Proceedings Paper:
Wesselkamp, V., Fouad, I., Santos, C. et al. (3 more authors) (2021) In-Depth Technical and Legal Analysis of Tracking on Health Related Websites with ERNIE Extension. In: WPES '21: Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society. CCS '21: 2021 ACM SIGSAC Conference on Computer and Communications Security, 15 Nov 2021, Seoul, Korea. ACM , pp. 151-166. ISBN 9781450385275

https://doi.org/10.1145/3463676.3485603

In-Depth Technical and Legal Analysis of Tracking on Health Related Websites with ERNIE Extension

Vera Wesselkamp, Imane Fouad, Cristiana Santos, Yanis Boussad, Nataliia Bielova, Arnaud Legout

In-Depth Technical and Legal Analysis of Tracking on Health Related Websites with ERNIE Extension

Vera Wesselkamp
Inria
Sophia Antipolis, France

Imane Fouad
Inria
Sophia Antipolis, France

Cristiana Santos
Utrecht University
Utrecht, Netherlands

Yanis Boussad
Inria
Sophia Antipolis, France

Nataliia Bielova *
CNIL
Paris, France

Arnaud Legout
Inria
Sophia Antipolis, France

ABSTRACT

Searching the Web to find doctors and make appointments online is a common practice nowadays. However, simply visiting a doctors website might disclose health related information. As the GDPR only allows processing of health data with explicit user consent, health related websites must ask consent before any data processing, in particular when they embed third party trackers. Admittedly, it is very hard for owners of such websites to both detect the complex tracking practices that exist today and to ensure legal compliance.

In this paper, we present Ernie, a browser extension we designed to visualise six state-of-the-art tracking techniques based on cookies. Using Ernie, we analysed 385 health related websites that users would visit when searching for doctors in Germany, Austria, France, Belgium, and Ireland. More specifically, we explored the tracking behavior before any interaction with the consent pop-up and after rejection of cookies on websites of doctors, hospitals, and health related online phone-books. We found that at least one form of tracking occurs on 62% of the websites before interacting with the consent pop-up, and 15% of websites include tracking after rejection. Finally, we performed a detailed technical and legal analysis of three health related websites that demonstrate impactful legal violations.

This paper shows that while, from a legal point of view, health related websites are more privacy-sensitive than other kinds of websites, they are exposed to the same technical difficulties to implement a legally compliant website. We believe Ernie, the browser extension we developed, to be an invaluable tool for policy-makers and regulators to improve detection and visualization of the complex tracking techniques used on these websites.

CCS CONCEPTS

• **Security and privacy → Pseudonymity, anonymity and untraceability**; **Web application security**; **Privacy protections**; • **Social and professional topics → Governmental regulations**; **Patient privacy**.

KEYWORDS

tracking; browser extension; GDPR; health data; explicit consent

ACM Reference Format:
Vera Wesselkamp, Imane Fouad, Cristiana Santos, Yanis Boussad, Nataliia Bielova, and Arnaud Legout. 2021. In-Depth Technical and Legal Analysis of Tracking on Health Related Websites with ERNIE Extension. In *Proceedings of the 20th Workshop on Privacy in the Electronic Society (WPES '21), November 15, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3463676.3485603

1 INTRODUCTION

Health data is known to be one of the most sensitive types of data (Article 35(3)(b) of the General Data Protection Regulation) [57]. Massive health data leaks are recognized to be of particularly high severity to the users' privacy, according to the European Data Protection Board [16]. Searching for doctors online has become an increasingly common practice among Web users since telemedicine peaked in 2020 during the global Covid-19 pandemic [55]. However, the mere visit to a doctor's website can reveal a lot about the visitor: one can infer which diseases a visitor has or is interested in. Whenever health websites integrate third-party trackers, they *expose their potential patients' medical secrets to third parties*[1]. Moreover, when *users take appointments on doctor's websites*, one can reasonably infer health data from an individual's list of appointments [60] regarding the medical specialty the user is interested in. In 2021, a data breach that occurred on the platform `doctolib.de` – which allows booking appointments with doctors in Germany–, demonstrates this problem at scale: data about 150 million booked appointments was publicly accessible for several months [83]. *From a legal perspective*, when providing services or monitoring user's behavior in the EU, health related websites integrating third-party trackers are in breach with the GDPR because processing of sensitive health data (derived from a visit to a website) is *generally forbidden*, unless allowed by several exceptions (Article 9(2) GDPR).

As a result, Data Protection Officers of the health related websites as well as Data Protection Authorities have the urgent need to be able to detect tracking and advanced cookie synchronization techniques on their website in order to determine whether the included third parties may be leaking patients' health data. Whereas

*The views, opinions and positions expressed in this article are those of the author and are not endorsed by its institution. The work has been carried out by Nataliia Bielova while she was at Inria until August 2021.

[1]The French Code of Public Health [25, Article L1110-4]states that medical secret covers "all information about the person coming to the knowledge of the professional, of any member of the staff of these organizations (...)".

WPES '21, November 15, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8527-5/21/11…$15.00
https://doi.org/10.1145/3463676.3485603

some browser extensions visualize known tracking third parties or third-party cookies [32, 35, 48, 62, 66], *no browser extension exists to visualize sophisticated forms of cookie synchronization and sharing of user's identifiers* [45, 68, 69] across third parties. Therefore, owners of health-related websites cannot identify tracking and complex cookie syncing included in their websites.

Moreover, since processing health data without user consent is forbidden by the GDPR, health website owners must implement a specific consent mechanism called *explicit consent*, to make such processing lawful for all third parties included in the website. But even for a basic consent to be legally valid, it has to comply with at least 22 different fine-grained requirements [74]. Whereas websites generally implement consent pop-ups [2] to comply with the legal requirement of consent, recent work made evident that in practice websites often do not contain any consent pop-ups, or pop-ups that do not respect the user's choice [63, 67, 68, 73]. Therefore, doctors and hospitals need to ensure that if their websites contain tracking including any form of sophisticated cookie syncing, a *valid and explicit consent must be collected* before any tracking is performed.

In this paper, we make the first in-depth study on tracking in health-related websites in five EU countries: France, Germany, Belgium, Austria, and Ireland. We designed a new Firefox browser extension called Ernie that detects and visualizes sophisticated forms of tracking and ID sharing on visited websites, based on 6 different categories of third-party tracking from Fouad et al. [45].

Instead of relying on categorisation services [73, 75], we carefully selected 385 websites that Web users would find if they were searching for 10 popular doctors specialties in the capitals of the analyzed EU countries. With Ernie, we visited these websites and detected all 6 categories of tracking techniques before interacting with the consent pop-up (if it was present) and after rejecting consent (if it was possible) across two widely used browsers: Google Chrome and Mozilla Firefox. Finally, we performed a detailed legal analysis together with a legal expert, co-author of this paper, of 3 case studies that depict the most impactful legal violations.

In summary, we make the following contributions.

(1) We propose the first browser extension Ernie[3] that visualizes complex cookie syncing and ID sharing tracking techniques. Ernie detects 6 categories of such tracking behaviors – Basic tracking, basic tracking initiated by another tracker, first to third party cookie syncing, third to third party cookie syncing, third party cookie forwarding, and third party analytics– following the state-of-the-art methodology from Fouad et al. [45].

(2) We perform a legal and technical analysis of consent collection on 385 health related websites. We identified 3 practices potentially violating the GDPR and ePrivacy Directive:

1- Tracking before interaction: We found that 62% of the websites track users before any interaction with the banner, thus violating the *explicit and prior consent* requirements.

2- Not possible to reject: We found that 40% of websites do not display a consent pop-up, thus violating an *explicit* consent request, and 29% of the websites provide a cookie banner without a reject button, hence violating the *freely given and unambiguous* consent requirements.

3- No respect of user's choice: We show that the *user choice is not respected* on health related websites: 59 (15%) websites still contain tracking after cookie rejection, infringing the *lawfulness principle*.

(3) We analyse 3 case studies to provide an in-depth technical and legal analysis on health related websites. We observed that these websites do not comply with the legal requirements for explicit consent, as demanded by the GDPR and ePD. We concluded that the website's cookies performing cross-site tracking are related to advertising purposes, which, according to Data Protection Authorities, raises serious privacy concerns, since it is possible to build and enrich unique user profiles based on sensitive health data [33, 58].

Paper	Sensitive websites	Consent pop-ups	Detection of tracking techniques
Libert [61]	Health websites	×	×
Vallina [81]	Porn websites	✓	BT, FTCS, TTCS
Matic [75]	Health websites	×	BT
Sanchez [73]	Health websites	✓	BT
Matte. [63]	×	✓	Disconnect list
Papadogiannakis. [68]	×	✓	First party ID leaking (TA & FTCS), TTCS
Fouad [45]	×	×	All
Our paper	Health websites	✓	All

Table 1: Overview of related work. The abbreviations of tracking techniques are described in Section 3.1.2.

2 RELATED WORK

In this section, we provide an overview of work related to the interaction with consent pop-ups and to the detection of tracking on sensitive websites. Table 1 summarizes the related work.

Fouad et al. [45] were the first to distinguish first to third party cookie syncing and third to third party cookie syncing, differentiating between a total of 6 tracking techniques. We adopted their classification of tracking to build our extension Ernie.

Analysis of sensitive websites. Previous work explored the tracking behaviors in sensitive websites. Libert et al. [61] analyzed health related websites by taking the top 50 Bing results for 1,986 common diseases. They found that 91% of pages include third party content while 71% use cookies. Vallina et al. [81] analyzed a set of 6,843 pornographic websites. They found that 72% of the websites include basic tracking and 58% of the top 100 porn websites contain cookie syncing. Matic et al. [75] built a classifier that identifies sensitive URLs. They found that 40% of the cookies used on 20K detected health related websites are persistent third party cookies and 5% were set by trackers known from the Disconnect [32] and Ghostery [48] filter lists. Sanchez et al. [73] performed a manual

[2]Cookie pop-ups is a generic term we use in this paper to refer to any kind of user interface related to cookie information, selection, or rejection within a website.

[3]The main goal of this extension is to provide an easy-to-use tool for the privacy experts, such as DPOs, DPAs and the research community, as well as NGOs and legal experts knowledgeable in technology to visualize complex tracking. We will make the Ernie available and open-source upon acceptance of this paper.

analysis of 2000 websites. They found that only 4% of websites offer an easy way to reject the consent pop-up. They also looked at websites by category and found that more than 50% of health websites do not have a consent pop-up while still performing tracking, and 40% even create more cookies upon rejection.

Whereas previous works [73, 75] only investigated the presence of identifying third party cookies on health related websites, we detected complex cookie syncing techniques from Fouad et al. [45].

Analysis of consent pop-ups. Previous work studied the impact of the user's choices in the consent pop-up on the tracking behavior in a website. Matte et al. [63] studied the consent stored behind the IAB Europe's Transparency and Consent Framework (TCF) and found that 10% of websites stored a positive consent before interaction of the user with the cookie banner. They also analyzed the presence of third-party trackers on the websites using the Disconnect list [32] and found that refusing cookies increased the number of third-party trackers. Recently, Papadogiannakis et al. [68] studied the effect of user interaction with the banner on first-party ID leaking (not differentiating between third party analytics and first to third party cookie syncing), and third-party ID synchronization (third to third party cookie syncing in our work). They found that 52% of the websites were engaged in first-party ID leaking, and 24% in third-party ID synchronization before interaction with a banner.

Whereas previous work provided a quantitative study of the impact of interaction of consent pop-ups, in our paper, we combine that impact with detailed case studies and their legal implications.

Browser extensions. There are several browser extensions that use filter lists to block trackers and preserve user's privacy [32, 35, 48, 56]. Disconnect [32] shows third party inclusion chains, whereas uBlock Origin [56] shows which part of a URL is responsible for tracking. The Lightbeam extension [66] visualizes which third parties are included on which websites. All these extensions only provide a very limited overview of the tracking on a website. Website scanners [28, 43, 71, 82] allow a user to see what cookies are set on a website in order to determine if the website is compliant with the GDPR. The EDPS Inspection Software [77] gives information about web traffic caused by a website, as well as trackers based on the EasyPrivacy filter list. The tool closest to our extension Ernie is CNIL's Cookieviz 2 [62], which visualizes the third party domains that occur on websites on a sequence of visits. It also shows if the domains set a third party cookie and if that cookie is listed in an ads.txt file, indicating that it is used for advertisement.

Our extension Ernie is the first browser extension to visualize several types of cookie synchronization techniques and the cookies enabling the tracking. It also shows the origin of cookie syncing requests, providing a live overview of tracking on websites.

3 METHODOLOGY

In this section, we first describe the architecture of the browser extension Ernie (Section 3.1). We then describe the collection of data on the tracking behavior of health related websites (Section 3.2).

3.1 Ernie Extension

The browser extension Ernie is designed to detect the sophisticated cookie-based tracking mechanisms described by Fouad et al. [45]. Ernie detects six categories of tracking (see Section 3.1.2).

Ernie collects all first-party and third-party HTTP(S) requests and responses during a page visit in a specific browser tab. A page visit can be triggered by entering a new URL in the navigation bar, clicking a URL, clicking the forward/backward browser buttons, reloading a page, or a redirection event. All requests sent and responses received in that tab after the page visit and before the next one are considered part of the current page visit. Ernie provides a visualization that attributes these HTTP(S) requests and responses and the corresponding cookies to one of six studied categories.

3.1.1 Detection of ID cookies and ID sharing.

Detection of ID cookies. The Ernie extension implements a standard approach to detect cookies that are likely to identify a user [13, 36, 37, 45] by comparing cookies between two different users. Ernie simulates a different user by opening a hidden tab in a separate container for each page visit, which is only used by the extension. We discuss the limitation of this technique in Section 3.1.3. To create the container in Firefox, the extension uses the Firefox API `contextualIdentities` [3]. *Contextual identities* are containers within a browser profile that have a separate cookie storage, localStorage, indexedDB, HTTP data cache, and image cache. To create a container in Chrome, the extension uses the incognito mode, which also maintains its own stores as listed above. In order to achieve the same behavior as the contextual identities in Firefox, we took the following steps: by default, the Chrome browser blocks third-party cookies in incognito mode. This functionality has to be disabled in order that the same tracking behavior occurs as in a regular browser session. Additionally, in incognito mode, all stores are cleared upon closing the incognito window. To maintain a profile that simulates a user in the shadow tab, upon closing the window, we save the cookies from the incognito session into the extensions database. Upon restarting the extension, these cookies are loaded back into the incognito window.

In the following, we refer to the hidden tab as the *shadow tab*. If the cookies with the same key and domain have different values for the main and shadow tab, Ernie concludes that the cookie is "user-specific", we refer to such cookies as *ID cookies*. The extension displays and analyzes all (first- and third-party) ID cookies set in the browser (via HTTP(S) requests, responses, or Javascript). If the value of a cookie is the same in the main and shadow tab, the cookie is labeled as *safe* and is saved in a local database of the extension.

Detection of ID Sharing. To identify if an ID cookie is shared, the extension implements an ID sharing algorithm inspired by prior work [13, 36, 45]. All cookie values and URL parameters are split using as delimiters any character not in [a-zA-Z0-9-_.]. Fouad et al. considered three additional ways to share an identifier in the parameters: Google Analytics (GA) sharing, base64 sharing, and encrypted sharing. The extension implements these detection methods as well, and extends GA sharing to all the domains listed on the privacy policy of Google [6], because we observed this type of sharing not only on `google-analytics.com`, but also on `doubleclick.net` and `google.com` owned by Google. To reduce the chance of coincidental matches, after splitting, we don't consider values that are shorter than 4 characters, and *true* or *false* values.

All the requests, responses, and corresponding cookies where ID sharing is detected are stored in an external database located on the same device for later analysis.

3.1.2 Tracking detection. By detecting ID cookies and ID sharing, the Ernie extension can identify six types of tracking behaviors presented by Fouad et al. [45]. In order to identify a tracking behavior, the extension first finds the initiator of the request, that is, the resource which caused the request as follows.

(1) If the request is caused by a 30x HTTP redirect, the initiator is the source of the redirection. Ernie labels the request that caused the redirection as the initiator.
(2) If there is no redirection, but the HTTP-*Referer*-header of the request is set, Ernie labels as the initiator the previous request with the same URL as the one in the Referer-header. If the Referer is set to a domain, the whole domain is considered the initiator. This is always the case in Chrome, which by default sets the Referer to the domain.
(3) For requests whose initiator cannot be found by either (1) or (2), Ernie considers that the initiator is the first party.

Once the initiator of a request is identified, Ernie detects whether the request is responsible for one of the six tracking behaviors presented below.

Basic tracking (BT) is the most common tracking technique. To detect Basic tracking, the extension checks whether a third-party ID cookie is sent in a third-party request/ set in a third-party response.

Basic tracking initiated by another tracker (BTIT) occurs when (1) a basic tracker initiates a third-party request to another third-party domain and (2) this other third-party domain sets or sends an ID cookie. To detect the Basic tracking initiated by another tracker, the extension performs algorithm 1.

Algorithm 1: Detection of BTIT in website *site*

```
Let C be the set of ID cookies detected in site;
for Every request r in site do
    if r is sent to a third party: Tracker1 then
        Extract all cookies sent/received by Tracker1 and
          put them in set C1;
        Extract initiator of Tracker1: Tracker2;
        Extract cookies sent/received by Tracker2 and put
          them in set C2;
        if C1 ⋂ C ≠ ∅ and C2 ⋂ C ≠ ∅ then
            Tracker1 and Tracker2 are performing Basic
              tracking initiated by another tracker
        end
    else
        Continue to the next request;
    end
end
```

First to third party cookie syncing (FTCS) occurs when (1) a first-party ID cookie is shared with a third-party domain via the request URL (either in the key or value of the parameter, or the path of the URL - see Section 3.1.1), and (2) the third-party domain sets or sends its own ID cookie (See Figure 1). To detect the first to third party cookie syncing, the extension performs algorithm 2.

Third to third party cookie syncing (TTCS) occurs when an ID cookie of a third party is shared in the request URL of another third-party request, either in the key or value of the parameter, or in the path of the URL (see the ID sharing section above). The third-party request additionally sets its own ID cookie. We detect the sharing of the cookie through the whole initiator chain.

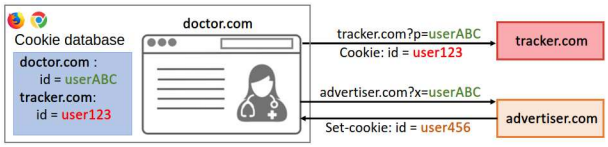


Figure 1: Two examples of first to third-party cookie synchronization: either the third-party cookie is already present in the browser and hence automatically sent to a third party (case of tracker.com) or is actively set by a third-party domain (case of advertiser.com).

Algorithm 2: Detection of FTCS

```
Let's note C_site the set of identifier cookies set by site.;
if C_site ≠ ∅ then
    for Every request r in site do
        if r is sent to a third party: Tracker1 then
            Extract the chain of initiators to Tracker1: T_1 ...
              T_n with n the length of the chain;
            while j <= n do
                if ∃ c in C_site shared with T_j, and T_j
                  received/set its own third-party ID cookie
                  then
                    First-party cookie is synchronized with
                      T_j
                end
            end
        else
            Continue to the next request;
        end
    end
end
```

Third party cookie forwarding (TF) occurs when an ID cookie of a third party is shared in the request URL of another third-party request, either in the key or value of the parameter, or in the path of the URL. Unlike the case of third to third party cookie syncing, the third-party request does not set its own ID cookie. We detect the sharing of the cookie through the whole initiator chain.

Third party analytics (TA) occurs when an ID cookie of the first party is shared in the request URL of a third-party request, either in the key or value of the parameter, or in the path of the URL. The third-party request does not set its own ID cookie.

3.1.3 Limitations of the Ernie extension. The limitation of using a *shadow tab* to simulate a different user is that, even if requests on the shadow tab are sent with different cookie values, they are still sent from the same IP address and the same device. If the website uses browser fingerprinting to recognize users, the requests from the shadow tab will likely be recognized as being from the same user as the original requests. Also, using the Referer header has some limitations. If a third party makes a request to another third

party, the Referer is often still set to the URL of the first party. Additionally, due to privacy concerns, the Referer is often not set by the websites that serves the request. We therefore may miss the initiators and label them as first-party.

3.2 Experimental setup

To be able to identify the type of visited websites, find contact information, and analyse the content of consent pop-ups, we collected websites from five European countries. Those countries are Austria, Belgium, France, Germany, and Ireland, where either French, German, or English, languages the authors speak fluently, is an official language. Figure 2 presents an overview of our experimental setup. We first selected health related websites from the five European countries (Section 3.2.1). Next, we setup the browser (Section 3.2.2) and collected data upon different interaction modes (Section 3.2.3).

3.2.1 Website Selection.

Extraction of most frequently visited doctors specialties. To extract the doctors specialties most frequently searched by users, we looked at the most popular *aggregator websites* of the countries we visit. We define an aggregator health website as a website that acts as an online phone book for doctors. This category of websites often provides the list of the most popular searched specialties. To find the most popular aggregator site for each country, we took the following steps.

- First, we categorized the list of the 10K Alexa top global websites [1] using the McAfee service [64]. This service uses various technologies, such as link crawlers, and customer logs to categorize websites. It is used by related work [80]. A description of the reported McAfee categories can be found in the McAfee reference guide [65]
- For every country, we extracted the list of websites with the corresponding country code top-level domain (e.g., we extracted the list of websites ending with `.fr` for French websites).
- Then we extracted the list of websites categorized as health websites, and manually checked starting from the most popular websites if (1) the website is an aggregator website, and (2) if the website provides statistics about the frequently searched doctors specialties.

Using this method, we found that `doctolib.fr` and `jameda.de` are the most popular aggregator websites for France and Germany, respectively. Both websites provide statistics on which doctors specialties are most frequently searched for by users. For the other three countries, we were unable to find aggregators, and thus the most frequently searched specialties. To build a consistent list of the ten most popular specialties, we took the union of the top 8 specialties from `doctolib.fr` and the top 7 specialties from `jameda.de`. With the overlap in specialties in both aggregators, we obtained a list of 10 unique most popular specialties. We present the English translations of the specialties in Table 2. The German and French translations can be found in Table 7 in the appendix.

Simulating a user in a city searching for a doctor. To retrieve health related websites of interest to real users, we simulated the behavior of a user in a given city interested in finding a doctor of a certain specialty in the same city.

Specialties	
Dentist	Dermatologist
General practitioner	Osteopath
Paediatrician	Physiotherapist
Gynaecologist and obstetrics	Orthopaedist
Ophthalmologist	Neurologist

Table 2: Top doctors specialties.

We simulated users in the capitals of France, Germany, Belgium, Austria, and Ireland: Paris, Berlin, Brussels, Vienna, and Dublin. For each city-specialty pair, we imitated a user that makes a Google search of the specialty in the corresponding city. We searched for ⟨`city`⟩ ⟨`specialty`⟩ on the country-specific search engine of Google, using a VPN based in that city. To search for health related websites in France for instance, we used `google.fr` and a Paris-based VPN. As a VPN-provider we chose PrivateVPN [72] because it provides a city-specific VPN for studied cities. We then automatically extracted the top 10 URLs of the results of each search, while skipping duplicates of domains between the searches. We extracted the URLs with Pupeteer version 5.8 [8] running on Chromium 88.0. As a result, we have a list of 500 URLs. This process is represented in the top-left corner of Figure 2.

Further analysis of collected websites. By manually analyzing the content of each of the websites resulting from the Google search, we categorized each site as one of the following.

- Aggregator: A website where a user can search for doctors in an area and potentially book appointments. An example is the aforementioned `doctolib.fr`.
- Personal: A doctor's personal website.
- Hospital/Joint office: A hospital website, or a website of a joint office of more than one doctor.

We exclude sites that do not fall into one of those categories or that are not reachable, resulting in a set of 434 websites.

In early experiments, we found that many personal and hospital/joint office websites contain subpages of the following types:

- *contact page*, where potential patients can find phone number, address or other contact information.
- *appointment pages*, where users can book an appointment.

From a legal perspective, these pages are of special interest, as a user accessing either of them indicates an intent. Therefore, we navigated to these pages and analyzed their tracking behavior.

To identify contact and appointment pages, we defined a list of keywords that must be contained in the menu item or button that is clicked to navigate to the page. To gather the keywords, we extract a random subset of 50 pages from the collected URLs for each language (From the Austrian and German URLs for German, from the Belgian and French websites for French, and from the Irish websites for English). Each URL is visited by two of the authors, who extract the menu item of the page that they identified as contact or appointment page. If there is a disagreement, we visited the website again and checked the proposals of the two authors. In the cases where one author clearly missed the menu item, the item is added to the list. As a result, only the menu items where there is a consensus between the two authors were retained. From the

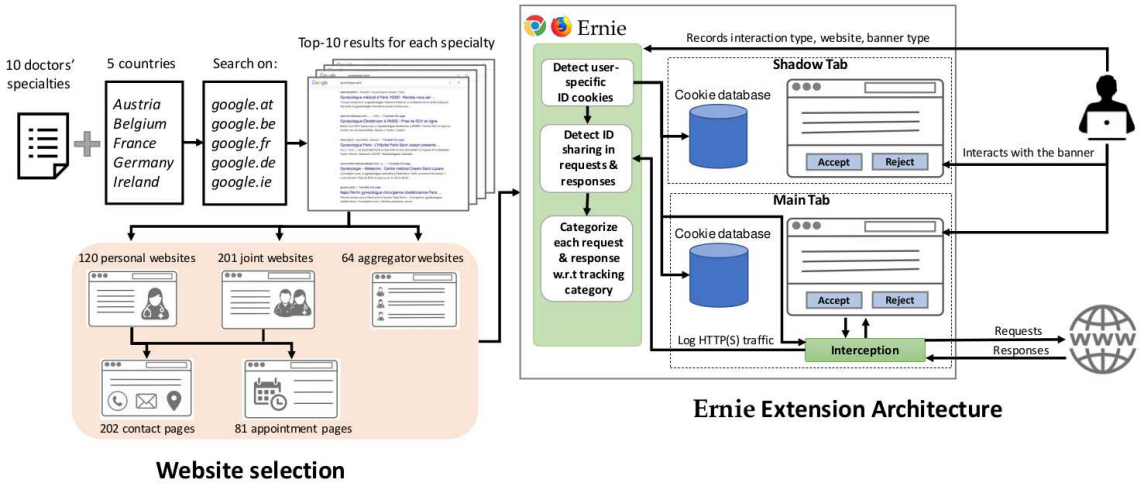


Figure 2: High-level overview of our experimental setup. In Section 3.2, we describe the website selection process as well as browser setup and website analysis. We present the Ernie extension architecture in detail in Section 3.1.

resulting menu items, we extracted frequently occurring keywords to cover different inflections and phrasing. For instance, from the phrases "Make an appointment" and "Request an appointment", we retained the keyword "appointment". The list of keywords for each language can be found in the appendix in Table 8.

3.2.2 Browser setup.

Browser settings. To analyse the tracking on a health website as experienced by users, we performed our analysis for two popular browsers: Google Chrome (version 91.0) and Firefox (version 89.0). We assume that most people use their browsers out of the box, leaving the default settings untouched. We thus also used the browser's default setting to simulate a real user. In Firefox, *Enhanced Tracking Protection* is enabled by default, meaning that Firefox already blocks some cross-site and social media trackers based on the *Disconnect* list [5]. In Chrome, we allowed 3rd party cookies to be set in incognito mode, which is needed by Ernie. This does not affect the tracking ability of a website loaded in the main tab.

Simulation of a base browsing profile. To simulate a real user, we created a base profile per country that we installed in the browser before visiting health related websites. This way, the cookie storage of the browser already contained cookies.

To build the base browsing profiles, we collected the top 100 global websites from the Alexa top list [1]. Then, for each country, we built a country-specific profile by visiting each of the top 100 websites from a VPN based in the capital of the country, without interacting with the websites. We set a timeout of 60 seconds for each website visit. The full list of unique websites visited to build the profile is publicly available [2]. We then visited each health related website collected in Section 3.2.1 with the base browsing profile in place. Each time we visited a new website, we reset the base profile to its initial state, that is we performed *stateless* crawls with a common base profile (per country). We built all base profiles and visited all health related websites in June 2021.

3.2.3 Data Collection. With the base browsing profile in place, we visited each of the collected websites and logged the tracking behavior that the extension found. For all websites, we reloaded the page once after the initial visit because after interacting with a cookie banner, some websites include additional content only on the next page load.

Interactions with the consent pop-ups. Previous work explored the interaction with the consent pop-up [30, 73]. However, automated interaction with consent pop-up remains challenging: Srdjan *et al.* [75, Sec. 3.1] report that only 4.4% of websites contain a cookie banner that we can automatically interact with via advanced tools like Consent-O-Matic [27, 67].

To ensure that all consent pop-up are correctly labeled and interacted with, we decided to manually label the type of pop-up. The EU legislation requires consent before setting or sending tracking cookies. We therefore evaluated the types of consent pop-ups and changes in the tracking behavior based on the choice made by the user in the pop-ups. We interacted with the pop-ups in two ways and recorded each interaction type in our dataset.

- **No Interaction.** We don't interact with the cookie banner, but still visit the website and the contact or appointment subpages on `Personal` and `Hospital/Joint office` websites.
- **Reject All.** We reject as many cookie categories and vendors as proposed in the banner interface. This is not possible on all websites that have consent pop-ups, because many pop-ups only describe their use of cookies and other tracking technologies, but do not offer a possibility to reject them.

We repeated the same process for every website in both Firefox and Chrome. Before rejection, not all contact and appointment pages were accessible to a user because of *consent walls* [53, 74] that block access unless a user interacts with the consent pop-up. To ensure not to introduce bias in our study, we only considered the pages both accessible before interaction and after rejection.

Data collection from manual analysis. The Ernie extension saved all collected data to a local database. The database contains data related to page visits (described in Section 3.1) as well as data

about manual analysis of the website content. We manually collected the following data for each visit: (1) the site type (`Personal`, `Aggregator`, or `Hospital/Joint office`), (2) whether the website contained a consent pop-up, and (3) whether rejection was possible.

3.2.4 Limitations of the experimental setup. A limitation of the method we used for site selection is that the search results may be biased because we rely on results from Google. Additionally, due to issues when loading the base profile, we had to exclude the collected data of 49 of the totally visited 434 websites, resulting in a final dataset of 385 visited websites.

4 REGULATORY COMPLIANCE

In this section, with a legal expert and co-author, we present the legal requirements for online tracking on health websites.

Our legal analysis is based on the General Data Protection Regulation (GDPR) [47] and the ePrivacy Directive (ePD) [38], as well as in its recitals (which help the interpretation of rules in a specific context, though they are not mandatory for compliance). The GDPR applies to the processing of personal data [39] and requires organizations to choose a legal basis to lawfully process personal data (Article 6(1)(a)). In case this legal basis is consent, the GDPR also defines the requirements for a valid consent. The ePD provides *supplementary* rules to the GDPR in particular for the use of tracking technologies. We have additionally consulted the guidelines of both the European Data Protection Board (an EU advisory board on data protection) and the Data Protection Authorities. Even if these guidelines are not enforceable, they are part of the data protection EU framework which we apply in this work to evaluate the compliance of tracking on health related websites.

Legal requirements for online tracking. To comply with the GDPR and the ePD, websites must obtain a valid *consent* from users located in the EU when monitoring users' behavior (Article 5(3) ePD) through cookies and other tracking technologies. A common method to obtain consent is through the use of consent pop-ups. For consent to be valid, it must be prior to any data collection, freely given, specific, informed, unambiguous, readable, accessible, and revocable (Articles 4(11) and 7 GDPR) [74]. Though consent is generally needed for tracking purposes, some purposes are exempted (e.g, functional cookies, see Recital 66 ePD). In fact, the only way to assess, with certainty, whether consent is required is to analyse the *purpose* of each tracking technology on a given website [17, 46].

Data concerning health As indicated by Recitals 51 and 53, data concerning health deserves higher protection, as the use of such sensitive data may have significant adverse impacts for data subjects. The EU data protection framework acknowledges a broad definition of 'data concerning health', both at the GDPR [79] and at the EU Court of Justice levels [29]. According to the GDPR, *data concerning health* means personal data related to the past, current, or future physical or mental health of a person. This includes all data pertaining to the health status of a data subject (Article 4(15)). Recital 35 thereto is an example of the amplitude of this concept as it lists information on *"disease risk"* as data concerning health, *"independent of its source"*. For data to qualify as health data it is not always necessary to establish 'ill health' [57]. The EDPB states that even if visits to websites providing information on special categories of data do not directly disclose these categories for the visitors, there is a high impact on those visitors' privacy if they are labelled as being interested in such information [12]. The EDPB [20] further states that information could become health data because of its usage in a *specific context*. In this paper, we consider that visiting a health related website could configure such specific context: the information related to the types of health related websites is shared with third parties together with unique identifiers that are associated with users. In other words, a third party with whom this data is shared could potentially and easily argue that, for example, a user visited a gynecologist website. When health websites integrate third-party trackers, they expose their users sensitive data to third parties. Considering the large number of users a single third party can follow, the collected information may provide detailed insights from a very large number of users. While this information might not be accurate, it is well-accepted that information does not have to be true or false to be personal data [39].

Legal requirements for online tracking on health websites. The processing of data concerning health is *forbidden* by the GDPR. However, there are exceptions to this prohibition contained in Articles 9 (2)(a-j). For the purposes of online tracking in health related websites, only the *explicit consent* exception seems to be the applicable legal basis to process this special category of data [47, Article 9(2)(a)]. An *explicit consent* request should abide to the following requirements [9, 14, 33]: i) include double confirmation or verification from the user, ii) consist of a separated request from any other consents [41] (Recital 43 GDPR), iii) specify the nature of the special category of data through a specific legend. This additional effort is justified *to remove all possible doubt and potential lack of evidence in the future* [42]. Simply put, the processing of sensitive personal data places a higher compliance burden on the controller as there is a qualitative difference between "regular" consent and the explicit consent provided for under Article 9(2)(a) GDPR. Without explicit consent from users, tracking on health websites may therefore be found to infringe the lawfulness principle (Article 9 (2)(a) GDPR), rendering any subsequent processing *unlawful*. Consequently, such websites might be subject to administrative fines up to 20,000,000 EUR, or in the case of an undertaking, up to 4 % of the total worldwide annual turnover of the preceding financial year, whichever is higher (Article 83 (5)(a) GDPR).

Online tracking and advertising purposes Cabañas et al. [21] explored the link between tracking, advertising purposes, and inferences. They show that a visit to a website reveals sensitive related health data. The Irish DPA reported that health related websites using advertising and targeting cookies, including cookies set by DoubleClick that is owned by Google, share details of illnesses with third parties through profiles based on unique identifiers, without a lawful basis [34]. The French DPA refers that health data can be derived from crossing data allowing inferences on health status or health risk of a person [24].

5 RESULTS

In this section, we describe the tracking and consent collection we observed on health websites (see Section 3.1 for the full set of tracking categories that Ernie detects).

For the sake of simplicity, we consider the following two main categories of tracking: *analytics* that corresponds to Third party

	# Websites	# Contact	# Appointment
Personal	120 (53%)	73 (58%)	27 (59%)
Hospital/Joint office	201 (65%)	129 (55%)	54 (56%)
Aggregator	64 (66%)	_	_
Total	385 (62%)	202 (56%)	81 (57%)

Table 3: Visited websites before interaction by type. *The full list of 385 analysed websites can be found in the supporting material [7].* Between parenthesis, we present the percentage of websites including cross-site tracking before interaction. *#Websites*, includes the cross-site tracking on contact or appointment pages.

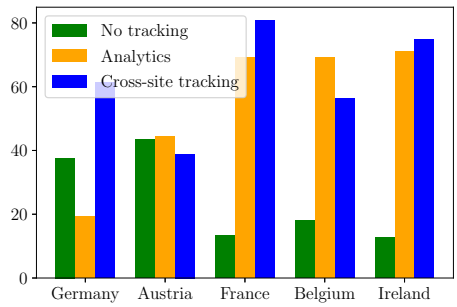


Figure 3: Percentage of websites performing cross-site tracking and analytics before interaction per country.

analytics, and *cross-site tracking* that corresponds to Basic tracking (BT), Basic tracking initiated by another tracker (BTIT), First to third party cookie syncing (FTCS), Third to third party cookie syncing (TTCS), and Third party cookie forwarding (TF). Unlike cross-site tracking, analytics only recognizes users within the same website and hence does not allow to recreate the browsing history.

We consider that a domain is performing analytics (resp. cross-site tracking) on a website *site.com* if it is performing analytics (resp. cross-site tracking) on *site.com* in either Firefox or Chrome.

5.1 Websites analysis

In this section, we analyze the 385 health related websites. First, we present cross-site tracking by category of websites, then the specificity of contact and appointment pages within a health related website, and finally, the impact of geographic distribution on cross-site tracking and analytics. All results reported in this section occur before interaction with a consent pop-up.

Cross-site tracking per website category. We successfully visited a total of 385 websites, out of which, as a result of our manual labelling (Section 3.2.1), 120 are categorized as `Personal`, 201 as `Hospital/Joint office`, and 64 as `Aggregator` (Table 3). As shown in Table 3, cross-site tracking is present on all studied website categories. Overall, `Aggregator` sites are the most common to include cross-site tracking. In fact, 66% of the visited websites in that category include cross-site tracking before interacting with a consent pop-up.

Specificity of contact and appointment pages. On `Personal` and `Hospital/Joint office` websites, we make an additional visit to contact and appointment pages if they exist in the website. As a result, we visited 202 contact pages and 81 appointments pages before interaction with a consent pop-up (Table 3).

Cross-site tracking is widely deployed on contact and appointment pages. When only considering the visits to the initial page of `Personal` and `Hospital/Joint office` websites, 41% and 55% of websites include cross-site tracking, compared to the aforementioned 53% and 65% when including contact and appointment pages. Overall, we found cross-site tracking on 56% of contact and 57% of appointment pages. A visit including a contact page therefore is more likely to result in tracking of a user.

Google Maps is commonly included in contact pages to provide the location of the doctor's office. We found that 76 (67%) of contact pages include Google Maps as a cross-site tracker, compared to 25% of initial pages (Section 3.2.3). When visiting `google.com` directly, we found that no cookies are set in the user's browser before interaction. However, upon our visit to the base profile websites, `google.com` content was loaded in a third-party context, and the NID cookie was stored in the user's browser by `google.com` as part of the HTTP response. According to `google.com`'s policy, the NID cookie expires 6 months from a user's last use [52]. When accessing a website containing Google content, the NID cookie is automatically attached with every request to the `google.com` sub-domain. In all 76 contact pages that include Google Maps as a cross-site tracker, the cookie is never set by Google when the map is loaded from `maps.google.com`, `maps-api-ssl.google.com` or `google.com/maps/`, but is automatically sent to Google because it is attached in the cookie header of the HTTP request. In total, 12 of 24 domains that perform cross-site tracking on contact pages never actively set an identifying cookie.

Geographical distribution. We present the distribution of analytics and cross-site tracking across the 5 studied countries in Figure 3. We remind that the reported behavior occurs before interaction with the cookie banner. We observe that France, Ireland and Belgium show a similar tracking distribution, with France having the highest percentage of websites performing cross-site tracking.

Summary. We observed a higher amount of cross-site tracking when including contact and appointment pages (53% for `Personal` and 65% for `Hospital/Joint office`) than on the websites homepage alone (41% and 55%, respectively). This fact raises privacy concerns, as visiting a contact or appointment page discloses the intent of the visitor more sharply rather than visiting a home page. We also found that the tracking by over half of the domains performing cross-site tracking on contact pages is due to the HTTP protocol mechanism. Namely, on 76 contact pages, the NID cookie is never explicitly set, but is sent by default with every request to google.com when the browser fetches Google Map included in the website.

5.2 Consent pop-ups

We manually grouped the consent pop-ups encountered in the visited health websites into the following categories that correspond to consent infringements:

Banner	# Websites	Before	After
No banner	155	97 (63%)	-
No rejection	112	82 (73%)	-
Rejection possible	118	58 (49%)	59 (50%)
Total	385	237 (62%)	59 (15%)

Table 4: Overview of types of consent-pop-ups and cross-site tracking on studied websites before and after rejection.

- **No banner**: No consent pop-up is present on any of the visited pages of the website. The absence of *any* method set forth to collect the user's explicit consent required to process health data, renders any forthcoming tracking unlawful due to the lack of legal basis (Article 9(2)(a) GDPR).
- **No rejection**: The website includes a consent pop-up, however, it does not provide a possibility to reject. Such consent pop-ups are in breach of the "configurable" and "balanced choice" consent requirements (Articles 4 (11), 7(3) GDPR) [10, 53, 74], which are compulsory for a valid unambiguous and freely given consent collection that is in line with the principle of "data protection by design and by default" (Article 25 GDPR).
- **Rejection possible**: The rejection of cookies is possible in the consent pop-up.

As a result of our manual labelling, we found that out of the 385 visited websites across all the website categories (`Personal`, `Joint` and `Aggregators` - Section 3.2.1), 155 (40%) do not have a consent pop-up. Out of the remaining 230 websites that include a consent pop-up, 112 (43%) websites do not offer a reject option (Table 4).

5.2.1 Tracking before interaction. Tracking occurs on websites with no banner. In fact, 97 websites (63% of the 155 websites with no consent pop-up) include cross-site tracking.

Out of 112 websites that display a consent pop-up with no option to reject, we detected cross-site tracking on 82 (73%) websites and analytics on 70 (64%) websites. We further analyzed the 118 websites where it's possible to reject: 58 (49%) include cross-site tracking before interaction and 37 (31%) include analytics.

Figure 4 presents the top 10 domains performing cross-site tracking and analytics before any interaction with the consent pop-up. The domain `google.com` is the top cross-site tracking domain, tracking users on websites with all types of consent pop-ups. In total, 41 domains track users cross-site before interaction even though rejection is possible on the website. As expected, `google-analytics.com` is the top domain performing analytics. However, to our surprise, `doubleclick.net` [4] is the second most popular domain performing analytics before interaction followed by `facebook.com`. We explain in details why `doubleclick.net` appears as an analytics service in Section 5.3.2. On the 24 websites where `facebook.com` appears as an analytics domain, it receives the first-party cookie *_fbp*. According to `facebook.com`, this cookie identifies browsers for the purposes of providing advertising and site analytics services and has a lifespan of 90 days [44]. Interestingly, on all of these websites, at least one of the HTTP responses by `facebook.com` contains an empty Set-Cookie header.

5.2.2 Tracking after rejection. In this section, we focus on the 112 websites where consent pop-up contains a possibility to reject, in order to compare tracking before interaction and after rejection. We observe cross-site tracking on 59 websites compared to 58 websites before interaction. We further analyzed the 59 websites with tracking after rejection, and we found that

- the number of domains performing cross-site tracking remains the same before and after rejection for 39 (66%) sites.
- the number of tracking domains even grows after rejection in either Firefox or Chrome for 15 (25%) websites. The most common trackers newly appearing after rejection are again `google.com` and `doubleclick.net` on 2 websites each.
- out of these 15 websites, 6 websites did not include any tracking before interaction.

In total, we detect 43 domains that perform cross-site tracking after rejection, and 23 domains that perform analytics. Figure 4 presents the top domains that are tracking users after rejection. The top domain that tracks users cross-site after rejection is `google.com` on 42 websites. The top domain performing analytics is `google-analytics.com` on 26 websites.

Summary. We observe cross-site tracking on health websites regardless of the presence of a consent pop-up. In fact, we found that 63% of health related websites with no consent pop-up include cross-site tracking. This practice breaches the requirements for an explicit and prior consent request. Moreover, we also found that 50% of the websites that include a consent pop-up with a possibility to reject still enable cross-site tracking after rejection. This practice violates the lawfulness principle (Article 6(1)(a) GDPR).

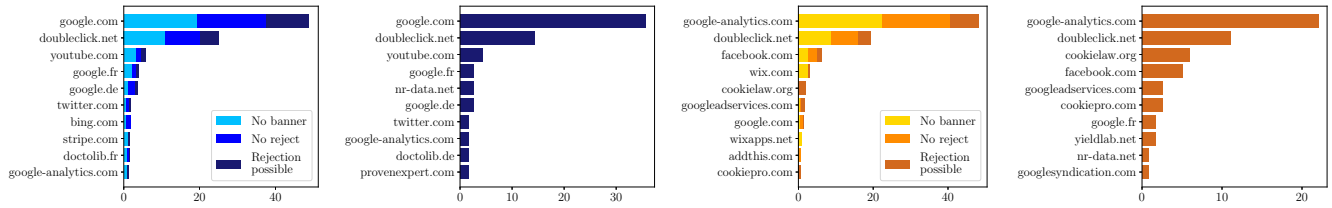
5.3 Cross-site tracking on health websites

In this section, we present the different cross-site tracking behaviors detected on the studied 385 health related websites using Ernie. We report the results from both studied browsers: Firefox and Chrome. We used the browsers default setting, therefore, Enhanced Tracking Protection (ETP) was enabled in Firefox (see Section 3.2.2).

5.3.1 Tracking before interaction. Before interaction with the consent pop-up, we observed that all studied tracking categories occur on health websites. As presented in Section 5.2, 62% out of the studied 385 websites perform at least one type of cross-site tracking behavior. Figure 5 depicts an overview of the tracking behaviors both in Firefox and Chrome. Note that a single website can include multiple tracking behaviors.

Basic tracking. We found that basic tracking is the top tracking category on health websites before interaction, as it is present on 219 websites in Firefox and 214 websites in Chrome (57% and 56% of all websites). Basic tracking initiated by another tracker is present on 25 websites in Firefox and 30 in Chrome. The most common tracking domain initiating requests to other tracking domains is `youtube.com`. It redirects to `google.com` on 18 websites both in Firefox and Chrome and additionally redirect to `doubleclick.net` on 20 websites in Chrome.

Cookie syncing. We study all cookie syncing tracking categories (*First to third party cookie syncing* (FTCS), *Third to third party cookie syncing* (TTCS), and *Third party cookie forwarding* (TF)) performed on websites before any interaction with the banner.



(a) Cross-site tracking before interaction **(b) Cross-site tracking after rejection** **(c) Analytics before interaction** **(d) Analytics after rejection**

Figure 4: Top 10 domains that perform cross-site tracking or analytics before interaction depending on the consent pop-up.

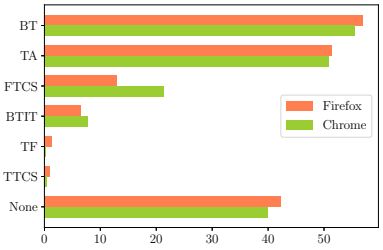


Figure 5: Percentage of websites including each tracking behavior before interaction. See Section 3.1.2 for the explanations of the abbreviations.

	# Firefox	# Chrome
google.com	41	39
doubleclick.net	0	73
bing.com	7	7
quantserve.com	3	3
nr-data.net	2	2
Total	50	82

Table 5: Top 5 domains performing first to third party cookie syncing by number of websites they occur on in either Firefox and Chrome before interaction.

Using Ernie, we detected cookie synchronization on 53 websites in Firefox and 82 websites in Chrome before interaction. This cookie synchronization is performed by 20 and 23 distinct third-party domains, respectively. We detected instances of first to third party cookie syncing on 50 websites in Firefox and 82 websites in Chrome. Table 5 presents the top domains performing first to third party cookie syncing. First to third party cookie syncing is more common in Chrome, where it occurs on 21% of the visited websites. In fact, the top domain performing first to third party cookie syncing in Chrome, `doubleclick.net`, is never synchronizing first party cookies in Firefox. The reason is the Firefox ETP, which is enabled by default in Firefox to ensure protection against tracking techniques at the level of the browser. The deployment of such protection mechanism can block identifying cookies and consequently impact the tracking behaviors experienced by the user. Here, ETP is blocking identifying cookies from `doubleclick.net`, most notable the *IDE* cookie used for tracking in Chrome. As a result, the synchronization attempts by `doubleclick.net` are categorized as analytics behavior in Firefox (the first party cookie is still shared with `doubleclick.net`, but no ID cookie is set or sent).

Additionally, we found that on all 75 websites where `doubleclick.net` is performing either first to third party cookie syncing or third party analytics in Chrome and Firefox, the website includes `google-analytics.com` as well. In all these cases, both `google-analytics.com` and `doubleclick.net` receive at least one identical first party cookie, which is always either *_ga* or *_gid*. These two first party cookies shared with `doubleclick.net` on the 75 websites belong to `google-analytics.com` [50].

We additionally detected instances of third to third party cookie syncing on 4 websites in Firefox and 3 websites in Chrome, as well as third party cookie forwarding on 5 and 2 websites respectively.

5.3.2 Tracking after rejection. In this section, we focus on the 118 websites where it is possible to reject, to compare tracking before and after rejection. We consider the websites pages that were both accessible before interaction and after rejection.

To our surprise, after rejection, the number of websites where tracking occurs is higher than before interaction. Table 6 shows tracking behavior detected on the studied 134 websites.

Basic tracking. Basic tracking and basic tracking initiated by another tracker occurred more often after rejection in Firefox. The domains that appeared additionally in those tracking categories in both Firefox and Chrome are `cookiefirst.com`, `clarity.ms` and `gigya.com`. The first two of these domains were also present before interaction, but only started tracking after rejection. One of the ID cookies used by `clarity.ms`, *MUID*, is used for advertising [22]. The last domain, `gigya.com`, was only loaded after interaction. Both it's ID cookies, *gmid* and *ucid*, are used for user identification [49].

Cookie Syncing. After rejection, we detected that no website stopped first to third party cookie syncing after rejection in Firefox, whereas in Chrome, 4 websites stopped performing first to third party cookie syncing to `google.com` and `doubleclick.net`. Additionally, 1 website started first to third party cookie syncing only after rejection in both Firefox and in Chrome.

Summary. We detected cookie syncing behavior on 53 websites in Firefox and 82 in Chrome before interaction. We found that `doubleclick.net` is performing first to third party cookie syncing only in Chrome. In Firefox, identifying cookies by `doubleclick.net`

	# Firefox		# Chrome	
	before	after	before	after
BT	54	57	52	54
TA	36	33	35	32
FTCS	7	8	18	15
BTIT	7	8	8	7
TTCS	1	2	0	1
TF	2	2	0	0
None	64	61	62	62

Table 6: Number of websites (out of 118 where it is possible to reject) that include tracking before and after rejection.

are blocked by ETP, which results in this behavior becoming analytics. Tracking before interaction infringes the requirement of prior consent, as tracking is deemed unlawful if carried out before consent is requested (Article 6(1)(a) GDPR) [40].

6 CASE STUDIES

In this section, we perform an in-depth technical and legal analysis of three different health related websites. We chose three cases across our dataset that are of special interest from a legal point of view. For each case, we selected the website that includes the most domains performing cross-site tracking. We analysed the type of pages and identified the legal violations per case study. We contacted the owners of these websites, informing them of the violations on their website, but have not yet received answers.

6.1 Appointment page on a personal doctors website with no possibility to reject

The website dermatologie-weissensee.de is the personal website of a Berlin-based dermatologist [31]. When visiting the website's appointment page in Firefox, an iframe by `doctena.com` is loaded. This iframe allows users to book an appointment. It also includes a consent pop-up informing users that cookies are used, but does not give a possibility to reject.

On this page, 6 domains are classified as cross-site trackers by Ernie. Notably, `google-analytics.com` and `doubleclick.net` are performing third party analytics receiving the *_ga* and *_gid* cookies from the first party `dermatologie-weissensee.de`. Both *_ga* and *_gid* cookies are used to distinguish users, and respectively have a lifetime of 2 years and 24 hours [51]. Using these first-party cookies, the two domains can track the user within the same website. Additionally, `google-analytics.com` and `doubleclick.net` perform third party cookie forwarding and receive the *_ga* and *_gid* cookie, this time from `doctena.com`. This behavior enables tracking across websites. We also found that `google.com` is synchronizing both the first-party cookie *_ga* of `dermatologie-weissensee.de` and the third-party cookie *_ga* of `doctena.com` with its own third-party cookie *NID* using first to third party cookie syncing and third to third party cookie syncing.

Legal analysis. According to Google's privacy policy [52] the *NID* cookie is used for advertising purposes, thus requiring user consent. *Type of page.* A GP appointment in a personal doctor website, in isolation, might not tell anything about a person's health, as it may be a check-up or screening appointment. However, as the UK DPA states, one could *reasonably infer health data from an individual's list of appointments* [60]. A person may be associated with cookie identifiers which may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them (Recital 30 GDPR). Accordingly, cookies enable users to be 'singled out' with an online identifier (even if their real names are not known) [70]. A personal health related website embedding cross-site tracking may disclose users' sensitive medical data (such as the illness or health condition that justifies booking an appointment) for advertising purposes with an ecosystem of advertisers, without the knowledge of the user. Even if it is the personal website of a doctor, it is not exempted of compliance obligations regarding health data processing, in particular, the obligation to protect user's privacy by design and by default (Article 25 of the GDPR).
Breaches. This website does not offer any *possibility to reject.* which conflicts with requirements of "configurable banner" and "balanced choice" (Articles 4 (11), 7(3) GDPR) [10], which are compulsory for an unambiguous consent of a user; and the principle of "data protection by design and by default" that demands the most privacy-friendly default settings to be used (Article 25 GDPR). *Cross-site tracking.* Cookie syncing breaches the following principles. Fairness principle, because it disregards the legitimate expectations of the data subject at the very time of data collection. Any disclosure to third parties of sensitive data is out of any user reasonably expectations (Article 5(1)(a) GDPR). *Transparency principle*: users should be informed of the existence of cookie syncing operations and its purposes (Recital 60 GDPR), of the extent, risks, and consequences of cookie syncing (Recital 39), including *profiling* and the rights and safeguards they are afforded with (Articles 13(2)(f), 22(1)(4) GDPR).

6.2 Aggregator website after rejection

The aggregator website 118000.fr is an online phone book to find contact of professionals for various categories. We analyzed a specific subpage of `118000.fr` to search for specialists of gynecology [11]. The website presents to users a consent pop-up with a possibility to reject. To reject cookies, the user first needs to choose the option "Adjust your preferences" in the consent pop-up, then at a second level, the user needs to choose the option "Reject all". We rejected all cookies and found, using Chrome, that 6 domains are performing cross-site tracking on this website after rejection: `le118000.fr`, `doubleclick.net`, `pagesjaunes.fr`, `consentframework.com`, `mediakiosque.com`, and `tribalfusion.com`. We analyzed the policies of these domains, and we found that 4 do not provide a description of the cookie purpose, and 2 domains (`doubleclick.net` [4], and `tribalfusion.com` [78]) state that they are using cookies for advertising purposes.

Legal analysis. Both the consent pop-up and the policies refer to at least one purpose (personalized ads profile) that requires consent. Cross-site tracking in this aggregator website allows third parties to build more detailed user profiles relying on user's browsing history (related to the type of health specialties the user is interested in) in order to serve more relevant advertising. The EDPB [15] illustrates that targeted ads based on profiling might have significant effects on users depending on the particular characteristics of the case, such as:

i) the intrusiveness of the profiling process, including the tracking of individuals across different websites, devices and services (which is the case of cross-site tracking); ii) the expectations and wishes of the individuals concerned; or iii) using knowledge of the vulnerabilities of the data subjects targeted (such as health related vulnerabilities). We argue that the amount and variety of personal health collected for advertising purposes across the above-mentioned websites is against their expectations and wishes, and might have a significant effect on users regarding their concrete health status.

Type of page. Even if all websites have a responsibility to protect the privacy of their visitors and comply with existing laws, a particularly severe case is the one of an aggregator website that shares granular and health data with third parties for advertising purposes. This aggregator website lists more than 80 partners that may process the user's personal and sensitive health data. As claimed by the EDPB, *combination* of personal data (containing health data) is more sensitive than a single piece of personal data [16]. As an aggregator website is operated by a company offering a privacy-sensitive service, visitors might reasonably expect greater privacy protection. Both DPAs (Irish, UK) showed serious concerns about tracking for advertising purposes in health websites, since the chances to build and enrich unique profiles of users is bigger [33, 58], which can reveal health risks of these same users. Having regard to aggregator websites, the possibility to build users profile is then augmented. Moreover, the processing of health data on a large scale (as operated by cross-site tracking in aggregator websites) makes plausible to discern different categories of health data. Such processing at large scale results in a *high risk*, and therefore requires, prior to the processing, for the data controller to carry out a data protection *impact assessment* (DPIA) [18] of the envisaged processing operations on the protection of personal data (Article 35(3)(b)GDPR).

Breaches. This website aggregator triggers the following issues: *Tracking after rejection.* This practice breaches the *lawfulness principle* (Articles 5(1)(a), 6(1)GDPR, 5(3) ePD) since consent to the use of cookies is required. The EDPB [40] specified that if the user decided against consenting, any data processing that had already taken place would be unlawful due to lacking legal basis for processing.

6.3 French website containing first to third party cookie syncing

The aggregator website `starofservice.com` is an online phone book to search for the contact of professions on different categories. We analyzed a specific subpage used to search for a pediatrician [76]. When visiting the cited URL, no consent pop-up is presented to the user. Using Ernie, we detected that three domains performed first to third party cookie syncing: `google.com` [52], `bing.com` [19], and `doubleclick.net` [4]. All three domains are using advertising cookies. We found that both `google.com` and `doubleclick.net` are synchronizing the first-party cookie *_ga* set by google-analytics.com with their third-party cookies (NID and IDE). The *_ga* cookie normally used for analytics is in that case synchronized with third-party cookies and can therefore help link user's activity within the website with her activity across websites. Figure 6 shows a screenshot of Ernie on this website.

Legal analysis. The EDPB [17] refers that analytics are exempted from user consent insofar they are limited to first party (website owner) anonymized and aggregated statistical purposes, as these are not likely to create a privacy risk.

Type of page. This aggregator website is using Google analytics and its cookie is then synced with doubleclick for advertisement purposes without explicit user consent. Some DPAs [54, 58] are stringent by declaring that Google Analytics, used for web analytics purposes, require the user's prior consent [59] asserting these technologies are not considered *strictly necessary* for a website to provide a functionality explicitly requested by the user, because the user can access the website when such cookies are rejected. Relatedly, the French DPA [23] fined the Europe's largest hypermarkets chain called Carrefour since it placed Google Analytics cookies on the users' devices without consent. It ruled that integrating Google Analytics with Google Ads allowed advertisers to merge data to identify their most interesting segments and then engage those users with personalized messages. Consequently, these cookies are not strictly necessary for the provision of the service and require consent (Article 5(3)ePD). Combined data, according to the EDPB, enable to draw a conclusion about the actual health status or health risk of a person. [57]. Regarding the use of analytics, the French DPA [26] in general takes a moderate position. It states that consent is hence required whenever tracers allow the overall monitoring of the navigation of the person using different applications or browsing different websites, or when data stemming from such tracers are combined with other processing operations or transmitted to third parties, these different operations not being necessary for the operation of the service. In the context of health related websites, we see that analytics services are not the subject of a consensus of DPAs on the need for user consent. However, considering the privacy implications of the collected data (that might disclose health related information) and considering that inclusion with third parties might leak health related information, we argue that analytics services must be subject of explicit user consent on health related websites.

7 CONCLUSION

In this paper we have gleaned robust evidence of tracking technologies deployed on health-related websites (before user consent interaction and also after rejection). Our open-source browser extension Ernie can be used to collect further evidence and demonstrate cookie-based tracking technologies and sophisticated cookie syncing techniques employed on websites. We hope that the Ernie extension can be beneficial to both policy-makers, to advance the enforcement of EU Privacy and Data Protection law, and to DPO's of health websites that so far had no access to such visualisation tools. We have further contacted the website owners that we mention in our case studies and we are willing to help them change their practices towards improving the afforded protection of privacy and health data of Web users.

ACKNOWLEDGMENTS

We would like to thank our colleague Damian Clifford for his valuable input on Section 4. This work has been partially supported by the ANR JCJC project PrivaWeb (ANR-18-CE39-0008), the ANSWER project PIA FSN2 No. P159564-2661789/ DOS0060094 between Inria and Qwant, and by the Inria DATA4US Exploratory Action project.

REFERENCES

[1] Alexa Top Sites. https://www.alexa.com/topsites.

[2] Alexa websites visited. https://www.dropbox.com/sh/nwjw7ggcx08o1x7/AACYrHqsxo7DcZjbVArE5Fxua?dl=0.

[3] Contextual Identities. https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/API/contextualIdentities.

[4] Cookie Guide. https://developers.google.com/authorized-buyers/rtb/cookie-guide.

[5] Enhanced tracking protection in Firefox for desktop. https://support.mozilla.org/en-US/kb/enhanced-tracking-protection-firefox-desktop.

[6] Google Cookie Types. https://policies.google.com/technologies/types.

[7] List of websites visited. https://www.dropbox.com/s/l1ebx791ipp12xn/visited_urls.txt?dl=0.

[8] Pupeteer. https://github.com/puppeteer/puppeteer.

[9] Guidance on the use of cookies and similar technologies, 2019. https://ico.org.uk/media/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies-1-0.pdf.

[10] Judgment in Case C-673/17 Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV v Planet49 GmbH, 2019. http://curia.europa.eu/juris/documents.jsf?num=C-673/17.

[11] 118000.fr website. https://www.118000.fr/v_paris_75/c_gynecologue-obstetricien-medecin-specialiste-en-gynecologie-obstetrique.

[12] 29WP Opinion 03/2016 on the evaluation and review of the ePrivacy Directive (2002/58/EC), Adopted on 19 July 2016,. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2016/wp240_en.pdf.

[13] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juárez, Arvind Narayanan, and Claudia Díaz. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pages 674–689, 2014.

[14] Guide on use of cookies, 2021. https://www.aepd.es/sites/default/files/2021-01/guia-cookies-en.pdf.

[15] Article 29 Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251 rev.01).

[16] Article 29 Working Party. Guidelines on Personal data breach notification under Regulation 2016/679. https://ec.europa.eu/newsroom/article29/items/612052/en.

[17] Article 29 Working Party. Opinion 04/2012 on Cookie Consent Exemption (WP 194).

[18] Article 29 Working Party. WP 248 Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.

[19] Bing policy. https://www.timeshighereducation.com/cookie-policy.

[20] 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. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202003_healthdatascientificresearchcovid19_en.pdf.

[21] J. Cabañas, Ángel Cuevas, and R. C. Rumín. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. In *USENIX Security Symposium*, 2018.

[22] Microsoft Clarity Cookie List. https://docs.microsoft.com/en-us/clarity/cookie-list, accesed 16. July 2021.

[23] CNIL: Délibération de la formation restreinte n° san-2020-008 du 18 novembre 2020 concernant la société CARREFOUR FRANCE. https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000042563756. Accessed on 19 March, 2021.

[24] Qu'est-ce ce qu'une donnée de santé ? https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante. Accessed on 18 May 2021.

[25] Code de la santé publique, version in effect as of February 27, 2021. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000036515027/. Translated with DeepL https://www.deepl.com on February 27, 2021.

[26] Commission Nationale de l'Informatique et des Libertés (French DPA). French guidelines on cookies: Deliberation No 2020-091 of September 17, 2020 adopting guidelines relating to the application of article 82 of the law of January 6, 1978 amended to read and write operations in a user's terminal (in particular to "cookies and other tracers"), 2020. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042388179.

[27] Consent-O-Matic browser extension. matic/mdjildafknihdffpkfmmpnpoiajfjnjd.

[28] Cookiebot. Cookie Scanner for GDPR/ePR and CCPA Compliance. https://www.cookiebot.com/en/cookie-scanner/.

[29] Court of Justice of the EU. C-101/01, LINDQUIST, 6.11.2003 ECLI:EU:C:2003:596. https://curia.europa.eu/juris/liste.jsf?num=C-101/01.

[30] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We Value Your Privacy... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy.

[31] Dermatologie-weissensee.de website. https://www.dermatologie-weissensee.de/termine/.

[32] Disconnect Official website. https://disconnect.me/.

[33] Guidance note on the use of cookies and other tracking technologies, 2020. https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Guidance%20note%20on%20cookies%20and%20other%20tracking%20technologies.pdf.

[34] Report by the Data Protection Commission on the use of cookies and other tracking technologies, 2016. https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Report%20by%20the%20DPC%20on%20the%20use%20of%20cookies%20and%20other%20tracking%20technologies.pdf.

[35] EFF. Privacy Badger. https://privacybadger.org/.

[36] Steven Englehardt and Arvind Narayanan. Online Tracking: A 1-million-site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security ACM CCS*, pages 1388–1401, 2016.

[37] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. Cookies That Give You Away: The Surveillance Implications of Web Tracking. In *Proceedings of WWW 2015*, pages 289–299, 2015.

[38] Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws (Text with EEA relevance). Directive 2009/136/EC, 2009. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009L0136.

[39] European Data Protection Board. Opinion 4/2007 on the concept of personal data (WP 136), adopted on 20.06.2007. https://ec.europa.eu/justice/article-29/documentation/opinionrecommendation/files/2007/wp136_en.pdf.

[40] European Data Protection Board. Guidelines 05/2020 on consent, Version 1.1, adopted on 4 May 2020, 2020. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

[41] European Data Protection Board (EDPB). Opinion 2/2010 on online behavioural advertising, 22 June 2010, WP171, p. 10.

[42] European Data Protection Board (EDPB). Guidelines 05/2020 on consent under Regulation 2016/679, 2020.

[43] EZIGDPR. GDPR Website Compliance Check. https://www.ezigdpr.com/products/gdpr-website-compliance-checker.

[44] Facebook Privacy Policy. https://www.facebook.com/policies/cookies.

[45] Imane Fouad, Nataliia Bielova, Arnaud Legout, and Natasa Sarafijanovic-Djukic. Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels . *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2020, 2020. Published online: 08 May 2020, https://doi.org/10.2478/popets-2020-0038.

[46] Imane Fouad, Cristiana Santos, Feras Al Kassar, Nataliia Bielova, and Stefano Calzavara. On Compliance of Cookie Purposes with the Purpose Specification Principle. In *2020 International Workshop on Privacy Engineering, IWPE*, 2020. https://hal.inria.fr/hal-02567022.

[47] 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/EC (General Data Protection Regulation) (Text with EEA relevance). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32016R0679.

[48] Ghostery Official website. https://www.ghostery.com/.

[49] SAP Advanced Cookie Reference. https://help.sap.com/viewer/8b8d6fffe113457094a17701f63e3d6a/GIGYA/en-US/41419ee070b21014bbc5a10ce4041860.html, accessed 16. July 2021.

[50] Google analytics solutions. https://www.google.com/analytics.

[51] Google.com cookie usage. https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage.

[52] Google policy. hhttps://policies.google.com/technologies/cookies?hl=en-US.

[53] Colin Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damien Clifford. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *ACM CHI 2021*, 2020. https://arxiv.org/abs/2009.10194.

[54] Greek DPA (HDPA). Guidelines on Cookies and Trackers, 2020. http://www.dpa.gr/APDPXPortlets/htdocs/documentSDisplay.jsp?docid=84,221,176,170,98,24,72,223.

[55] Harward Business Review. What Patients Like — and Dislike — About Telemedicine. https://hbr.org/2020/12/what-patients-like-and-dislike-about-telemedicine accessed on 27 February 2021.

[56] Raymond Hill and Contributors. uBlock Origin. https://github.com/gorhill/uBlock/.

[57] Information Commissioner's Office. Article 29 WP, Annex 2015), 2015. https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf.

[58] Information Commissioner's Office. Update report into adtech and real time bidding. https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf, accessed on 2019.07.10, 2019.

[59] Information Commissioner's Office. Hellenic Data Protection Authority guidance on the use of cookies (and similar technologies), 2020. https://iapp.org/resources/article/cookie-guidance-from-greece/.

[60] Information Commissioner's Office. ICO guide on special category of data, 2020. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/what-is-

special-category-data/#scd5.
[61] Timothy Libert. Privacy implications of health information seeking on the web. *Communications of the ACM*, 58(3):68–77, 2015.
[62] LINC. Cookieviz 2: new features to observe hidden web practices. https://linc.cnil.fr/fr/cookieviz-2-new-features-observe-hidden-web-practices.
[63] Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *IEEE Symposium on Security and Privacy (IEEE S&P 2020)*, 2020.
[64] McAfee categorization service. https://www.trustedsource.org/.
[65] Description of McAfee categories. https://www.trustedsource.org/download/ts_wd_reference_guide.pdf.
[66] Mozilla. Lightbeam 3.0. https://addons.mozilla.org/en-GB/firefox/addon/lightbeam-3-0/.
[67] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *CHI*, 2020.
[68] Emmanouil Papadogiannakis, Panagiotis Papadopoulos, Nicolas Kourtellis, and Evangelos P. Markatos. User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users. In *Proceedings of WWW 2021*, 2021. https://arxiv.org/abs/2102.08779.
[69] Panagiotis Papadopoulos, Nicolas Kourtellis, and Evangelos P. Markatos. Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 1432–1442, 2019.
[70] Article 29 Data Protection Working Party. Opinion 2/2010 on online behavioural advertising, Adopted on 22 June 2010. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp171_en.pdf.
[71] Piwik. Free Online Cookie Scanner. https://piwik.pro/cookie-scanner/.
[72] PrivateVPN website. https://privatevpn.com/.
[73] Iskander Sánchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control. In *Proceedings of the ACM Asia Conference Computer and Communications Security*, pages 340–351, 2019.
[74] Cristiana Santos, Nataliia Bielova, and Célestin Matte. Are cookie banners indeed compliant with the law? Deciphering EU legal requirements on consent and technical means to verify compliance of cookie banners, journal=Technology and Regulation. pages 91–135, 2020.
[75] Matic Srdjan, Iordanou Costas, Smaragdakis Georgios, and Nikolaos Laoutaris. Identifying Sensitive URLs at Web-Scale. In *ACM Internet Measurement Conference (ACM IMC 2020)*, 2020.
[76] Starofservice website. https://www.starofservice.com/annubis/ile-de-france/paris/paris/pediatrie.
[77] European Data Protection Supervisor. EDPS Inspection Software. https://edps.europa.eu/press-publications/edps-inspection-software_en.
[78] Tribalfusion policy. https://www.havaianas-store.com/fr/Cookies+policy.html.
[79] Christopher Uner, Lee A. Bygrave, Christopher Docksey, Laura Drechsler, and Luca Tosoni. The EU General Data Protection Regulation: A Commentary/Update of Selected Articles. 2021. https://global.oup.com/academic/product/the-eu-general-data-protection-regulation-gdpr-9780198826491?cc=pt&lang=en&.
[80] Tobias Urban, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. Beyond the Front Page: Measuring Third Party Dynamics in the Field. In Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen, editors, *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pages 1275–1286. ACM / IW3C2, 2020.
[81] Pelayo Vallina, Álvaro Feal, Julien Gamba, Narseo Vallina-Rodriguez, and Antonio Fernández Anta. Tales from the porn: A comprehensive privacy analysis of the web porn ecosystem. In *Proceedings of the Internet Measurement Conference*, pages 245–258, 2019.
[82] Webcookies. Web Cookies Scanner. https://webcookies.org/.
[83] Eva Wolfangel. Ist Ihr Arzttermin sicher?, Jun 2021.

APPENDIX

Doctors specialities. We extracted health related websites by searching for 10 doctors specialties in the capital cities of the studied countries in Google. We used the most commonly searched specialities as given by the aggregator websites `doctolib.fr` and `jameda.de`. We searched for ⟨ city ⟩ ⟨ speciality ⟩. We used French terms for France and Belgium, English terms for Ireland, and German terms for Germany and Austria. Table 7 shows the English, French and German translation of the specialities.

French	English	German
Chirurgien-dentiste	Dentist	Zahnarzt
Médecin généraliste	General practitioner	Allgemeinmediziner
Pédiatre	Paediatrician	Kinderarzt
Gynécologue médical et obstétrique	Gynaecologist and obstetrics	Gynäkologe und Geburtshilfe
Ophtalmologue	Ophthalmologist	Augenarzt
Dermatologue	Dermatologist	Hautarzt
Ostéopathe	Osteopath	Osteopath
Masseur-kinésithérapeute	Physiotherapist	Physiotherapeut
Orthopédiste	Orthopaedist	Orthopäde
Neurologue	Neurologist	Neurologe

Table 7: Top doctors specialties extracted from https://www.doctolib.fr/ and https://www.jameda.de/, translated into French, English and German.

Contact and appointment pages. On `Personal` and `Hospital/Joint office` websites, we additionally visited *contact* and *appointment* pages, which often contain third party content and whose visit at the same time discloses a users intent more sharply. To build consensus between the two authors doing the manual selection of contact and appointment pages, we defined a list of keywords for each language. The keywords are listen in Table 8. Only menu items or buttons containing these keywords are selected.

	French	English	German
Contact	Contact(er/ez), Nous trouver, Pour venir	Contact, Get in touch, How to find us	Kontakt, Anfragen, Anfahrt, Adresse
Appointment	Reservez, RDV, Rendez-vous	Appointments, Book	Termin

Table 8: Keyword list to identify contact and appointment pages for each language.

Ernie extension. Ernie helps to get a quick understanding of complex tracking behaviors on a given website. Figure 6 shows a screenshot of Ernie on `starofservice.com`. The extension shows the number of requests from each tracking category. The request to `doubleclick.net` is opened, showing in yellow the *_ga* and *_gid* cookies forwarded in the URL parameters, as well as its own ID cookie *IDE*.

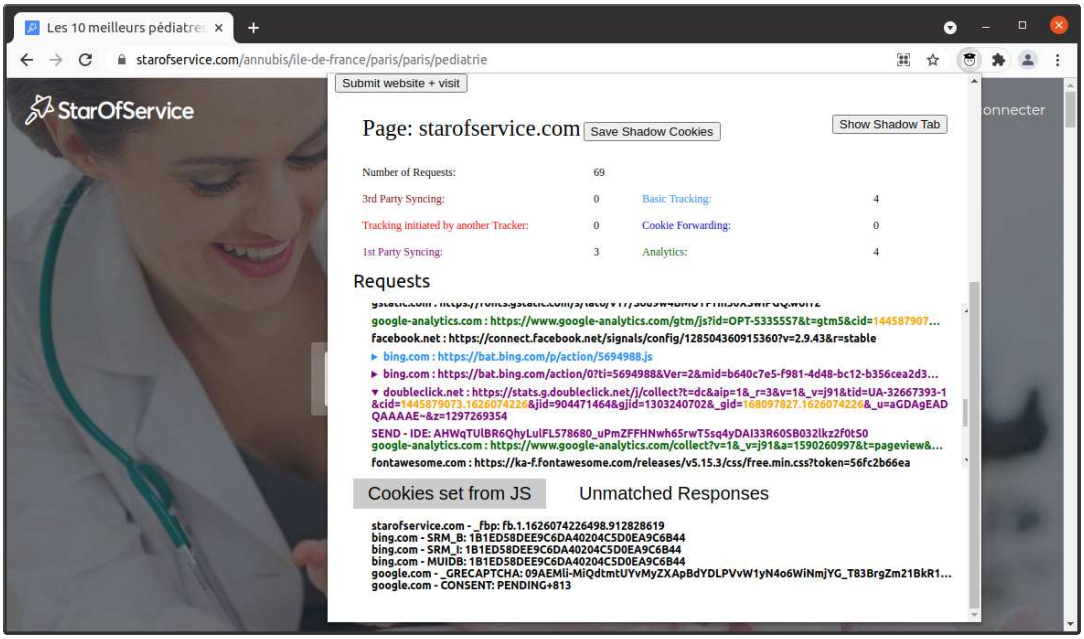


Figure 6: Screenshot of Ernie on `starofservice.com`.