 140, 40, p. 11]) and disclosing this data to third parties increases the risk level of processing. Thus, any other interpretation would diminish the individual's right to data protection and privacy.

5 Discussion

We have presented a study on data leaks among Finnish online pharmacies. Out of the studied 163 online pharmacies, 57 (35.0 %) leaked a specific prescription medicine name connected with identifying personal data on the customer. Pharmacies employ analytics services to gain insights into the usage of their websites, but at the same time third party analytics providers can get valuable information of the user's medical status in exchange. For instance, although many concerns have been voiced about Google's data collection practices in the EU area [6], Google was present on 134 out of 163 (81.6%) pharmacy websites, giving it a comprehensive view on the use of Finnish online pharmacies. This is especially concerning given Google's ability to often easily combine the user's action to their real name. We have also seen that the platforms used by online pharmacies often have a clear connection to the used third-party analytics services and what kind of sensitive personal data they collect. While these findings give reason for great concern, the fact that the vast majority of these leaks appears to be fixed already is a positive outcome.

Sensitive information such as intended medicine orders should have special protection. This data can be utilized to deduce what kinds of medical conditions a person is suffering from. When the third party has an opportunity to witness several purchases from the same customer over the course of time, or when the customer orders several medicines, details on a person's medical history can be effectively revealed. It is important to note we cannot make any claims regarding whether the analytics providers really store and use the sensitive data they receive. Still, the fact that data is shared with them is unacceptable to begin with. Even though third parties receiving the leaked personal data on prescription medicine orders may not seek to further utilize the data, collection and combining large amounts of data, including highly sensitive information, exposes individuals' critical vulnerabilities for unlawful and unethical exploitation such as malicious targeting or profiling and, thus, should be taken seriously.

In the light of our findings, it is clear that software developers and data protection officers should be more mindful of the personal data sent to third parties from their websites. Such data flow analysis should be an important part of web service design as well as security testing. An analysis similar to one described in this study can be used. Developers should carefully evaluate the privacy practices of the used platforms and third-party services. The use of each third-party analytics service should be thoroughly justified and these tools should not be used on pages which process delicate personal data. When analytics are considered essential, the collected data can be stored locally on the pharmacy's own trusted servers. This allows the pharmacy to control the data and avoids sharing it with a third party. This is possible by using free open source analytics solutions such as Matomo [64, 65].

Letting sensitive health related data leak to third-party services is not likely to be intentional, but the repercussions can be very serious in the case of online pharmacies. From the legal perspective, understanding of what constitutes "data concerning health" and sensitive data in the digital age is of the essence in granting appropriate protection to it. In this study we provided an analysis of how data protection law, case-law and legal literature address the limits of health data notion from the perspective of data concerning online purchases on prescription medicines, advocating broad interpretation of health data notion in the context of online pharmacies. The developers may not immediately realize that in terms of privacy, an online pharmacy cannot be considered an average online store. However, as we have seen, there are also software platforms specifically built for online pharmacies, which have still failed to adequately protect customer privacy. The use of third-party analytics services has become commonplace. Analytics services help track and measure conversions, such as completed sign-ups and purchases. The collected data can provide valuable insights on the effectiveness of marketing strategies and optimizing the website design. Developers of online pharmacy websites are probably happy to provide these kinds of extra features with their products without fully realizing the possible consequences on a website handling sensitive data. Many platforms used by developers also offer easy ways to add third-party analytics as plugins, and they are considered an essential part of modern websites. If third-party services are delivered by outside developers as a part of the package, most pharmacies probably usually do not have resources or expertise to assess

the potential downsides of these services. When building websites for healthcare providers and companies involved in the provision of pharmaceutical products, for example, an external privacy audit would not be a bad idea. Developers should also familiarize themselves with the application area to be aware of the risks associated with processing health related data.

Online pharmacies should also pay attention to making privacy policies transparent and clear. Our results indicate that online pharmacies have not succeeded in writing transparent and truthful privacy policies. Unfortunately, this is not rare for web services [20, 66]. One apparent reason for this oversight is the fact that the pharmacies themselves are not aware of the accidental data leaks. In the current study, the privacy policy documents did not contain appropriate information about data processing activities. In the majority of the studied privacy policies, the data concerning health and the third-party services collecting the data were not mentioned at all. Using a standardized template as a basis when creating a privacy policy document would probably aid in making them more comprehensive and understandable [67]. Of course, even if the data processing were to be transparent, sharing data on the intended order of prescription medicines with a third party can be deemed unnecessary and unethical (cf. [68]).

6 Conclusion

As a vital component of healthcare delivery, pharmacies have a fundamental responsibility to protect their customers' privacy onsite and in their web services. Unfortunately, our results show this has not been the case when it comes several Finnish online pharmacies. At the same time, our study has already had a significant positive impact on the presence of third-party analytics and data leaks on the Finnish online pharmacy websites. We hope that in the near future, all the studied online pharmacies will successfully fix any remaining data leaks. Still, many other fields of business and sectors processing sensitive data are likely to have web services where the state of privacy is equally unsatisfactory. It should also be further studied and clarified what constitutes sensitive data under the GDPR in the digital contexts. Studying online pharmacies in other countries and comparing results to the situation in Finland is also a topic for future research.

The findings presented in this study should also act as a reality check for web developers and data protection officers responsible for designing and maintaining essential web services in healthcare industry and other sectors where sensitive data is being processed. It is pivotal for providers of essential services to understand their accountability for safeguarding their customers' privacy. In web services handling sensitive data concerning health and used by many people who are already in a vulnerable position from the start, it is particularly important to adopt the data protection by design approach. On online pharmacy websites, the use of any third-party service that may collect personal data and possibly transfer it outside Europe without the user's knowledge, let alone several of such services, is hard to justify. The goal for developers of online pharmacies should be to create web services that are as trustworthy and confidential as conventional brick-and-mortar pharmacies.

Author contributions Conceptualization: Sampsa Rauti, Sini Mickelsson, Robin Carlsson; Methodology: Sampsa Rauti, Robin Carlsson, Timi Heino, Sini Mickelsson; Data collection: Robin Carlsson, Timi Heino; Data analysis: Sampsa Rauti, Sini Mickelsson, Robin Carlsson, Timi Heino; Writing - original draft preparation: Sampsa Rauti, Sini Mickelsson, Robin Carlsson; Writing - review and editing: Sampsa Rauti, Sini Mickelsson, Tuomas Mäkilä, Robin Carlsson, Timi Heino, Ville Leppänen, Elina Pirjatanniemi; Funding acquisition: Ville Leppänen, Elina Pirjatanniemi; Supervision: Ville Leppänen, Elina Pirjatanniemi, Tuomas Mäkilä.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital). This research has been funded by Academy of Finland project 327397, IDA - Intimacy in Data-Driven Culture.

Availability of data The research data is available on request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. Public Health. 2015;129(6):611–20.
2. Somenahalli S, Shipton M. Examining the distribution of the elderly and accessibility to essential services. Procedia Soc Behav Sci. 2013;104:942–51.
3. Almeida F, Santos JD, Monteiro JA. The challenges and opportunities in the digitalization of companies in a post-Covid-19 world. IEEE Eng Manage Rev. 2020;48(3):97–103.

4. Hattingh HL, Emmerton L, Ng Cheong Tin P, Green C. Utilization of community pharmacy space to enhance privacy: a qualitative study. Health Expect. 2016;19(5):1098–110.
5. Anderson C, Blenkinsopp A, Armstrong M. Feedback from community pharmacy users on the contribution of community pharmacy to improving the public's health: a systematic review of the peer reviewed and non-peer reviewed literature 1990–2002. Health Expect. 2004;7(3):191–202.
6. Quintel D, Wilson R. Analytics and privacy. Inf Technol Libr. 2020;39(3).
7. Wambach T, Bräunlich K. The evolution of third-party web tracking. In: Camp O, Furnell S, Mori P, editors. Information Systems Security and Privacy. Springer, Cham: Switzerland; 2017. p. 130–47.
8. 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. 2022. New York: ACM; p. 197–11.
9. Friedman AB, Bauer L, Gonzales R, McCoy MS. Prevalence of third-party tracking on abortion clinic web pages. JAMA Intern Med. 2022;182(11):1221–2.
10. 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. New York: ACM; 2022. p. 1–6.
11. Thompson N, Ravindran R, Nicosia S. Government data does not mean data governance: Lessons learned from a public sector application audit. Gov Inf Q. 2015;32(3):316–22.
12. Carlsson R, Rauti S, Mickelsson S, Mäkilä T, Heino T, Pirjatanniemi E, Leppänen V. Several online pharmacies leak sensitive medical information to third parties. Accepted to World Conference on Information Systems and Technologies, WorldCIST'23.
13. Long CS, Kumaran H, Goh KW, Bakrin FS, Ming LC, Rehman IU, Dhaliwal JS, Hadi MA, Sim YW, Tan CS. Online pharmacies selling prescription drugs: systematic review. Pharmacy. 2022;10(2):42.
14. Alwon BM, Solomon G, Hussain F, Wright DJ. A detailed analysis of online pharmacy characteristics to inform safe usage by patients. Int J Clin Pharm. 2015;37(1):148–58.
15. Fincham JE. Negative consequences of the widespread and inappropriate easy access to purchasing prescription medications on the internet. Am Health Drug Benefits. 2021;14(1):22.
16. Orizio G, Merla A, Schulz PJ, Gelatti U, et al. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res. 2011;13(3):1795.
17. Kuzma J. Web vulnerability study of online pharmacy sites. Inform Health Soc Care. 2011;36(1):20–34.
18. Vaas L. GoodRx stops sharing personal medical data with Google, Facebook. 2020. https://nakedsecurity.sophos.com/2020/03/03/goodrx-stops-sharing-personal-medical-data-with-google-facebook/. Accessed 24 Jun 2023.
19. Zheutlin AR, Niforatos JD, Sussman JB. Data-tracking among digital pharmacies. Ann Pharmacother. 2022;56(8):958–62.
20. Kuzma J, Dobson K, Robinson A. An examination of privacy policies of global on-line e-pharmacies. European Journal of Research and Reflection in Management Sciences. 2016;4(6):23–8.
21. Brown SD, Levy Y. Towards a development of an index to measure pharmaceutical companies' online privacy practices. Online Journal of Applied Knowledge Management (OJAKM). 2013;1(1):93–108.
22. Linardon J, Rosato J, Messer M. Break binge eating: Reach, engagement, and user profile of an internet-based psychoeducational and self-help platform for eating disorders. Int J Eat Disord. 2020;53(10):1719–28.
23. Santin O, McShane T, Hudson P, Prue G. Using a six-step co-design model to develop and test a peer-led web-based resource (PLWR) to support informal carers of cancer patients. Psychooncology. 2019;28(3):518–24.
24. 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. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4469981.
25. 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, 50(1), 1–9.
26. 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. https://journals.library.ualberta.ca/ojs.cais-acsi.ca/index.php/caisasci/article/view/627.
27. Huesch MD. Privacy threats when seeking online health information. JAMA Intern Med. 2013;173(19):1838–40.
28. Masters K. The gathering of user data by national medical association websites. Int J Med Inform. 2012;6(2).
29. 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). New York City, United States: IEEE; 2022. p. 278–86.
30. Friedman AB, Merchant RM, Maley A, Farhat K, Smith K, Felkins J, Gonzales RE, Bauer L, McCoy MS. Widespread third-party tracking on hospital websites poses privacy risks for patients and legal liability for hospitals. Health Aff. 2023;42(4):508–15.
31. Kes I, Heinrich D, Woisetschlager DM. Behavioral targeting in health care marketing: uncovering the sunny side of tracking consumers online. In: Let's Get Engaged! Crossing the Threshold of Marketing's Engagement Era: Proceedings of the 2014 Academy of Marketing Science (AMS) Annual Conference. Cham, Switzerland: Springer; 2016. p. 297.
32. Zheutlin AR, Niforatos JD, Sussman JB. Data-tracking on government, non-profit, and commercial health-related websites. J Gen Intern Med. 2022;37(5):1315–7.
33. Martínez D, Calle E, Jové A, Pérez-Solà C. Web-tracking compliance: websites' level of confidence in the use of information-gathering technologies. Comput Secur. 2022;122:102873.
34. Liu Y, Song HH, Bermudez I, Mislove A, Baldi M, Tongaonkar A. Identifying personal information in internet traffic. In: Proceedings of the 2015 ACM on Conference on Online Social Networks. COSN '15. New York, NY, USA: Association for Computing Machinery; 2015. p. 59–70.
35. Finck M, Pallas F. They who must not be identified-distinguishing personal from non-personal data under the GDPR. International Data Privacy Law. 2020;10(1):11–36.
36. Purtova N. The law of everything. Board concept of personal data and future of EU data protection law. Innovation and Technology. 2018;10(1):40–81.
37. Schäfke-Zell W. Revisiting the definition of health data in the age of digitalized health care. Int Data Priv Law. 2022;12(1):33–43.
38. Malgieri G, Comandé G. Sensitive-by-distance: quasi-health data in the algorithmic era. Inf Commun Technol Law. 2017;26(3):229–49.
39. Taka A-M. A deep dive into dynamic data flows, wearable devices, and the concept of health data. Int Data Priv Law. 2023;13(2):124–40.
40. Article 29 Data Protection Working Party: Opinion 4/2007 on the concept of personal data. Adopted on 20th June. WP 136.
41. Case C-582/14, Patrick Breyer v. Bundesrepublik Deutschland [2016] ECLI:EU:C:2016:779, paragraph 49.
42. C-434/16, Peter Nowak v Data Protection Commissioner [2017] ECLI:EU:C:2017:994, paragraph 34–35.

43. Bygrave L, Tosoni L. Article 4(1). Personal data. In: Kuner C, Bygrave L, Docksey C, Drechsler L, editors. The EU General Data Protection Regulation: A Commentary. Oxford, United Kingdom: Oxford University Press; 2020.
44. Belgian supervisory authority, Litigation Chamber: Decision of 2 February 2022. https://edpb.europa.eu/system/files/2022-03/be_2022-02_decisionpublic_0.pdf. Accessed 20 Dec 2023.
45. Swedish supervisory authority: Decision of 30 June 2023. https://edpb.europa.eu/system/files/2023-10/se_2023-06_decisionpublic_redacted.pdf. Accessed 20 Dec 2023.
46. C-319/22, Gesamtverband Autoteile-Handel eV v Scania CV AB [2023] ECLI:EU:C:2023:873.
47. Case T-557/20, Single Resolution Board (SRB) v. European Data Protection Supervisor (EDPS) [2023] ECLI:EU:T:2023:219.
48. C-252/21, Meta Platforms and Others [2023] ECLI:EU:C:2023:537.
49. Mishra V, Laperdrix P, Vastel A, Rudametkin W, Rouvoy R, Lopatka M. Don't count me out: On the relevance of IP address in the tracking ecosystem. In: Proceedings of The Web Conference 2020. New York: ACM; 2020. p. 808–15.
50. C-311/18, Data Protection Commissioner v Facebook Ireland Ltd and Maximillian Schrems [2020] ECLI:EU:C:2020:559.
51. Austrian supervisory authority: Decision of 22 December 2021. https://noyb.eu/sites/default/files/2022-01/E-DSB%20-%20Google%20Analytics_DE_bk_0.pdf. Accessed 28 Jun 2023.
52. Austrian supervisory authority: Decision of 22 April 2022. https://noyb.eu/sites/default/files/2022-04/Bescheid%20geschw%C3%A4rzt%20EN.pdf. Accessed 27 Jun 2023.
53. French supervisory authority: Decision of 2 March 2022. https://noyb.eu/sites/default/files/2022-04/20220302_CNIL_101-complaints-decision-two_Redacted.pdf. Accessed 26 Jun 2023.
54. European Data Protection Board: Binding Decision 1/2023 on the dispute submitted by the Irish SA on data transfers by Meta Platforms Ireland Limited for its Facebook service (Art. 65 GDPR). Adopted on 13 April 2023.
55. C-184/20, Vyriausioji tarnybinės etikos komisija [2022] ECLI:EU:C:2022:601, paragraphs 122–127.
56. Bygrave L, Tosoni L. Article 4(15). Data concerning health. In: Kuner C, Bygrave L, Docksey C, Drechsler L, editors. The EU General Data Protection Regulation: A Commentary. Oxford, United Kingdom: Oxford University Press; 2020.
57. European Court of Human Rights: Z v Finland, Appl. No 22009/93, judgment of 25 February 1997, paragraph 95.
58. European Data Protection Board: Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak. Adopted on 21 April 2020.
59. Case C-101/01, Criminal proceedings against Bodil Lindqvist [2003] ECLI:EU:C:2003:596, paragraph 50.
60. Article 29 Working Party's letter of 2015: Annex - health data in apps and devices.
61. European Data Protection Board: Guidelines 3/2019 on processing of personal data through video devices, paragraph 62–64. Adopted on 29 January 2020.
62. Case T-105/03, Triantafyllia Dionyssopoulou v Council of the European Union [2005] ECLI:EU:T:2005:189, paragraph 33.
63. T-343/13, CN v European Parliament [2015] ECLI:EU:T:2015:926, paragraph 50.
64. Chandler A, Wallace M. Using Piwik instead of Google analytics at the Cornell University Library. Ser Libr. 2016;71(3–4):173–9.
65. 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. 2021;13(2):133–53.
66. Mulder, Trix, Health Apps, their Privacy Policies and the GDPR (June 3, 2019). European Journal of Law and Technology, 2019, University of Groningen Faculty of Law Research Paper No.15/2020, https://ssrn.com/abstract=3506805.
67. Rowan M, Dehlinger J. A privacy policy comparison of health and fitness related mobile applications. Prog Comput Sci. 2014;37:348–55.
68. Schwartz PM. Privacy, ethics, and analytics. IEEE Secur Priv. 2011;9(3):66–9.