In-depth technical and legal analysis of Web tracking on health related websites with Ernie extension

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

Vera Wesselkamp*, Imane Fouad*, Cristiana Santos, Nataliia Bielova, and Arnaud Legout

In-depth technical and legal analysis of Web tracking on health related websites with Ernie extension

Abstract: Searching for doctors online has become an increasingly common practice among Web users. However, when health websites owned by doctors and hospitals integrate third-party trackers, they expose their potential patients' medical secrets to third parties, thereby violating the GDPR which only allows the processing of sensitive health data with the explicit consent of a user. While previous works detected sophisticated forms of cookie syncing at scale, no tool exists as of today that would allow owners of health websites detecting complex tracking practices and ensure legal compliance. In this paper, we develop Ernie - a browser extension that visualises six tracking and complex cookie syncing state of the art techniques. We report on the analysis with Ernie on 176 websites of medical doctors and hospitals that users would visit when searching for doctors in France and Germany. At least one form of tracking or cookie syncing occurs on 64% websites before interacting with the consent banner, and 76% of these websites fail to comply with the GDPR requirements on a valid explicit consent. Furthermore, an in-depth analysis of case study websites allowed us to provide comprehensive general explanations of why tracking is embedded: for example, in all 45 webpages, where doctors include a Google map to help locating their office, tracking occurs due to the Google's cookie already present in the user's browser which is attached to a request that fetched the Google map useful content.

Keywords: online tracking, cookie syncing, browser extension, GDPR, explicit consent, health data

DOI Editor to enter DOI
Received ..; revised ..; accepted ...

1 Introduction

Health data is known to be one of the most sensitive types of data, and massive health data leaks is recognized to be of particularly high severity to the users' privacy, according to the French Data Protection Authority (CNIL) [22]. Searching for doctors online has become an increasingly common practice among Web users since telemedicine peaked in 2020 during the global Covid-19 pandemic [51]. However, the mere visit to a doctor's website can reveal a lot about its 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]. When providing services or monitoring user's behaviour in the EU, health related websites integrating third-party trackers are in breach with the General Data Protection Regulation (GDPR) [43] because processing of sensitive health data (derived from a visit to a website) is *generally forbidden*, unless allowed by several exceptions therein considered (Article 9(2) GDPR).

In the last decade, research has been focused on quantifying the prevalence of tracking based on cookies or lists of known tracking domains [15–17, 32, 33, 54, 56, 57, 71, 72], while several recent studies detected sophisticated forms of cookie syncing and ID sharing [42, 66, 67]. These studies were performed with customized large-scale crawlers and hard to replicate for non-experts. Moreover, quantitative studies measure the prevalence of various tracking techniques, but rarely explain *the reason why tracking is included*. This question is particularly important for health related websites that, differently from commercial websites, do not have an incentive to include targeted advertisement.

As a result, owners of health related websites, such as doctors and hospitals, have the urgent need to be able to detect tracking and advanced cookie synchroni-

***Co-first Author: Vera Wesselkamp:** Inria - Institut National de Recherche en Informatique et en Automatique
***Co-first Author: Imane Fouad:** Inria - Institut National de Recherche en Informatique et en Automatique
Cristiana Santos: Utrecht University [Utrecht]
Nataliia Bielova, Arnaud Legout : Inria - Institut National de Recherche en Informatique et en Automatique

1 According to the French Code of Public Health [23, Article L1110-4], medical secret covers "all information the person coming to the knowledge of the professional, of any member of the staff of these establishments, services or organizations and of any other person in relation, by virtue of his activities, with these establishments or organizations. It applies to all professionals working in the health care system".

sation techniques on their website in order to determine whether the included third parties may be leaking their patients' health data. While some browser extensions visualise known tracking third parties or third party cookies [28, 31, 44, 61, 64], *no browser extension exists as of today that is able to visualise sophisticated forms of cookie synchronisation and sharing of user's identifiers* [42, 66, 67] across third parties. Therefore, owners of health-related websites are in a difficult position where it is close to impossible to determine tracking and complex cookie syncing included in their websites.

Moreover, since processing health data is forbidden by the GDPR, health website owners can only rely on one exception and implement a specific type of consent mechanism, called *explicit consent*, to make such processing lawful for all third parties included in the website. However, even for a basic consent to be legally valid, it has to comply with at least 22 different fine-grained requirements [75]. While general websites implement cookie banners to comply with the legal requirement of consent, recent works made evident that in practice websites often do not contain any cookie banners, or contain banners that do not respect the user's choice [63, 65, 66, 74]. Therefore, doctors and hospitals need to ensure that if their health websites contain tracking or any form of sophisticated cookie syncing, a *valid and explicit consent must be collected* before any of such activities are included.

In this paper, we perform the first in-depth qualitative study of third-party tracking, including complex cookie syncing and ID sharing techniques on health related websites, that are mostly owned by doctors and hospitals in two EU countries: France and Germany. We designed a new Firefox browser extension called Ernie that performs a state-of-the-art detection and visualizes sophisticated forms of tracking and ID sharing on a visited website, based on 6 different categories of third-party tracking from Fouad et al. [42].

Instead of relying on categorisation services [74, 76], we carefully selected 176 websites that Web users would find whenever searching for particular doctors in 2 major French and German cities, and manually visited them with Ernie extension. With Ernie, we monitored and recorded all 6 categories of tracking techniques before and after interacting with the cookie banner. Finally, we performed a detailed legal analysis together with a legal expert, co-author of this paper, to understand when each technique is potentially violating the GDPR.

Unlike previous works that measured tracking *quantitatively* on a large scale, we opted for *a deep technical and legal qualitative analysis* of one case study website for each type of potential violation. This analysis helped us (1) to uncover the mechanisms used by trackers that circumvent Firefox's Enhanced Tracking Protection [40] used in our experiments; and (2) to identify the reasons why tracking is included in otherwise unsolicited health websites. This approach demonstrates the usefullness of Ernie browser extension that is a first prototype of an extension that can be further used by non-expert users[2].

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 to the state-of-the-art methodology from Fouad et al. [42].

(2) We perform a legal and technical analysis of consent collection on 176 health related websites and identify practices potentially violating the GDPR and the ePrivacy directive. We found that 64% of the websites track users before any interaction with the banner. Moreover, 76% of these websites fail to comply with the legal requirements for a valid explicit consent: out of 176 studied websites, 46% do not display a cookie banner, and 75% thereof still contain tracking, thus violating the *explicit consent* legal requirement; 26% of the websites provide a cookie banner without a reject button, and 86% of these websites include tracking, hence violating the requirement to give users *the possibility to reject tracking*. Moreover, we show that the *user choice is not respected* on health related websites: 33 (19%) websites still contain tracking after cookie rejection.

(3) We analyse in depth 5 case study websites, one per each type of tracking and legal violation, to provide a comprehensive explanation of why tracking is happening on health related websites. Such in depth analysis helped us to conclude which techniques companies use to

2 We will make the Ernie available and open-source upon acceptance of this paper.

3 The main goal of this extension is to provide an easy-to-use tool for the non-experts, such as doctors, the end users and research community, NGOs and legal experts to visualise complex tracking and the regulatory authorities to evaluate compliance.

deploy tracking even in privacy-friendly browsers, such as Firefox ETP [40]. We found that in every 45 webpages wherein doctors include a Google map to help locating their office, tracking occurs. While Google maps doesn't explicitly track users, tracking happens because of the `NID` cookie of `google.com` that is already present in the user's browser, and the HTTP standard [53] requires cookies to be automatically attached to every outgoing HTTP(S) request. Moreover, we found that such practice not only enables tracking with Google map content, but it also enables explicit tracking on 84 (47.73%) websites.

2 Related Work

In this section we provide an overview of previous works related to the interaction with cookie banners and also related to detection of tracking on sensitive websites. Table 1 summarizes related works. Fouad et al. [42] were the first to differentiate between first to third party cookie syncing and third to third party cookie syncing; they made a categorization of 6 different tracking techniques. We adapted their classification of tracking to build our extension Ernie. This extension is designed to facilitate research studies of third party tracking. Differently from related works, we perform the first *qualitative study on health related websites*. Using Ernie we analyze complex tracking and cookie syncing techniques and we study the impact of user interaction with cookie banner in depth. As a result, we identified 5 different cases of privacy violations, and we performed a detailed legal analysis of each of these cases.

Analysis of sensitive websites. Previous works explored the tracking behaviors in sensitive websites. Vallina et al. [78] 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. [76] 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 [29] and Ghostery [45] filter lists. Sanchez et al. [74] performed a manual analysis of 2000 websites. They found that only 4% of websites offer an easy way to reject in the cookie notice. They also looked at websites by category and found that more than 50% of health websites do not have a banner while still performing tracking, and 40% even create more cookies upon rejection.

Paper	Analysis of sensitive websites	Analysis of consent banners	Detection of tracking techniques
Vallina et al. [78]	Lists and manual labelling of porn websites	✓	BT, Cookie syncing (FTCS & TTCS)
Matic et al. [76]	Content classifier	×	BT
Sanchez et al. [74]	Symantec RuleSpace DB	✓	BT
Matte et al. [63]	×	✓	Disconnect list
Papadogiannakis et al. [66]	×	✓	First party ID leaking (TA & FTCS), TTCS
Fouad et al. [42]	×	×	TA, TTCS, FTCS, BT, BTIT, TF
Our paper	User simulation for health websites	✓	TA, TTCS, FTCS, BT, BTIT, TF

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

Our work analyzes health related websites collected by simulating real users search behaviour. While previous works [74, 76] only investigated the presence of identifying third party cookies on health related websites, we detected complex cookie syncing techniques from [42].

Analysis of consent banners. Previous works studied the impact of the user's choices in the cookie banner 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 [29], and found that refusing cookies increased the number of third-party trackers. Recently, Papadogiannakis et al. [66] studied the effect of user interaction with the banner on first-party ID leaking (they did not differentiate third party analytics and first to third party cookie syncing and unites them into one category), and third-party ID synchronization (we call it third to third party cookie syncing). 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.

We made the first in depth analysis of different tracking behaviors deployed on health websites using the identifier cookies, moreover, we provided a comple-

mentary legal analysis, and described different alleged violations detected in these websites.

While previous works provided a quantitative study of the impact of interaction of cookie banners, in our paper, we combine that impact with detailed case studies and their legal implications.

Browser extensions. There are several popular browser extensions that use filter lists to block trackers and preserve user's privacy [28, 31, 44, 52]. Disconnect [28] additionally shows third party inclusion chains, while uBlock Origin [52] shows which part of a URL is responsible for tracking. The Lightbeam extension [64] 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 [26, 39, 68, 79] 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 detailed 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 [61], which visualizes which third party domains occur on which websites on a sequence of visits. It also shows if the domains dropped 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 tool that visualises several types of cookie synchronization techniques, and additionally shows which cookies and identifiers trigger tracking. It also shows the origin of Cookie Syncing requests and thus allows a detailed live overview of the tracking on a given website.

3 Methodology

3.1 Ernie Extension

The browser extension Ernie has been designed to detect the sophisticated cookie based tracking mechanisms described by Fouad et al. [42]. 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 send and responses received in that tab after the page visit and before the next one are considered part of the current page visit. As a result, Ernie provides a visualization that attributes to one of the six considered categories the HTTP(S) requests and responses, and the corresponding cookies.

3.1.1 Detection of ID cookies and ID sharing

Detection of ID cookies Ernie extension implements a standard approach to detect cookies that are likely to identify a user [10, 33, 34, 42] by comparing cookies between two different users. Ernie simulates a different users by opening a hidden tab in a separate container for each page visit, which is only used by the extension. To create the container, the extension uses the Firefox API `contextualIdentities` [4]. *Contextual identities* are containers within a browser profile which have a separate cookie storage, localStorage, indexedDB, HTTP data cache, and image cache. In the following, we refer to the hidden tab as *shadow tabs*. If the cookies with the same key and domain have different values for the two users, Ernie concludes that the cookie is "user-specific", we call in the following such cookies *ID cookies*. The extension displays and analyses all (first-party and third-party) ID cookies set in the browser (via HTTP(S) requests, HTTP(S) responses, or Javascript).

If the value of a cookie is the same between the main and shadow tabs, then the cookie is categorized as *Safe* and is simply saved in a local database of the extension.

Detection of ID Sharing To recognize if an ID cookie is shared via a URL parameter, the extension implements an ID sharing algorithm inspired by prior works [10, 33, 42]. All cookie values and URL parameters are split using as delimiters any character not in [a-zA-Z0-9-_.]. Differently from [42] and in order to reduce the chance of coincidental matches, after splitting, we don't consider values that are shorter than 4 characters or that are only the value *true* or *false*. Fouad et al. [42] 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.

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

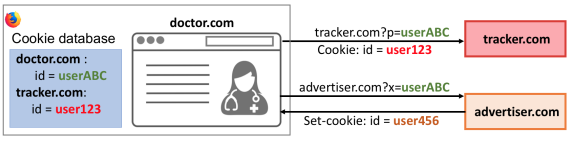


Fig. 1. Two examples of first to third-party cookie synchronisation: either the third-party cookie is already present in the browser and hence automatically sent to a third party (case of tracker.com) or is actively set by a third-party domain (case of advertiser.com).

3.1.2 Tracking detection

While detecting ID cookies and ID sharing, the ERNIE extension can identify six types of tracking behaviours presented by Fouad et al. [42]. In order to identify a tracking behavior, the extension first needs to discover the initiator of the request, that is, the resource which caused the request. ERNIE finds the initiator as follows.

1. If the request is caused by a 30x HTTP redirect, the initiator is the source of the redirection. ERNIE labels the previous 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.
3. For requests whose initiator cannot be found by either of the two previous steps, 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 behaviours 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 or 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.

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 for details), and (2) the third party domain sets or sends its own ID cookie (See Figure 1).

Algorithm 1: Detection of Basic tracking initiated by another tracker 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
```

To detect the first to third party cookie syncing, the extension performs algorithm 2.

Algorithm 2: Detection of First to third party cookie syncing

```
Let C be the set of ID cookies Detected in site;
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_i with i the length of the
             chain;
            while j <= i 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 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 all the initiators chain.
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 all the initiators 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 recognise users, the requests from the shadow tab will likely be recognized as being from the same user as the original requests.

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 Referrer header is often not set at all by the websites that serves the request. We therefore may miss some of the initiators and label them as first-party. As a result, our method may miss some of the tracking categories.

3.2 Experimental setup

Figure 2 presents an overview of our experimental setup. We first select health related websites (Section 3.2.1). Next, we setup the browser (Section 3.2) and collect data upon different interaction modes (Section 3.2.3).

English	French	German
gynaecologist	gynécologue	Frauenarzt
urologist	urologue	Urologe
infectologist	infectiologue	Infektiologe
oncologist	oncologue	Onkologe
cardiologist	cardiologue	Kardiologe
endocrinologist	endocrinologue	Endokrinologe
psychiatrist	psychiatre	Psychater
neurologist	neurologue	Neurologe
orthopaedist and traumatologist	chirurgien orthopédiste et traumatologue	Orthopäde und Traumatologe
pulmonologist	pneumologue	Pneumologe

Table 2. Doctors professions in English, French and German.

3.2.1 Websites Selection

Simulating user search for a doctor in a city. Recent work have shown that classifiers need to be used to detect whether a given website belongs to a sensitive category, such as health, automatically [76]. We instead have opted for a method that closely simulates a user that is interested to find information about a given medical profession in a given city. We decided to simulate typical users in two EU countries: France and Germany.

Notably, the Germany and French Data Protection Authorities are allocated with the highest tech specialists in Europe to face GDPR infringements [18]. Authors are fluent in both French and German, so they are able to analyse the type of visited website, find contact information and analyse the content of cookie banners.

Table 2 shows the list of 10 doctor professions that we have built from a list of long term illnesses that are fully covered by the French health insurance due to their severity [1]. We then simulate users in two major cities in France ("Paris", "Marseille") and Germany ("Berlin", München"). For each of the studied doctor professions, we pretend to be a user that makes a Google search of one doctor in one city. Specifically, we make the following search ⟨ city ⟩ ⟨ doctor ⟩ on `google.fr`, using a French VPN for French cities and doctors' professions in French, and on `google.de`, using a German VPN for German cities and doctors' professions in German.

We then automatically extract the URL links of the top 5 results of each search with Pupeteer version 5.4.1 [8] running on Chromium 87.0.4272.0. As a result, we have a list of 200 URLs. This process is represented in the top-left corner of Figure 2.

Further analysis of collected websites. By manually analysing content of each of the 200 websites, we

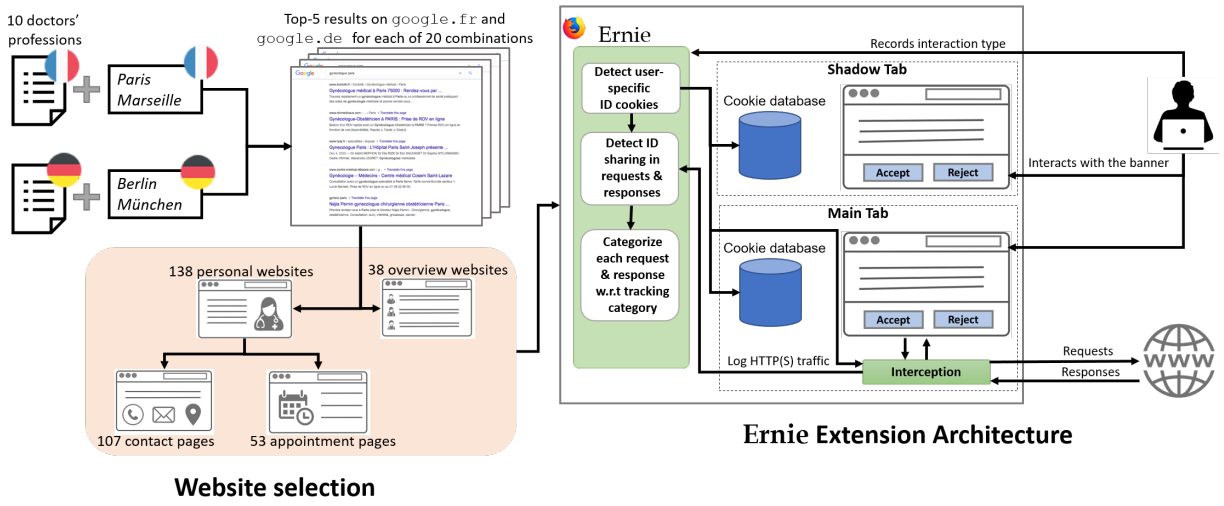


Fig. 2. High level overview of our experimental setup. Website selection process as well as browser setup and website analysis is described in the remainder of this section. ERNIE extension architecture is presented in details in Section 3.1.

categorise each site as either as doctor's `Personal` website, or an `Overview` website, where a user can search for doctors in an area and potentially book appointments. An example for an overview website is `doctolib.fr`. We found that many of `Personal` websites have

- *contact pages*, where potential patients can find phone number or other contact information. These pages are usually visible via "Contact"/"Where to find us" link of menu item.
- *appointment pages*, which include external content to book an appointment. These pages are found via searching for "Book an appointment" information on the website.

We have therefore added contact and appointment subpages to each visit to a `Personal` doctors websites. We imitate a user's behaviour and access these two subpages only by navigating within the visited `Personal` website.

During our manual analysis of 200 websites, we removed sites that are not related to our interest, such as news websites, PDF documents, job offerings and websites of doctors unions. After removal of these websites, we obtained 176 websites in our dataset. Table 3 presents the list of visited websites that are also shown in an orange box of Figure 2.

Personal	**138**
with Contact	**107**
with Appointment	**53**
Overview	**38**
Total websites	**176**

Table 3. Visited websites by type. We successfully visited at least one subpage of 176 websites, among them 138 are `Personal` and 38 are `Overview` websites. Out of the 138 `Personal` websites, 107 include a contact subpage and 53 include an appointment subpage. The full list of 176 analysed websites can be found in support materials [7].

3.2.2 Browser setup

Browser settings. We use Firefox version 78.4.1 on Debian, which has *Enhanced Tracking Protection* activated by default, meaning that Firefox already blocks some cross-site and social media trackers based on the *Disconnect* list [5]. Additionally the *Web Page Language Settings* are set to the languages that authors are fluent with: English [en], German [de], French [fr] and English (United States) [en-us] to be able to analyse the visited websites and their policy.

Simulation of a base browsing profile. Instead of visiting websites with a clean browser, we simulate real users by install generic browsing profile to insures that their profile already has common cookies set when visiting health related websites.

To build the base browsing profile, we first collect a list of popular websites globally, in France and Germany, by combining the top-30 global, the top-30 websites in France, and top-30 websites in Germany from the Alexa top list [2]. To build the user profile, we visited the 90 collected websites on the 13th of November 2020. The full list of unique websites visited to build the profile can be found at [3].

We then visit each health related website collected in Section 3.2.1 with the browsing profile in place, but the follow up visiting is *stateless*, that is we don't keep the state between two websites. We visited health related websites with the browsing profile between the 13th and 17th of November, 2020.

Reachable websites. If a website times out 3 times with the standard browser settings, it is defined as unreachable for both the browser profile collection as well as the visits of health related websites. This occurred only once for `microsoftonline.com`.

3.2.3 Data Collection

With the browsing profile in place, we visit each of the collected websites with version 2.1 of our extension and log the tracking behaviour that the extension finds. For all websites, we reload the page once after the initial visit. We do reloading because after interacting with a cookie banner, some websites include additional content only on the next page load.

Interactions with the cookie banners. Previous works explored the interaction with the cookie banners [27, 74]. However, automated interaction with banners remains challenging: Matic *et al.* [76, Sec. 3.1] report that only 4.4% of websites contain a cookie banner we can automatically interact with via advanced tools like Consent-O-Matic [25, 65].

Given the relatively small number of websites included in our study, we decided to manually label the type of banners and interactions. The EU legislation requires consent before setting or sending tracking cookies. We therefore evaluate the types of banners and changes in the tracking behaviour based on the choice made by the user in the cookie banner. We interact with the banners in three ways, and also record 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` websites. This is not possible on every website, as cookie banners sometimes block the access to a website until the user has made a choice in the cookie banner.

`Accept All` We accept all cookie preferences the cookie banner suggests to us. Most of the time, that means clicking the "Accept All" button. This is only possible on websites that have a cookie banner.

`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 cookie banners, because many banners only describe their use of cookies and other tracking technologies, but do not offer a possibility to reject them.

For each type of interaction, we visit as many page types as possible. This means that we have at least two page visits (initial visit and reload) and at most 12 page visits per website (three interaction types on a maximum of four page types).

Data collection from manual analysis. The Ernie extension saves all collected data to a local database on the same device with which we visit the health related websites. 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 collect the following data upon each manual visit to a health related website:

- the URL and the country of the website (France or Germany depending on which search has lead to the website - see Section 3.2.1);
- the site type (`Personal`, `Overview` - see Table 3),
- whether the website contains a banner, and the type of consent banner the website employs (according to the classification of banners by Degeling et al. [27]),
- URLs of contact and appointment subpages for `Personal` websites.

3.2.4 Limitations of the experimental setup

The methods we used to select websites and interact with them have some limitations. First, our site selection may be biased because we rely on search results from `google.de` and `google.fr`. Secondly, to imitate French and German users, we used the VPN of a German and a French institution. These IP addresses might be recognized as not belonging to a private household, which might introduce bias in the content being served, as shown in [81].

In our experiments we used a Firefox browser with Enhanced Tracking Protection on, however users of other browsers without any tracking protection, such as Google Chrome, could experience much more tracking that Ernie extension is also able to detect.

4 Results

In this section we present the main findings regarding consent collection and potential illegal tracking occur-

ring on health websites where we observed, at least, one type of tracking (see Section 3.1 for the full set of tracking categories ERNIE detects). We say that *a website includes tracking* if we detect, at least, one type of tracking behavior on, at least, one page of a website. We refer to domains that participate in tracking as *tracking domains*.

Distinctly, we found that before any interaction with the website, tracking occurs on 65% of the 176 visited websites. Notice that we include Third party analytics category in these findings because it requires consent according to several Data Protection Authorities [19, 50, 55] and to the European Data Protection Board (EDPB) [35].

We present each finding firstly with a technical description, followed by a legal analysis and alleged violations triggered by tracking practices, alongside with a case study demonstrating such violations. The legal analysis is performed together with a legal expert co-author of this paper.

Legal requirements for online tracking. To comply with the GDPR and the ePrivacy Directive (ePD), websites must obtain *consent* from users located in the EU when monitoring users' behavior (Article 5(3) ePD). A common method to obtain consent is through the use of ubiquitous consent banners. For consent to be legally valid, it must be prior to any data collection, freely given, specific, informed, unambiguous, readable and accessible and finally, should be revocable (Articles 4(11) and 7 GDPR) [75].

Though consent is generally needed for tracking, some types of trackers are exempted of consent, and the only way to assess with certainty whether consent is required, is to analyse the *purpose* of each tracking technology on a given website [14]. To determine a purpose of each tracking cookie in our case study, we analyse privacy policies of third parties that set such cookie.

Data concerning health status. Health status of users are particularly sensitive by their nature, and under the GDPR [43, Article 9], merits specific protection, as their processing could create *significant risks to the fundamental rights and freedoms* of users (Recital 51 GDPR). *Data concerning health* means personal data related to the physical or mental health of a person, including the provision of health care services, which reveal information about her health status (Article 4(15), Recital 35 GDPR). When a user visits a health related website, this mere visit surely reveals information about the health condition of this visitor. It might be argued that this information is not 100% certain. However, when health websites integrate third-party trackers, they expose their potential patients' health condition to third parties. Considering the large number of websites and the large number of users a single third party can follow, the collected information will undoubtedly be very informative on the health condition of a very large number of users.

Legal requirements for online tracking on health websites. The processing of data concerning health is *forbidden* by the GDPR, unless allowed by several exceptions (Article 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 [43, Article 9(2)(a)]. An *explicit consent* request should abide to the following requirements [12, 30, 55]: i) include double confirmation or verification from the user, ii) consist of a separated request from any other consents [37] (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* [38].

Without explicit consent from users, tracking on health websites infringes the lawfulness principle (Article 9 (2)(a) GDPR), rendering any forthcoming processing *unlawful*, and consequently such websites will 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).

Methodology used for the legal analysis and case study. In the legal analysis of the following subsections we take a double approach. First, we analyse *straightforward violations* independently of whether tracking requires or is exempted of consent. Then, we additionally analyse further violations related to the presence of trackers that definitely require (not exempted of) consent. Pursuant to this, we analyse the *purpose* of each cookie to determine whether consent is needed. We name this later analysis as *violations depending on the purpose of the cookie*. We then report in a case study only cookies responsible for tracking or syncing that definitely require consent.

4.1 No consent banner and tracking

Technical description and prevalence. By manually analyzing the studied health related websites, we found that out of the 176 visited websites across all the website categories (both `Personal` and `Overview`- see Section 3.2.1), 81 (46.02%) do not have any cookie ban-

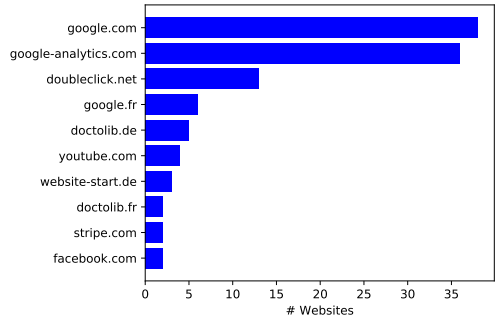


Fig. 3. Top 10 trackers on websites that do not include a consent banner. *In total, we detected 81 websites that do not display a cookie banner and include tracking.*

ner, and 61 (75.31%) thereof include at least one of the studied tracking categories before interaction. Figure 3 presents the top 10 domains performing at least one of the studied tracking behaviors (see Section 3.1.2) on the 61 websites where no banner is displayed.

Legal analysis. *Straightforward violation*: The absence of *any* method set forth to collect the user's explicit consent renders any forthcoming tracking unlawful due to the lack of legal basis (Article 9(2)(a) GDPR), hence, allegedly violating the lawfulness principle. *Violations depending on the purpose of the cookie*: if the purpose of all the cookies used in a website does not require consent, the absence of a banner would not entail any legal violation. However, if the purpose of at least one cookie requires consent, then the violations would consist of: i) lawfulness principle, due to lack of any method to collect the user's consent; ii) prior consent, as tracking becomes unlawful if carried out before consent is requested (Article 6(1)(a) GDPR) [36].

Case study. We analyzed in depth the health related website `logicrdv.fr` [62]. `Logicrdv.fr` is an intermediate french website between doctors and patients that is specialized in the management of appointments. Through this website, the user can search for doctors of a given profession in a specific region, and set an appointment. The particular page we have visited provides a list of cardiology doctors near Marseille. When we first visited the website we found that no banner was included, and there were no means for the user to express her privacy preferences regarding tracking. The website moreover did not have any privacy policy.

With Ernie extension we detected tracking from 3 different third party domains: `stripe.com`, `google.com` and `google-analytics.com`. On a further analysis, we found that `google.com` is responsible for 89 tracking requests, while `stripe.com` and `google-analytics.com` exhibit only 2 tracking requests each on this website. Moreover, all tracking by `google.com` is Basic tracking (see Section 3.1.2) caused by its own cookie named `NID`.

We found that the `NID` cookie is never set by `google.com` on the visited website, but it was always sent as part of the request. In fact, the `NID` tracking cookie was first set on the user's browser when we built the user profile and visited `google.com` website (see Section 3.2). Once stored in the user's browser, the cookie was automatically sent with every request to `google.com`'s sub-domains as part of the management mechanism of the HTTP cookie standard [53]. As a result, when we visited `logicrdv.fr` – which includes Google maps to indicate the doctors location–, the browser automatically sent a request to `google.com` to fetch the content and automatically attached the `NID` cookie with every request. All tracking request sent to `google.com` from `logicrdv.fr` were used to fetch the google map. We never consented on the use of cookies neither on our visit to `logicrdv.fr`, nor on `google.com`. Google privacy policy states that "The NID cookie contains a unique ID we use to remember your preferences and other information, such as your preferred language, how many search results you prefer to have shown on a results page [..]", and at the very same time claims that "'NID' is used for these [advertising] purposes to show Google ads in Google services for signed-out users" [46]. Therefore, according to the purpose of this cookie, it requires consent since it is used, among other purposes, for advertising. As stipulated by regulatory guidance, such purpose is subject to the legal basis of consent [14, 19, 30, 55].

Findings. In our dataset, we detected 45 contact pages that include Google maps, and in all these websites tracking occurs because of the management mechanism of the HTTP cookie standard [53]. When the user first visits `google.com`, the `NID` is automatically set by `google.com`. Upon visits to websites containing Google maps, `NID` cookie is automatically attached with every request to a sub-domain of `google.com` to fetch the Google map. The impact of this practice is particularly severe for users' privacy because `google.com` is the default page visited upon installation of all major browsers: Google Chrome browser (used by 2.65 billion users in 2020 [20]), Safari browser (446 million users [73]), and Firefox browser (250 million users [41]).

Banner	Accept	Reject	# of websites
No Option			6
Confirmation	✓		40
Binary	✓	✓	26
Slider	✓	(✓)	0
Checkbox	✓	(✓)	14
Vendor	✓	(✓)	7
Other			2
Total			95

Table 4. Overview of banner types, and if they allow rejecting and accepting. (✓) means that it is allowed for some categories in that banner, but not for others, e.g., one can reject cookies for some vendors in a "Vendor" banner, but not for all vendors.

Fig. 4. Interface of the "Ramsay Santé Hôpital privé Résidence du Parc" private hospital website. Captured on 18th February 2020 from https://hopital-prive-residence-du-parc-marseille.ramsaygds.fr/vous-etes-patient-pourquoi-choisir-notre-etablissement/urologie-22.

4.2 No possibility to refuse in a consent banner and tracking

We found that 95 (53.98%), out of the 176 studied health related websites, include a consent banner. However, some of these banners are not designed to provide an unambiguous and freely given choice to the user, rendering unlawful such consent collection [36, 49]. Using the categorization of consent banner design proposed by Degeling et al. [27], we grouped the cookie banners detected in the visited health websites into 6 categories, which we depict in Table 4.

We further analyzed the 95 websites that include a cookie banner, and we found that on 49 (52%) websites this banner implements a reject button ("Binary", "Checkbox", "Vendor" and "Other" banner types). We found that rejection is actually possible on 44, out of 49 websites, and after rejection of cookies, 33 websites still include trackers. On 20 (60%) out of these 33 websites, the number of tracking domains before and after rejection of cookies remains the same. Therefore, even the presence of reject options is often ineffective.

Technical description and prevalence. Out of the 95 websites that include a cookie banner, 46 (48%) thereof display a cookie banner that either (i) is only informative ("No Option") and the user doesn't have any option; or (ii) only includes a confirmation button ("Confirmation") without giving the user any possibility to reject. Using Ernie, we detected at least one tracking behavior category on 40 (86.96%) out of 46 websites that display cookie banner without a possibility to reject.

Legal analysis. *Straightforward violation*: Considering the legal requirements for explicit consent, one should contend that the lack of any possibility to confirm a user rejection – as to make evident the user's choice regarding the processing of her sensitive data – would render such consent request unlawful (Article 9 (2)(a) GDPR). *Violations depending on the purpose of the cookie*: if the purpose of cookies does not require consent, the absence of a rejection button in a cookie banner does not seem to entail any legal violation. However, if the purpose of a cookie requires consent, then such practice allegedly is conflicting with the following consent requirements and data protection principles: i) requirements of "configurable banner" and "balanced choice" (Articles 4 (11), 7(3) GDPR) [9], which are compulsory for an unambiguous consent of a user; and ii) the principle of "data protection by design and by default" which demands the most privacy-friendly default settings to be used (Article 25 GDPR).

Case study. `Ramsaygds.fr` is a website of a private hospital in Marseille, France. The particular page we have visited [69] explains to patients why they should choose this hospital when they have health problems related to urology. When we first visited the website, we noticed that it presents a cookie banner to the user. However, the banner contains only one button "I understood", and does not include any reject button (see figure 4). Before interacting with the banner, Ernie detected tracking behaviors from 4 distinct domains. We found that `ramsaygds.fr` includes analytics performed by `google-analytics.com` and `doubleclick.net` and cross-site tracking behaviors by `google.com` and `google.fr`. We further analyzed the tracking behaviors on `ramsaygds.fr` after clicking on "I understood" button, and found that the tracking domains before interaction and after acceptance are identical. The website includes a privacy policy [70], however, they only state the use of Google analytics cook-

ies for analytics purposes and do not mention the usage of other tracking forms detected on the website. In their policy they state that the user can manage and reject cookies in her browser, and block them using their browser storage according to the advice by the French Data Protection Authority (CNIL) on how to manage cookies [21]. The provided CNIL website is in fact a recommendation to users on how to protect their privacy in the web, and can not in any case replace the implementation of a reject button in the website cookie banner.

Findings: We found that cookie banners that do not provide a possibility to reject are only informative and do not affect the number of trackers. We compared the number of tracking domains before interaction and after accepting cookie on the 42 (23.86%) websites where there is no reject option and we successfully accepted cookies. We found that on 40 (95.24%) out of the 42 websites, the number of trackers remained the same before and after clicking the accept button. Moreover, 33 out of 44 websites that propose reject option still include trackers after rejection. Hence, cookie banners are not effective on these websites.

4.3 Cookie Syncing before interaction or after rejection

To create a more complete profile of the user, domains need to merge user's data they have collected on different websites. One of the most known techniques to do so is cookie syncing. In this section, we study all cookie syncing tracking categories (*First to third party cookie syncing, Third to third party cookie syncing, and Third party cookie forwarding*) performed on websites before any interaction with the banner or after rejection is selected on the banner.

Technical description and prevalence. Using Ernie, we detected cookie synchronization on 17 websites before interaction. This cookie synchronization is performed by 8 distinct third-party domains. Before interacting with the banner, we didn't detect any instance of Third to third party cookie syncing nor Third party cookie forwarding. The only synchronization activity we detected before interaction is First to third party cookie syncing, where `google.com` is the top domain that performs such syncing on 11 websites.

After rejection, to our surprise, we detected cookie synchronization on 8 websites performed by 3 distinct third party domains. We found that `google.com` is simultaneously performing First to third party cookie syncing and Third to third party cookie syncing on 4 and 1 websites respectively.

Senders	Receivers
Before interaction	
ramsaygds.fr	google.com
psychologies.com	facebook.com
rdvmedicaux.com	facebook.com
jameda.de	ioam.de
pagesjaunes.fr	facebook.com
After rejection	
jameda.de	ioam.de
pagesjaunes.fr	facebook.com
institutpaolicalmettes.fr	facebook.com
118000.fr	facebook.com
atos-kliniken.com	google.com

Table 5. Cookie syncing. Top 5 senders and receivers of cookie synchronization before interaction and after rejection. *All presented domains perform First to third party cookie syncing.*

Legal analysis. *Straightforward violations*: Cookie syncing potentially breaches the following principles: *Lawfulness principle*: the absence of the user's explicit consent for cookie syncing, before interaction and after rejection, breaches this principle (pursuant to Article 9 (2)(a) GDPR). *Fairness principle*: cookie syncing disregards the legitimate expectations of the data subject at the very time of data collection. Any (extensive) disclosure to third parties of sensitive data is out of any user reasonably expectations (Article 5(1)(a) GDPR). *Transparency principle*: in both scenarios users should be informed of the existence of cookie syncing operations and its purposes (Recital 60 GDPR), and should be made aware their personal data are shared with other third-parties. Moreover, users should be informed of the extent, risks and consequences of cookie syncing (Recital 39). In particular, considering the extent of data being sharing with third-parties, users should be informed of the existence of *profiling* and the rights and safeguards they are afforded with (Articles 13(2)(f), 22(1)(4) GDPR). The violation of these transparency obligations breaches the transparency principle and renders processing unlawful. *Minimization principle*: this practice contradicts expressly the minimization principle which requires personal data to be collected and processed limited to what is necessary, proportional and relevant to fulfil the data controller purpose (Article 5(1)(c) GDPR). *Violations depending on the purpose of the cookie*: if the purpose of cookies would not require consent, then no further breaches are accounted. However, if the purpose of a cookie requires consent, then such practice allegedly violates the following consent re-

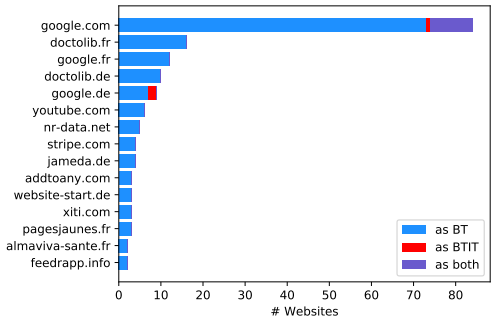
quirements: *Prior consent*: cookie syncing becomes unlawful if carried out before the request for consent due to the lack of a legal ground (Articles 4 (11), 6(1)(a) GDPR). *Informed consent*: users should be informed about third parties with whom the cookies are shared with – an obligation prescribed in the Court of Justice of the EU case law [9] and in the GDPR (Articles 4 (11), 13 (1)(e) GDPR). Users should also be informed about the purposes for which sensitive data will be collected for (Article 13 (1)(c) GDPR.

Case study. `Lefigaro.fr` is a phone book website that allows users to search for a doctor and make an online appointment by providing the user's phone number and address. The specific page that we visited [58] list Endocrinology doctors in Marseille. We noticed that no banner was displayed when we directly visited the subpage, however, the website does include a cookie banner in its home page. Due to this behavior users directly accessing the subpage through a Google search can not provide their consent. A cookie banner should be available through all website pages. Using Ernie, we detected first to third party cookie syncing between `lefigaro.fr` and two third parties before interaction: `google.com` and `acpm.fr`. We detected that `lefigaro.fr` shares the first party cookie that has as key `measure` with `acpm.fr` as part of the URL path. `Acpm.fr` then sets it's own cookie on the user's browser. `Acpm.fr` is a third party domain that provides to media websites a certification of the distribution, attendance, measuring of the audience by making it more visible to media agencies and advertisers [11]. `Lefigaro.fr` declares collaboration with `acpm.fr` in their policy [59] and they state that they are using `acpm.fr` cookies to measure audience in the website, but they do not provide information regarding cookies sharing.

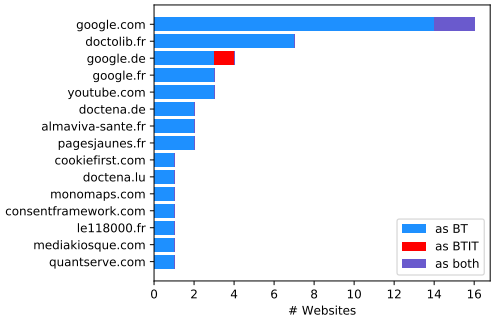
Findings. First to third party cookie syncing is a common practice we detected before interaction on 17 (9.96%) websites with the Firefox ETP [40] protection activated. This practice was shown before by Fouad et al. [42], but it didn't receive much attention. In this paper, we show that it still happens. We contacted the Firefox team and shared results for them to improve their tracking protection.

4.4 Explicit Tracking before interaction or after rejection

In this section, we analyse two categories of tracking together - Basic tracking and Basic tracking initiated by another tracker (see Section 3.1.2) – that we call *Explicit Tracking* in this section.



(a) No interaction



(b) After rejection

Fig. 5. Explicit tracking. Receivers of explicit tracking before interaction and after rejection. *BT: Basic tracking, BTIT: Basic tracking initiated by another tracker*

Technical description and prevalence Using Ernie, we studied explicit tracking on the 176 websites, where at least one subpage is successfully visited. Before interacting with the banner, we found that explicit tracking occurs on 116 (66%) of the visited health related websites by 43 distinct domains. Figure 5 shows that `google.com` is the top domain performing explicit tracking, it is responsible of explicit tracking on 84 (47.73%) of the visited websites before any interaction. Moreover, after rejection, 29 (66%) out of 44 websites that provide possibility to reject, are explicitly tracking the user. Such tracking is performed by 24 distinct domains.

Legal analysis. *Straightforward violations*: We observe that explicit tracking before interaction and after rejection on health websites violates the following principles. *Lawfulness principle*: the absence of explicit consent for this tracking category in both scenarios, breaches the lawfulness principle (pursuant to Article 9 (2)(a) GDPR). *Fairness principle*: after rejecting tracking,

users do not expect still to be tracked. Accordingly, such practice seems to infringe the fairness principle (Article 5(1)(a)). *Violation depending on the purpose of the cookie*: if the purpose of cookies does not require consent, then no further breaches are accounted. However, if the purpose of a cookie requires consent, then such practice allegedly is in breach of the *prior* consent requirement (Articles 4 (11), 6(1)(a) GDPR).

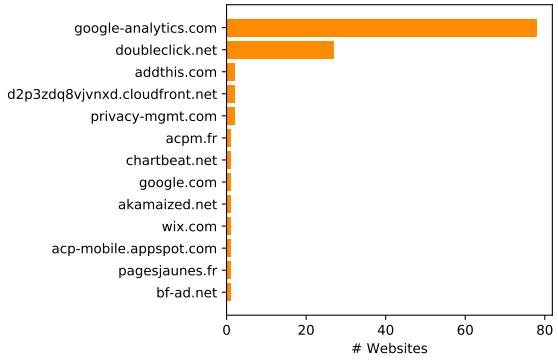
Case study. `Ameli.fr` is a major health website in France: it allows any French resident to access different health insurance services such as consulting reimbursements, downloading certificates, obtaining European card, etc. We analyzed a specific subpage of `ameli.fr` [13] that helps users search for doctors and medical institution using the doctor or institution name, profession or the required service. This `ameli.fr` website displays a banner, but no choice can be made by the user. The banner is only used to inform the user that if she continues browsing the website than she accepts the usage of cookies. Using Ernie, we detected Basic tracking before interacting with the website from the third party domain `xiti.com`. `Xiti.com` define themselves as a web traffic measurement website [80]. We detected that `xiti.com` is performing Basic tracking on the `Ameli.fr` website using the following cookie: `idrxvr`, `atidx`, and `atid`. These cookies are classified as analytics cookies used to provide measurement on the website [60]. However, differently from standard first-party cookies used for analytics, these analysed cookies are third-party cookies, and therefore differently from analytics services, they can be used for cross-site tracking.

Finding. We found that all explicit tracking performed by `google.com` before interaction on the 84 (47.73%) visited websites were a result of the management mechanism of the HTTP cookie standard [53]. In fact, when we first visited `google.com` website upon profile creation (see Section 3.2), `google.com` set an ID cookie `NID` in the user browser. This cookie was then sent with every request to `google.com` in the 84 websites before interaction, thus following the same mechanism as described in Findings of Section 4.3.

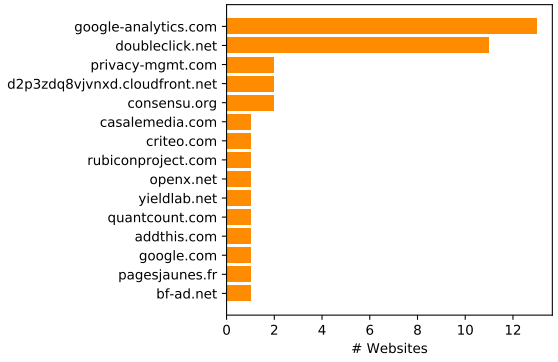
4.5 Third-party Analytics before interaction or after rejection

As of today, website developers tend to use third party analytics services to measure audience in their websites. These analytics services provide report on the website traffic by measuring the number of repeated visits, the most popular pages, etc. Such practice allows tracking only within the same website. According to the ePrivacy Directive (Article 5(3)) websites owners are bound to request user consent before performing such tracking practices on their websites.

Technical description and prevalence. We analyzed the prevalence of the third-party analytics behavior in health related websites before interaction and after rejection of cookies. We found that analytics behavior is simultaneously performed on 81 websites before any interaction and 16 websites after rejection. `google-analytics.com` is the top domain responsible of third-party analytics on health related websites without user's consent (see figure 6). It is tracking users on 77 websites before interaction and on 13 after rejection. It is followed by `doubleclick.net` that performs analytic behavior on 26 websites before interaction and on 11 after rejection.



(a) No interaction



(b) After rejection

Fig. 6. Third-party analytics. Receivers of analytics tracking before interaction and after rejection.

Legal analysis. *Straightforward violations*: We observe that third-party analytics before interaction or after re-

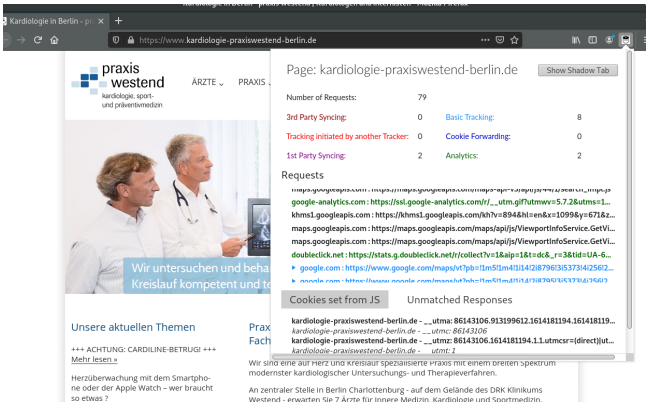


Fig. 7. Detection of third party analytics behavior on kardiologie-praxiswestend-berlin.de website using ERNIE extension.

jection on health websites violates the lawfulness principle due to lack of explicit consent. *Violation depending on the purpose of the cookie*: Is it already determined that using third-party analytics requires consent from users. This stance is upheld by several Data Protection Authorities [19, 50, 55] that assert these technologies are not considered *strictly necessary* for a website to provide a functionality explicitly requested by the user, because the user can access all the functionalities provided by the website when such cookies are rejected. The French DPA [24] adds further that consent is 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.

Case study. `kardiologie-praxiswestend-berlin.de` is a joint medical office of several cardiologists. The website does not have a cookie banner. In their privacy policy they explain that their website uses `google-analytics.com` and `googleadservices.com`, and the data collected by `google-analytics.com` will not be linked to other data from Google.

Before interaction and after rejection, we detected analytics behavior on the studied website using the ERNIE extension (see Figure 7). We found that `google-analytics.com` first receives the *___utma* first party cookie as part of the request, then `google-analytics.com` makes a redirection to `doubleclick.net` and shares the first party cookie *___utma* with it. According to google's policy [48], the *___utma* cookie is used to distinguish users. The two requests sent to `google-analytics.com` and `doubleclick.net` are categorized as analytics. `Doubleclick.net` then redirects to `google.com`, which again redirect to `google.de`. The first party cookie is shared with `google.com` and `google.de`, moreover, the browser automatically attaches the `NID` cookie set in the browser in our base profile. These two requests are therefore first to third party cookie syncing-requests, effectively allowing the linking of the *___utma* first party cookie to the `NID` cookie.

Findings. Due to the redirection inclusion process, third party domains track users on websites where they were not initially included. Moreover, using this redirection, trackers share first party identifiers and link them with third party IDs. We found that on 25 websites out of the 26 websites where `doubleclick.net` is performing analytics, `google-analytics.com` is included as well, and both `google-analytics.com` and `doubleclick.net` receive the same first party identifier. We detected that all first party cookies `_ga`, `_gid` and `__utma` shared with `doubleclick.net` on these 25 websites belong to `google-analytics.com` [47]. Therefore, we suspect that `google-analytics.com` is responsible of including and sharing the first party ID with `doubleclick.net`.

5 Conclusion

In this paper we have gleaned robust evidence of tracking technologies deployed on health-related websites (before user consent interaction, and also after accepting and rejecting). 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 ERNIE extension can be beneficial to both policy-makers, to advance the enforcement of EU Privacy and Data Protection law, and to owners of health websites, such as doctors and hospitals 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 changing their practices towards improving the afforded protection of privacy and health data of Web users.

References

[1] Affection longue duree. https://www.ameli.fr/assure/droits-demarches/maladie-accident-hospitalisation/affection-

longue-duree-ald/affection-longue-duree-ald.
[2] Alexa top sites. https://www.alexa.com/topsites.
[3] Alexa websites visited. https://www.dropbox.com/sh/nwjw7ggcx08o1x7/AACYrHqsxo7DcZjbVArE5Fxua?dl=0.
[4] Contextual identities. https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/API/contextualIdentities.
[5] Enahnced 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/sh/96pcfgj1qruow90/AABzQmH3CCLCMDYOBHaKxeG9a?dl=0.
[8] Pupeteer. https://github.com/puppeteer/puppeteer.
[9] 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.
[10] 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.
[11] Acpm.fr website. https://www.acpm.fr/Adherer/Pourquoi-adherer-a-l-ACPM.
[12] Guide on use of cookies, 2021. https://www.aepd.es/sites/default/files/2021-01/guia-cookies-en.pdf.
[13] Ameli.fr website. http://annuairesante.ameli.fr/professionnels-de-sante/recherche/fiche-detaillee-AbE1mjY2MDCw.html.
[14] Article 29 Working Party. Opinion 04/2012 on Cookie Consent Exemption (WP 194).
[15] Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William K. Robertson, and Christo Wilson. How tracking companies circumvented ad blockers using websockets. In *Internet Measurement Conference 2018*, pages 471–477, 2018.
[16] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *Proceedings of the 25th USENIX Security Symposium*, Austin, TX, August 2016.
[17] Muhammad Ahmad Bashir and Christo Wilson. Diffusion of User Tracking Data in the Online Advertising Ecosystem. In *Proceedings on Privacy Enhancing Technologies (PETS 2018)*, 2018.
[18] Brave. Europe's governments are failing the gdpr – brave's 2020 report on the enforcement capacity of data protection authorities. https://brave.com/wp-content/uploads/2020/04/Brave-2020-DPA-Report.pdf, 2020.
[19] Bundesbeauftragter für den Datenschutz und die Informationsfreiheit (BfDI, German DPA). Guidance from german authorities for telemedia providers (translation). https://deutschland.taylorwessing.com/de/documents/get/1820/guidance-from-german-authorities-for-telemedia-providers-partial-translation.PDF_show_on_screen, accessed on 2020.01.21.
[20] Number of chrome users. https://www.statista.com/statistics/543218/worldwide-internet-users-by-browser/.
[21] ookies : les outils pour les maîtriser. https://www.cnil.fr/fr/cookies-les-outils-pour-les-maitriser.
[22] Cnil: Violation de données de santé. https://www.cnil.fr/fr/violation-de-donnees-de-sante-la-cnil-rappelle-les-obligations-des-organismes-la-suite-dune-fuite-de.
[23] 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.
[24] 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.
[25] Consent-O-Matic browser extension. matic/mdjildafknihdffpkfmmpnpoiajfjnjd.
[26] Cookiebot. Cookie scanner for gdpr/epr and ccpa compliance. https://www.cookiebot.com/en/cookie-scanner/.
[27] 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.
[28] Disconnect. Disconnect. https://disconnect.me/.
[29] Disconnect Official website. https://disconnect.me/.
[30] 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.
[31] EFF. Privacy badger. https://privacybadger.org/.
[32] Steven Englehardt, Jeffrey Han, and Arvind Narayanan. I never signed up for this! privacy implications of email tracking. In *Privacy Enhancing Technologies*, 2018.
[33] 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.
[34] 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.
[35] European Data Protection Board. Working document 02/2013 providing guidance on obtaining consent for cookies, adopted on 2 october 2013. https://www.pdpjournals.com/docs/88135.pdf.
[36] 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.
[37] European Data Protection Board (EDPB). Opinion 2/2010 on online behavioural advertising, 22 june 2010, WP171, p. 10.
[38] European Data Protection Board (EDPB). Guidelines 05/2020 on consent under regulation 2016/679, 2020.

[39] EZIGDPR. Gdpr website compliance check. https://www.ezigdpr.com/products/gdpr-website-compliance-checker.
[40] Firefox Now Available with Enhanced Tracking Protection by Default Plus Updates to Facebook Container, Firefox Monitor and Lockwise . https://blog.mozilla.org/blog/2019/06/04/firefox-now-available-with-enhanced-tracking-protection-by-default/.
[41] Number of Firefox users. https://consent.yahoo.com/v2/collectConsent?sessionId=3_cc-session_85b2e8b2-05db-4bd0-8ace-208266886510.
[42] 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.
[43] 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.
[44] Ghostery. Ghostery. https://www.ghostery.com/.
[45] Ghostery Official website. https://www.ghostery.com/.
[46] Google.com privacy policy. https://policies.google.com/technologies/cookies?hl=en-US.
[47] Google analytics solutions. https://www.google.com/analytics.
[48] Google.com cookie usage. https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage.
[49] 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.
[50] 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.
[51] 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.
[52] Raymond Hill and Contributors. ublock origin. https://github.com/gorhill/uBlock/.
[53] HTTP cookie standard. https://tools.ietf.org/html/rfc6265.
[54] Muhammad Ikram, Hassan Jameel Asghar, Mohamed Ali Kaafar, Anirban Mahanti, and Balachandar Krishnamurthy. Towards seamless tracking-free web: Improved detection of trackers via one-class learning. In *Privacy Enhancing Technologies*, 2017.
[55] Information Commissioner's Office. 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.
[56] Costas Iordanou, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris. Tracing cross border web tracking. In *ACM Internet Measurement Conference (IMC)*, 2018.
[57] Tobias Lauinger, Abdelberi Chaabane, Sajjad Arshad, William Robertson, Christo Wilson, and Engin Kirda. Thou shalt not depend on me: Analysing the use of outdated javascript libraries on the web. In *Network and Distributed System Security Symposium, NDSS*, 2017.
[58] lefigaro.fr website. https://annuaire.lefigaro.fr/annuaire/ville/marseille-1er-arrondissement-13/endocrinologue.
[59] lefigaro.fr privacy policy. http://mentions-legales.lefigaro.fr/page/infos-cookies.
[60] Letour.fr privacy policy. https://www.letour.fr/en/privacy-policy.
[61] LINC. Cookieviz 2: new features to observe hidden web practices. https://linc.cnil.fr/fr/cookieviz-2-new-features-observe-hidden-web-practices.
[62] Logicrdv.fr visited website. https://www.logicrdv.fr/cardiologue/13006-marseille-6eme.html.
[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] Mozilla. Lightbeam 3.0. https://addons.mozilla.org/en-GB/firefox/addon/lightbeam-3-0/.
[65] 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.
[66] 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.
[67] 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.
[68] Piwik. Free online cookie scanner. https://piwik.pro/cookie-scanner/.
[69] Ramsaygds.fr website. https://hopital-prive-residence-du-parc-marseille.ramsaygds.fr/vous-etes-patient-pourquoi-choisir-notre-etablissement/urologie-22.
[70] Ramsaygds.fr privacy policy. https://ramsaygds.fr/mentions-l%C3%A9gales.
[71] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, and Phillipa Gill. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In *Network and Distributed System Security Symposium, NDSS*, 2018.
[72] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2012*, pages 155–168, 2012.
[73] Number of Safari users. https://www.statista.com/statistics/543218/worldwide-internet-users-by-browser/.
[74] 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.

[75] 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. *Technology and Regulation*, pages 91–135, 2020.

[76] Matic Srdjan, Iordanou Costas, Smaragdakis Georgios, and Nikolaos Laoutaris. Identifying sensitive urls at web-scale. In *ACM Internet Measurement Conference (ACM IMC 2020)*, 2020.

[77] European Data Protection Supervisor. Edps inspection software. https://edps.europa.eu/press-publications/edps-inspection-software_en.

[78] 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.

[79] Webcookies. Web cookies scanner. https://webcookies.org/.

[80] Xiti.com website. https://www.xiti.com/en/.

[81] David Zeber, Sarah Bird, Camila Oliveira, Walter Rudametkin, Ilana Segall, Fredrik Wollsén, and Martin Lopatka. The representativeness of automated web crawls as a surrogate for human browsing. In *Proceedings of The Web Conference 2020*, pages 167–178, 2020.