

Changes in Third-Party Content on European News Websites after GDPR

Authors: Timothy Libert, Lucas Graves and Rasmus Kleis Nielsen

Introduction

This factsheet compares the prevalence of third-party web content and cookies on a selection of European news websites one month before and one month after the introduction of the EU General Data Protection Regulation (GDPR). In order to begin to understand how news organisations may be adapting to the new privacy framework, prominent news websites in seven countries (Finland, France, Germany, Italy, Poland, Spain, and the UK) were analysed during the months of April and July 2018 using a purpose-built software tool, webXray, which traces the network of outside parties that load content — and potentially track users — on a given web site.

This study builds on an earlier report, *Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement*, which compared third-party content on news sites with other popular websites during the first three months of 2018. Our prior measurements revealed that news websites tend to have dramatically higher volumes of third-party content and cookies than other popular websites. In advance of GDPR we identified several steps news websites could take to improve user privacy, such as migrating third-party services (like social media sharing tools) to function on a first-party basis.

The current investigation finds that news websites in most countries studied are setting substantially fewer cookies without user consent post GDPR. Based on analysis of homepages from over 200 news sites loaded with the webXray software

platform in April and July, a total of 10,168 page loads, nearly 1 million content requests, and 2.7 million cookies were captured and analysed. On this basis, we find that the overall number of third-party cookies on news sites is down 22%, including significant drops in advertising and marketing (14%) and social media (9%) cookies, and a seven percentage point drop in the number of news sites that host third-party social media content, such as sharing buttons from Facebook or Twitter. (All results presented here reflect site activity prior to obtaining consent; the picture may change dramatically once the user provides the affirmative opt-in GDPR requires.) These changes suggest that some news organisations are responding to GDPR either by obtaining consent for third-party tracking or by curbing the use of outside cookies in general.

While there is no change in the overall percentage of pages from news providers which contain some form of third-party content (99%) or third-party cookies (98%) and the average number of third-parties found on each news site has remained fairly static, we find a number of specific changes between April (pre-GDPR) and July (post-GDPR):

- The number of third-party cookies per page dropped by 22% across all news sites. German news sites, which had the second-lowest number of cookies in April, exhibit the smallest change with 6% fewer cookies. UK news sites, which had the most cookies per page in April, had 45% fewer by July.

- The percentage of news sites hosting third-party social media content, such as sharing buttons from Facebook or Twitter, dropped significantly, from 84% in April to 77% in July.
- Decreases in third-party cookies between April and July vary by the type of content setting the cookie. Across our sample, on average, the number of cookies from design optimisation tools is down 27%, advertising and marketing cookies down 14%, and social media cookies down 9%.
- The US-based technology companies Google (96%), Facebook (70%), and Amazon (57%) remain present on the highest number of the news sites in our sample; of these, only Facebook has seen a significant drop in reach after GDPR (down five percentage points). But most of the other companies with the widest presence in April have seen significant drops in their post-GDPR reach, in many cases of ten percentage points or more.

Background: Third-Party Web Content

Modern websites are rarely self-contained and typically include a mix of first- and third-party content. First-party content is defined as material downloaded from the address a user sees in their browser window, which is typically the outlet or organisation the user intends to visit. For example, a website at the address 'http://example.com' may contain a first-party image downloaded from 'http://example.com/newsimage.jpeg'.

In contrast, third-party content is downloaded from a different address, and in many cases a different company, than the website a user is visiting. If 'http://example.com' includes a video which is hosted at the address 'http://video-hosting.com/newsvideo.mp4', this means 'video-hosting.com' is a third-party content host, and 'newsvideo.mp4' is third-party content. When third-party content is present on a website, data about the user's browsing habits may be transferred to third parties, representing a potential impact on privacy.

Websites use third-party content for a variety of purposes. Advertising and marketing are the best known, and such content often relies on 'tracking' users' web browsing in order to display advertisements tailored to users' interests. Websites also use third parties to assist with a variety of other functions such as measuring the size and nature of the audience, optimising website design, facilitating social media

sharing, recommending related articles, and hosting content such as videos. Depending on how data is collected and used, such services may fall under the GDPR.

Although the GDPR is a new law, it is an evolution of well-established approaches to privacy regulation which give users control over how personal information is collected and used. The GDPR places substantive limits on how various categories of data may be processed and requires affirmative 'opt-in' from users in many cases, and can result in heavy fines for those who violate the law. Therefore, websites and third parties which collect user data for purposes such as tailoring advertising may be required to obtain user consent before processing data collected on a given website.

Website Selection and Third-Party Content Measurement

Seven countries were selected for this study to represent a mix of population sizes and media markets in the EU. The countries studied are Finland, France, Germany, Italy, Poland, Spain, and the United Kingdom. In each country, we included a selection of prominent news sites, chosen on the basis of prior work measuring their reach and significance (Newman et al., 2017). Please see the methods appendix for additional details on the site selection.

To measure the presence and nature of third-party content on the selected websites, we used the open-source software tool webXray.¹ This tool analyses a page by opening it in the Chrome web browser and creating a new user profile which has no cookies or history. The software then loads the page in Chrome, during which time all requests for third-party content are monitored. At no point is the browser interacted with in any way, and no cookie or tracking consent buttons are clicked. After waiting 30 seconds, webXray extracts all cookies from the internal Chrome database, records them, and closes the browser. webXray may miss some content requests due to a variety of factors, such as websites which block automated browsers. The measures produced by webXray are therefore low-bound measures and the true number of third parties on a given page may be higher.

Two sets of data are compared in this factsheet, drawn from measurements taken during the months of April and July 2018. These months are chosen to

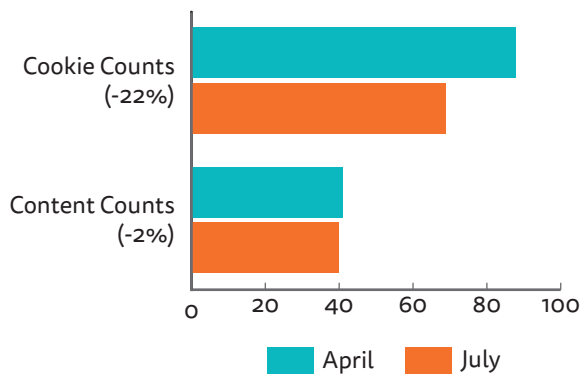
¹ Those who wish to know more about webXray may visit the project website (<https://webxray.org>) and download the software (<https://github.com/timlib/>).

represent samples one month before and one month after the introduction of the GDPR in the EU. For more details on the sample and methods, please see the appendix.

Top-Level Findings

In our previous factsheet, based on data collected over the first three months of 2018, we found that 99% of news websites included some form of third-party content and 99% set at least one third-party cookie (Libert and Nielsen, 2018). These broad measures are unchanged between April and July — the use of third-party content and cookies remains effectively universal after GDPR. However, digging deeper into the data reveals a number of significant changes.

Figure 1. Third-party domains and cookies per page (April-July change in parenthesis)



Per Figure 1, the average number of cookies is down 22% from April to July, though the number of third parties found on a given page load fell from 41 to 40 in aggregate, a marginal change at best.

However, on a per-country basis, shown in Figure 2,

we find wide variation. Five of the seven countries have experienced a drop in the average number of third parties on news sites. The figure fell by 16% in France, 13% in the United Kingdom, 12% in Spain, 8% in Finland, and 4% in Italy; Germany, where news sites in April already had far fewer third-party domains than most countries covered, experienced no change. This suggests that while adjustments are happening in some countries, such changes are uneven and may reflect varying interpretations of GDPR.

The average number of third-party cookies per page has dropped by 22% over all per Figure 1, with wide variation across countries per Figure 3. German news sites, which had the second-lowest number of cookies in April, exhibit the smallest change with 6% fewer cookies in July. UK news sites, which had the most cookies in April, have 45% fewer cookies per page in July, placing them in fourth place among the seven countries examined. Spain, France, and Italy all have over 30% fewer cookies. Once again it is important to emphasise that these are cookies set without clicking on any cookie notifications; users who accept cookies will likely have more set.

While nearly all countries have either stayed the same or experienced drops, Poland has experienced a 29% rise in third parties per page and a 20% rise in third-party cookies in aggregate. This is largely due to major increases in four of the 29 websites examined. We may not rule out that these sites may have changed in a way which has impacted our measurement tool, and when excluding these four sites we find the average number of cookies across the 25 remaining sites to be static, hewing more closely to trends in other countries.

At present it is not possible to state with certainty why the changes we have observed are happening, and some shifts may be unrelated to GDPR. However, it is

Figure 2. Third-party domains per page by country (April-July change in parenthesis)

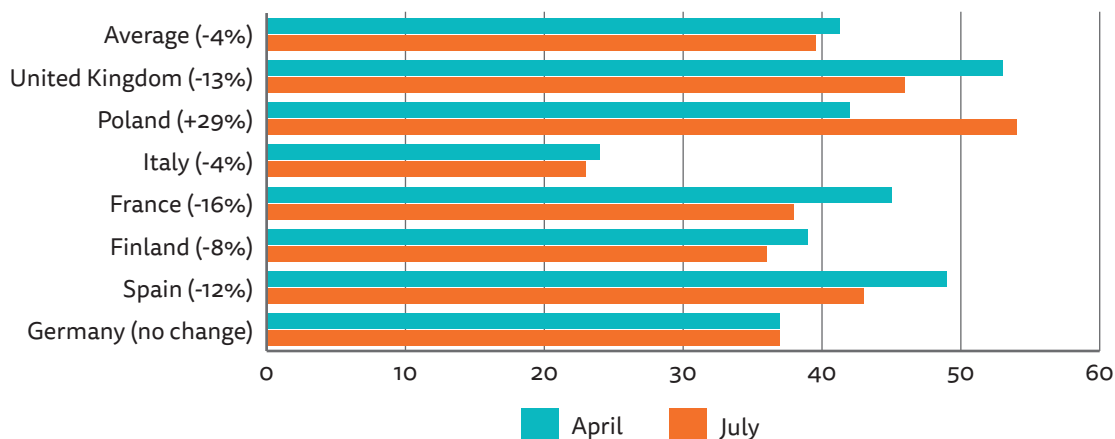
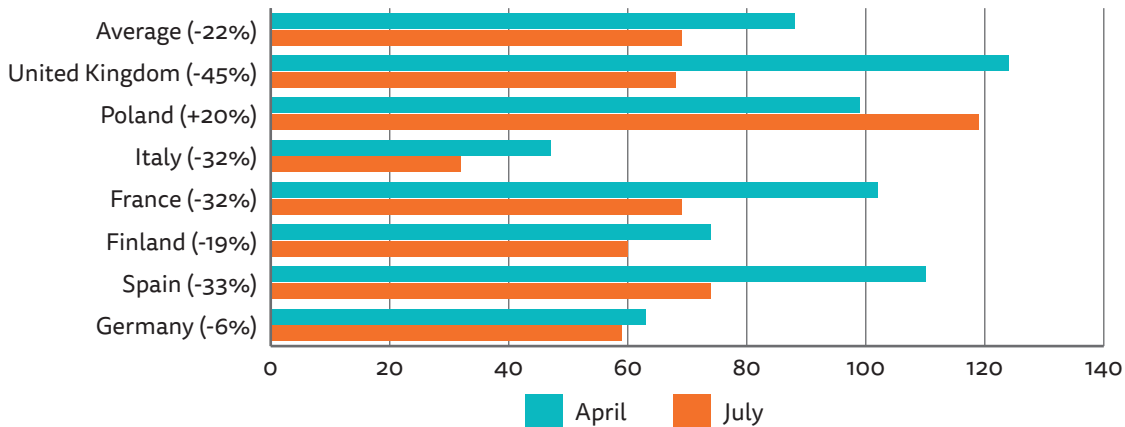


Figure 3. Third-party cookies per page by country (April-July change in parenthesis)

worth noting at least two likely factors. First, due to the GDPR's requirements for consent, news organisations may simply be deferring some tracking cookies until after a user clicks to accept the site's terms on a pop-up consent dialogue. This could also mean that, depending on the preferences of a given user, the number of cookies ultimately set may be similar — but is then based on affirmative opt-in, as required for many kinds of data collection and processing under GDPR.

Second, we may be observing a kind of "housecleaning" effect. Modern websites are highly complex and evolve over time in a path-dependent way, sometimes accumulating out-of-date features and code. The introduction of GDPR may have provided news organisations with a chance to evaluate the utility of various features, including third-party services, and to remove code which is no longer of significant use or which compromises user privacy. A closer look at the types of content being removed provides some insight into this factor.

Types of Third-Party Content

As noted above, third-party content is used for different purposes, and the top-level findings obscure more fine-grained shifts within common categories relied on in the news industry. Here again it is useful to distinguish between the presence and the prevalence of different kinds of content. As shown in Figure 4, we saw almost no change in the percentages of pages with at least one instance of third-party advertising, audience measurement, content recommendation, design optimisation, and hosting.

The only sizable shift was a fall of 7 percentage points (8%) in the share of sites with any third-party social media content — in other words, many news sites don't include even a single instance of content loaded from a social media firm — and a 6% decline in the use of third-party content recommendation systems. This change is in line with our prior report, which suggested removing third-party social media content as one possible step to reduce potential issues with GDPR compliance.

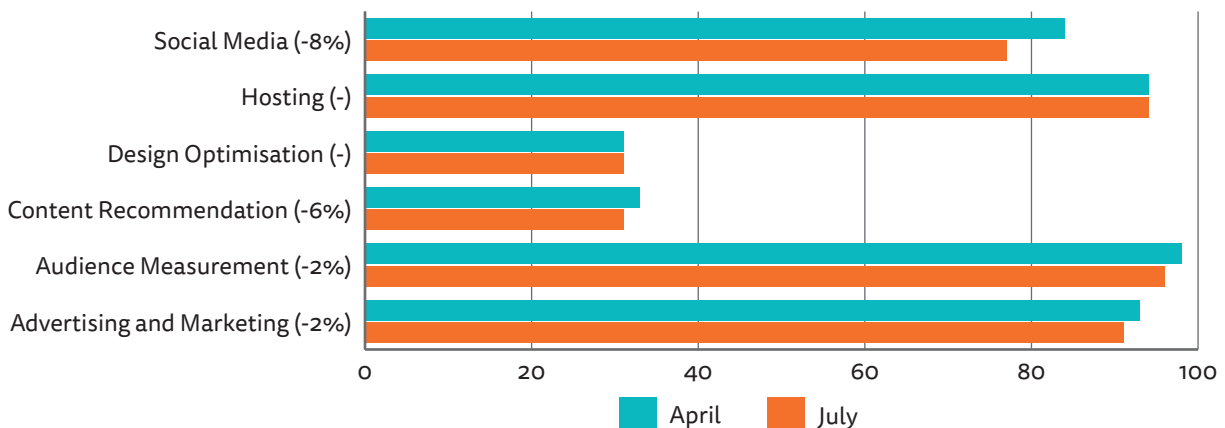
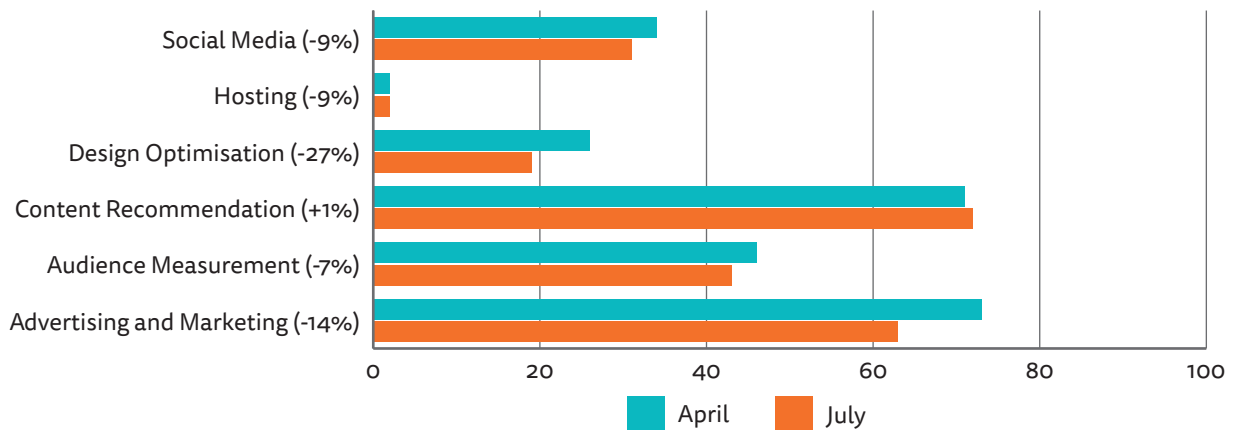
Figure 4. Content cookie use (April-July change in parenthesis)

Figure 5. Content cookie use (April-July change in parenthesis)

Even though the types of content present have not changed dramatically overall, many types of content are now setting cookies without user consent at lower rates. As shown in Figure 5, the percentage of instances of content which set a cookie has decreased for four of six types of content, with advertising and marketing cookies down by 14%, design optimisation down by 27%, and social media down by 9%.

Prevalence of Specific Companies

The software used for this study, webXray, is able to identify nearly 500 different companies and services associated with third-party content. In both April and July, Google, Facebook, and Amazon have the widest presence across the European news sites analysed.

Google, which was present on 97% of all the news sites covered in April and 96% in July, hosts a variety of services, and as Table 1 shows, has seen some slight decreases in reach when looking at specific services.

Table 1. Percentage of sites with content from Google subsidiaries and services

Service	April	July
DoubleClick	89	87
Google Analytics	88	86
Google Tag Manager	82	80
AdSense	79	72
Google APIs	70	69
YouTube	13	11
Google App Engine	4	4

Beyond Google, the overall reach of many companies

appears to have declined, as can be seen from comparing Table 2 (showing the companies which tracked the highest percentage of pages in April) and Table 3 (with the same data for July).

Table 2. Percentage of news sites with content from company, April 2018

Company	Owner country	Percent pages tracked
Google (Alphabet)	US	97
Facebook	US	75
Amazon	US	59
Oath (Verizon)	US	57
AppNexus	US	56
Rubicon Project	US	56
Oracle	US	53
AdForm	DK	53
comScore	US	51
WPP	UK	50

Table 3. Percentage of news sites with content from company, July 2018

Owner	Owner country	Percent pages tracked
Google (Alphabet)	US	96
Facebook	US	70
Amazon	US	57
comScore	US	55
AppNexus	US	50
Oath (Verizon)	US	44
Rubicon Project	US	44
AdForm	DK	39
The Trade Desk	US	37
Criteo	FR	35

Among the top three, only Facebook has seen a significant drop in reach, but the reach of many other companies has declined more substantially. In April all of the top ten companies tracked at least 50% of pages, whereas in July only five companies do so. It is still the case that 8 of the top 10 companies are US-based, though Oath, AppNexus, Rubicon Project, and Oracle (as well as the Danish advertising technology company AdForm and the UK-based WPP) have all fallen from above to below the 50% mark.

As noted above, the prevalence of third-party social media services on news sites has fallen substantially. A closer examination shows that declines vary between social media services. Facebook's presence across news pages examined has dropped from 75% to 70%, Twitter's from 31% to 29%, and that of AddThis from 20% to 10%. The drop in AddThis usage has had a particularly strong effect on parent company Oracle, which has dropped from 53% to 32% of sites tracked overall.

Conclusion

In our prior factsheet we noted that many websites in Europe contained large volumes of third-party content and cookies which could cause potential issues with GDPR compliance and more broadly raise privacy concerns. Likewise, we found that news websites could face significantly greater challenges under GDPR than other popular websites because of their heavy reliance on third parties. We recommended migrating some third-party content to function on a first-party basis, suggesting for example that social media content be prioritised for migration.

In this factsheet we compare third-party content and cookies in pre-GDPR (April) and post-GDPR (July) and find many changes have occurred during this time period. While outside content and cookies are still found on virtually all news sites, we see somewhat less third-party content and significantly fewer third-party cookies, with large variation by country and the largest drops in the UK. Some of the biggest declines have occurred in advertising and marketing cookies as well as social media content, indicating that news sites may have recognised the potential compliance risks posed by some of this content and removed it or tied it to affirmative opt-in from users. In sum, we find that the introduction of GDPR has been followed by significant reductions in the volume of third-party cookies set without consent on many European news sites.

References

Libert, T. 2018. An Automated Approach to Auditing Disclosure of Third-Party Data Collection in Website Privacy Policies. In *Proceedings of WWW 2018: The 2018 Web Conference* (International World Wide Web Conference Committee), 207-16.

Libert, T., and Nielsen, R. K. 2018. *Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement*. Oxford: Reuters Institute for the Study of Journalism.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., and Nielsen, R. K. 2017. *Reuters Institute Digital News Report 2017*. Oxford: Reuters Institute for the Study of Journalism.

Methods Appendix

For this study there are two main methodological considerations: developing lists of sites to study, and measuring privacy impacts. These steps are detailed below.

SITE SELECTION

For each country, a list of news sites and popular sites were selected for analysis. For news, prior work conducted by the Reuters Institute for the Study of Journalism was used to identify 30 news sites in Germany, 33 in Spain, 20 in Finland, 30 in France, 31 in Italy, 29 in Poland, and 31 in the UK.

MEASURING PRIVACY IMPACTS

Once the list of pages was assembled, privacy impacts were measured for the months of April and July 2018. To do so, the open-source software tool webXray was used. This tool has been used extensively for academic research (e.g. Libert, 2018). As noted above, for this study webXray was configured to use the Chrome web browser. This browser was chosen as it is popular with users and it may be instrumented to run in an automated environment.

To ensure that measurement reflected what users would see in the European Union, a computer based at the University of Oxford in the United Kingdom was used. This is particularly important in the context of cookies as users in the EU have different legal protections than users in other regions, such as the US.

ABOUT THE AUTHORS

Timothy Libert is Special Faculty Instructor, Carnegie Mellon University and former Research Fellow at the Reuters Institute for the Study of Journalism, University of Oxford.

Lucas Graves is Senior Research Fellow at the Reuters Institute for the Study of Journalism, University of Oxford.

Rasmus Kleis Nielsen is the Director of Research at the Reuters Institute for the Study of Journalism and Professor of Political Communication, University of Oxford.