

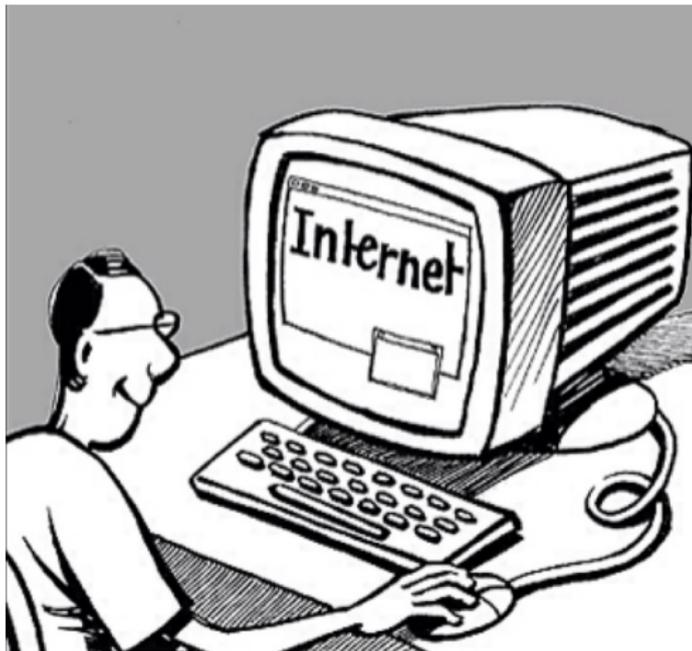
# Measuring Web Cookies in Governmental Websites

Matthias Gotze, **Srdjan Matic**, Costas Iordanou,  
Georgios Smaragdakis and Nikolaos Laoutaris.



*Appeared in: proceedings of the ACM WebSci'22*

We browser the Web every day...



Google

YouTube

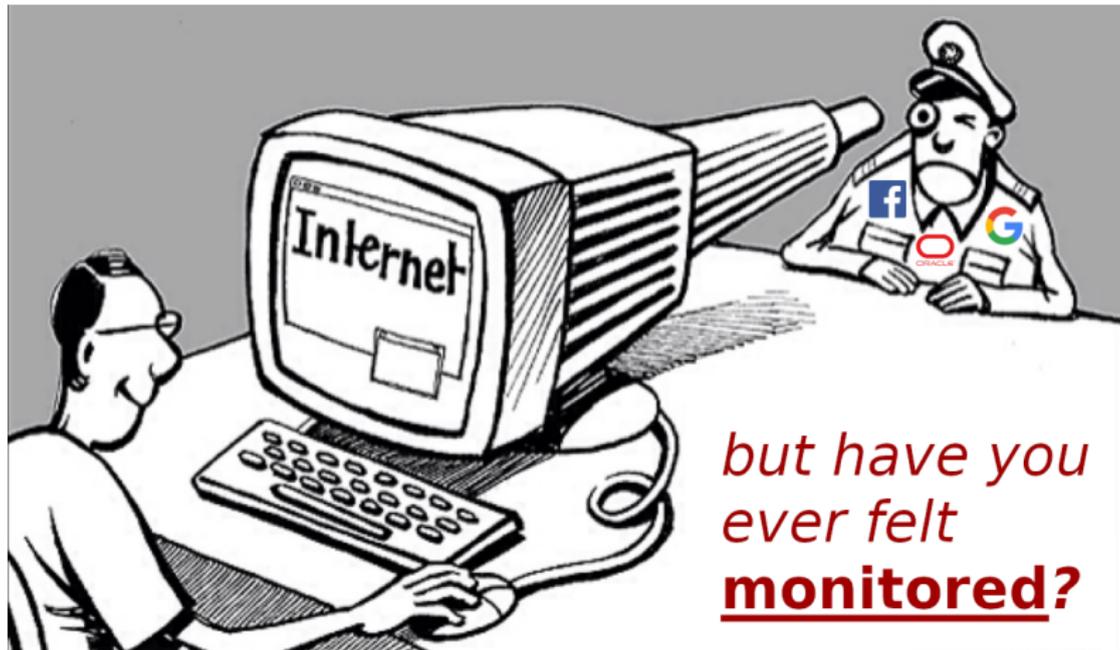
EL PAÍS

amazon

as

ÚLTIMAS NOTICIAS

We browser the Web every day...



*but have you  
ever felt  
monitored?*

# Example of Cross-site Tracking

First, access an *online banking website*

The screenshot shows the Sabadell online banking website. The browser address bar displays "Personal - BANCO S.A. | bancsabadell.com/cs/Satellite/SabAtl". The website features a blue header with the Sabadell logo and navigation tabs for Personal, Business, and Private Banking. Below the header, there are sub-navigation tabs for Home, Accounts, Cards, Saving, Investment, Financing, Insurance, and Renting. The main content area includes a "Personal Access" login form with fields for ID and Password, a "Remember user" checkbox, and an "Enter" button. To the right of the login form is a large promotional banner for "Open your account with everything" featuring a woman looking at her phone. The banner highlights "0 fees conditions (0% TIN, 0% TSE)", "1/6" (Exclusive for new customers), and "We'll pay you back 1% APR on your account balance of up to €30,000 the first year". Below the banner is a "Commitment to Sustainability" section with an image of a wind turbine. At the bottom, there are three columns: "Moments" with links like "Welcome to Spain and to Banco Sabadell", "I need to" with links like "Online banking" and "Open an account designed to cover my needs", and "Corporate Information" with details like "SAB share" and "17,398,071,020,322".

# Example of Cross-site Tracking

Then read articles a *news website*

The screenshot shows the homepage of the Corriere della Sera website. At the top, there is a navigation bar with the site name, location (Milano 17°), and a search bar. Below the navigation bar is the main header with the newspaper's name "CORRIERE DELLA SERA" and the date "domenica 09 ottobre 2022 - Aggiornato alle 12:07".

A large advertisement for Cherry Bank is featured, with the headline "Il conto che frutta è di Cherry Bank. Fino al 3,20%". The ad includes a "FAPRI IL CONTO" button and a small text box stating: "Tasso lordo annuo. Il conto deposito è riservato ai titolari del conto corrente online Cherry Bank. Messaggio pubblicitario con finalità promozionale. Per le condizioni contrattuali fare riferimento ai fogli informativi disponibili nella sezione Documenti e Trasparenza del sito (cherrybank.it)".

Below the ad, there is a row of navigation links: "Temi caldi di oggi: F1 GIAPPONE", "UCRAINA RUSSIA", "MINISTRI", "COMO", and "UCRAINA UN AIUTO SUBITO".

The main content area is titled "CORRIERE TV - I video del giorno" and features several video thumbnails. The primary article is titled "Gp Giappone: Verstappen campione! Gara sospesa, Leclerc penalizzato, il caos dei punti" by Davide Sparico. The article text includes: "Gara sospesa al 2° giro per pioggia. Sbirri a muro. Dopo 2 ore di attesa al via, Verstappen vince il titolo di Max Verstappen".

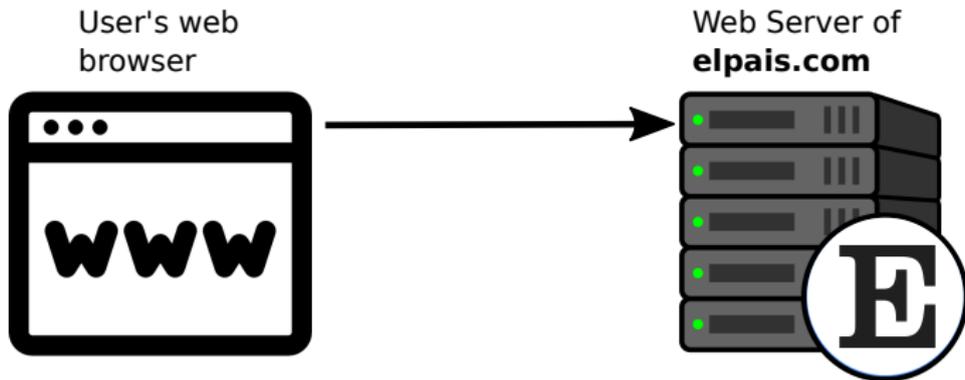
Other visible elements include a "Cuenta 100% Online, todo con tu móvil" advertisement and a "Cuenta Online Sabadell" advertisement.

# Example of Cross-site Tracking

Suddenly you start receiving **personalized ads**.

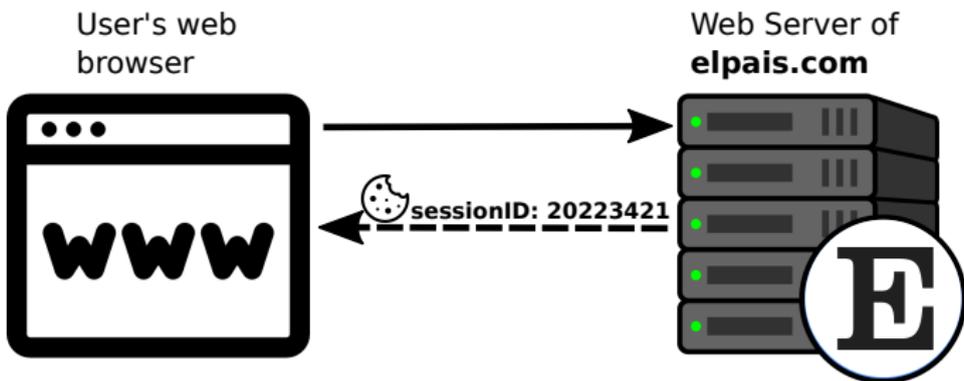
The screenshot shows the homepage of the Italian newspaper Corriere della Sera. A prominent advertisement for Cherry Bank is circled in pink, with the text "Il conto che frutta è di Cherry Bank. Fino al 3,20%". Below it, a pink arrow points to the text "Banking Ads". Another pink arrow points from the Cherry Bank ad to a Sabadell advertisement at the bottom of the page, which says "Cuenta 100% Online, todo con tu móvil". The Sabadell ad is also circled in pink. The browser's address bar shows "corriere.it".

# HTTP Cookies



- 1) The user visits a webpage

# HTTP Cookies



- 1) The user visits a webpage
- 2) The web server sends a response with a "Set-Cookie" header

# HTTP Cookies



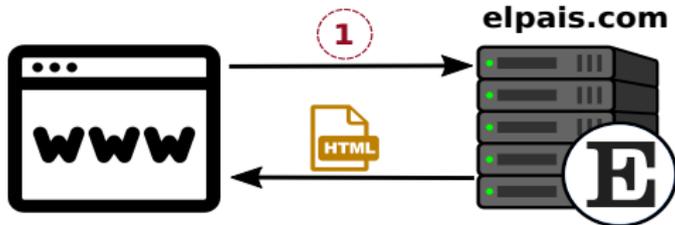
- 1) The user visits a webpage
- 2) The web server sends a response with a "Set-Cookie" header
- 3) The web browser automatically includes the cookie in *future requests* to `www.corriere.it`

# HTTP Cookies in Real-world



HTML contains:

- hyperlinks
- images
- scripts

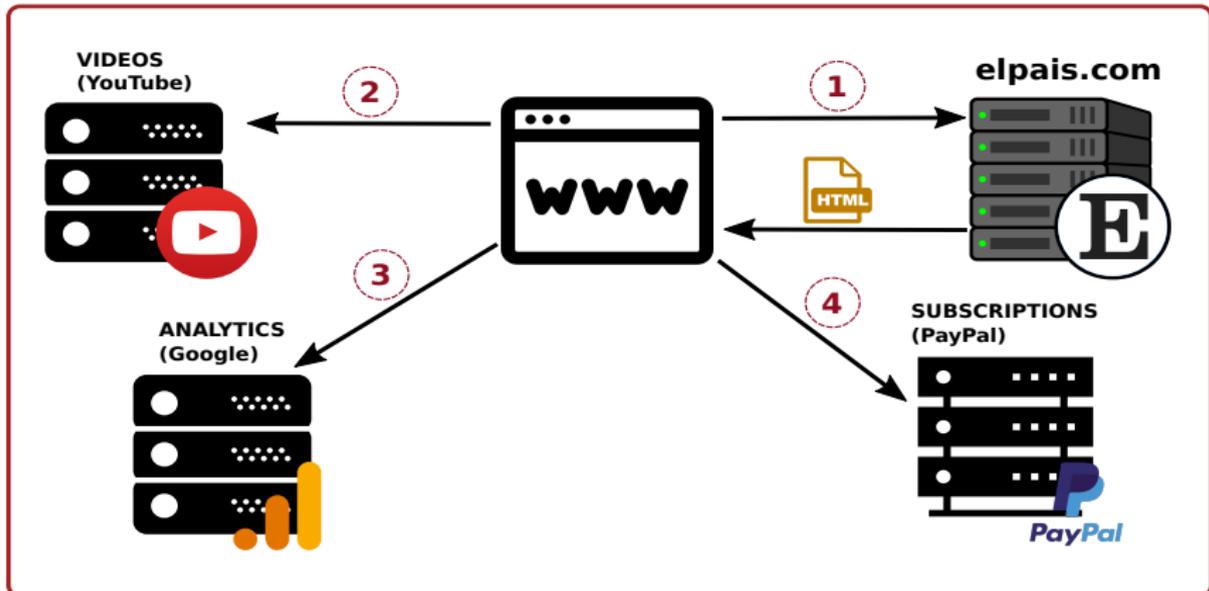


# HTTP Cookies in Real-world



HTML contains:

- hyperlinks
- images
- scripts

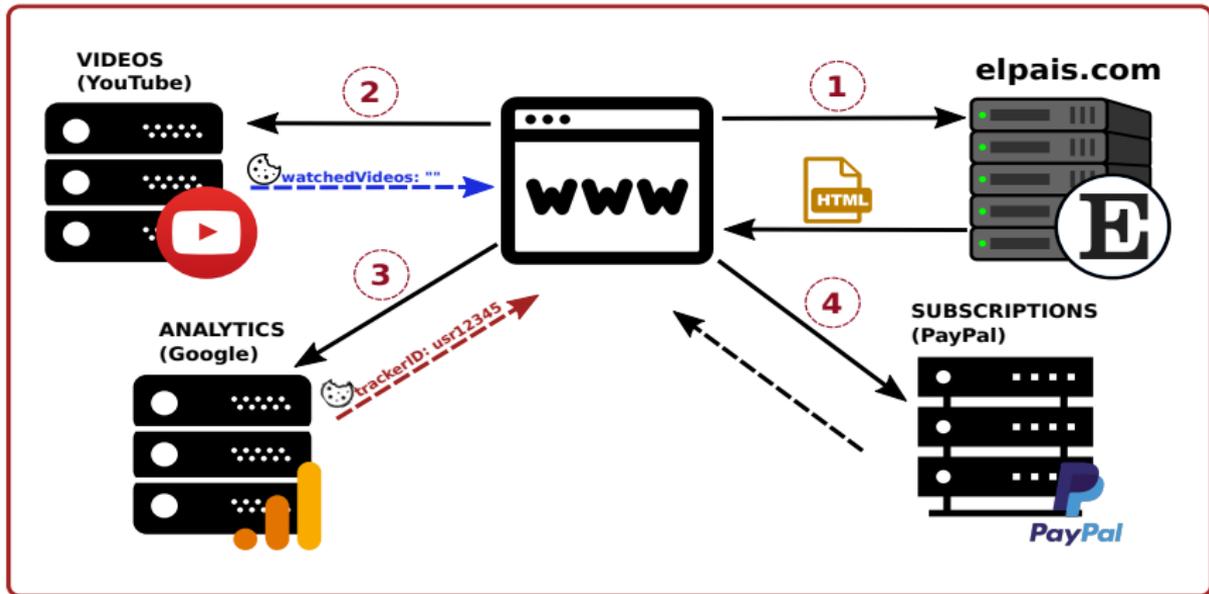


# HTTP Cookies in Real-world



HTML contains:

- hyperlinks
- images
- scripts

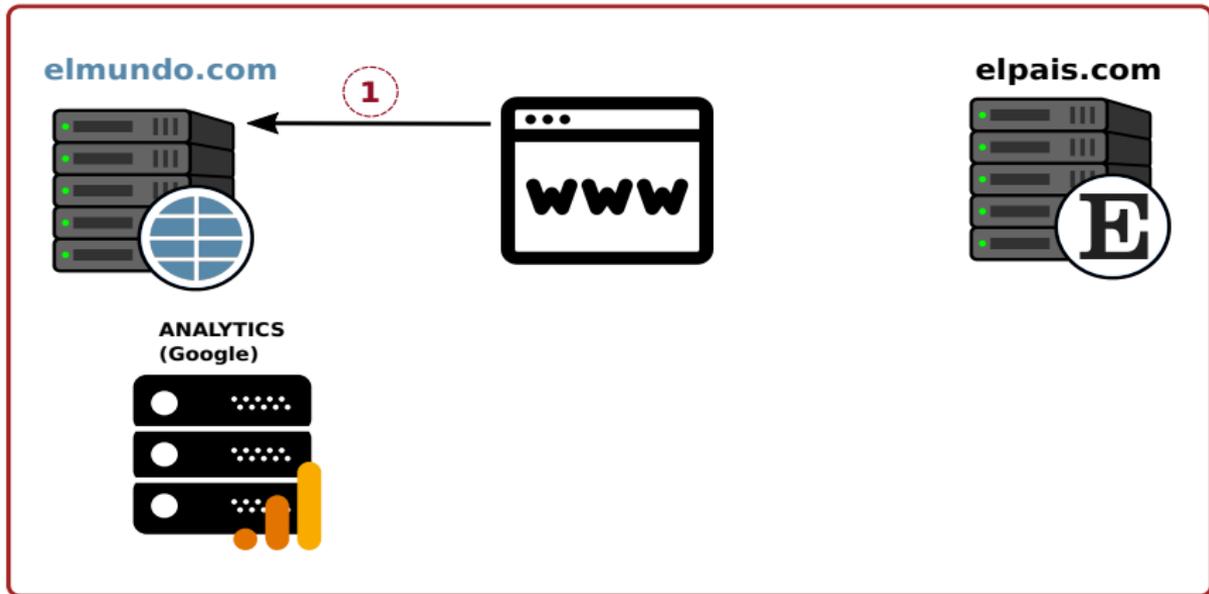


# HTTP Cookies in Real-world



HTML contains:

- hyperlinks
- images
- scripts

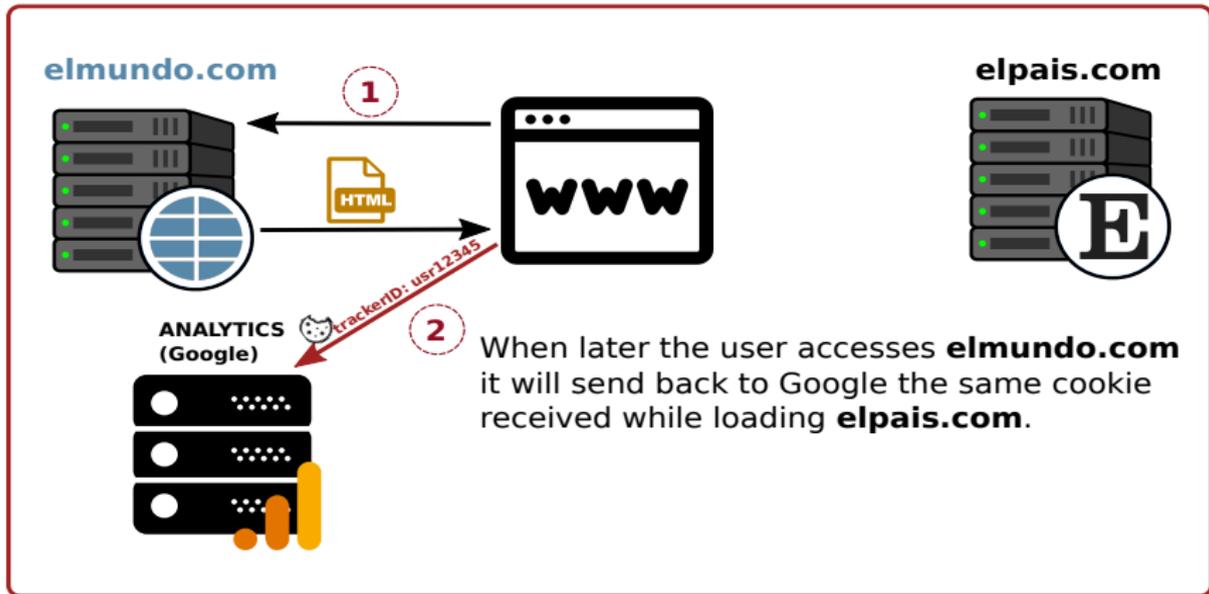


# HTTP Cookies in Real-world



HTML contains:

- hyperlinks
- images
- scripts

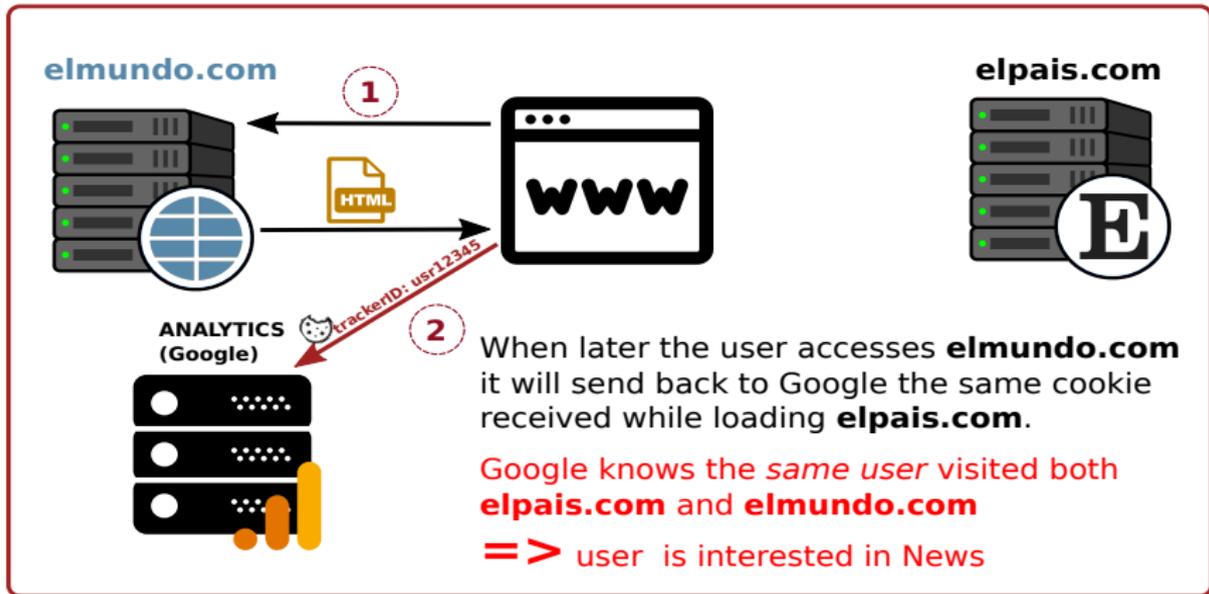


# HTTP Cookies in Real-world



HTML contains:

- hyperlinks
- images
- scripts



# Digital Profiles



# Digital Profiles



idealista



idealista



Autocasión

# Digital Profiles



job X: \$



job Y: \$\$\$



job X: \$

*“Women less likely to be shown ads for high-paid jobs on Google, study shows”<sup>1,2</sup>*

1. Guardian article: <https://tinyurl.com/omen-less-likely-ads>

2. PoPETS paper: [https://petsymposium.org/2015/papers/18\\_Datta.pdf](https://petsymposium.org/2015/papers/18_Datta.pdf)

# Abusing Profiles Information



1) Create new interests



2) Endorse products and goods



3) Influence voting

# Laws and Regulations

2002 (revised in 2009)



2018



2020



2020



# Laws and Regulations

2002 (revised in 2009)



2018



1. ePrivacy: regulates the processing of personal data in electronic communications
  - Obtain consent **before using cookies\***  
(e.g., preferences, statistics, marketing)
2. ePrivacy 2009: Consent  $\Rightarrow$  **explicit** and **informed** consent
3. GDPR: **Extraterritorial** scope

## 4 Years of EU Cookie Law: Results and Lessons Learned **2019**

Martino Trevisan, Stefano Traverso, Eleonora Bassi, and Marco Mellia  
Politecnico di Torino

## *We Value Your Privacy ... Now Take Some Cookies:* **2019** Measuring the GDPR's Impact on Web Privacy

Martin Degeling\*, Christine Utz\*, Christopher Lentzsch\*, Henry Hosseini\*, Florian Schaub†, and Thorsten Holz\*  
\*Ruhr-Universität Bochum, Germany  
†University of Michigan, Ann Arbor, MI, USA

## Do Cookie Banners Respect my Choice? **2020** Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework

Célestin Matte                      Nataliia Bielova                      Cristiana Santos  
Université Côte d'Azur, Inria      Université Côte d'Azur, Inria      Research Centre for Justice and Governance

Fines for GDPR violations and on-compliance<sup>1</sup>: **30k** Vueling (2019),  
**2.2M** Carrefour (France 2020), **50M + 50M** Google (2019, 2020)

---

<sup>1</sup>SOURCE: <https://www.cookie-law-info.com/major-gdpr-fines-2020-2021/>

# Audit Governmental and Official Websites

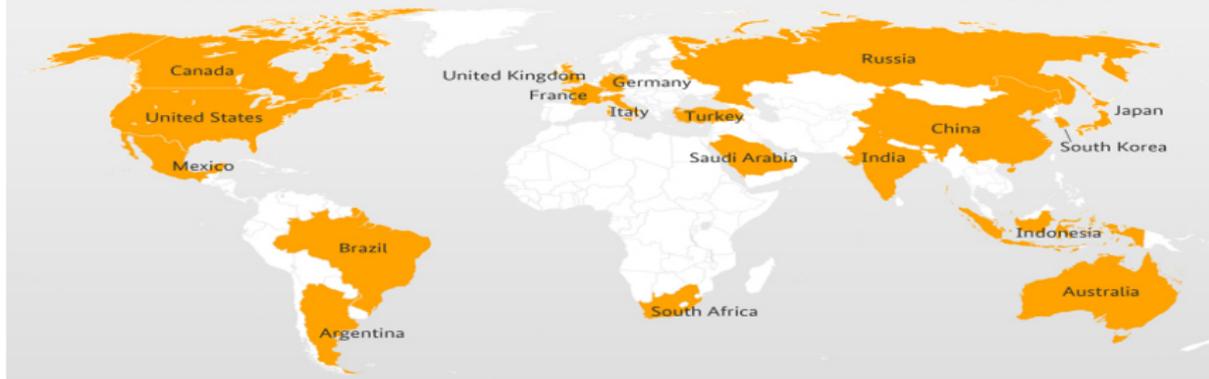


1. Allows for-large scale **surveillance**
2. Often the **only point** of interaction
3. Potential **break of trust** among citizens and authorities



# Our Study

## G20 members



## International Organizations



### # Websites:

- G20: 5,563
- Organizations: 242
- Covid-19: 131



# Example #1

URL: <https://european-union.europa.eu/>

The screenshot shows a web browser window displaying the 'Portal de la Unión Europea' (European Union portal). The browser's developer tools are open to the 'Application' tab, specifically the 'Cookies' section. The table below lists the cookies found on the page.

Name	Value	Domain	P.	Expires / Max-Age	S.	H	S	S	S	P	P
_pk_ses.c951359...	*	european-union.europa.eu	/	2022-10-13T20:02:25.000Z							
_pk_id.c9513597-...	67b4802e4...	european-union.europa.eu	/	2023-11-10T19:32:25.000Z							
cck1	%7B%22cm...	.european-union.europa.eu	/	2023-04-11T19:32:25.000Z							

# Example #1

URL: <https://european-union.europa.eu/>

The screenshot shows a web browser displaying the European Union website. The browser's Application tab is open, showing a table of cookies. The table has columns for Name, Value, Domain, P., Expires / Max-Age, and several status columns (S., H, S, S, S, P, P). Three cookies are listed:

Name	Value	Domain	P.	Expires / Max-Age	S.	H	S	S	S	P	P
_pk_ses.c951359...	*	european-union.europa.eu	/	2022-10-13T20:02:25.000Z							
_pk_id.c9513597...	67b4802e4...	european-union.europa.eu	/	2023-11-10T19:32:25.000Z							
cck1	%7B%22cm...	.european-union.europa.eu	/	2023-04-11T19:32:25.000Z							

Below the table, there are two buttons: "Aceptar todos los cookies" and "Aceptar solo las cookies esenciales". Both buttons are circled in red.

# Example #1

URL: <https://european-union.europa.eu/>

The screenshot shows the Application tab in a browser's developer tools. The 'Cookies' section is expanded, showing a list of cookies for the URL 'https://european-union.europa.eu/'.

Name	Value	Domain	P.	Expires / Max-Age	S	H	S	S	S	P	P
_pk_ses.c9513597...	*	european-union.europa.eu	/	2022-10-13T20:02:25.000Z							
_pk_id.c9513597...	67b4802e4...	european-union.europa.eu	/	2023-11-10T19:32:25.000Z							
cck1	%7B%22cm...	.european-union.europa.eu	/	2023-04-11T19:32:25.000Z							

A red bracket on the right side of the table is labeled "cookies stored on the device".

At the bottom of the browser window, a cookie consent banner is visible with two buttons: "Aceptar todos los cookies" and "Aceptar solo las cookies esenciales". These buttons are circled in red. A red arrow points from this area to the text "We did not give any consent!".

# Example #1

URL: <https://european-union.europa.eu/>

The screenshot shows the Application tab in a browser, displaying a list of cookies for the URL <https://european-union.europa.eu/>. The cookies table is as follows:

Name	Value	Domain	Expires / Max-Age
_pk_ses.c951359...	*	european-union.europa.eu	2022-10-13T20:02:25.000Z
_pk_id.c9513597...	67b4802e4...	european-union.europa.eu	2023-11-10T19:32:25.000Z
cck1	%7B%22cm...	european-union.europa.eu	2023-04-11T19:32:25.000Z

A red bracket on the right side of the table indicates that these cookies are stored on the device. A red arrow points from the 'Aceptar todos los cookies' button in the cookie consent banner to the text 'We did not give any consent!'.

cookie origin = visited domain  $\implies$  **First-Party (FP)** cookie

# Example #1

URL: <https://european-union.europa.eu/>

Name	Value	Domain	Expires / Max-Age	S	H	S	S	S	P	P
_pk_ses.c951359...	*	european-union.europa.eu	2022-10-13T20:02:25.000Z							
_pk_id.c9513597...	67b4802e4...	european-union.europa.eu	2023-11-10T19:32:25.000Z							
cck1	%7B%22cm...	european-union.europa.eu	2023-04-11T19:32:25.000Z							

**Domain:** the domain that set the cookie

**Expires / Max-Age:** how long the cookie will last

**cookies stored on the device**

**We did not give any consent!**

cookie origin = visited domain  $\implies$  **First-Party (FP)** cookie

# Example #2

URL: <https://government.ru/>

The screenshot shows a web browser displaying the website 'The Russian Government' (government.ru). The browser's developer tools are open, showing the 'Application' tab. The 'Cookies' section is expanded, displaying a table of cookies. The table has columns for Name, Value, Domain, P., Expires / Max-Age, and a set of checkboxes for S., H, S, S, P, P.

Name	Value	Domain	P.	Expires / Max-Age	S.	H	S	S	P	P
_ym_isad	2	.government.ru	/	2022-10-14T15:36:20.000Z						
_ym_d	1665689780	.government.ru	/	2023-10-13T19:36:20.000Z						
_ym_uid	166568978...	.government.ru	/	2023-10-13T19:36:20.000Z						
ymex	169722578...	.yandex.ru	/	2023-10-13T19:36:20.509Z						
i	vjQTQikmu...	.yandex.ru	/	2023-11-17T19:36:20.509Z						
yandexuid	679104067...	.yandex.ru	/	2023-10-13T19:36:20.508Z						
yuidss	679104067...	.yandex.ru	/	2023-10-13T19:36:20.509Z						
yabs-sid	184830146...	mc.yandex.ru	/	Session						

# Example #2

URL: <https://government.ru/>

The screenshot shows a web browser displaying the Russian Government website. The Application tab is open, showing a list of cookies. A red arrow points to the 'All News' link in the bottom left corner, which is circled in red. Another red arrow points to the cookies table, with a red bracket and the text 'cookies stored on the device' next to it.

Name	Value	Domain	P.	Expires / Max-Age	S.	H	S	S	S	P	P
_ym_isad	2	.government.ru	/	2022-10-14T15:36:20.000Z							
_ym_d	1665689780	.government.ru	/	2023-10-13T19:36:20.000Z							
_ym_uid	166568978...	.government.ru	/	2023-10-13T19:36:20.000Z							
ymex	169722578...	.yandex.ru	/	2023-10-13T19:36:20.509Z							
i	vjQTQikmu...	.yandex.ru	/	2023-11-17T19:36:20.509Z							
yandexuid	679104067...	.yandex.ru	/	2023-10-13T19:36:20.508Z							
yuidss	679104067...	.yandex.ru	/	2023-10-13T19:36:20.509Z							
yabs-sid	184830146...	mc.yandex.ru	/	Session							

# Example #2

URL: <https://government.ru/>

The screenshot shows the Application tab in a browser. The cookies table is as follows:

Name	Value	Domain	P.	Expires / Max-Age	S.	H	S	S	S	P	P
_ym_isad	2	.government.ru	/	2022-10-14T15:36:20.000Z							
_ym_d	1665689780	.government.ru	/	2023-10-13T19:36:20.000Z							
_ym_uid	166568978...	.government.ru	/	2023-10-13T19:36:20.000Z							
ymex	169722578...	yandex.ru	/	2023-10-13T19:36:20.509Z							
i	vjQTQIkmu...	yandex.ru	/	2023-11-17T19:36:20.509Z							
yandexuid	679104067...	yandex.ru	/	2023-10-13T19:36:20.508Z							
yuidss	679104067...	yandex.ru	/	2023-10-13T19:36:20.509Z							
yabs-sid	184830146...	mc.yandex.ru	/	Session							

Annotations in the image:

- A red oval highlights the "All News" link in the footer.
- A red arrow points from the text "Missing consent banner" to the cookies table.
- A red bracket on the right side of the cookies table is labeled "cookies stored on the device".
- Text below the table reads: "Third-party: set from external domains".

cookie origin  $\neq$  visited domain  $\implies$  **Third-Party (TP) cookie**

# Example #2

URL: <https://government.ru/>

The screenshot shows the Application tab in a browser, displaying a table of cookies. The table has columns for Name, Value, Domain, P., Expires / Max-Age, and S. The cookies listed are:

Name	Value	Domain	P.	Expires / Max-Age	S.	H	S	S	S	P	P
_ym_isad	2	.government.ru	/	2022-10-14T15:36:20.000Z							
_ym_d	1665689780	.government.ru	/	2023-10-13T19:36:20.000Z							
_ym_uid	166568978...	.government.ru	/	2023-10-13T19:36:20.000Z							
ymex	169722578...	.yandex.ru	/	2023-10-13T19:36:20.509Z							
i	vjQTQIkmu...	.yandex.ru	/	2023-11-17T19:36:20.509Z							
yandexuid	679104067...	.yandex.ru	/	2023-10-13T19:36:20.508Z							
yuidss	679104067...	.yandex.ru	/	2023-10-13T19:36:20.509Z							
yabs-sid	184830146...	mc.yandex.ru	/								

Annotations in the image:

- A red bracket on the right side of the table groups the cookies from `.yandex.ru` and `mc.yandex.ru` with the text "cookies stored on the device".
- A green box highlights the `ymex` and `i` cookies, with the text "Third-party: set from external domains".
- An orange box highlights the `Session` cookie, with the text "Session: the cookie will be deleted when the browser is closed".
- A red oval at the bottom left of the page content is circled, with an arrow pointing to it and the text "Missing consent banner".

cookie origin  $\neq$  visited domain  $\implies$  **Third-Party (TP) cookie**

# Approach

## 1. Visit websites

- collect cookies with an instrumented web browser
- 10 URLs per website
  - only HTML content and clickable hyperlinks
- NO consent given (or any action performed)



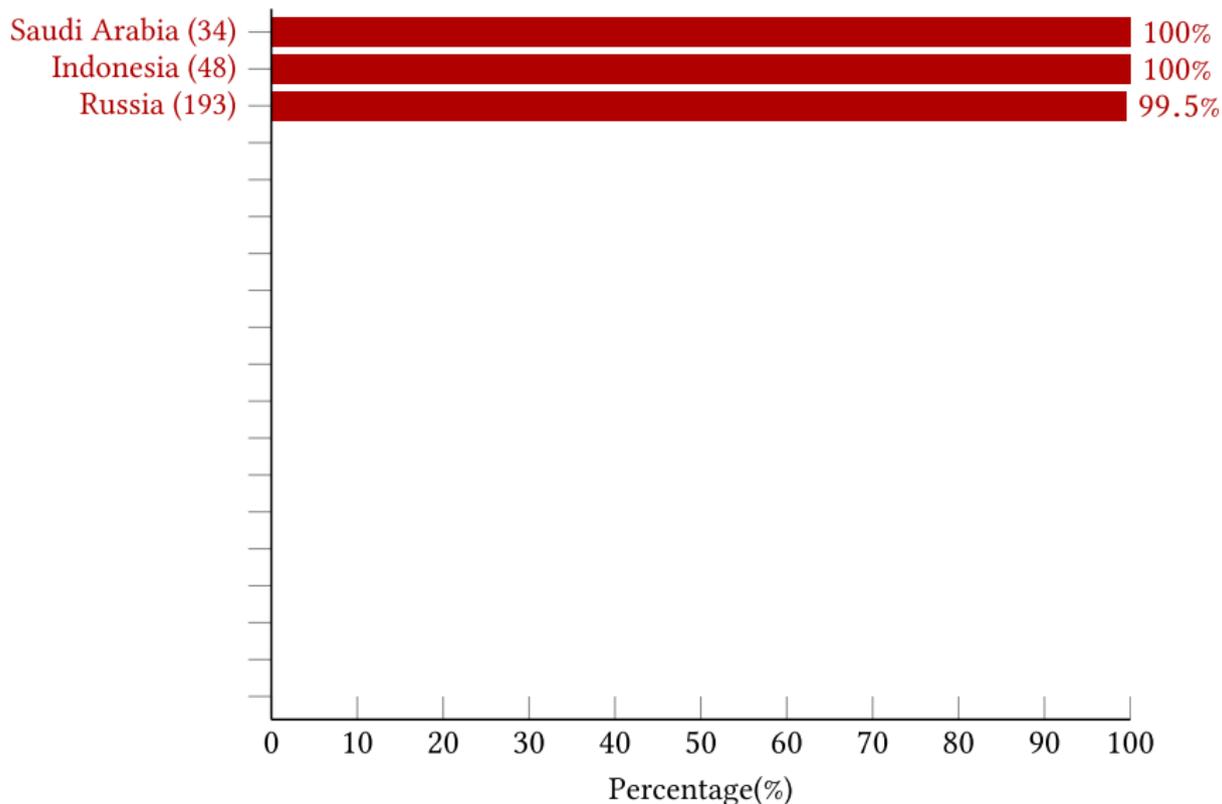
## 2. Evaluate cookies

- *Origins*: First-Party (**FP**) vs Third-Party (**TP**)
  - Third-Party Trackers (**TPT**)
- *Lifetime*: session vs persistent



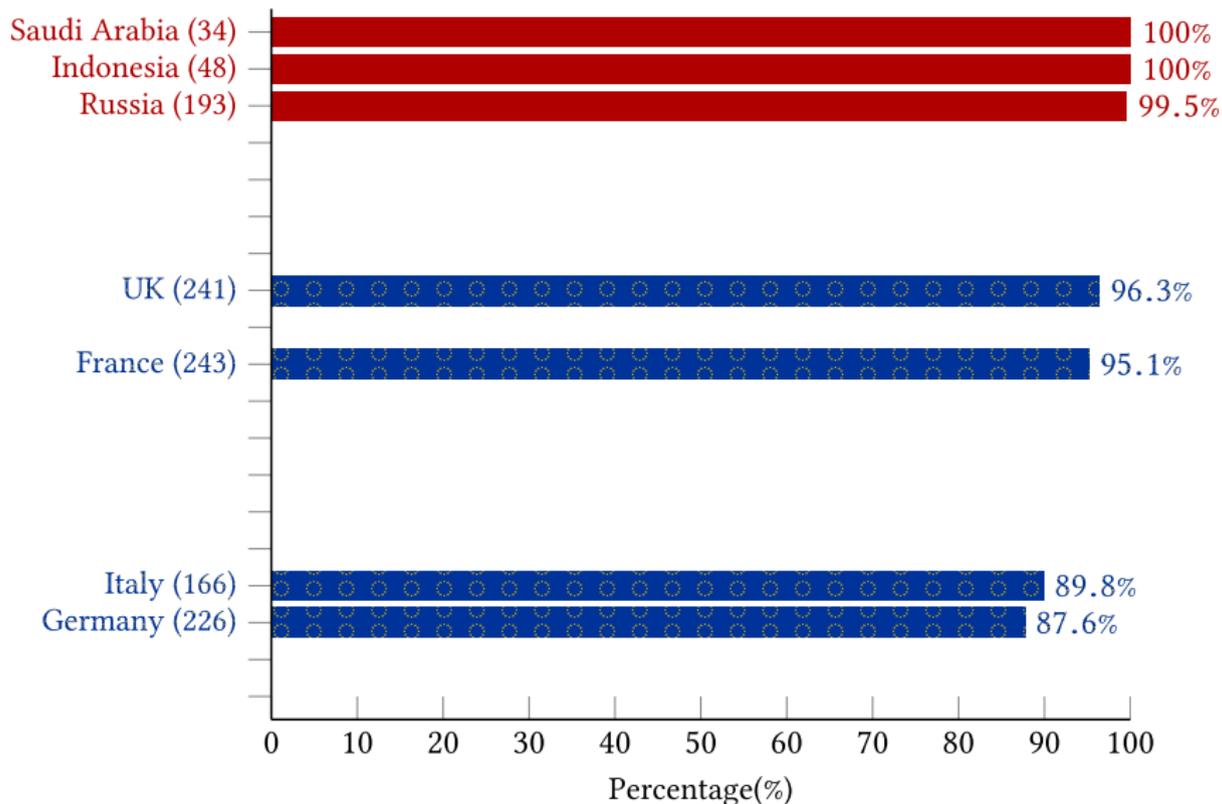
# Do G20 Websites Set Cookies?

## Top-3



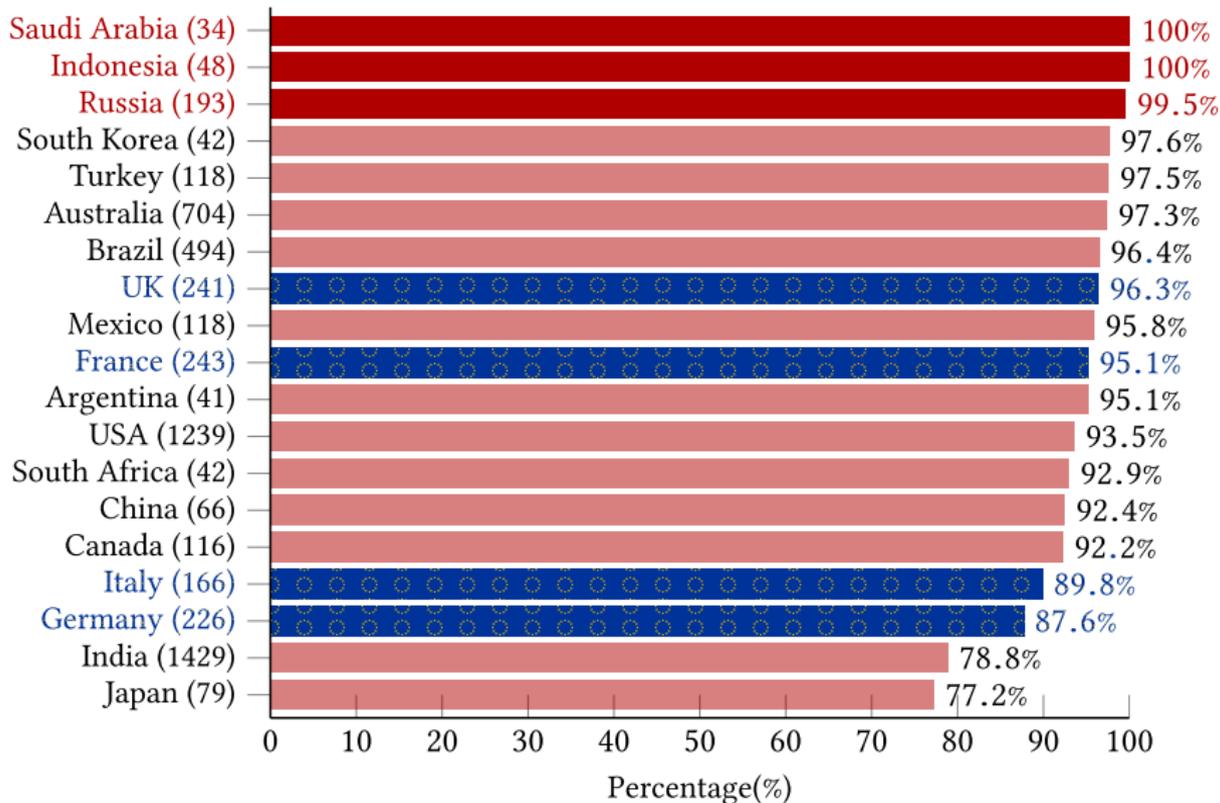
# Do G20 Websites Set Cookies?

## Top-3 & EU



# Do G20 Websites Set Cookies?

All



# Do G20 Websites Set Cookies?

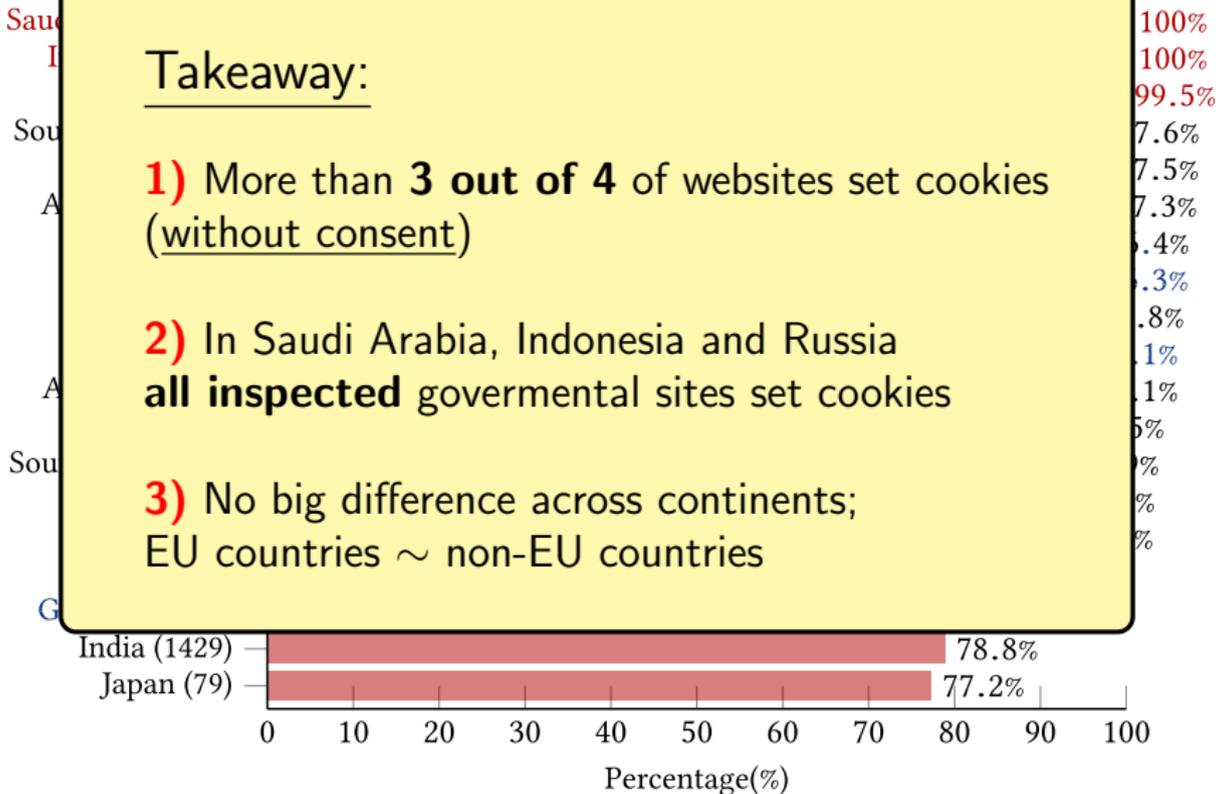
All

## Takeaway:

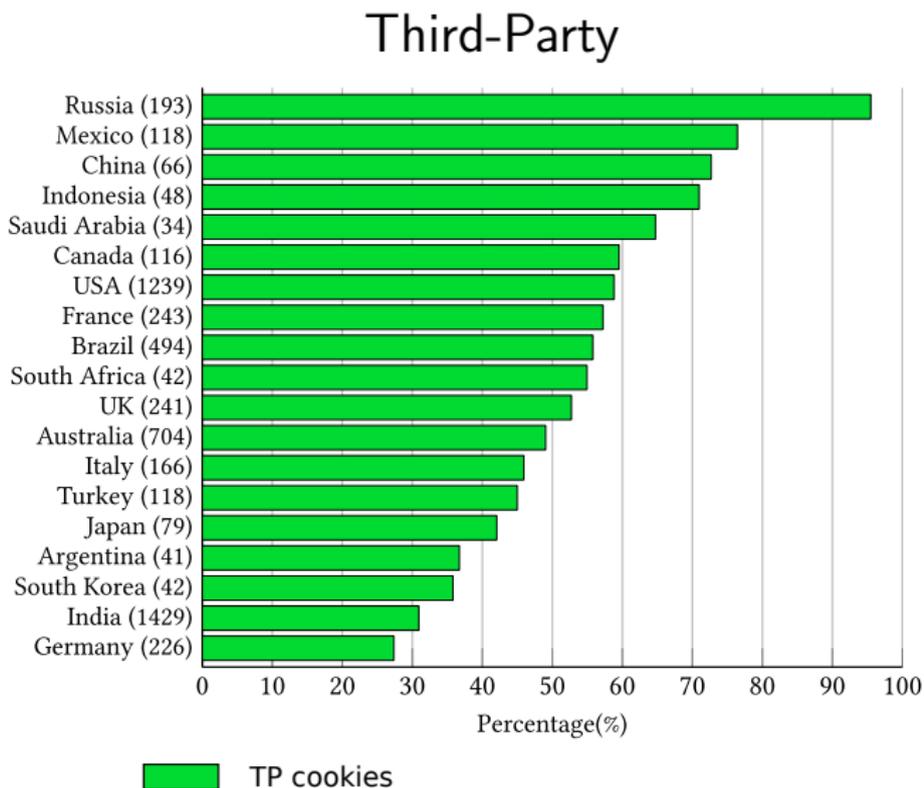
1) More than **3 out of 4** of websites set cookies  
(without consent)

2) In Saudi Arabia, Indonesia and Russia  
**all inspected** governmental sites set cookies

3) No big difference across continents;  
EU countries ~ non-EU countries

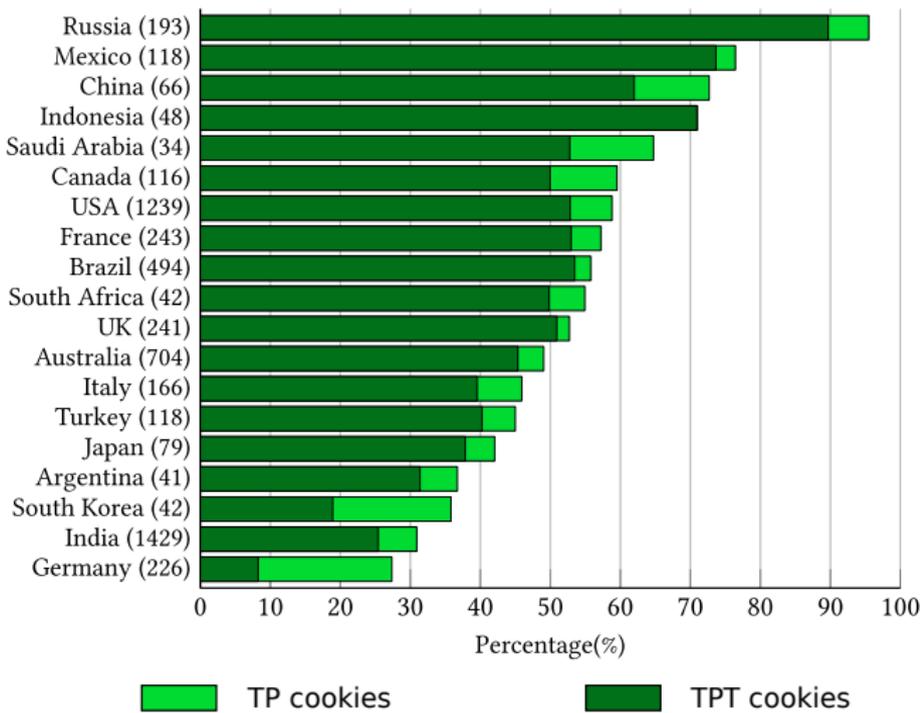


# Do G20 Websites Set Third-Party Cookies?



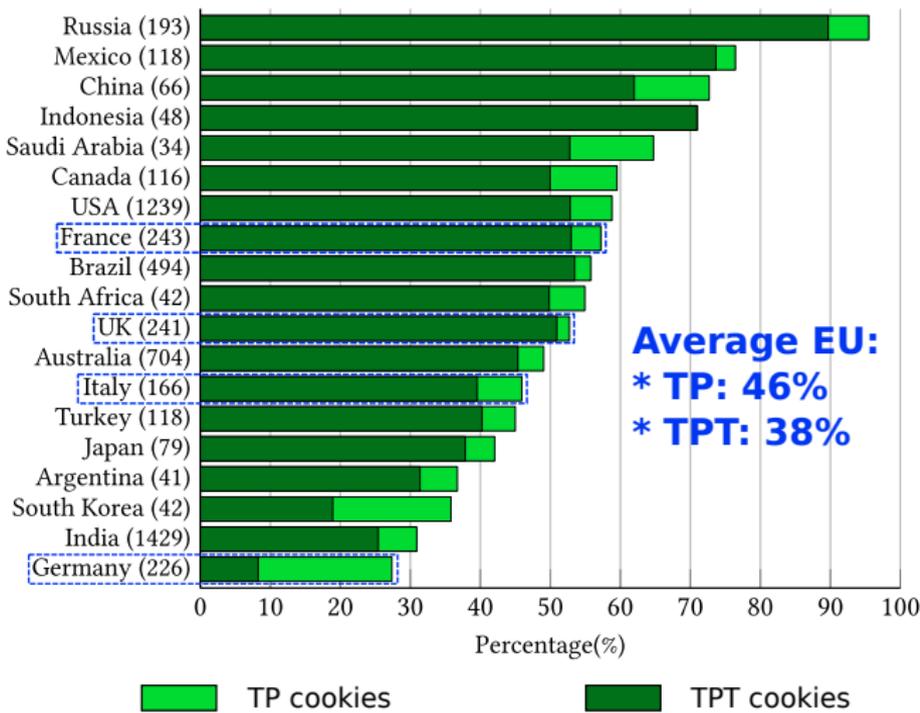
# Do G20 Websites Set Third-Party Cookies?

## Third-Party & Third-Party Trackers



# Do G20 Websites Set Third-Party Cookies?

## Third-Party & Third-Party Trackers

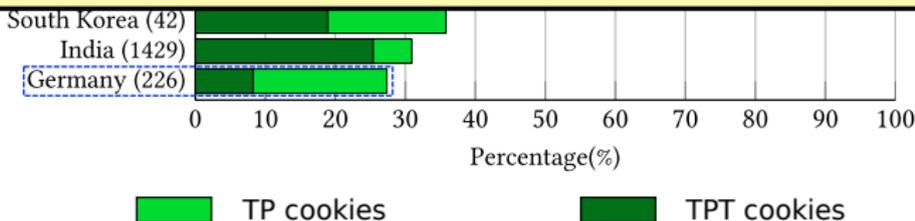


# Do G20 Websites Set Third-Party Cookies?

## Third-Party & Third-Party Trackers

### Takeaway:

- 1) More than **1\* out of 3** of websites set Third-Party cookies for 11 out of 19 countries this ratio is  $\geq 50\%$
- 2) Website with TP cookies  $\implies$  they are from **known trackers**
- 3) Germany as the most privacy-oriented (EU) country

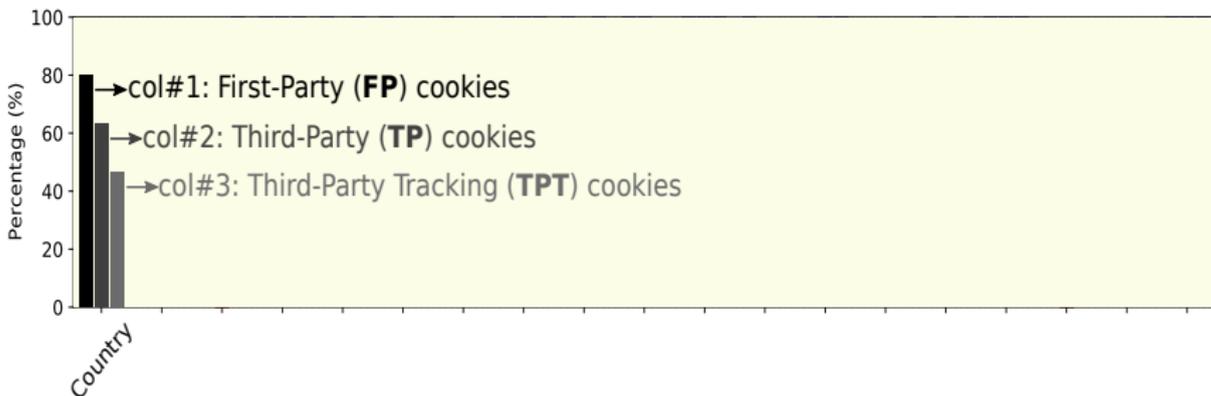


# How Long do Cookies Last?

*Cookie origin:*

**FP** = First-Party

**TP** = Third-Party    **TPT** = Third-Party Tracker



# How Long do Cookies Last?

## Cookie origin:

**FP** = First-Party

**TP** = Third-Party    **TPT** = Third-Party Tracker

## Cookie duration:

Session

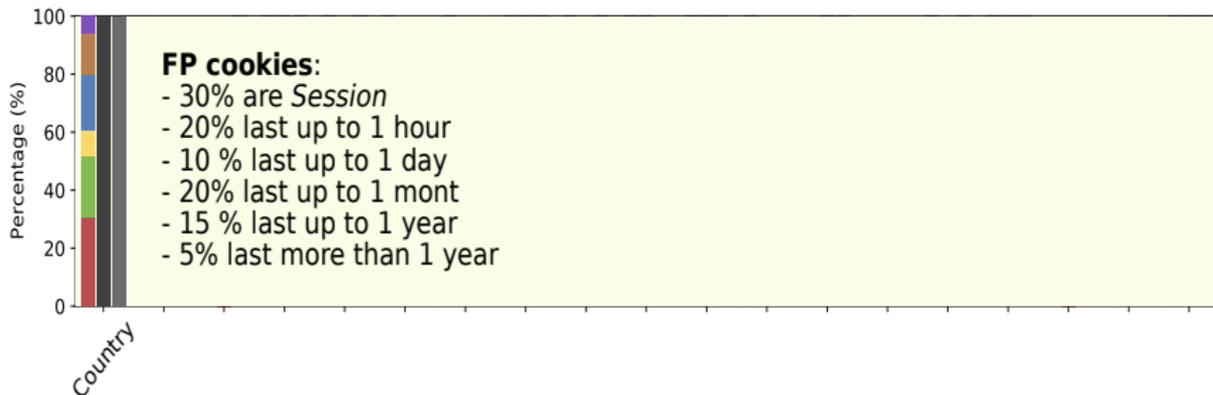
1 hour

1 day

1 month

1 year

> 1 year



# How Long do Cookies Last?

## Cookie origin:

**FP** = First-Party

**TP** = Third-Party    **TPT** = Third-Party Tracker

## Cookie duration:

Session

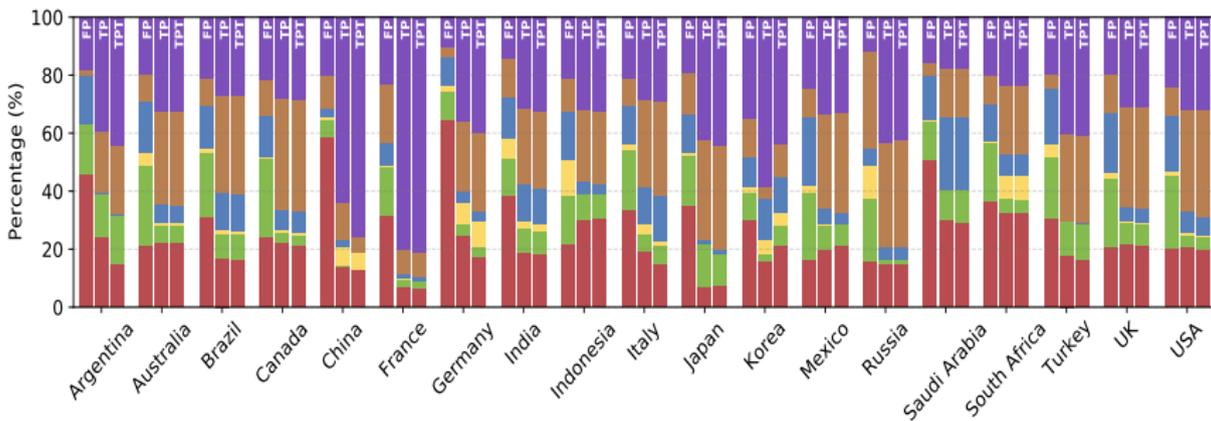
1 hour

1 day

1 month

1 year

> 1 year



# How Long do Cookies Last?

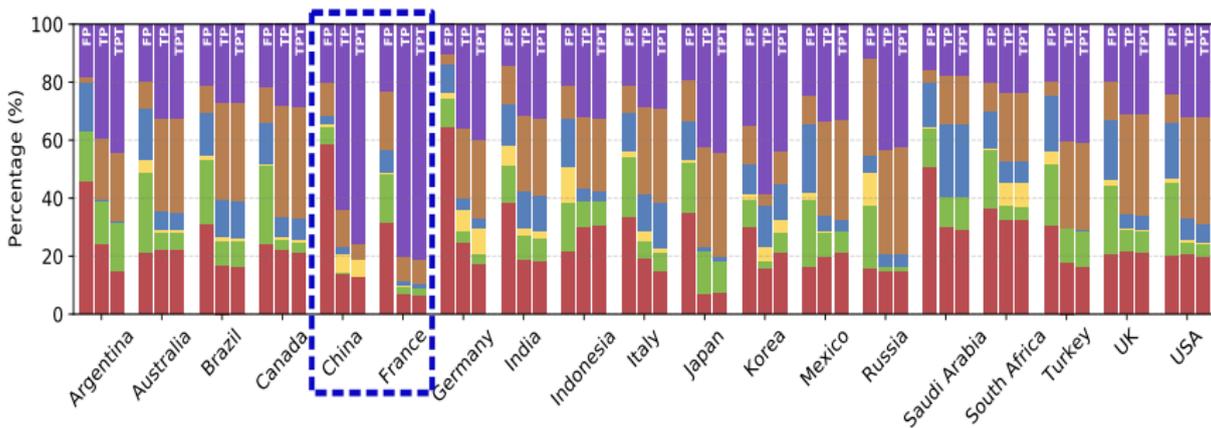
*Cookie origin:*

**FP** = First-Party

**TP** = Third-Party    **TPT** = Third-Party Tracker

*Cookie duration:*

Session    1 hour    1 day  
1 month    1 year    > 1 year



# How Long do Cookies Last?

## Cookie origin:

**FP** = First-Party

**TP** = Third-Party    **TPT** = Third-Party Tracker

## Cookie duration:

Session

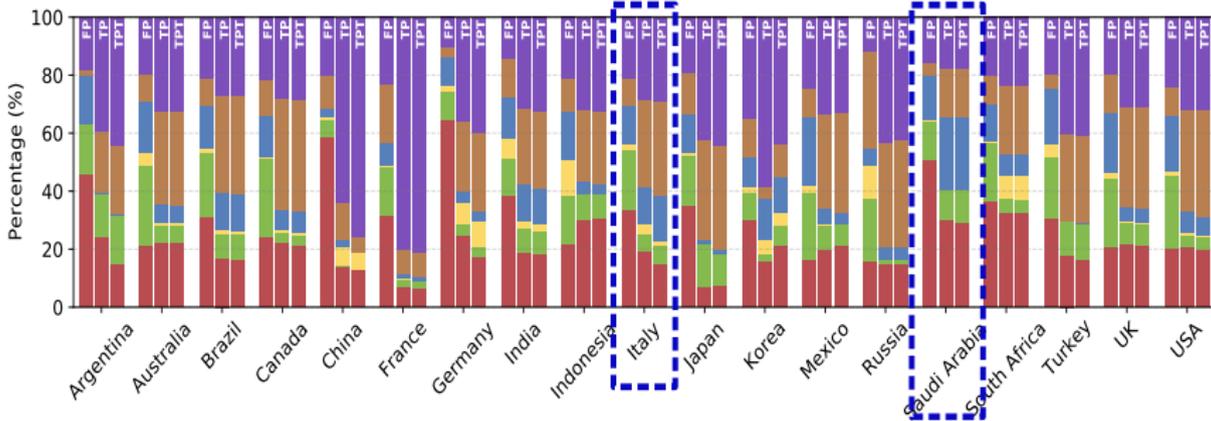
1 hour

1 day

1 month

1 year

> 1 year



# How Long do Cookies Last?

*Cookie origin:*

**FP** = First-Party

**TP** = Third-Party    **TPT** = Third-Party Tracker

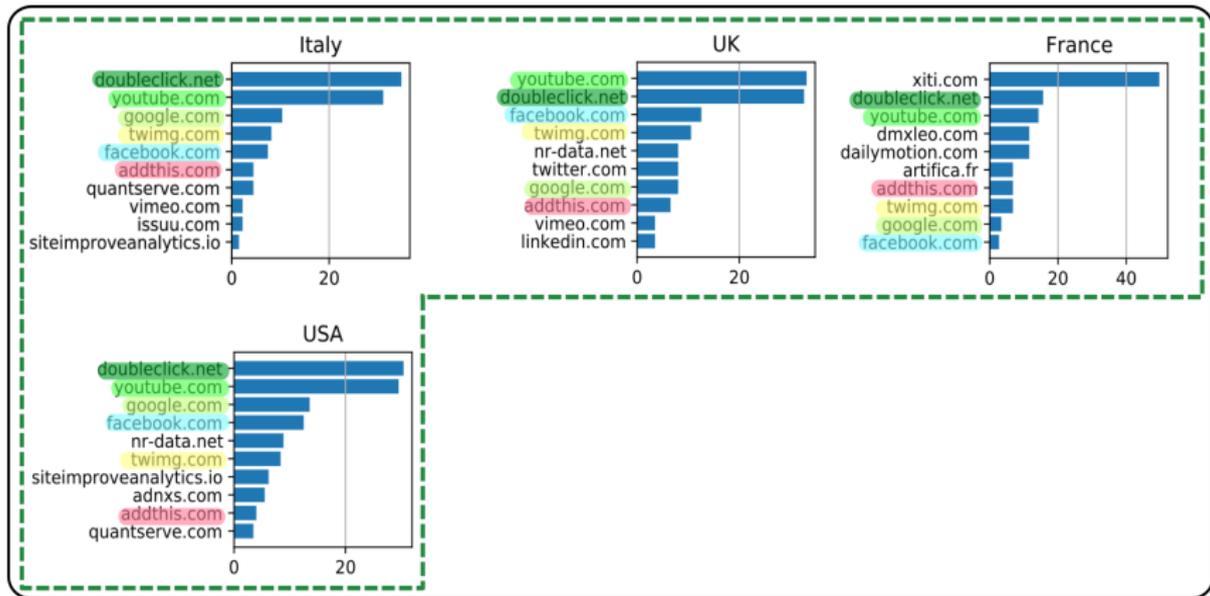
*Cookie duration:*

Takeaway:

- 1) Cookies are either very-short or long-lived
- 2) TP and TPT cookies last longer the FP
- 3) More than **50%** of TP and TPT cookies lasts 1+ year(s)

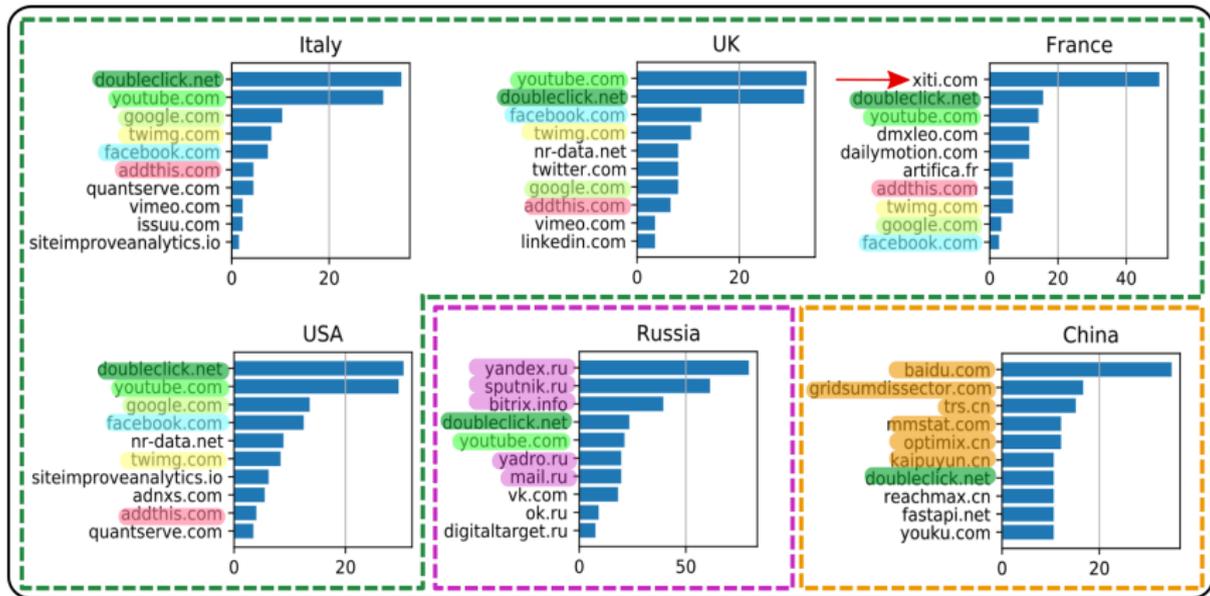


# Which are the Most Popular Trackers?



- 1) Google trackers the dominant ones
- 2) Facebook, Twitter and Oracle are also widespread

# Which are the Most Popular Trackers?



- 1) Google trackers the dominant ones
- 2) Facebook, Twitter and Oracle are also widespread
- 3) Presence of regional trackers for some countries

# Why are Trackers There?

**Governmental (and official websites) do not receive special attention, and they suffer of the exact same issues as commercial websites.**



## Cause of issues:

- Hyperlinks to social networks
- Use of videos
- Popular analytics tools
- Third-party software libraries

# Why are Trackers There?

Governmental (and official websites) do not receive special attention, and they suffer of the exact same issues as commercial websites.



## Cause of issues:

- Hyperlinks to social networks
- Use of videos
- Popular analytics tools
- Third-party software libraries

## Recommendations:

- Responsible development
- Dedicated clouds
- Website audits
- Education on web tracking

# Questions and Acknowledgments

## Questions?



## Acknowledgments:

This work was supported in part by the Atracción de Talento grant (Ref. 2020-T2/TIC-20184) funded by Madrid regional government, the TV-HGGs project (OPPORTUNITY/0916/ERC-CoG/0003) co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation, the European Research Council (ERC) Starting Grant ResolutioNet; “Resolving the Tussle in the Internet: Mapping, Architecture, and Policy Making” (ERC-StG-679158), and the EU H2020 Research and Innovation programme under grant agreements No. 871370 (Pimcity).



Horizon 2020  
European Union Funding  
for Research & Innovation



## 1. G20 Governmental websites:

- no crawling + inferring *if* the website is governmental
- official sources (e.g, [https://administracion.gob.es/pag\\_Home/atencionCiudadana/SedesElectronicas-y-Webs-Publicas/websPublicas.html](https://administracion.gob.es/pag_Home/atencionCiudadana/SedesElectronicas-y-Webs-Publicas/websPublicas.html))

## 2. International organizations:

- agencies from the EU and UN websites
- major recognized international organization

## 3. Covid-19:

- URLs from the US Center for Disease Control and Prevention
- URLs from the EU Center for Disease Prevention and Control
- SimilarWeb to find the covid-related websites for each one of the G20 countries

# Identifying Trackers

Use of **filter lists**:

1. Disconnect (Firefox tracking protection)
2. Privacy Badger (from EFF)
3. EasyList, EasyPrivacy and Adblock (community maintained)

**NOTE:** Conservative approach; our results as lower bound:

- bias toward well-known Western trackers
- tracker *IIF* “*the full domain appears in a filter list*”
- majority vote on trackers

# Results for Int. Orgs. and Covid-19 Websites

## Cookie origin:

**FP** = First-Party

**TP** = Third-Party

**TPT** = Third-Party Tracker

## Cookie duration:

 Session

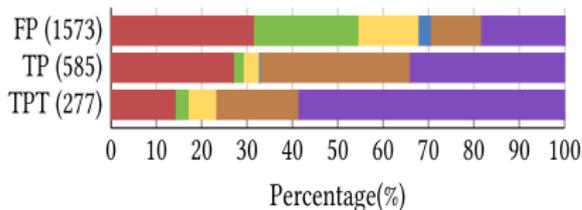
 1 hour

 1 day

 1 month

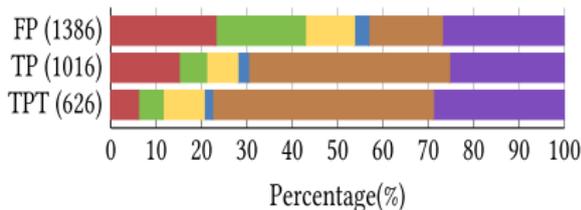
 1 year

 > 1 year



Hostname	#Trackers
www.wfp.org	36
icsid.worldbank.org	13
www.itu.int	13
www.worldfishcenter.org	13
irena.org	12

### Top-5 trackers



Hostname	#Trackers
coronavirus.jhu.edu	7
www.landlaeknir.is	6
covid19.min-saude.pt	4
www.spkc.gov.lv	4
deputyprimeminister.gov.mt	3

### Top-5 trackers

## International Organizations

## Covid-19 websites