# Understanding the Effect of GDPR-like Policies for Cookies in Web Applications

Mike Zeng, Frank Forliano
mzeng5@uiowa.edu, frank-forliano@uiowa.edu

## Abstract

*Majority of websites use cookies, as a direct consequence of needing data. In the past, they have caused many violations to privacy laws like GDPR or CCPA. The EU has made many lawsuits to companies for breaking GDPR laws, and by using a web scraper, we found out that these policies and enforcements have been working sufficiently well at reducing cookie usage in web applications, and causing the overall amount of cookies that load initially to decrease as compared to places that do not have such GDPR-like policies.*

## 1. Introduction

With conflict between many privacy-focused laws such as GDPR/CCPA and the uses of cookies, it is very important to understand how cookie use is being affected by laws, and if the laws should be stricter or lesser on cookies.

Cookies are generally persistent pieces of data stored by web browsers so that reloads, navigation, does not reset the website state. For example, logins would reload the page, and then without persistent data, it would not know that it logged in. There are two kinds of parties, first and third party cookies. First party cookies are planted by the website itself to help it function, and third parties cookies are planted by organizations that are not part of the website's organization, that provide services to it, which needs the cookies to run.

An example of first-party cookies is the notion of a shopping cart on a website. When the user hits the add shopping cart button  website a request is sent to the server, in addition to the server's response, there will be a bit of data that is stored in the cache of the browser from the client (in this case a request to buy some items), the server then will store this, and as the client goes through different pages of the website, due to the first-party cookies that state will be maintained. Once the user checks out, there should be some expiration date for the cookie, and it.

An example of third-party cookies are online ad trackers. When you visit a website with ads, a request goes to a different server to grab the ad content, and a small piece of data, a third-party cookie, comes back with it and gets stored in your browser. This cookie tracks what you do on different sites within the same ad network, creating a sort of user profile. This info helps personalize ads based on your interests. As you move across websites, the third-party cookie sticks around, making sure you see ads that match your preferences. These cookies might have an expiration date, and you can usually control or block them for privacy. In a nutshell, third-party cookies play a key role in tailoring online ads to your liking and following your online journey across various sites.

However, third party cookies have run into many problems with privacy focused laws, such as GDPR (General Data Protection Regulation), CCPA (California Consumer Privacy Act). These laws, particularly the GDPR, have heavy implications on the use of cookies. For example, Google Analytics, one of the most commonly used Analytics, had to rebuild their analytics system so it used less cookies. Furthermore, many companies have been fined, such as Meta and Amazon, providing them incentives to use less tracking cookies.

In the 88 page long document that GDPR is, there is only one paragraph where it discusses cookies, it states: **"Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them".**

The implications of this means that cookies that identify your personal data are protected under GDPR. Companies have a right to this data if they receive consent, or have a legitimate interest in the data. The regulations regarding cookies to be compliant under the GDPR guidelines are stated as:

- You must obtain the users' consent before you use any cookies **except** strictly necessary cookies
- Document and store consent received from users.

- Allow users to access your service even if they refuse to allow the use of certain cookies
- Make it as easy for users to withdraw their consent as it was for them to give their consent in the first place.
- Provide accurate and specific information about the data each cookie tracks and its purpose in plain language before consent is received.

The central goal of this paper will be to now see how these policies have impacted the internet. While Europe is a big part of the world, its total population is not even a billion. Other countries like China, India, Russia, Brazil, and the United States all have quite a large population and still use the internet daily. Our plan is to see how the cookies differ between countries using web scraping technology Selenium, and the Python programming language.

## 2 Design and Process

To answer our questions, we constructed a web scraper using Selenium and Python. The selenium constructs a working browser simulator using chrome and a chrome webdriver, and simulates the loading of a page. The chrome driver was run in a headless state, though this should not affect initial cookie loading. Once the page has loaded, it collects all cookies that have been placed, while also looking for any. It repeats this twice, one for a list of websites that the GDPR applies to, EU countries, and one other which it does not.

In our case, the scraper gathered cookies from 40 websites, 20 in non-EU countries and 20 in the EU. For each website, it adds the number of cookies to the cookie counter for that country (EU or not EU), and at the end of the process, it returns both the counters.

While we believe that this web scraper was optimal for balancing accuracy and our resources, there were some limitations to it. For example, certain web applications, such as YouTube, load ads initially. However, after logging in, there is a significantly larger amount of cookies. Another example may be a biased URL list. If for example, we chose only nonprofit organization websites, these will have a significantly less amount of cookies due to less marketing overall. These impact the significance of our results, because it may not accurately represent the true nature of websites that happens during user experience. In addition, if we use a biased URL list, then it will follow that it does not accurately represent the websites that a user visits.

To minimize the effect of these limitations, we were careful to keep our URL list unbiased, basing it off the most popular websites to accurately represent a consumer. In addition, we have provided a variety of websites from different countries in order to potentially see if any trends exist. We included websites with .br, .kr, ru, cn, etc. top level domain extensions. Another thing we considered was including a variety of different websites, if we only considered news article websites, or video streaming websites, we would have more biased data.

In addition, we believe that because this is to look at the difference between EU and US websites, it is okay to count necessary cookies because they are going to end up on both sides of the equation in a comparison.

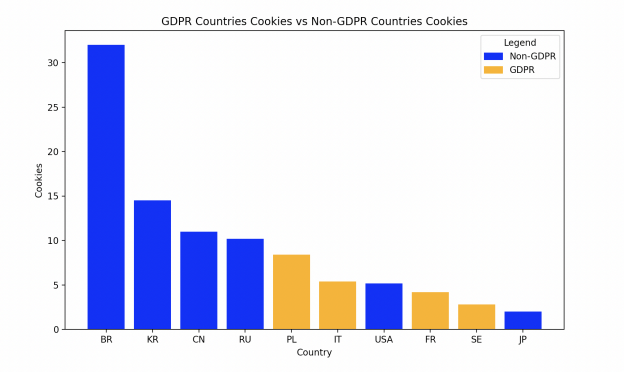
Other methods of collecting cookie data have been considered. One method that we considered was to scrap sites and look for third party scripts. This is an extremely accurate way to look for cookies, as if something did have a script, it is almost guaranteed to collect cookies. This also ensures that emerging cookies, that is, cookies that appear as you continue to use the site would be counted.

However, we did not choose this method because it does not account for the fact that a site may collect more than others, as this can be specified in analytic settings. Instead, we chose our method because it considers how many cookies are used in general, weighting sites that use a lot of unnecessary cookies more, and assuming that our URL list is equal, the more cookies on a web application that does the same thing means more is being collected, so it means that it is more likely to interfere with GDPR.

Overall, while there were some limitations in our method of gathering data, we believe that it was the best and most accurate method of capturing data with the resources we had.

## 3. Results and Analysis

All 20/20 EU websites responded, however, only 19/20 responded and got cookies. However, despite that, the total amount of cookies from the non-EU countries was 180, and the total amount from the EU was 108, which is roughly about 40% less.




A graph representing cookie differences across countries.

This is a significant amount less, and shows that there are overall less cookies in the EU, the GDPR-affected region, then the region that is not GDPR. Subtracting them off, it was found that despite having more working links, the EU still had less cookies. To ensure that it was not mere luck of the websites, we ran this several times with a subsection of 5 of each URL using our data. We noted that if the highs of the EU list were chosen while the lows of the normal list were chosen, it was possible for the cookie count of the EU list to be higher than the normal list, but after running it for 5 times, the EU cookie amount was never noted. We noted the result for various sample sizes of 5, 10, and 15.

Number of Cookies for Sites based on Sample Sizes

| | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| EU Cookies | 23.5 | 57.6 | 77.3 | 108 |
| Non-EU Cookies | 54.2 | 93.2 | 132.3 | 180 |

A table representing the number of cookies based on sample sizes.

We deduced that on average with random sets of various sizes, the EU has a significant lower amount of cookies than non-EU sites.

Finding that there is a lower amount of cookies on EU sites, we came to the conclusion that the EU's GDPR policy has significantly lowered the amount of cookies that are used upon websites in the European Union. What this implies, is that as time passes and more measurements are done, it is likely that the amount of cookies will decrease. As more countries adopt GDPR-like policies or the EU enforces GDPR internationally, many web applications will likely decrease in the amount of cookies that are used. Does this necessarily mean less data is being collected? Most likely not, as demand in data analytics platforms such as Google Analytics have only increased, but what this implies is that data collection in the future is most likely going to use less cookies. This data supports

## 4 Conclusion

Our findings have been very interesting, and it seems to be accurate towards our predictions. Countries like Brazil can have websites globo.com, which have over 30 cookies as you run it, whereas websites in Sweden such as their daily news services have 0 cookies, unless you accept them. Of the countries not in the GDPR, the USA has the lowest cookies but this could be because the EU actually has sued them over privacy invading cookies before. If these types of policies start to increase throughout the world, we can expect a web where cookies have to be much more secure, and purposeful than ever. The lower cookies found in the GDPR countries show that it has been very effective in reducing the amount of cookies, and it likely improves the privacy of the users. For the foreseeable future though, this is only the beginning. Europe is moving towards a new ePrivacy Directive which is planning to build on GRP, and many other countries will likely follow in their footsteps, and we will hopefully have a more secure web for everyone.