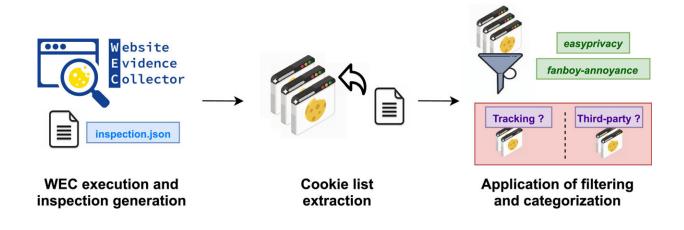# The most visited websites do not comply correctly with privacy laws and actively track their users, finds Spanish study

March 9 2023

Steps of the cookies detector algorithm. Credit: *Computers & Security* (2022). DOI: 10.1016/j.cose.2022.102873

Only a small percentage of the 500 most visited websites in Spain (which include everything from government sites to streaming and adult content platforms) correctly fulfill the requirements set out in the General Data Protection Regulation (GDPR). This is one of the main findings of a study involving researchers from the Universitat Oberta de Catalunya (UOC), the University of Girona and the Center for Cybersecurity Research of Catalonia (CYBERCAT).

The results, which are published in *Computers & Security* under a

Creative Commons license, were reached using novel automated methods for analyzing web-tracking techniques and compliance with internet privacy regulations.

In addition to the incorrect and non-consensual use of cookies, these analysis algorithms detected the use of web-tracking techniques that are little known to the average user, such as web beacons and technologies based on the browser's digital fingerprint.

Widespread non-compliance with privacy laws

The European Parliament's approval of the General Data Protection Regulation in 2016 was set to forever change how companies, websites and digital platforms manage users' personal data. The European regulation, which was transposed in Spain as the Organic Law on the Protection of Personal Data and Guarantee of Digital Rights in 2018, was supposed to mark a turning point in the protection of citizens' privacy. However, six years later, the actual implementation of this regulation is progressing at a faltering pace.

"We found that websites still have a long way to go to correctly implement the requirements set out in the General Data Protection Regulation," explained Cristina Pérez-Solà, who took part in analyzing this issue as a researcher at the UOC's Faculty of Computer Science, Multimedia and Telecommunications. She said, "Many of the websites analyzed inform users of the use of cookies, but either do not wait for their consent to use them or acquire this consent improperly."

For this study, the team of researchers developed several algorithms to analyze the 500 most visited websites in Spain according to the Alexa ranking. The results revealed a high percentage of sites that lack an appropriate form to obtain users' consent for the use of cookies and other data collection tools.

The [analysis tools](#) also detected the use of nearly 7 tracking cookies on average per website and 11 web beacons, which are small pieces of code embedded in the site to invisibly collect certain types of information from web traffic. In addition, 10% of the sites analyzed in the study use browser fingerprinting techniques, which are also difficult to detect.

According to Pérez-Solà, an expert in web security and privacy, "The purpose of all these techniques is usually to track the online behavior of web users in order to create profiles that can then be used to adjust the advertising that will be shown or the prices that will be offered for services or products." The analysis carried out by the researchers from the UOC (Pérez-Solà and Albert Jové) and the University of Girona (David Martínez and Eusebi Calle) shows that only 8.91% of websites that obtain users' consent as required apply this consent successfully in practice.

New algorithms to analyze compliance with the GDPR

Beyond the analysis results, the importance of this research lies in the algorithms used to study compliance with online privacy laws. The sheer number of pages and platforms on the internet makes it imperative to automate the process, as studying each case manually would be infeasible.

Additionally, some of the web-tracking techniques used are extremely hard to detect, with no clear markers to indicate their presence. To overcome these challenges, the researchers developed a proprietary method involving four algorithms and a measure—the Websites Level of Confidence—to assess the state of regulatory compliance.

"Our method uses a combination of automation and manual inspection. The implemented algorithms automatically browse the analyzed websites and take screenshots that are then manually inspected," said Pérez-Solà.

"In order to detect web-tracking techniques, we also used a tool developed by the European Data Protection Supervisor called the Website Evidence Collector. This tool is designed to perform privacy inspections on websites and makes it possible to detect the use of cookies, web beacons and browser fingerprinting tools."

- Each of the algorithms used by the researchers has a well-defined function:
- The Consent Inspector Algorithm (CIA) captures clear images of the website's cookie banners and identifies buttons that should allow users to customize the use of these tracking elements.
- The Website Evidence Collector (WEC) gathers information on the different web-tracking techniques being used on each website.
- The Cookies Detector Algorithm (CDA) categorizes the cookies that websites use in the browsers without user consent, based on the data provided by the WEC.
- The Web Beacons Detection Algorithm (BDA) not only extracts web beacons detected by the WEC, but also identifies browser fingerprinting techniques.

"Our study focuses on analyzing compliance with the General Data Protection Regulation by the most visited websites in Spain," Pérez-Solà added. "We selected the 500 most visited websites according to the Alexa ranking and analyzed their use of these web-tracking techniques as well as the information they give to users and the alternative options they provide them with. Finally, we compiled the results of this analysis into a measure, the Websites Level of Confidence, which makes it possible to assess the current state of compliance."

"Understanding the details of the regulations that apply at any given time and knowing how to tell what techniques a website is using are beyond the grasp of most users," she concluded; "Our proposed Websites Level

of Confidence (WLoC) measure provides users with insight into the compliance status of the most popular websites and lets them see how it changes over time without the need for legal or technical knowledge."

More information: David Martínez et al, Web-tracking compliance: websites' level of confidence in the use of information-gathering technologies, *Computers & Security* (2022). [DOI: 10.1016/j.cose.2022.102873](#)

Provided by Universitat Oberta de Catalunya