

# CSI-COP

**Citizen Scientists Investigating Cookies and App GDPR compliance**

---

## **Deliverable D5.1 | D18**

Taxonomy of Digital Cookies and Online Trackers

**Delivery date: 29 March 2023**

**LEAD Partner for deliverable: University of Patras (UPAT)**

---

**Deliverable Authors: Shah, H., Gialelis, Y., Konstantinopoulos, I., Lantavou, K., Karakonstanti, A., Rigler, D., Stepankova, O., Ignat, A. and Celentano, U.**

**Version: 7.2.1**

**Project funded from the European Union's Horizon 2020 SwafS research and innovation programme under grant agreement N° 873169  
44 months January 2020-August 2023 (with 2<sup>nd</sup> Amendment)**

### Dissemination Level

	Public	
PU	Public	X
R	Report, <b>DEM:</b> Demonstrator, pilot, prototype, plan designs, <b>DEC:</b> Websites, patents filing, press & media actions, videos, etc., <b>OTHER:</b> Other (Database, online tools, questionnaires, etc)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	
CI	Classified, information as referred to in Commission Decision 2001/844/EC.	



## Version control

Version	Date	Authors	Institution	Change and where applicable reason for change
1	02.11.22	Shah, H.; Gialelis, Y.	CU, UPAT	Outline
2	05.01.23	Gialelis, Y.	UPAT	Executive Summary, Introduction
3	10.01.23	Gialelis, Y.; Karakonstanti, A.; Konstantinopoulos, I.;	UPAT	Background Information, Web and app tracking technologies
4	20.01.23	Konstantinopoulos, I.; Karakonstanti, A.; Lantavou, N.	UPAT	Web and app tracking technologies, CSI-COP investigation tool, Privacy by design
5	03.02.23	UPAT from CSI-COP partners translations	UPAT	Adding translated column headings from CSI-COP Citizen Scientist web and app investigation tool
6	17.02.23	Gialelis, Y.; Konstantinopoulos, I.; Karakonstanti, A.; Lantavou, N.	UPAT	CSI – COP initial taxonomies and some statistical analysis of the investigation findings
7	21.02.23-19.03.23	Shah, H.	CU	Rewriting content to scope into D5.1 deliverable requirement. {Removed investigation tool column heading translations: required for T5.2 repository, incomplete statistical analysis extracted for journal paper preparation}.
7.1	21.03.23	Shah, H.	CU	Added review by Austrian Data Protection Authority of Google Analytics
7.2	25.03.23	Shah, H.	CU, NaTE, CTU, IB, UOULU	Added more content: zero-party text, TikTok references, partners' contributions in Section 3; CSI-COP's no-tracking website information in Section 5. Edited Summary and Introduction.
7.2.1	29.03.23	Shah, H.	CU	Final grammar and image/screenshots/table numbers check before submission.



## Table of Contents

Summary .....	4
Keywords.....	4
1. Introduction .....	5
2. Cookies.....	6
Origin of cookies .....	7
Cookie parameters.....	8
3. Taxonomy of Cookies .....	9
Website cookies .....	10
Advertising .....	14
Apps and App Permissions.....	17
Website cookie banners, pop-ups and privacy notices .....	21
4. Taxonomy of trackers .....	25
Fingerprinting.....	26
Email tracking.....	27
5. Privacy by Design .....	29
Privacy as the default.....	29
6. Further work .....	32
References .....	34



## Summary

This deliverable presents a classification of cookies and trackers and is aimed at the general public, including parents and teachers who can guide school pupils and children to safeguard their personal data. In this CSI-COP report, a taxonomy of the types and purposes of digital cookies and tracking techniques is detailed with examples from the project's citizen scientists' investigations of websites and apps. These include duration cookies (session; persistent) and cookies that help to measure the performance of a website or an app, as well as tracking tools for marketing and advertising purposes to observe our behaviour as we navigate across the internet. Readers are invited to review CSI-COP citizen scientists' investigations available as open-access databases from the project's website. Future work involves pulling together CSI-COP's citizen scientists and project team members' investigations into a searchable repository. This is currently under development and expected for launch in May 2023. The outcome of the taxonomy and the project is to instil an effort to avoid the temptation of 'accepting cookies' without knowing what these are.

## Keywords

App investigations; AdTech; AppTrackingTransparency, Apple; browser; cookies; first-party; FLoC; Google; internet; server; third-party; third-party requests; tracking; web investigations; world wide web; zero-party



## 1. Introduction

An awakening is stirring about **privacy as a human right**. According to DataGrail (2022), there is “..mounting body of evidence ..average person now very aware how companies track their online behavior ... they will cease doing business with companies and services that they do not trust”. In CSI-COP, members of the general public concerned about online data protection and privacy joined the project to investigate the websites they visit and apps they use.

The classifications in this public report detail the different types of tracking technologies, including cookies, uncovered by CSI-COP citizen scientists investigating beneath the websites they visited and apps they used. This is part of the project’s drive to find how far websites and apps comply with the principles of ‘transparency’, and ‘informed consent’ enshrined in the EU’s general data protection regulation (GDPR).

Two open-access searchable databases from the results of CSI-COP citizen scientists’ explorations are available for review to complement this document. The report here presents classifications of the techniques to make websites and apps work, and the digital artefacts that track us as we visit websites and use apps.

When we visit a website, we might notice a ‘cookie banner’ inviting us to “Accept All”. For convenience and timesaving, we might select this option to access the information we seek from the website we are visiting. But what exactly have we ‘accepted’? Similarly for mobile device applications, or apps on our smart, internet connected devices, such as a transport information app, or a health app, we might download without knowing what default permissions the app might have to our personal contacts, messages and pictures and videos. To investigate the extent of online tracking, and uncover how transparent websites and apps are about informing us before they extract our personal data, CSI-COP project funded by the EU Horizon2020 under the *science with and for society* (SwafS) theme, engaged the general public as citizen scientists to explore their Internet usage and record what they found. This deliverable report is the first of two outcomes from work package 5 (WP5): a) a taxonomy of trackers, and b) a searchable repository of trackers. The latter is under development and due for launch in May 2023. The first outcome presents a classification of trackers through a taxonomy of cookies and online trackers.

This deliverable report first describes **what cookies are and their origins in Section 2**. Next in **Section 3 a taxonomy** is presented with a classification of the **different types of cookies** and their purposes. A **taxonomy of trackers in Section 4** is followed by CSI-COP’s privacy by design, no-tracking approach in **Section 5**. The report concludes in **Section 6 with further work**.



## 2. Cookies

Digital **Cookies** allow websites to ‘remember’. For example, to remember what a consumer intends to buy online from an e-shopping website (such as Amazon.com), by electronically storing chosen purchases virtually placing them in a digital shopping basket. This digital memory enabled through a cookie enables people to continue shopping, or pay for their selected items. As reported by [Simon Hill](#) in [Digital Trends](#) (2015) “**cookies are just a fundamental part of how the Web works**”. The Web here is the world wide web (www) a “collection of web pages” found on the **Internet** “a global network” of connected computers” ([BBC Bitesize](#), n.d.). Cookies find their way on to your Internet-connected devices as you navigate to and across web pages. As [Dutko](#) (2018) explains “in their most basic form, computer **cookies are text strings that websites save to your hard disk**”.

A website will store a **cookie** as “**small pieces of information**” on the device you use to access it ([Behera](#), 2023). [GDPR.EU](#), a project co-funded under the Horizon2020 framework (2023) state that **cookies** are “**small text files that websites place on your device as you are browsing**”. Cookies “**can store a wealth of data, enough to, potentially identify you without your consent.**”. This digital artefact is “a primary tool that advertisers use to track your online activity so that they can target you with highly specific ads” (GDPR.EU). Cookies are ‘activated as we access the multitude of web pages on the internet from a computer, laptop, tablet or a smart mobile device using a **browser**. [Internet browsers](#) include:

- Brave
- Chrome
- DuckDuckGo
- Edge
- Firefox
- Internet Explorer
- Opera
- Safari
- Yandex

Google’s Chrome has the highest market share of the browsers with over 66 per cent ([Statcounter](#), 2023). Microsoft’s Edge has the next highest share at 10.84 per cent, just ahead of Apple’s Safari at 10.14 per cent. Mozilla’s Firefox has 6.84 per cent of market share, and Opera company’s Opera browser has 3.21 per cent of market share, according to [Statcounter](#) as at February 2023. Newer browsers like *DuckDuckGo* and *Brave* are building their brand on privacy. However, in 2022, *DuckDuckGo* had to reverse its decision when it was found that its “mobile browser permitted Microsoft” (a third-party) to load **tracker scripts** when users opened across its browser apps on Apple’s iOS and Google’s Android, while blocking those of Google, and Facebook (Sead Fadilpašić in [TechRadar](#), 2022). Google and Facebook are not without controversy either. Facebook “does not have a good track record when it comes to looking after people’s data” ([Wired](#), 2020). In April 2021, Natasha Lomas reported on Facebook’s **data breach** of 533million users’ data from its platform ([TechCrunch](#), 2021). Nayak and Rosenblatt (2021) reported that: “Google attempted to kill a lawsuit that it secretly scoops up troves of internet data even if users browse in “Incognito” mode” - supposedly in private using Chrome browser, “to keep their search activity private.”.

Additionally, in a civil action brought by the United States District Court-Southern District of New York on anti-trust grounds against Google filed in September 2021 it included the paragraph:



“175. Google presents a public image caring about privacy, but behind the scenes Google coordinates closely with the Big Tech companies to lobby government to delay or destroy measures that would actually protect users’ privacy”. (From [courtlistener.com](https://www.courtlistener.com))

Rashmita [Behera](#) (2023) reports that “since the advent of digital and programmatic advertising” cookies have become an important tool for the advertising technology (**AdTech**) sector. GDPR.EU (2023) inform that:

“Cookies are an important tool that can give businesses a great deal of insight into their users’ online activity. Despite their importance, the regulations governing cookies are split between the GDPR and the ePrivacy Directive.”

Why digital cookies were created is explained in the next section.

### Origin of cookies

Digital cookies emerged on the Internet from the mid-90s and are “a legacy of early web development” ([Jones](#), 2020). Jones recounts the stories of the origin of cookies, one in which an individual devised the digital cookie, and others in which different parties were involved in their creation:

1. The simple story refers to a browser used in the 1990s: Netscape. A computer science graduate, [Lou Montulli](#) at the University of Kansas “created cookies in order to provide” Netscape with the digital “shopping cart it has asked for”
2. Other tales are:
  - a. “a number of computer scientists and developers working in commercial settings and their own start-ups who wanted to expand the functionality of the web”;
  - b. “an international arrangement of participants seeking to capture the power of networked computing”, and
  - c. “diverse and negotiated notions of privacy”

The use of the term cookie, most famously known as a type of [sweet biscuit](#), for the digital cookie was named after the term ‘**magical cookie**’ for:

“Something passed between routines or programs that enables the receiver to perform some operation; a capability ticket or opaque identifier” ([Simon Hill in Digital Trends](#), 2015).

According to the [History of Information](#) (n.d.) in **1994**, Louis J. (**Lou**) [Montulli II](#) at [Netscape Communications Corporation](#), together with John Giannandrea “wrote the initial Netscape cookie specification”. It is reported that the first use of their cookie was for “checking whether visitors to the Netscape Web site had already visited the website” (History of Information, n.d.). Additionally, and relevant to the CSI-COP project concerned with data protection and online privacy, visitors to Netscape’s website were unaware of their visits becoming known through the use of Montulli and Giannandrea’s cookie. History of Information (n.d.) says “... **cookies were accepted by default ... users were not notified of the presence of cookies**”. They remind how the general public learned of the existence of cookies:



“Some people were aware of the existence of cookies ... in 1995, but **the general public learned about them after the *Financial Times* published an article about them in February 12, 1996**” (History of Information, n.d.).

Now the technology behind cookies has advanced to the level of *that* that when you visit a website, the **server** that the website sits on “can identify you and remember things about you through the cookies” ([Simon Hill in Digital Trends](#), 2015).

## Cookie parameters

Cookies have characteristics, or parameters. According to [Digital Shift](#) (2023) cookies have six parameters:

1. The cookie’s name
2. The value of the cookie, this is a computer programming exercise in cookie creation (set-cookie> name=value)
3. Expiration date (this is for the duration of the cookie in the user’s browser)
4. The path the cookie can take, a website not listed in the path cannot access the cookie
5. The domain the cookie is valid for (makes the cookie accessible to multiple servers in a domain)
6. Secure connection requirement (cookie needs to be used over a secure server connection).

Web and app developers use **cookie parameters** to set the objective of a cookie ([Mozilla](#), 2023):

1. **Remembering**: for example, remembering a games’ users score during playing a game.
2. Granting **individual choice**: for example, allowing a user to choose a ‘theme’ or change settings.
3. **Tracking**: recording and analysing a user’s online behaviour.

In the next section in CSI—COP’s **cookie taxonomy** we explain how the cookie parameters allow different types of cookies to act in a particular way, some convenient for us, some uncalled for (personalisation), and some undesired (tracking). We also provide examples of the different types of cookies found during CSI-COP investigations of websites and apps.





### 3. Taxonomy of Cookies

There are distinct cookies dependant on ‘party’. A website or app owner will be a **first party**. For example, the European Commission is the first-party for its official website at this Internet address (or URL): [https://commission.europa.eu/index\\_en](https://commission.europa.eu/index_en). A website might include **first-party cookies**, for example, to know how many visitors have navigated to their website, when they visited, what pages they looked at and for how long. In apps, as well as cookies, which are “highly effective in following consumers around various websites” ([Cookie Law Information](#), 2023) permissions to access features in a mobile device (such as the device owner’s contacts and messages) can allow app owners to extract information about you.

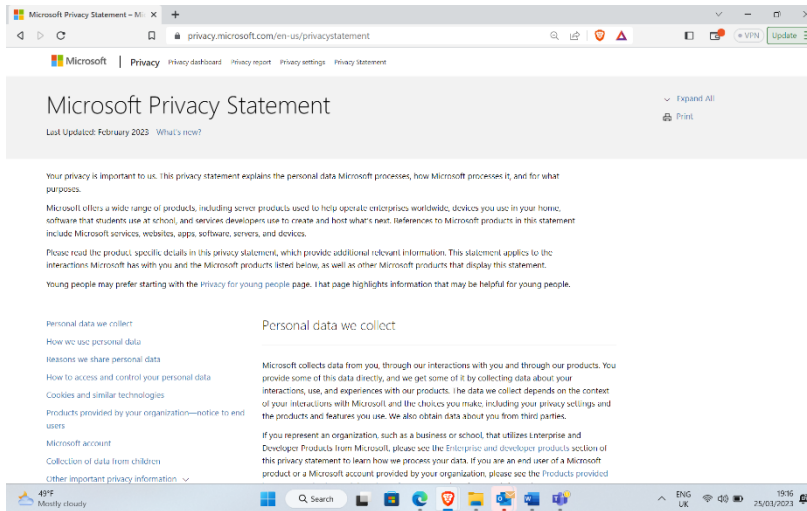
**First-parties** are the organisation whose website you visit. For example, if you visit the [European Commission](#)’s website ([https://commission.europa.eu/index\\_en](https://commission.europa.eu/index_en)), the first party for that website is the European Commission. All organisation’s need data. Whether it is the European Commission, an educational establishment (e.g., a university seeking new students or staff), or a business selling a product or service, the organisation will need data to survive so their website can be a platform to personalise information offered to suit visitors. In this way the website could encourage people to return, engage and ‘buy’ the service or product. A first-party website can collect particular data on its visitors, for example, collecting name and email to subscribe to the organisation’s newsletter. Name and email is **zero-party data** that a person shares to engage with a website. It is considered **up-to -date, accurate data** by the first-party.

If a person gives their name and address intentionally to purchase an item online, that data will be accurate if they want to receive the purchases. Hence zero-party data is “information that a customer freely provides to a company” – so the data is “not inferred from how a customer behaves” on a website ([Bloomreach](#), 2023). Additionally, “data given explicitly by a customer” could be “with the expectation that providing their data” a customer could be “rewarded with a better shopping experience” ([Bloomreach](#), 2023). According to [Marketing Labs](#) (2023) website, a first-party can ‘personalise’ its interaction with visitors in the following ways:

- “Utilise web analytics data [e.g., how long a person stayed on a web page] to improve page and content performance”.
- “Segment customers to deliver improved targeting and personalisation”, for example, a university website might want to segment students seeking postgraduate courses from information for students seeking undergraduate course information.
- “Sent targeted emails to customers based on their segments or the way they have interacted with the website, e.g., abandoned basket offers”. This could be the case where a person has started to shop for products on an e-commerce website but not completed the purchase by paying for the chosen items.
- “Use direct mail” to contact people who might have registered with a website. For example, an entertainment website such as Netflix might send a direct message to its customers informing of upcoming shows that could be of interest from previous show viewings.
- “Send messages or emails to customers who are local to businesses”.

First-party website owners might also seek information from an outside organisation. This would be a third-party. For example, in [Microsoft Privacy Statement](#) on their website (Screenshot1), in the section on ‘Personal Data We Collect’ Microsoft discloses (in its last sentence of the first paragraph): “We also obtain data about you from third parties” ([Microsoft](#), 2023). We look at third-parties next.





Screenshot 1: Microsoft Privacy Statement

**Third-parties** are companies that can be unrelated to a website or app owner (first party) that can access visitor information on someone else’s, first-party website through a **third-party cookie** or a **third-party request**. **First-party cookies, and third-party cookies and requests can track users**. However, under the EU’s general data protection regulation ([GDPR](#)) if a website or an app is likely to process personal data of an EU subject, regardless of whether the website or app owner is an EU or non-EU business, they are obliged to make it fully transparent what cookies are embedded and they must allow informed consent for a user to accept or decline cookies ([GDPR.EU](#), 2023).

In the next sub-sections, we explain ‘**cookies**’, beginning with website cookies.

### Website cookies

Website cookies are also known as HTTP cookies ([Target Internet](#), n.d.). A website as a first party can deploy various cookies. However, companies and organisations do not have to follow a format for their cookies. Some cookies are essential to make a website work, hence have a purpose. Other cookies record the amount of time a user is on a website: duration. Table 1 below presents cookies that might be explained in a website’s **privacy policy**.

Cookie type	Purpose		Duration	
	<i>Essential, or necessary</i>	Cookies that make a website work		<i>Session</i>
<i>Functional</i>	Store user preference			
<i>Performance</i>	Device statistics			
<i>Analytics</i>	Web navigation			
<i>Advertising</i>	Personalised ads		<i>Persistent</i>	Cookie that remains on a user’s device
<i>Marketing</i>	Targeted marketing			
<i>Behavioural</i>	Tracking online behaviour			
<i>Location</i>	Store location			
<i>Secure</i>	Encrypt data			

Table 1: Types of first-party website embedded cookies



### Duration cookies

Cookies that record the time a visitor spent on a website are session, and persistent (Table 1).

#### Session

Session cookies are temporary ones stored in a user’s browser .These types of cookies are deleted once the user closes the browser.

#### Persistent

Persistent cookies are stored on the user's computer for a specified period and are not deleted when the browser is closed (Table 1). These are explained below.

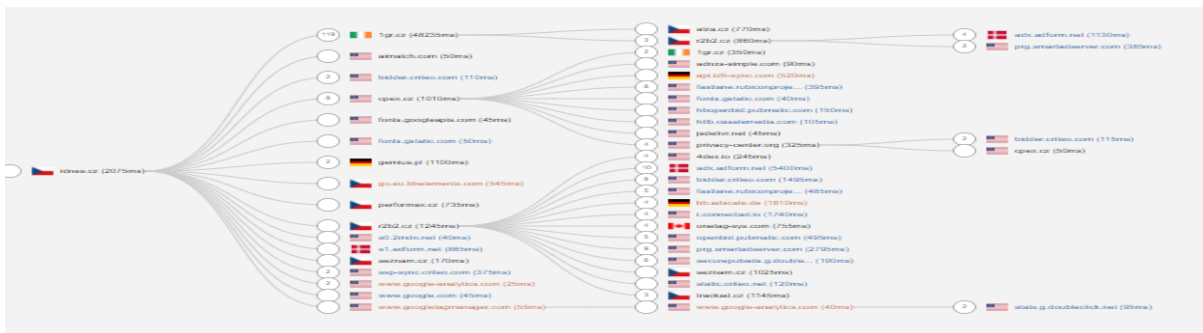
### Purpose cookies

There are several types of ‘cookies for purpose’. Below we detail the different purpose cookies.

#### Essential or Necessary

These cookies are essential for performing basic tasks on a website, such as maintaining an open session, accessing restricted areas, and preserving the contents of a shopping cart.

An example of a website from which you can see these types of ‘purpose cookies’ embedded is the Czech daily news website [iDnes/Zpravdajstvi](https://www.idnes.cz). A CSI-COP citizen scientist investigated this website (on 5.12.22) using a free online web audit tool: PageXray by Fou Analytics. Screenshot 2 shows the ‘data travel’ map from the first-party website across to third-parties, including to countries outside the European territory (country flags shown in the website audit map – Screenshot 2).



Screenshot 2: Web audit of iDnes/Zpravdajstvi using PageXray by Fou Analytics

Using a different free web audit tool, webbkoll in December 2022, the CSI-COP citizen scientist found that the Czech daily news website, [iDnes/Zpravdajstvi](https://www.idnes.cz), contained first-party and third-party cookies. This website also had third-party requests (Screenshots 3 and 4). An investigation by the first-author of this same website using [webbkoll](https://www.webbkoll.com) on 25.03.23 showed that the third-party requests were from advertising companies, as well as ‘big tech’ (Google), see Table 2.

Advertising third-party requests include	Other requests (tracking) include
Advertising (Teads.tv), Email, Advertising (Adform), Advertising (SAS)	FingerprintingGeneral, Email, Analytics (Google), FingerprintingGeneral, Advertising (Google). FingerprintingGeneral, Email, Advertising (Google)

Table 2: Third-party requests in iDnes/Zpravdajstvi website according to [webbkoll](https://www.webbkoll.com) free online web audit tool





✔ HTTPS by default

www.idnes.cz uses HTTPS by default.

Screenshot 3: WebbKoll free tool audit of iDnes/Zpravdajstvi Czech news website presenting number of cookies

### Third-party cookies (14)

Domain	Name	Value	Expires on	HttpOnly	Secure	SameSite
.adform.net	uid	3378260678620056211	2023-05-22 21:42:28Z	✗ No	✔ Yes	✔ Yes (None)
.adform.net	C	1	2023-04-23 21:42:28Z	✗ No	✔ Yes	✔ Yes (None)
.adscale.de	uu	31d129176b294fc98562...	2024-03-20 14:09:05Z	✗ No	✔ Yes	✔ Yes (None)
.adscale.de	cct	1679607745022	2024-03-20 14:09:05Z	✗ No	✔ Yes	✔ Yes (None)
.bidr.io	bitolsSecure	ok	2024-04-21 17:42:28Z	✗ No	✔ Yes	✔ Yes (None)
.bidr.io	bito	AAFYc07IOZAAACEDvCKQ...	2024-04-21 17:42:28Z	✗ No	✔ Yes	✔ Yes (None)
.bidswitch.net	tuuid_lu	1679607747	2024-03-22 21:42:28Z	✗ No	✔ Yes	✔ Yes (None)
.bidswitch.net	c	1679607747	2024-03-22 21:42:27Z	✗ No	✔ Yes	✔ Yes (None)
.bidswitch.net	tuuid	9d4034f1-d520-46ef-a...	2024-03-22 21:42:28Z	✗ No	✔ Yes	✔ Yes (None)
.criteo.com	uid	12f0c657-1717-4365-9...	2024-04-16 21:42:24Z	✗ No	✔ Yes	✔ Yes (None)
.hit.gemius.pl	Gtest	KIGb2MMGQMGGKVOFz8KZ...	2023-03-30 21:42:24Z	✗ No	✔ Yes	✔ Yes (None)
.rubiconproject.com	audit	1 hLZGFuTafB3MVLvtnS...	2024-03-22 21:42:25Z	✗ No	✔ Yes	✔ Yes (None)
.rubiconproject.com	khaos	LFLN32BE-A-MG7B	2024-03-22 21:42:25Z	✗ No	✔ Yes	✔ Yes (None)
.seznam.cz	sid	id=53169597973067217...	2023-04-22 23:42:38Z	✗ No	✔ Yes	✔ Yes (None)

Screenshot 4: WebbKoll free audit tool presenting third-party cookies in iDnes/Zpravdajstvi website



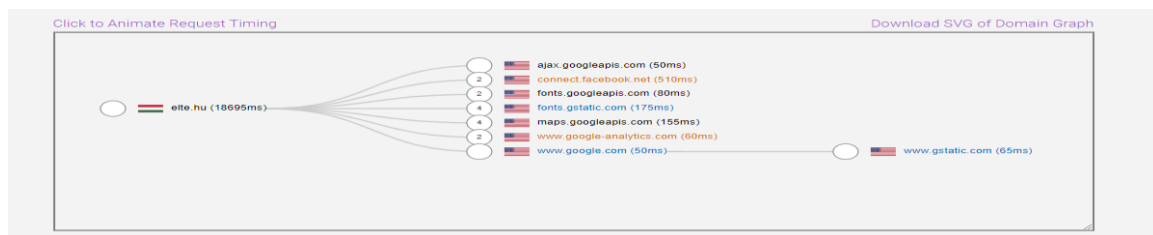
## Functional

This type of cookie stores user preferences, such as their preferred language, font size, and theme.

## Performance

Performance cookies, allow the website server to gather statistics on the performance of a user’s device and browser while they access and view a website. These cookies collect aggregated information, and in theory, cannot be associated with a specific individual.

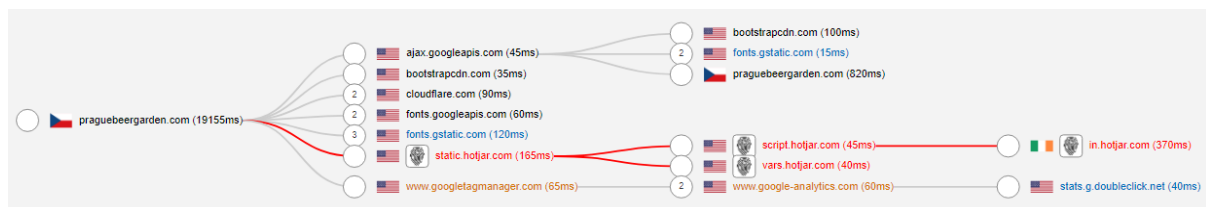
One of CSI-COP’s citizen scientists investigated the Hungarian website of Eötvös Loránd University’s Faculty of Economics (on 31.11.22). The free online web audit tool [PageXray by Fou Analytics](#) showed one performance cookie embedded in the website by Google Universal Analytics (see Screenshot 5).



Screenshot 5: ELTE GTK website audit using PageXray by Fou Analytics.

## Analytics

Analytics cookies collect information about how users interact with a website. The information gathered by these cookies is used to analyse website traffic and user behavior. They track things like the number of visitors to a site, the pages they visited, the length of time they spent on the site, and the sources of their traffic, i.e., which website the user visited previously, and which website they navigate to next. The aim of analytics cookies is for the website owners to improve the overall **user experience** on their website. Information from analytics cookies is aggregated and analysed giving website owners valuable insights into their audience and how they interact with the site. An example of an analytics cookie used by many websites is [Google Analytics](#). According to Google’s Privacy (point 7) in its [Terms of service](#) for the Google Analytics tool, organisations using this tool must not “assist or permit any third party, to pass information to Google that Google could use or recognise as personally identifiable information” (Google Marketing Platform, n.d.). Google Analytics would be a **third-party cookie** on someone else’s website. Using a free online web audit tool ([PageXray by Fou Analytics](#)) one of CSI-COP’s citizen scientists investigated [Prague Beer Garden website](#) (on 12.06.22) and found it embedded Google Analytics, among other digital artefacts (see Screenshot 6).



Screenshot 6: Prague Beer Garden website audit using PageXray by Fou Analytics

The above screenshot displays the flow of data from a website appearing to be in Prague, Czech Republic to other regions, according to the free online tool, PageXray by Fou Analytics (n.d.). It is to be noted that the **Austrian Data Protection Authority** “decided that the use of Google Analytics is

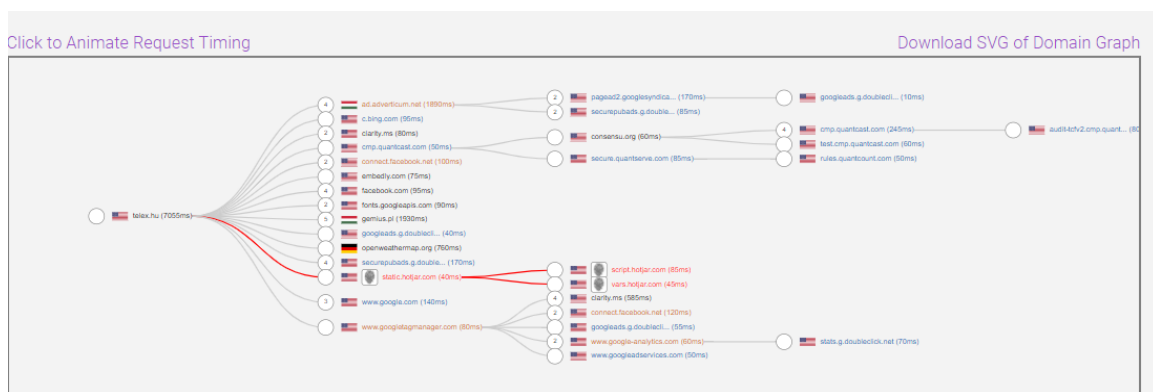


currently violating the GDPR” due to the transfer of data to the US (Stephan Winklbauer in [INPLP, 2022](#)).

### Advertising

These cookies are aimed at specific individuals in an attempt to uniquely identify users so that personalised advertisements can be displayed to them. The objective is to provide information more appealing to that user.

The Hungarian news portal [Telex](#) was explored by a CSI-COP citizen scientist (on 28.11.22) using [PageXray by Fou Analytics](#). As Screenshot 7 shows, there were several advertisement cookies embedded in the website set by providers such as Google, Facebook, Adverticum, Bing and the Polish market research platform Gemius. The duration of some of these cookies can be more than one year.



Screenshot 7: Telex website audit using PageXray by Fou Analytics

### Marketing

Marketing cookies allow organisations to gain a “better insight into their website’s visitors” (Demetra Edwards in [TechTarget, 2020](#)). For example, a university might want to understand how potential students are viewing the different pages on its website. This can lead to **personalisation** which allows an organisation “to display a consistent experience for individual users over time, across multiple sessions” ([Digital Shift, 2023](#)). Personalisation cookies can afford convenience for the consumer, because the cookie can store information that allows automatic-completion, or auto-fill details such as name and address, so that the consumer does not need to add this information again on the next visit.

An example of a website that embeds marketing cookies beneath it is the Romanian online pharmacy [Catena](#). This website was investigated using [webbkoll](#), a free website audit tool (on 5.2.23). Screenshots 8 and 9 show that Catena online pharmacy website in Romanian had first party cookies and also allowed third-party requests including by Google, and by Gemius. According to its website ‘About Us’ page, [Gemius](#) is an “international research and technology company” that provides “data as well as advanced tools for digital and traditional marketing activities such as web analytics, online campaign’s management and ad serving” (Gemius, 2023).





### Cookies

First-party cookies (9)

Domain	Name	Value	Expires on	HttpOnly	Secure	SameSite
catena.ro	_ga_CJ9I2H0B56	GS1.1.1679481108.1.0...	2024-04-25 10:31:49Z	No	No	No
catena.ro	_ga	GA1.2.1421792995.167...	2024-04-25 10:31:49Z	No	No	No
catena.ro	ao-fpgad	%7B%22fpcRequired%22...	2024-04-15 10:31:48Z	No	No	Yes (Lax)
catena.ro	_gat_UA-33138725-1	1	2023-03-22 10:32:48Z	No	No	No
catena.ro	_dc_gtm_UA-33138725-...	1	2023-03-22 10:32:48Z	No	No	No
catena.ro	_gid	GA1.2.1880492881.167...	2023-03-23 10:31:48Z	No	No	No
catena.ro	_gcl_au	1.1.383265393.167948...	2023-06-20 10:31:48Z	No	No	No
www.catena.ro	catena_session	UTUFOFMxAjhVfQp4ADsJ...	session	Yes	Yes	No
www.catena.ro	PHPSESSID	k382rbaq90hk5dqtccva...	session	No	No	No

Screenshot 8: First-party cookies beneath Catena online pharmacy according to webbkoll free web audit tool

### Third-party requests

44 requests (44 secure, 0 insecure) to 19 unique hosts.

A third-party request is a request to a domain that's not `catena.ro` or one of its subdomains.

Host	IP	Classification	URLs
10654980.fs.doubleclick.net	142.250.74.166	Fingerprinting(General), Email, Advertising (Google)	Show (2)
adro.hitgemius.pl	128.140.224.228	Advertising (Gemius)	Show (4)
adservice.google.com	2a00:1450:400f:80c:2002	Content (Google)	Show (1)
adservice.google.de	2a00:1450:400f:80c:2002	Content (Google)	Show (1)
ajax.googleapis.com	2a00:1450:400f:804:200a	Content (Google)	Show (1)
cdnjs.cloudflare.com	2606:4700:6811:180e		Show (1)
cloudflareinsights.com	2606:4700:6810:3965		Show (2)
consent.cookiebot.com	2001:2030:0:171:5f65:8541		Show (2)
consentcdn.cookiebot.com	2001:2030:4e:89-fd9		Show (1)
fonts.googleapis.com	2a00:1450:400f:804:200a	Content (Google)	Show (1)

Screenshot 9: Third party requests from Catena online pharmacy according to webbkoll free web audit tool

## Behavioural

Pat Walshe of [Privacy Matters](#) (2021) explains that the information arising from **behavioural data** online may include:

- A User's **web browsing** data – the websites you visit, the date and time you visit, the country you visited from (inferred from your IP address - a unique string of characters that identifies each device connecting to the internet and that is automatically sent when you visit a website). Also consider that when you leave a website, they'll be able to tell which site you are visiting next and the next website you visit may be able to tell which website you came from. All of this would be considered web browsing behavioural data.
- **'clickstream behaviour'** – data about an individual's interactions on a website, that can include what they click and scroll and tap on a touch screen
- **search engines** such as Google that may collect and use information about what people search for, what results you click on, your IP address, and that may use a unique cookie identifier to track you.



- **location** – the location and type of place you visit (supermarket, casino, place of worship, hospital), or where you used an app, the dates and times, route travelled, the frequency of a visit or the routes you travel. Location data can be very [revealing](#) and behavioural in nature.
- **purchase history** – this can include types of subscriptions (trade union membership, gym, newspapers etc), hotel or restaurant reservations that may have been made across [search, maps, smart assistants](#) or directly from retailers or third party services etc.
- **payment or ‘transactional’ data** – payments that reveal who/what organisation you paid (which can reveal the type of organisation - medical clinic, pharmacy, alcohol provider; food retailer, bookseller etc) and how much and when and how often. Tap and go card payments are a good example – think of that coffee you purchase at the start of a journey, the place, date and time you paid for it, and then payments you make later in the day with the same card.
- **streaming media** – “[you are what you stream](#)” and “[They know what You Watched Last Night](#)”. Streaming media generates **a lot** of behavioural data about:
  - the date and time you accessed a streaming music, audio or TV/movie service and non-precise location (country level or region level) you accessed it from
  - which profile accessed and used the service (a name + category e.g., child)
  - the category of music, audio book, TV/movie (e.g., political horror, adult)
  - searches for content
  - whether you paused a song or movie and for how long (including the date(s) and time(s))
  - whether you skipped/abandoned a song or a TV episode movie audio track
  - whether you shared content and who with and your interactions with others within the service
  - whether you rated a song, TV show or movie
  - playlists or ‘watch’ lists you create
  - the device used to access the service and IP address and device identifiers
- **Activity/Health data** – data about your use of activity apps such as cycling, running, walking or data about your health such as that obtained via dietary or fertility apps. This data can be very revealing and may often be connected with your location for example.
- **Social media graph** – data revealing interconnected social relationships between people and their nature and patterns of communications

### Location

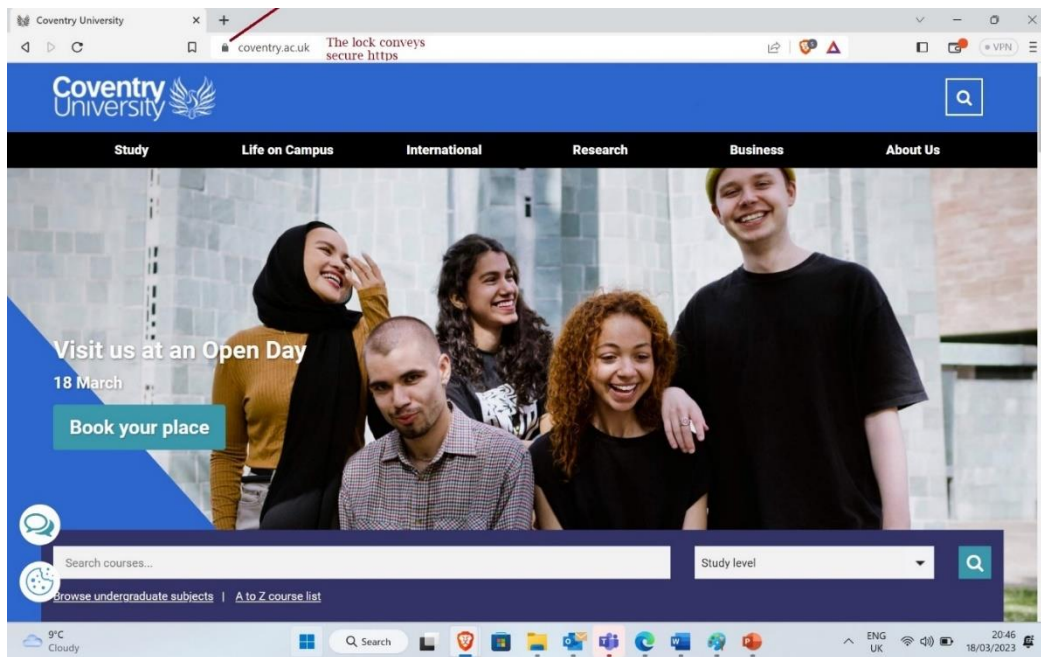
Location data is considered “not too personal” (AdExchanger, 2013). This type of cookie helps marketing companies to reach (locate) their consumers.

### Secure or HTTPS

Secure cookies can be session or persistent cookies, but which can only be transmitted via a secure, encrypted connection (for example, websites with http and ‘s’ in their internet address > **https**). Screenshot 10 illustrates a website using ‘https’. A ‘lock’ icon alongside the website’s address (URL) signifies it is using the https secure protocol (Screenshot 10).







Screenshot 10: CSI-COP Coordinator: Coventry University website showing lock in address bar for secure https.

In addition to the different types of cookies mentioned above, **malicious** cookies, and **zombie** cookies also exist.

### Malicious cookies

Malicious cookies can “store your preferences, and build profiles based on your interests” ([Digital Shift, 2023](#)). The purpose of malicious cookies is to “create a profile of website visitors with enough data collected from tracking cookies so that the information can be sold to different companies” ([Digital Shift, 2023](#)), without a consumer’s knowledge or consent.

### Zombie cookies

A zombie cookie places “data and code on a user’s device in a hidden location” ([Digital Shift, 2023](#)). Such cookies are “used to track a user’s browsing history across multiple sites” from inside the user’s computer.

### Apps and App Permissions

As mentioned earlier in this section, apps have ‘permissions’ that can be used to extract data from a smart phone device. Some permissions are required for an app to work. For example, a transport app will require access to **device location** to provide the owner, or user, to manage their time effectively so that people can leave their home, or other place, in sufficient time to catch the transport. Image 1 below depicts the top-half of a privacy audit report (from a free online tool, [Exodus Privacy](#)), for an app for ‘Bus Times’ (Image 1). According to the app investigation in 2021, the app was revealed to have 14 trackers, and 14 default permissions to features on the first author’s ‘phone’. The trackers were from parties unrelated to the company providing the bus timetable information through the app. The ‘third-parties’ included: Facebook Ads; Google Analytics; Google CrashLytics; Google Firebase Analytics;



Google Tag Manager; IAB Open Measurement; Integral Ad Science; Twitter MoPub; and Verizon Ads (according to [Exodus Privacy, 2021](#)).

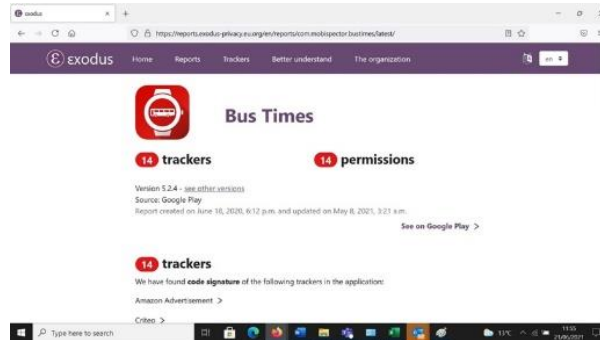
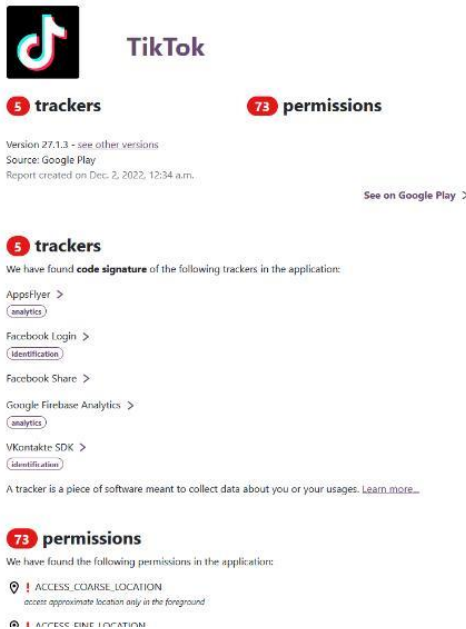


Image 1: Exodus Privacy audit of a transport app

[TikTok](#) a social media app, was investigated by a CSI-COP citizen scientist. TikTok, “a short-form video hosting service owned by a Chinese company ByteDance” has become controversial from governments banning it on their employees’ devices. In [February 2023 the European Commission](#) decided to “suspend the use of TikTok application on its corporate devices and on personal devices enrolled in the Commission mobile device service” (EU, 2023). This decision was made to “increase cybersecurity” by protecting “the Commission against cybersecurity threats and actions, which may be exploited for cyberattacks against the corporate environment of the Commission” (EU, 2023). The EU statement also added that the “security developments of other social media platforms will also be kept under constant review” (EU, 2023). On [16 March 2023, the UK government](#) announced it too would be banning TikTok on UK government devices “as part of wider app review” (UK, 2023). In the US, CNN



Business reported that “federal officials are demanding the app’s Chinese owners sell their stake in the social media platform, or risk facing a US ban of the app” ([CNN, 2023](#)).

Using the Exodus Privacy free online tool for Android apps, CSI-COP citizen scientist’s investigation uncovered that TikTok had 5 trackers and 73 permissions (Image 2). TikTok access to a device through its app, included access to the device’s location, the owner or user’s calendar, contacts, storage, microphone, camera and settings (see Image 2 left).

Image 2: Exodus Privacy audit of TikTok app.



To prevent this amount of tracking through an app by third-parties, app settings in a phone can be changed. Default settings can be altered to allow permission to the feature required for the app to function. For example, if photo-sharing is an activity shared with contacts in a smart phone, then a messaging app will need access to the camera and contacts. Table 3 shows the path to follow to change app settings in an Android device(left column), or in an Apple phone (right column). By following the instructions in Table 3, an app’s necessary permissions can be ascertained, and any other permissions can be blocked to preserve personal data in the device.

Android device (e.g., Samsung phone)	iOS 14 device (Apple phone)
On your <b>Android</b> device: open the Settings <b>app</b> .	Go to Settings >
Tap Apps & notifications	Select Privacy >
Tap the <b>app</b> you want to review	Tap a category of information, such as Calendars, Reminders, or Motion & Fitness.
Tap <b>Permissions</b>	
If a <b>permission</b> is turned off, the switch next to it will be grey	
Try using the <b>app</b> again.	For other iOS devices go to Apple support:
For more go to <a href="https://support.google.com/googleplay/answer/6270602?hl=en-GB#zippy=%2Csee-all-permissions-for-each-app">Google Support</a> : <a href="https://support.google.com/googleplay/answer/6270602?hl=en-GB#zippy=%2Csee-all-permissions-for-each-app">https://support.google.com/googleplay/answer/6270602?hl=en-GB#zippy=%2Csee-all-permissions-for-each-app</a>	<a href="#">Apple Support</a>

Table 3: Instructions to change app settings in a mobile device.

## Summarising cookies and permissions

### First-party and third-party cookies:

- a. Duration:
  - i. Session cookie
  - ii. Persistent Cookie
- b. Purpose:
  - i. Essential or necessary
  - ii. Functional



- iii. Performance
- iv. Analytics
  - v. Advertising
  - vi. Marketing
- vii. Location
- viii. Secure
- ix. Tracking

Permissions:

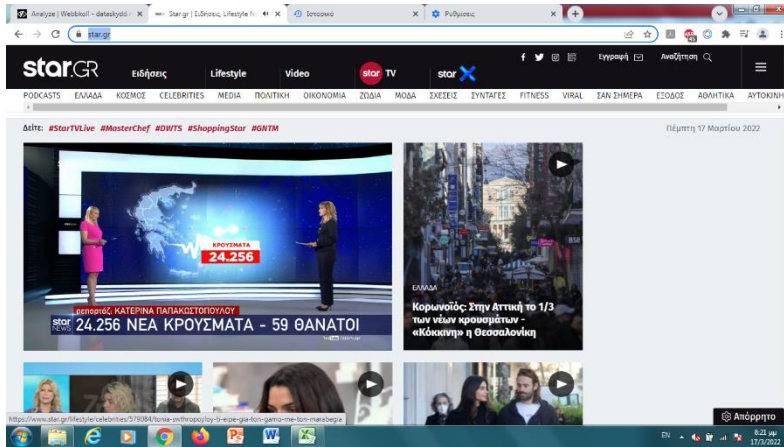
- c. Settings in apps that enable the application to access information from features in an internet-connected mobile device. Access to some features is necessary for the app to work.

In the next sub-section, we look at cookie banners and cookie notices on websites that at least make you aware of the existence of cookies in the website, but do they inform precisely what that cookie is for?



## Website cookie banners, pop-ups and privacy notices

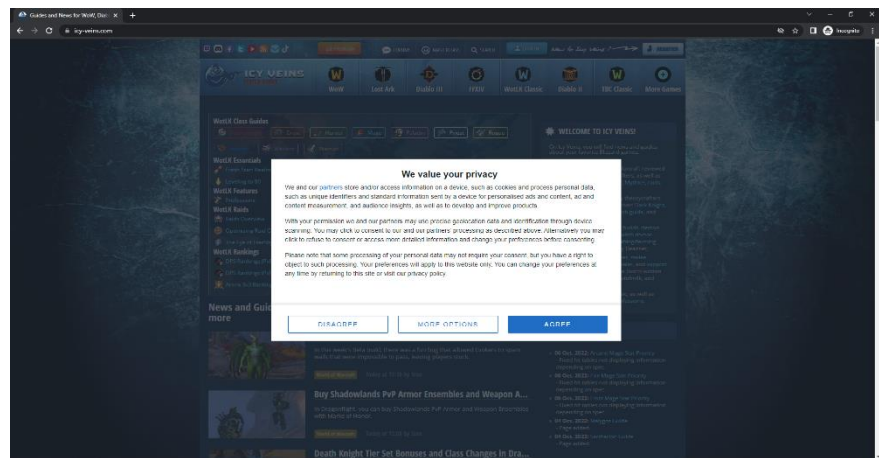
There is no standard for website cookie banners, pop-ups or privacy notices. This means organisations are free to craft their ‘cookie text’ in their own way and place on their websites informing about cookies and online privacy. Among investigations conducted by CSI-COP’s citizen scientists, they recorded



information on cookies from websites they visited. One of the eight hundred and fifty websites a CSI-COP citizen scientist investigated (Shah & Winter, 2022) was Star, a TV news channel in Greece (Screenshot 11). The citizen scientist reported that the website cookie notice was “Bottom of the page” and in their opinion “explicit, comprehensive”.

Screenshot 11: Star TV station – Greece website.

Icy-veins a games website investigated by another CSI-COP citizen scientist, the volunteer reported that the “First time the page is loaded, the user is presented with a banner, containing a privacy policy” (Screenshot 12). This ‘wall’ between website visitors and information on the website offers cookie information (Screenshot 12).

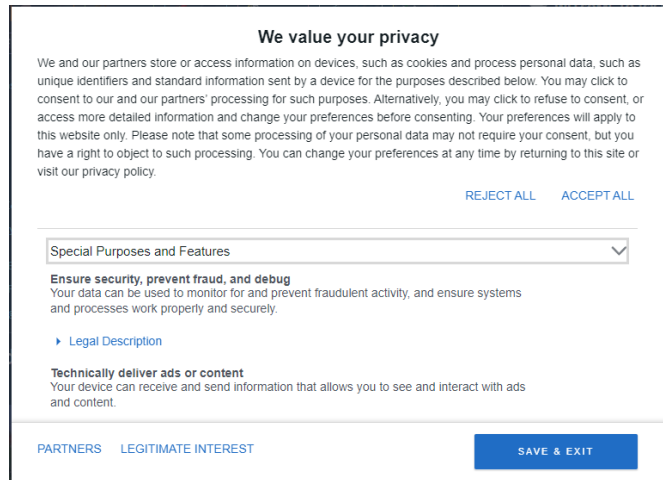


Screenshot 12: Icy-vein games website cookie wall

From Screenshot 13 of Icy-veins.com (overleaf) the visitor can observe two options beneath the top paragraph on the right of the wall:

REJECT ALL ACCEPT ALL.





Screenshot 13: Cookie information beneath Icy-vein games' website

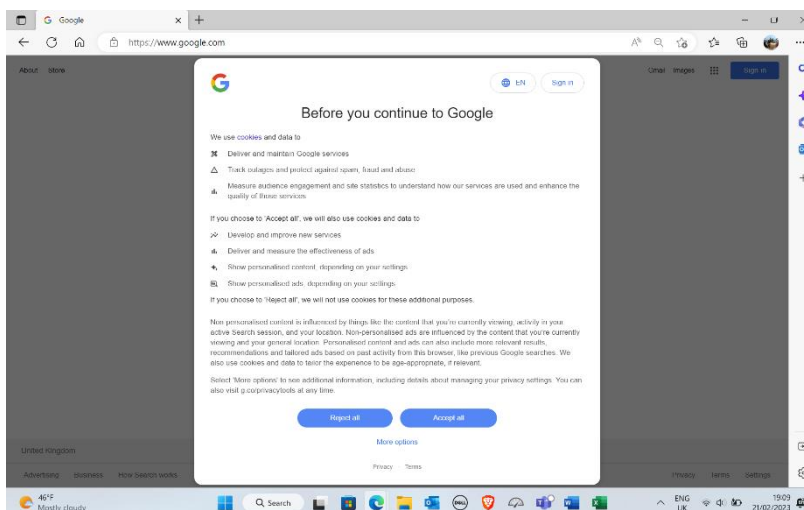
The paragraph in Icy-veins website's privacy policy opens with "We and our partners store or access information on devices, such as cookies and process personal data, such as unique identifiers" (Screenshot 14). The paragraph contains the statement "Please note that some processing of your personal data may not require your consent, but you have a right to object to such processing". It is not made transparent why some personal data can be processed without consent. At the bottom left of that privacy policy, two headers are presented:

PARTNERS    LEGITIMATE INTEREST.

Unless a visitor to Icy-vein's website checks who the 'Partners' are, and what 'Legitimate Interest' means, they might succumb to the third-party tracking.

### Website pop-ups

Screenshot14 shows the 'pop-up' that appears when you visit Google.com. Before you can access



Screenshot 14: Google.com 'cookie banner'

information about Google on Google.com you are presented with information about why Google uses cookies: to "Deliver and maintain Google services". Google informs that "If you choose to 'Accept all' then, among other services, Google will "Show personalised ads, depending on your settings" (Screenshot 15). How do Google create personalised ads? By tracking web users across the Internet.

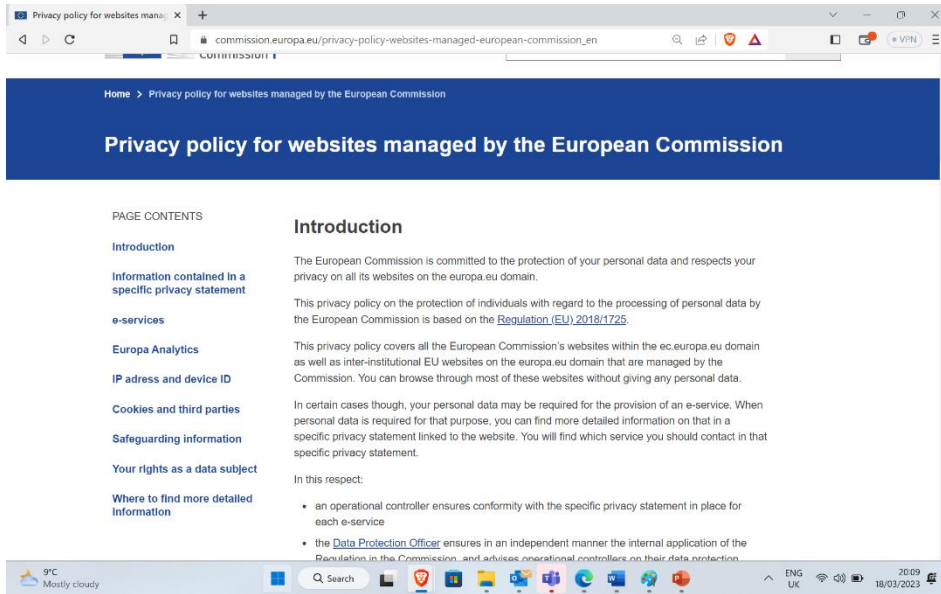




*Privacy Policies*

Privacy policies An example of a website privacy policy can be found on the [European Commission](#)'s official internet page (Screenshot 15):

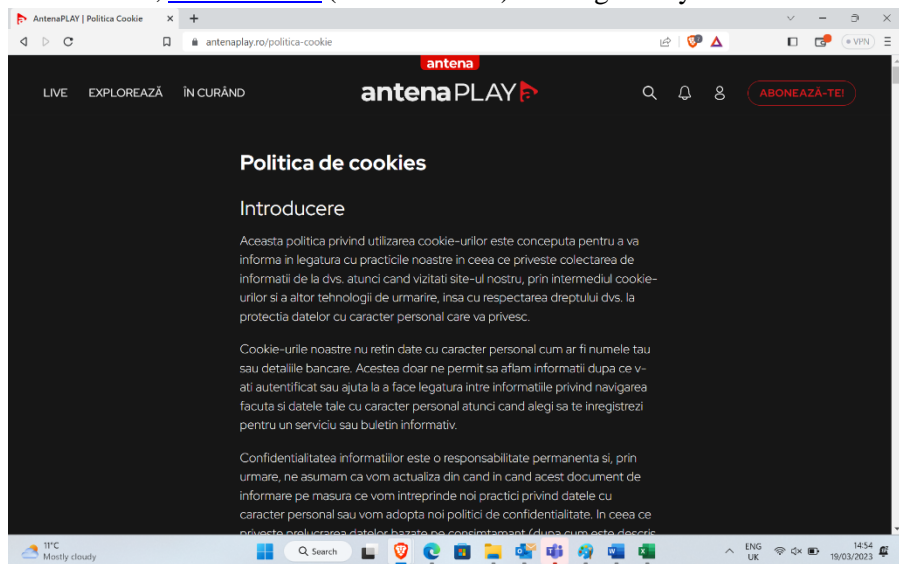
[https://commission.europa.eu/privacy-policy-websites-managed-european-commission\\_en](https://commission.europa.eu/privacy-policy-websites-managed-european-commission_en)



*Screenshot 15: European Commission websites' Privacy Policy*

An entertainment website in Romanian, [antenaPLAY](#) (Screenshot 16) investigated by another CSI-COP citizen scientist (on 22.11.22) was found to contain the same text for cookies and privacy: “The privacy policy is the same as the cookie policy”.

*Screenshot 16: antenaPLAY cookie-privacy notice*



CSI-COP’s citizen scientist here used another free online web audit tool, [webbkoll](#) – a tool that helps non-technical Internet users to check for free “what data-protection measures a site has taken to help exercise control over privacy) (webbkoll, n.d.). During the investigation, it emerged that antenaPLAY website allowed **47 requests to 17 unique hosts**, including:



- adservice.google.com;
- Content (Google);
- Social, FingerprintingGeneral (Facebook);
- FingerprintingGeneral, Email, Analytics (Google)

These 'requests' are third-party tracking discussed in the next section.





#### 4. Taxonomy of trackers

The myriad of cookie banners and cookie notices, at times hosting ambiguous text, succumb us to online tracking by known and unknown companies, sometimes without our informed consent. The principles in the EU's GDPR are clear about **transparency**, **informed consent**, and **purpose limitation**. Alongside the GDPR is the [ePrivacy Directive](#) (EU, n.d.) which is an “important legal instrument for privacy in the digital age” concerned with “confidentiality of communications and the rules regarding tracking and monitoring”. This means that e-businesses are required to comply with the legislation surrounding the use of cookies which could amount to surveilling people as they surf the web, and lead to inadvertently identifying who the people are, which in turn could lead to real harms as a consequence of a data breach.

Researchers (Sivan-Sevilla et al., 2020) reported in a study that “companies we may never have heard of are collecting data points on every aspect of our lives – our interests, purchases, health condition, locations, and more”. IAB (2019,) “These data points are then combined into exceptionally revealing behavioral profiles, exposing intimate parts of our identity and fuel the multi-billion-dollar advertising industry that claims to predict what we are likely to consume in order to target us with ads” (quoted in Sivan-Sevilla et al., 2020). They further reported that when advertisers cross information about users’ medical problems, educational interests, and news consumption habits they are in a position to better know when a user can be turned into a consumer and make purchasing decisions that advertisers would not be able to predict otherwise. Studies showed how data from different websites is aggregated and used to infer about the demographics and interests of users, exposing them to manipulative practices that try to make them click on the ‘right’ (personalized) advertisement at the ‘right’ (personalized) time ... The advertising industry had defined these moments as ‘**prime vulnerability moments of consumers**’ ...in which users are ‘uniquely receptive’ ...”. (Sivan-Sevilla et al., 2020)

How the advertising business works “has changed drastically over the last two decades” (Srinivasan (2020):

“Today, the largest category of advertising, online advertising, is rarely negotiated by people at all. Advances in technology allow ad space to be bought and sold electronically through centralized trading venues at high speeds, without people ever meeting face-to-face. When a user visits a website, the ad space on a page is instantly routed into one or more of these venues. There, the space is auctioned in real-time to the highest bidder. At the conclusion of these auctions, the advertisers’ ads return and display to the user in time for the page to load and before the user has noticed anything has occurred. The user just sees ads targeted to them, say one for Barclays bank.”

Cookies are not the only digital means that advertisers use to learn about us and our behaviour online, but they were a strong tool in an online advertiser’s armoury. Jon Waterman in Forbes (2020) reminded that:

“Online advertising was built on cookies. Third-party cookies track consumers' activity across the internet to gather data and generate information that is useful for **ad targeting** and **performance tracking**. Essentially, **cookies paint a picture of a consumer as they move across the web**. Without them, attribution gets fuzzy and brands are unable to figure out how to spend ad dollars effectively.”



However, [Google](#) announced that they would be phasing out cookies ([Forbes](#), 2020). Through its [Chromium blog](#), on 14 January 2020 Google pointed to a “A path towards making third party cookies obsolete”. The Blog ads:

“...with the web community, we [Google] are confident that with continued iteration and feedback, **privacy-preserving** and **open-standard mechanisms** like the Privacy Sandbox can sustain a healthy, ad-supported web in a way that will render third-party cookies obsolete. Once these approaches have **addressed the needs of users**, publishers, and advertisers, and we have developed the tools to mitigate workarounds, we plan to phase out support for third-party cookies in Chrome. Our **intention is to do this within two years**. But we cannot get there alone, and that’s why **we need the ecosystem to engage on these proposals**.”

A year after that Google blog post in 2020, Bennet Cyphers reported in the Electronic Frontier Foundation ([EFF](#), 2021) that **Google** had in fact begun an ‘origin trial; of its **Federated Learning of Cohorts**, or **FLoC**, its “new experimental technology for targeting ads”. This new technique means:

“A switch has silently been flipped in millions of instances of Google Chrome: those browsers will begin sorting their users into groups based on behavior, then sharing group labels with third-party trackers and advertisers around the web. A random set of users have been selected for the trial, and they can currently only opt out by disabling third-party cookies”.

EFF’s (2021) report states that:

“During the [Chromium] trial, trackers will be able to collect FLoC IDs *in addition to* third-party cookies. That means all the trackers who currently monitor your behavior across a fraction of the web using cookies will now receive your FLoC cohort ID as well. The cohort ID is a direct reflection of your behavior across the web. This could supplement the behavioral profiles that many trackers already maintain.”

Google’s plans for its FLoC have not yet materialised. Ron Amadeo in [ArsTechnica](#) (2021) reported:

“Everyone who isn’t an advertising company seems to have come out against FLoC. The EFF, Brave... DuckDuckGo have all put strong statements against the idea ... other browser vendors – like Apple, Opera, Mozilla and Microsoft – have flouted more tepid “no plans to implement” statements. Amazon is already blocking FLoC on Amazon.com ...”.

ArsTechnica reported Google revealing “it has received ‘substantial feedback’ from the web community after its first trial of FLoC, and now it’s going to take things a bit slower” (2021). According to Google, the delay strategy would “allow sufficient time for public discussion on the right solutions” and “continued engagement with regulators”. The expected disappearance of third-party tracking cookies notwithstanding, other tracking techniques do exist, these are explained next.

## Fingerprinting

Fingerprinting is a form of website tracking that **uses the attributes of the user’s device or browser to build a profile of a user** (see Table 2, Section 3). Information fingerprinters use include your device, the operating system you have on the device, screen resolution, browser and browser version, language, and time zone. Crawford (2020) states: “On its own, **each piece of information isn’t that valuable**.”



However, when it is **all put together, it provides an incredibly accurate way to identify users**. The **Electronic Frontier Foundation (EFF)** runs a site '[cover your tracks](#)' that “tests your browser to show how unique your fingerprint is in relation to others the site has tracked.”

### Email tracking

Email tracking **software places an invisible image pixel in your emails that can detect the exact time and date you opened an email**. The reason for email tracking is so that companies/retailers, etc., save time and learn whether a company's first email was interesting enough to you to open. If not, you might be unlikely to open future follow-up emails. By preventing unnecessary follow-up emails, email tracking saves time both for the sales rep and the email recipient. Similarly, if a company notices a contact is clicking on the links sent and viewing a cover letter or a proposal that was attached, the company knows that you are currently at the top of their minds. Reaching out to you at that point, when you are thinking about a company's proposal, say (e.g., purchase an item of clothing) makes the conversation much more relevant, and timely for the company/retailer.

### *Future of tracking and protection of personal data online*

It is not inconceivable that new and innovative means to track users online will emerge from the AdTech industry and Big Tech, though Apple has introduced new transparency rules for its store. Apple requires developers to use predefined privacy labels to disclose what data they use and why. Apple's new iOS [14.5](#) also requires developers to "get the user's permission before tracking their data across apps or websites owned by other companies for advertising, or sharing their data with data brokers." ([Apple, 2021](#)).

According to Apple, its new privacy function allows owners of Apple phones with this operating system to “tap the Privacy Report button to better understand how websites treat your privacy.” ([Apple, 2021](#)). Again, according to Apple, its **AppTrackingTransparency** function (**ATT**) will “require all apps to ask for explicit permission to track” and “Under Settings, users will be able to see which apps have requested permission to track, and make changes as they see fit.” ([Forbes, 2021](#)).

On the Mac OS, ('Big Sur') Apple has provided a privacy report tool that appears as an icon in the Safari browser. This lets users see what trackers are on a web page and being blocked. Apple's Safari browser is “giving you more ways to help protect your privacy” ([Apple, 2021](#)). Apple's privacy drive is a “game changer” according to Kate O'Flaherty ([in Forbes, 2021](#)). In its latest missive to developers, [Apple \(2023\)](#) states clearly:

“You must use the AppTrackingTransparency framework [ATT] if your app collects data about end users and shares it with other companies for purposes of tracking across apps and web sites. The AppTrackingTransparency framework presents an app-tracking authorization request to the user and provides the tracking authorization status.”.

The CSI-COP project is designed as a bottom-up approach to arm citizens with the knowledge to resist the temptation for convenience and reject accepting ‘All cookies’ option on websites. Changing browser settings to stop tracking, or using a privacy-by-design browser, could limit your personal data leaking to first and third-parties without your informed consent. For apps on mobile devices, taking time to check permissions, allowing only those that are necessary for the app to function, could prevent your



data being used unwittingly by others (see Table 3 in Section 3). Step 5 in CSI-COP's free informal education resource '[Your Right to Privacy Online](#)', available in English and twelve translations, provides further information on how to protect your personal data online.



## 5. Privacy by Design

CSI-COP’s taxonomy of cookies and trackers presented here follows the innovation of i) a no-tracking project website and ii) two databases created from CSI-COP citizen scientists’ website and app investigations.

At the top of the project’s internet home page it announces its ‘PRIVACY-BY DESIGN, NO TRACKING WEBSITE’ message (Table 4-left). The bottom of this page presents its cookie notice (Table 4-right):

- CSI-COP does not use any cookies that analyse traffic.

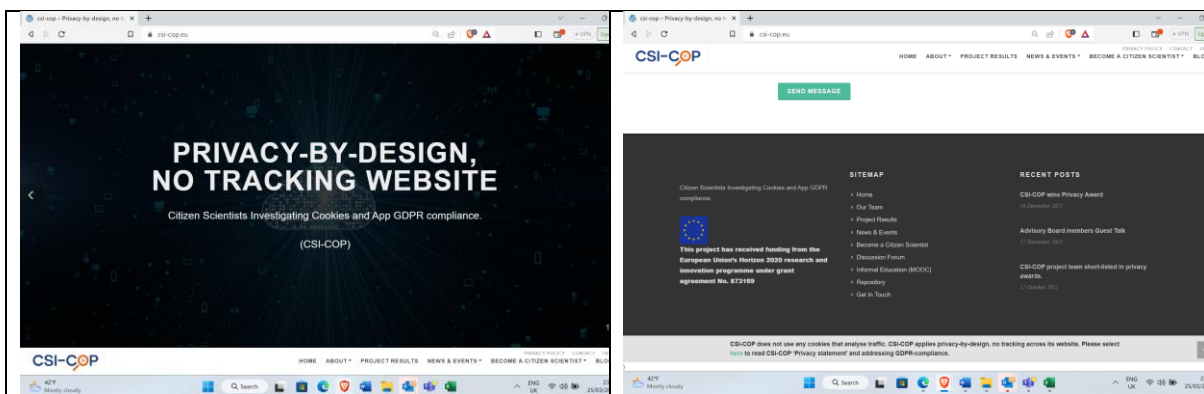


Table 4: Screenshots from top and bottom of CSI-COP project website home page: <https://csi-cop.eu/>

The cookie notice provides a link to CSI-COP Privacy Policy, which is designed to be easily understood without technical or legal knowledge. This sets a ‘privacy as the default’ for the project implementing as well as investigating GDPR compliance in websites and apps.

### Privacy as the default

Ann Cavoukian introduced the term “**privacy by design and by default**”, when she served as the Information and Privacy Commissioner of Ontario, Canada and wrote the seven foundational principles about privacy (Cavoukian, 2010). The **Privacy by Design (PbD)** approach focuses on prevention rather than resolution. It is a proactive rather than a reactive measure that aims to prevent privacy infractions before they happen. In contrast, **Privacy by Default** refers to the establishment of privacy-friendly default settings in technology, processes, and systems. This means that privacy is the default option unless the user chooses to alter the settings.

Barth, Ionita and Hartel (2022) produced comprehensive guidelines for PbD through a systematic review and leveraging invaluable insights of privacy practitioners. Their efforts were encoded in list of fifteen attributes. These attributes are:

1. **Accountability** refers to the capability of holding the service provider responsible for any privacy violations. This can be achieved through various means, such as a legally binding privacy policy, established legal precedents, regulation, etc.



2. **Anonymization** is the process of removing all identifiable markers from data, ensuring that it can never be traced back to a specific individual. This is achieved through methods such as high-level data aggregation. The ultimate goal of anonymization is to protect personal privacy.
3. **Collection** pertains to the identification of the specific data that is being gathered. Examples of collected data may include an IP address, phone number, credit card information, and others. A crucial distinction can be made between Personally Identifiable Information (PII), which relates to an identifiable living individual, and anonymous data. Additionally, different categories of personal data can be further differentiated. The principle of data minimization is also incorporated in the collection process, as it emphasizes the importance of only gathering the minimum amount of data necessary to provide the service.
4. **Control** refers to the requirement for obtaining the data subject's consent for the collection and processing of their data, and the extent to which the data subject can choose to opt-out of these activities. The cornerstone of control is the ability for the data subject to make a self-determined decision about what data to share and for what purpose. It also includes the capacity for the data subject to have an active influence on how their personal data is handled by the service provider. Obtaining informed consent, having the ability to request a copy of one's data, and the user-friendliness of privacy settings are all integral components of control.
5. **Correctness** pertains to the measures put in place to prevent and correct incorrect data. This can be achieved through mechanisms such as data request forms and the ability to edit collected data. The correctness of the data is the responsibility of both the service provider and the user, who should have the capability to correct inaccuracies. Correctness extends beyond control in that it involves the ability of the user to correct data about themselves that is no longer valid, even after it has been disclosed.
6. **Disclosure** deals with the approach of the service provider towards requests from law enforcement agencies. This can include disclosure upon request, only with a warrant, or only after a court order, among others. The attitude of the service provider towards disclosure is a crucial aspect of privacy protection and is closely linked to the jurisdiction of where the data is stored or processed, particularly in regard to data leaving the European Union (EU).
7. **Functionality** refers to the choice faced by the user between utilizing the full capabilities of the service and maintaining their privacy. This can be demonstrated through scenarios such as an application that will not run unless all permissions are accepted, a requirement for real names only, or the need to provide credit card details for a free trial, among others. The functionality of the service is impacted by the artificial restrictions the service provider imposes, which often require the provision of personal data.
8. **Purpose** concerns the utilization of the collected data. This can encompass the provision of the service, advertising, profiling, and more. The purpose of data collection is not limited to these uses and also includes the legal basis for processing, such as meeting legal requirements or serving a vital or public interest.
9. **Pseudonymization** involves the replacement of personally identifiable markers with artificial identifiers, or pseudonyms. This process enables data to be linked back to individual users only with the help of additional information. Examples of pseudonymization include replacing names with numbers, removing house numbers from addresses, or replacing birthdays with birth years, among others.
10. **Retention** refers to the length of time for which collected data is stored. It's an important aspect of privacy as it affects how long sensitive information about an individual can be accessed and potentially misused.



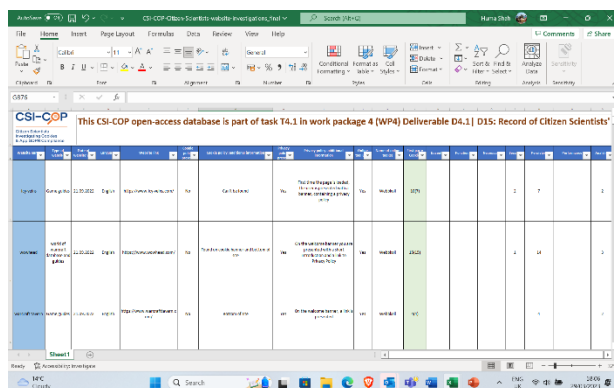


11. **The right to be forgotten** refers to a data subject's ability to request the removal of all their personal data. The implementation of this right can vary, with options ranging from hiding personal data to completely deleting it
12. **Sale** refers to the commercial practice of sharing user data with other organizations for financial gain. This involves the transfer of data from the service provider to a third party for financial compensation.
13. **Security** encompasses the technical measures implemented by the service provider to secure user data from unauthorized or malicious access. This includes the use of encryption, secure protocols, and other techniques to ensure the protection of sensitive information.
14. **Sharing** refers to the transfer of collected data from the ownership of the service provider to other entities, such as companies, advertisers, research institutions, etc. This can happen both intentionally and unintentionally and is sometimes called disclosure. The act of sharing does not involve a monetary exchange.
15. **Transparency** refers to the accessibility of information to users regarding how the service provider is handling their personal data. This can be achieved through various means, such as providing open-source code, availability of privacy policy, undergoing regular audits, etc. Transparency also encompasses the proactive provision of information to users before they give informed consent. Additionally, it encompasses the ability of the service provider to effectively demonstrate to both data subjects and regulators that they are implementing all relevant privacy attributes.

CSI-COP adheres to these privacy-by-design principles in order to build trust with visitors to the project's website, and with the citizen scientists engaged in the project. The citizen scientists' investigations were made through a bespoke tool created in Microsoft Excel by the Coordinating partner. The records collected included name of a website, or app, date of investigation, cookies and tracking found (Screenshot 17). The aim of a CSI-COP customised tool created in-house was to ensure no first or third-party personal data would be collected from an off-the-shelf app or tool. The investigation tool/ Excel spreadsheet can be viewed in the two databases of CSI-COP citizen scientists' investigations:

- a) citizen scientists' website investigations (Screenshot 17): <https://csi-cop.eu/project-results/citizen-scientists-website-investigations/>
- b) citizen scientists' app investigations: <https://csi-cop.eu/project-results/citizen-scientists-app-investigations/>

Screenshot 17: CSI-COP website investigation tool using Excel



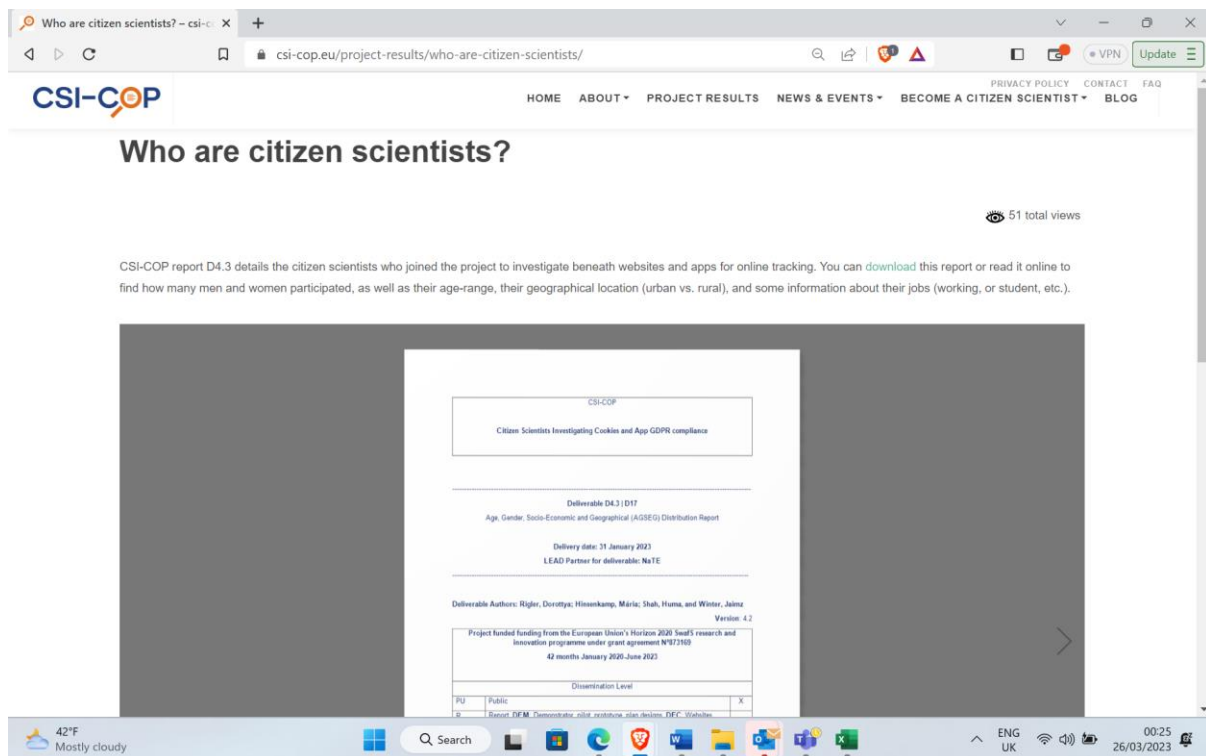
Website	Date	Investigator	Investigation Type	Results	Comments
Google	2023-01-15	John Doe	Website	Found cookies	...
Facebook	2023-01-16	Jane Smith	App	Found tracking	...
Amazon	2023-01-17	Mike Johnson	Website	Found cookies	...

Investigations of CSI-COP citizen scientists and the consortium team members will be detailed in the next project result from work package 5 (WP5): innovation of a repository of tracking found in CSI-COP web and app investigations.



## 6. Further work

To continue with CSI-COP’s privacy as the default approach, a no-tracking counter is featured on some of CSI-COP web pages, for example, in ‘Project Results’. This allows the project team to learn the number of views following the upload of a project report to the website (Screenshot 18). For example, the no-tracking counter on the web page for CSI-COP’s [anonymised report on the project’s citizen scientists](#) published at the end of January 2023 shows 51 total views as at 26.03.23 (Screenshot 18). The number of views assists the CSI-COP partners communication activities to dissemination and exploit project results.



Screenshot 18: CSI-COP web page for ‘Who are Citizen Scientists’ with 51 views as at 26.03.23

CSI-COP’s further work involves statistical analysis of the citizen scientists’ investigations. These will be detailed in scientific publications submitted for peer-review and summarised in the final project report (deliverable D6.6 due at the end of the project). Other work underway is combining the website and app investigations already completed. CSI-COP’s two generated databases are open-access for examination: has a website you visit featured in CSI-COP citizen scientists’ investigations? Does an app you use appear in the apps database? You can explore by selecting the links below to access the databases:

- Website investigations: <https://csi-cop.eu/project-results/citizen-scientists-website-investigations/>
- App investigations: <https://csi-cop.eu/project-results/citizen-scientists-app-investigations/>





The citizen scientists' investigations are being joined with the CSI-COP team members' investigations. The combined CSI-COP website and app investigations will be searchable from a free open-access repository of trackers under development at the time of this deliverable. The repository is due for first open demonstration in May 2023 during the project's main dissemination and exploitation activities in Brussels 23-26 May. Stakeholders (data protection professionals; policymakers; privacy researchers, web and app developers; academics and tech journalists) are invited to attend by first contacting [Dr. Huma Shah](#), leading the science in CSI-COP.



## References

- AdExchanger (2013). Location is the new cookie: here's how to get a bite. Accessed from: <https://www.adexchanger.com/the-sell-sider/location-is-the-new-cookie-heres-how-to-get-a-bite/>
- Apple (2021). iOS 14.5 delivers Unlock iPhone with Apple Watch, more diverse Siri voice options, and new privacy controls. Accessed from: <https://www.apple.com/newsroom/2021/04/ios-14-5-offers-unlock-iphone-with-apple-watch-diverse-siri-voices-and-more/>
- ArsTechnica (2021). Ron Amadeo article: Google delays FLoC roll out until 2023. Accessed from: <https://arstechnica.com/gadgets/2021/06/google-delays-floc-rollout-until-2023/>
- Barth, S., Ionita, D. and Hartel, P. (2022). Understanding Online Privacy—A Systematic Review of Privacy Visualizations and Privacy by Design Guidelines. *ACM Computing Surveys*, 55(3), pp.1–37. DOI: <https://doi.org/10.1145/3502288>
- BBC Bitesize (no date). What is the world wide web? Accessed from: <https://www.bbc.co.uk/bitesize/topics/zs7s4wx/articles/z2nbgk7>
- Behera, R. (2023). What are cookies? Different types of cookies Explained. AdPushUp Accessed from: <https://www.adpushup.com/blog/types-of-cookies/>
- Bloomreach (2023). The importance of Zero-party Data. Accessed from: <https://www.bloomreach.com/en/blog/2021/importance-of-zero-party-data>
- Chromium blog (2020). Building a more private web: A path towards making third party cookies obsolete. Accessed from: <https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html>
- CNN (2023). The US government is once again threatening to ban TikTok: What you should know. CNN Business. Accessed from: <https://edition.cnn.com/2023/03/18/tech/tiktok-ban-explainer/index.html>
- Cookie Law Information (2023). Cookies on Phone: How they Affect Mobile Security? Accessed from: <https://www.cookielawinfo.com/cookies-on-phone/>
- Cavoukian, A. (2010). Privacy by design - The 7 foundational principles. Accessible from: [https://iapp.org/media/pdf/resource\\_center/pbd\\_implement\\_7found\\_principles.pdf](https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf).
- Digital Shift (2023). What are cookies? Accessed from: <https://digitalshiftmedia.com/marketing-term/cookie/>
- Digital Trends (2015). Simon Hill article: Are cookies crumbling our privacy? We asked an expert to find out. Accessed from: <https://bit.ly/3Jjgwa4>
- Dutko, J. (2018). How Types of Computer Cookies Affect your Online Privacy. CruSolutions. Accessed from: <https://crusolutions.com/blog/how-types-of-computer-cookies-affect-your-online-privacy/>
- EFF (no date). The Electronic Frontier Foundation. The leading non-profit defending digital privacy, free speech, and innovation [for 30 years and counting!](https://www.eff.org/) Accessible from here: <https://www.eff.org/>



EFF (2021). B. Cyphers' article: Google's FLoC (Federated Learning of Cohorts) is a terrible idea. Electronic Frontier Foundation. Accessible from here:

<https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea>

EU (no date). European Data Protection Supervisor. Accessed from: [https://edps.europa.eu/data-protection/our-work/subjects/eprivacy-directive\\_en](https://edps.europa.eu/data-protection/our-work/subjects/eprivacy-directive_en)

EU (2023). EU Strengthens cybersecurity and suspends the use of TikTok on its corporate devices. EU press release. Accessed from: [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_23\\_1161](https://ec.europa.eu/commission/presscorner/detail/en/ip_23_1161)

Exodus Privacy (2023). Analyzes privacy concerns in Android applications. Accessed from: <https://exodus-privacy.eu.org/en/>

Forbes (2020). Will Google's Plan to Phase Out Cookies Choke Digital Advertising. Forbes: Leadership. Accessed from: <https://bit.ly/3n5Xo88>

Forbes (2021). Kate O'Flaherty article: Apple's Stunning iOS14 Privacy Move: a game-changer for all iPhone Users. Accessible from here: <https://bit.ly/3vpOq4v/>

Gemius (2023). Knowledge that supports business decisions. Accessed from: <https://www.gemius.com/homepage.html>

GDPR.EU (2023). Cookies, the GDPR and the ePrivacy Directive. Accessed from: <https://gdpr.eu/cookies/>

Google (no date). Google analytics Terms of Service. Google Marketing Platform. Accessed from: <https://marketingplatform.google.com/about/analytics/terms/us/>

History of Information (no date). Luis Montulli II invents the HTTP Cookie. Exploring the History of Information and Media through Timelines. Accessible from: <https://www.historyofinformation.com/detail.php?id=2102>

INPLP (2022). Stephan Winklbauer article: Use of Google Analytics Violates the GDPR – Recent Decision of the Austrian Data Protection Authority. International Network of Privacy Law Professionals (INPLP). <https://inplp.com/latest-news/article/use-of-google-analytics-violates-the-gdpr-recent-decision-of-the-austrian-data-protection-authority/>

Jones, M.L (2020). Cookies: A Legacy of Controversy. *Internet Histories: Digital Technology, Culture and Society*. Volume 4, issue 1: Legacy Systems: Internet histories of the abandoned, discontinued and forgotten. Accessed from: <https://www.tandfonline.com/doi/full/10.1080/24701475.2020.1725852>

Marketing labs (2023). How to deliver personalisation in a cookie-less world. Accessed from: <https://marketinglabs.co.uk/how-to-deliver-personalisation-in-a-cookie-less-world/>

Microsoft (2023). Microsoft Privacy Statement. Accessed from: <https://privacy.microsoft.com/en-us/privacystatement>

PageXray by Fou Analytics (no date). Free online web privacy audit tool: <https://pagexray.fouanalytics.com/>



Privacy Matters (2021). Co-creator 'Your Right to Privacy Online'. CSI-COP free informal education resource (MOOC). Accessible from: <https://csi-cop.eu/informal-education-mooc/>

Shah, H., and Winter, J. (2022). CSI-COP Project Report Accompanying Website Database. Deliverable D4.1|D15. Accessed from: <https://csi-cop.eu/project-results/citizen-scientists-website-investigations/>

Sivan-Sevilla, I., Chu, W., Liang, X. and Nissenbaum, H. (2020). Unaccounted Privacy Violation: A Comparative Analysis of Persistent Identification of Users Across Social Contexts. Federal Trade Commission (FTC) PrivacyCon 2020. Paper available online via this link: <https://news.cornell.edu/stories/2020/06/study-online-trackers-follow-health-site-visitors>

Srinivasan, D. (2020). Why Google Dominates Advertising Markets Competition Policy Could Lean on the Principles of Financial Market Regulation. 24 STAN. TECH. LAW REV. Accessed from here: <https://law.stanford.edu/publications/why-google-dominates-advertising-markets/>

Statcounter (2023). Desktop browser market share worldwide: Feb 2022-Feb 2024. Accessed from: <https://gs.statcounter.com/browser-market-share/desktop/worldwide>

Target Internet (n.d.) The Digital Marketing Guide to Web Cookies. Accessed from: <https://www.targetinternet.com/resources/the-digital-marketing-guide-to-web-cookies>

TechCrunch (2021). Natasha Lomas article: Answers being sought from Facebook over latest data breach. TechCrunch Media and Entertainment. Accessed from here: <https://tcrn.ch/3xfrTsE>

TechRadar (2022). Sead Fadilpašić article: DuckDuckGo rolls out new Microsoft Blockers after Backlash. Accessed from: <https://www.techradar.com/news/duckduckgo-rolls-out-new-microsoft-blockers-after-backlash>

UK (2023). TikTok banned on UK government devices as part of wider app review. Gov.UK. Accessed from: <https://www.gov.uk/government/news/tiktok-banned-on-uk-government-devices-as-part-of-wider-app-review>

Webbkoll (no date). Free online web privacy audit tool: <https://webbkoll.dataskydd.net/>

