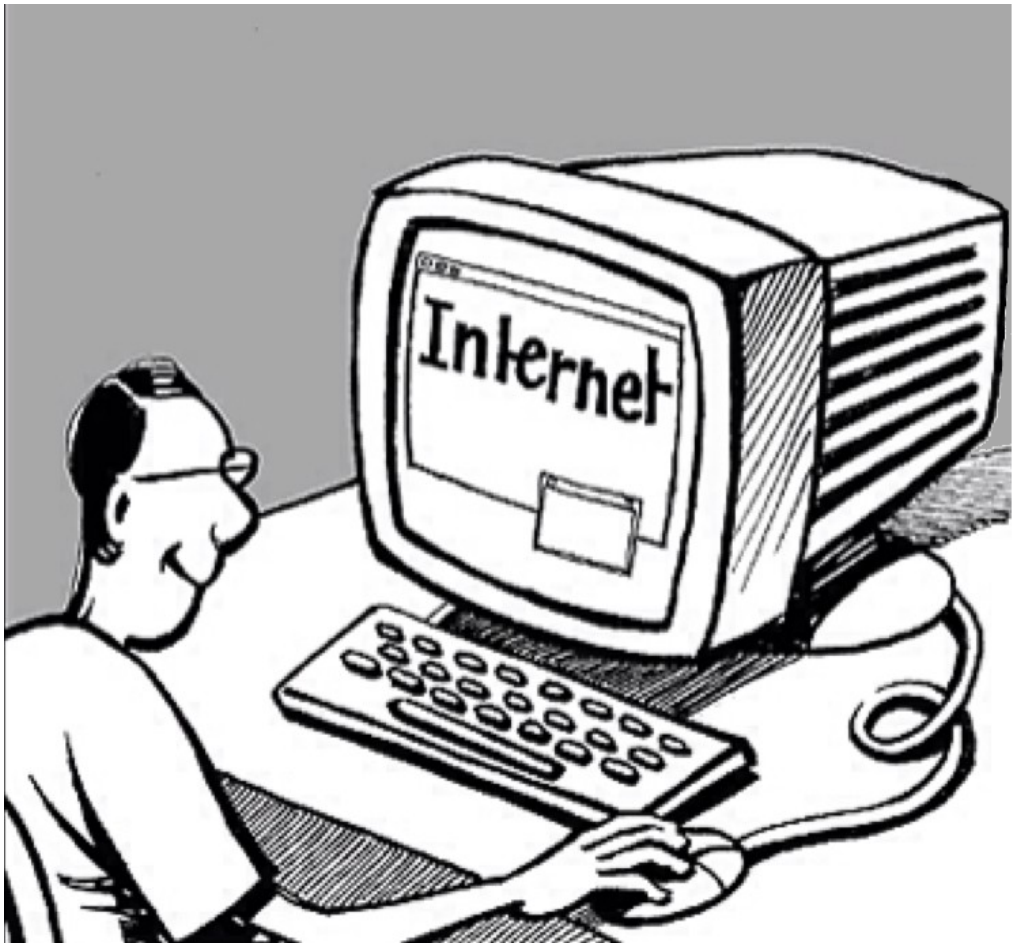
Measuring Web Cookies in Governmental Websites

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

Appeared in: proceedings of the ACM WebSci'22

Web Browsing

We browser the Web every day...



Google

YouTube

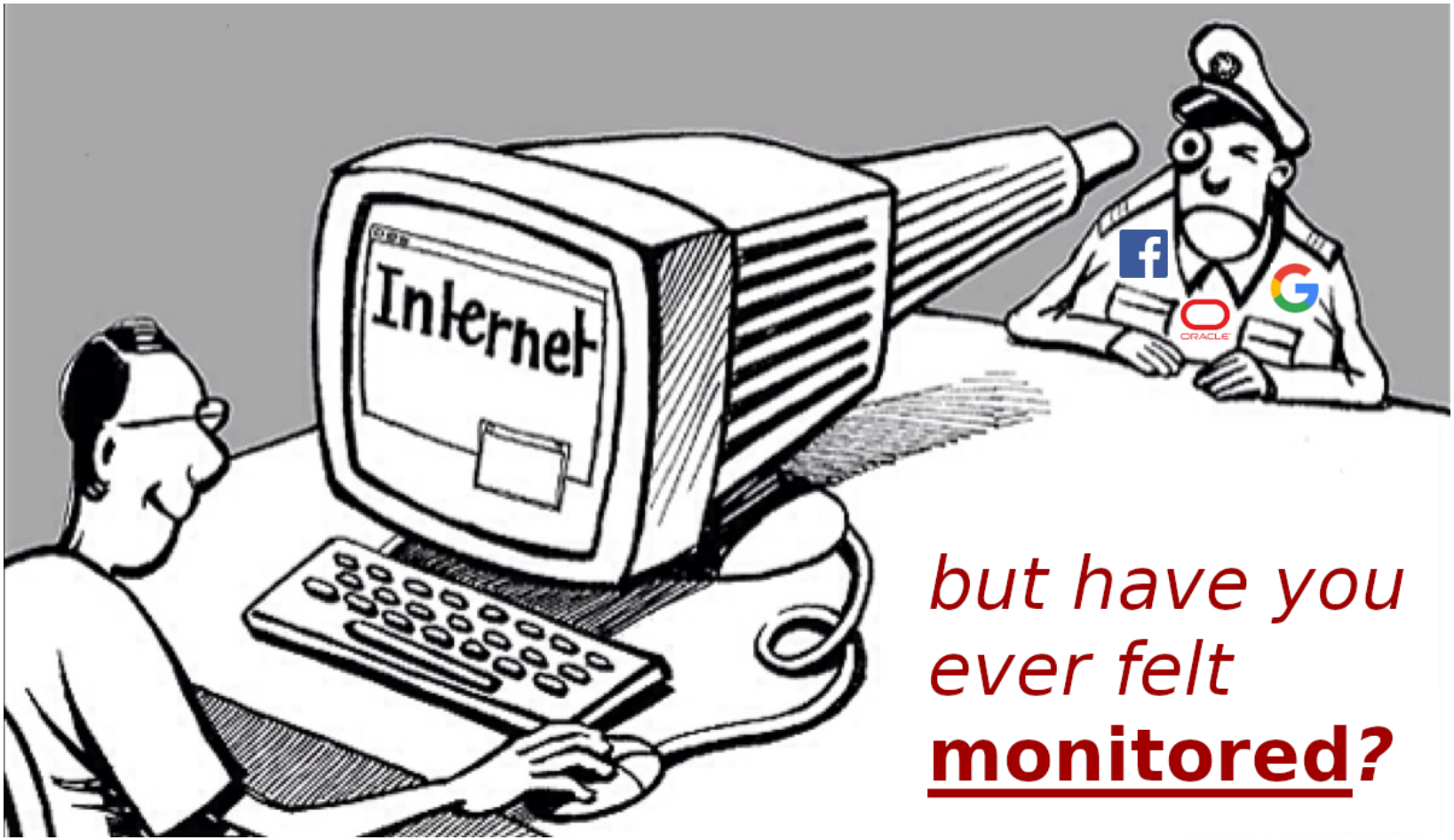
EL PAÍS

amazon

as
ÚLTIMAS NOTICIAS

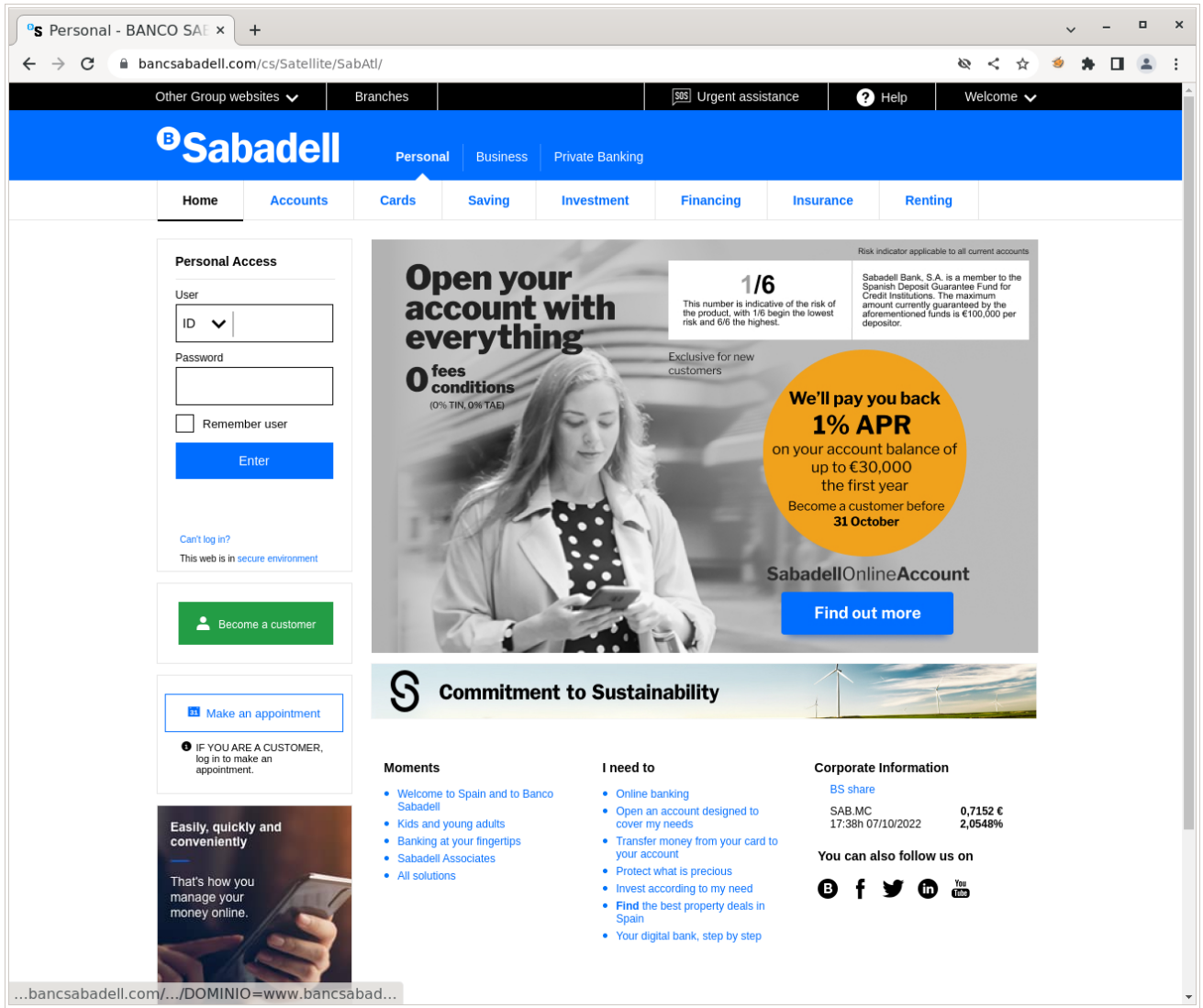
Web Browsing

We browser the Web every day...



Example of Cross-site Tracking

First, access an *online banking website*

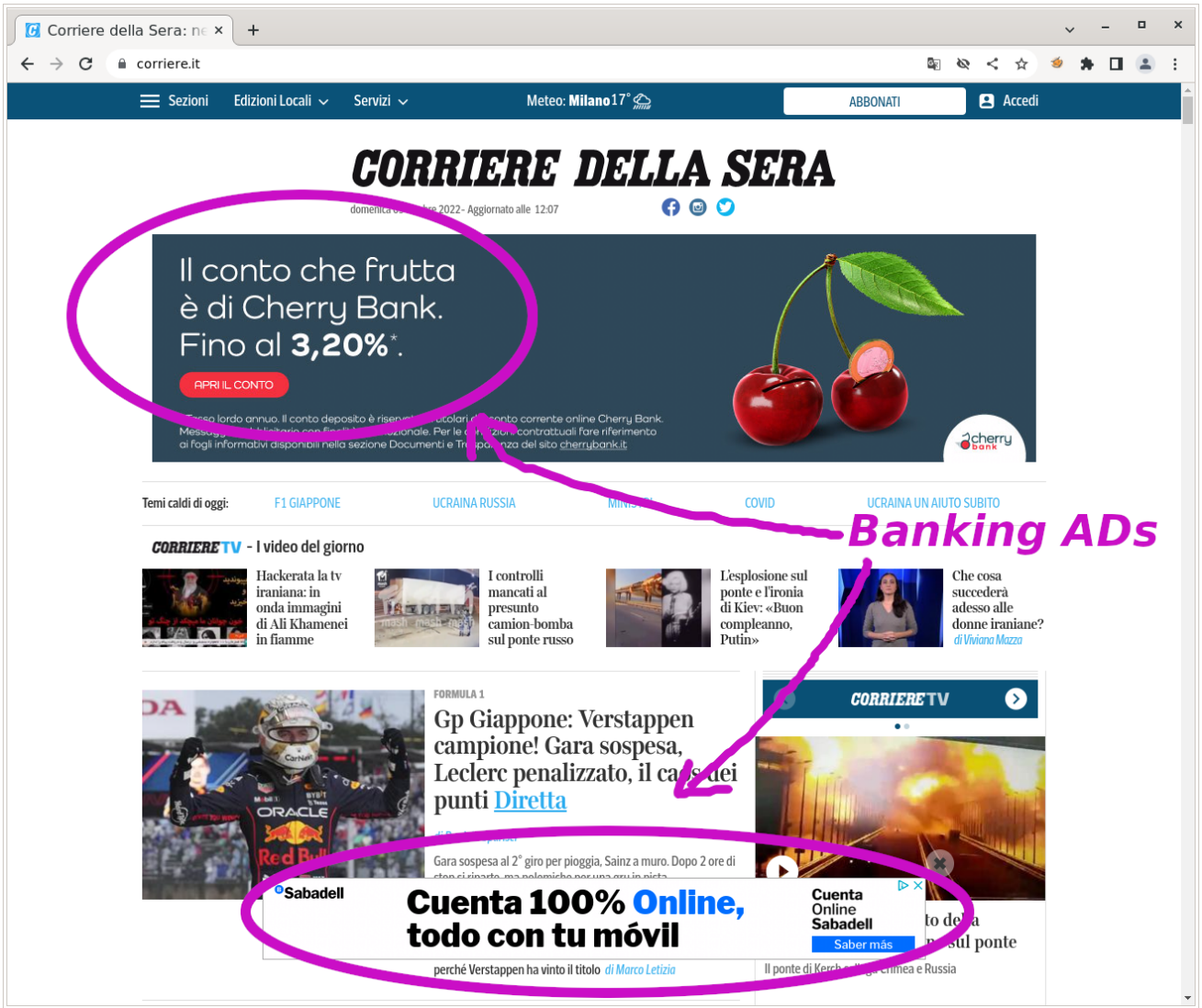


Example of Cross-site Tracking

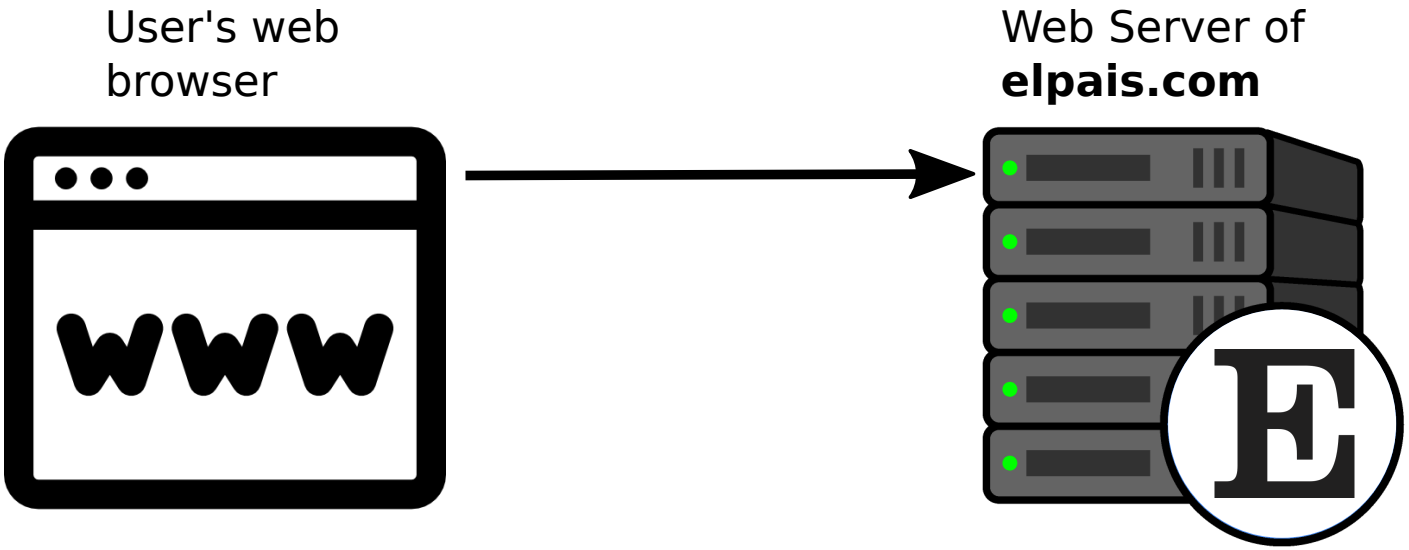
Then read articles a *news website*

Example of Cross-site Tracking

Suddenly you start receiving **personalized ads.**

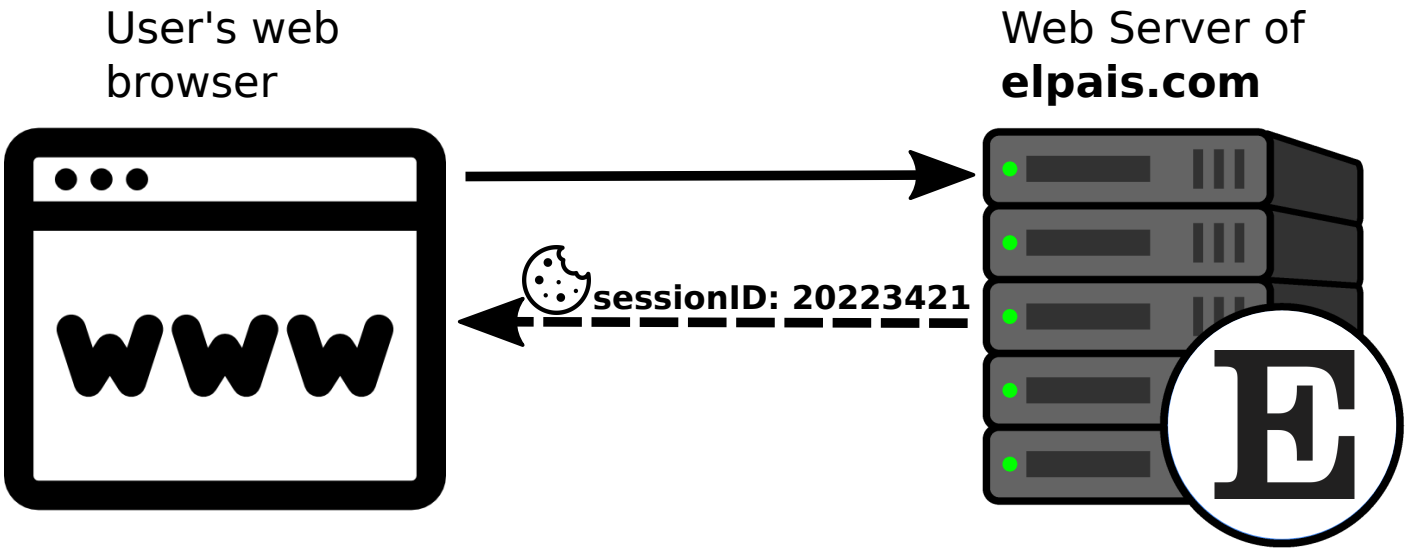


HTTP Cookies



1) The user visits a webpage

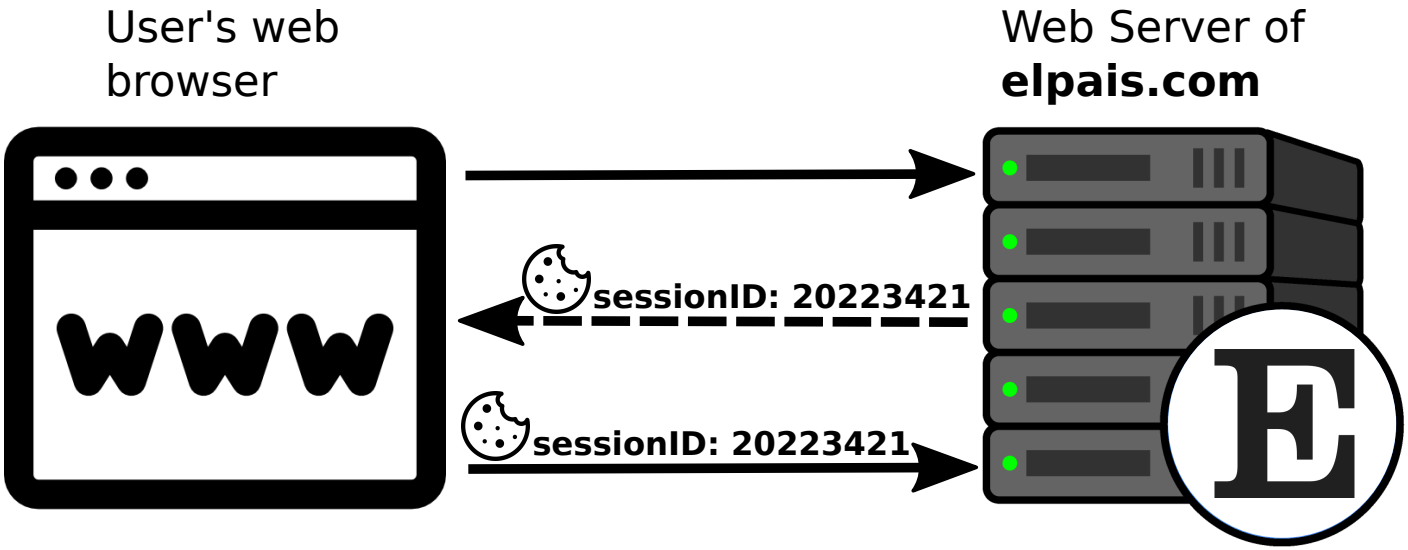
HTTP Cookies



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

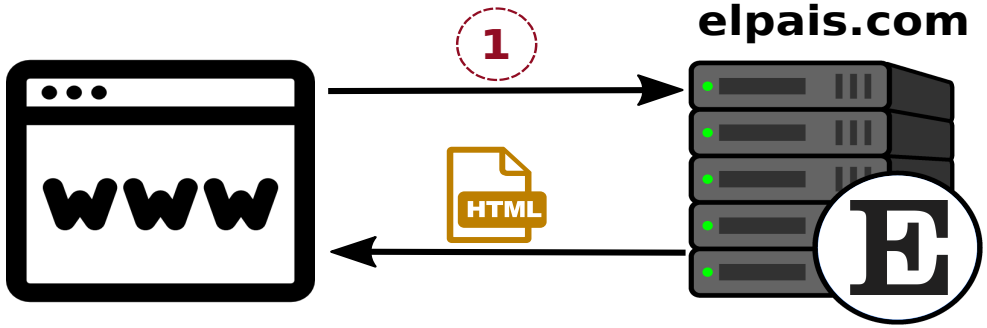
2) The web server sends a response with a "Set-Cookie" header

3) The web browser automatically includes the cookie in *future requests* to `www.corriere.it`

HTTP Cookies in Real-world

HTML contains:

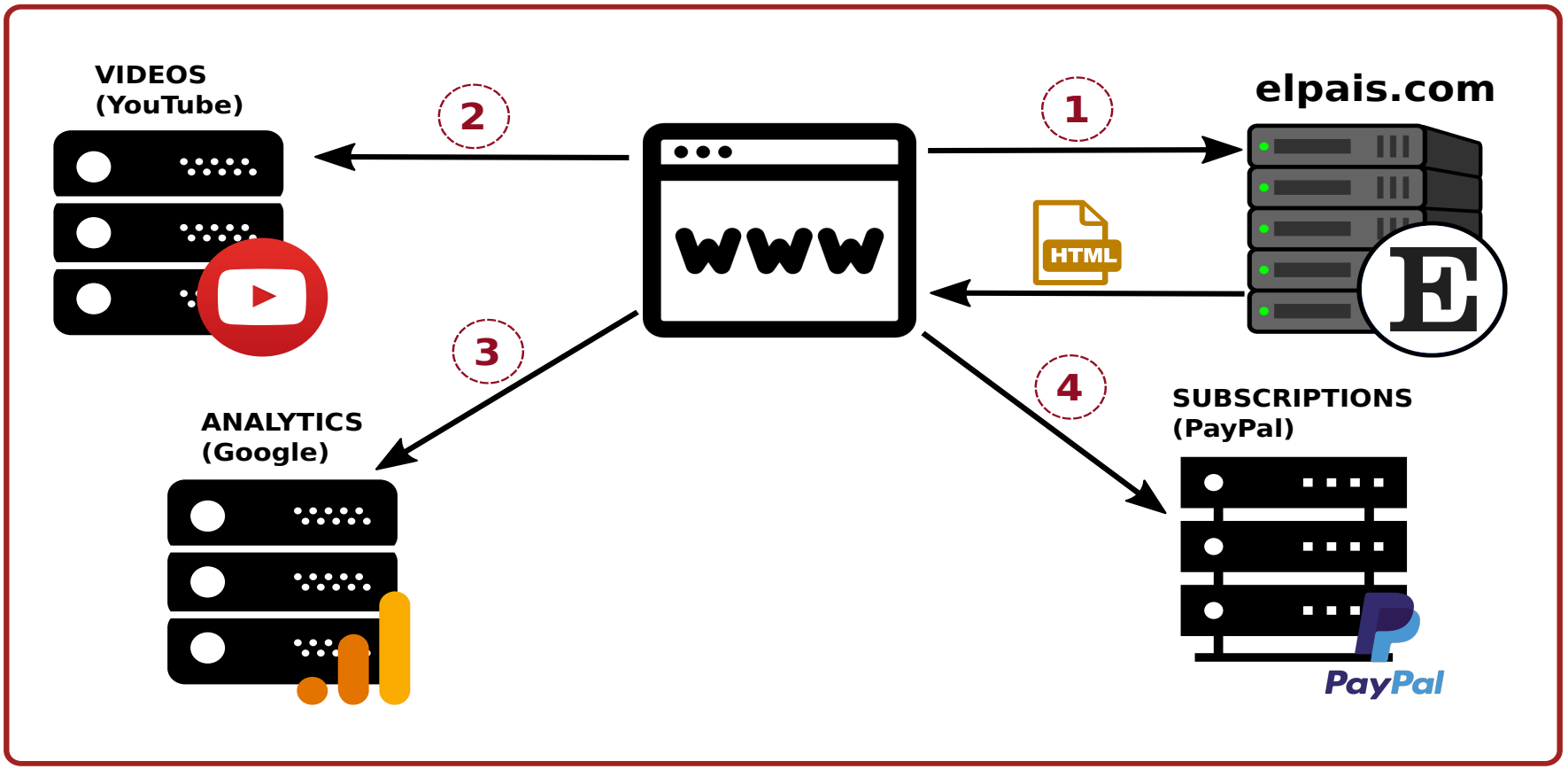
- hyperlinks
- images
- scripts



HTTP Cookies in Real-world

HTML contains:

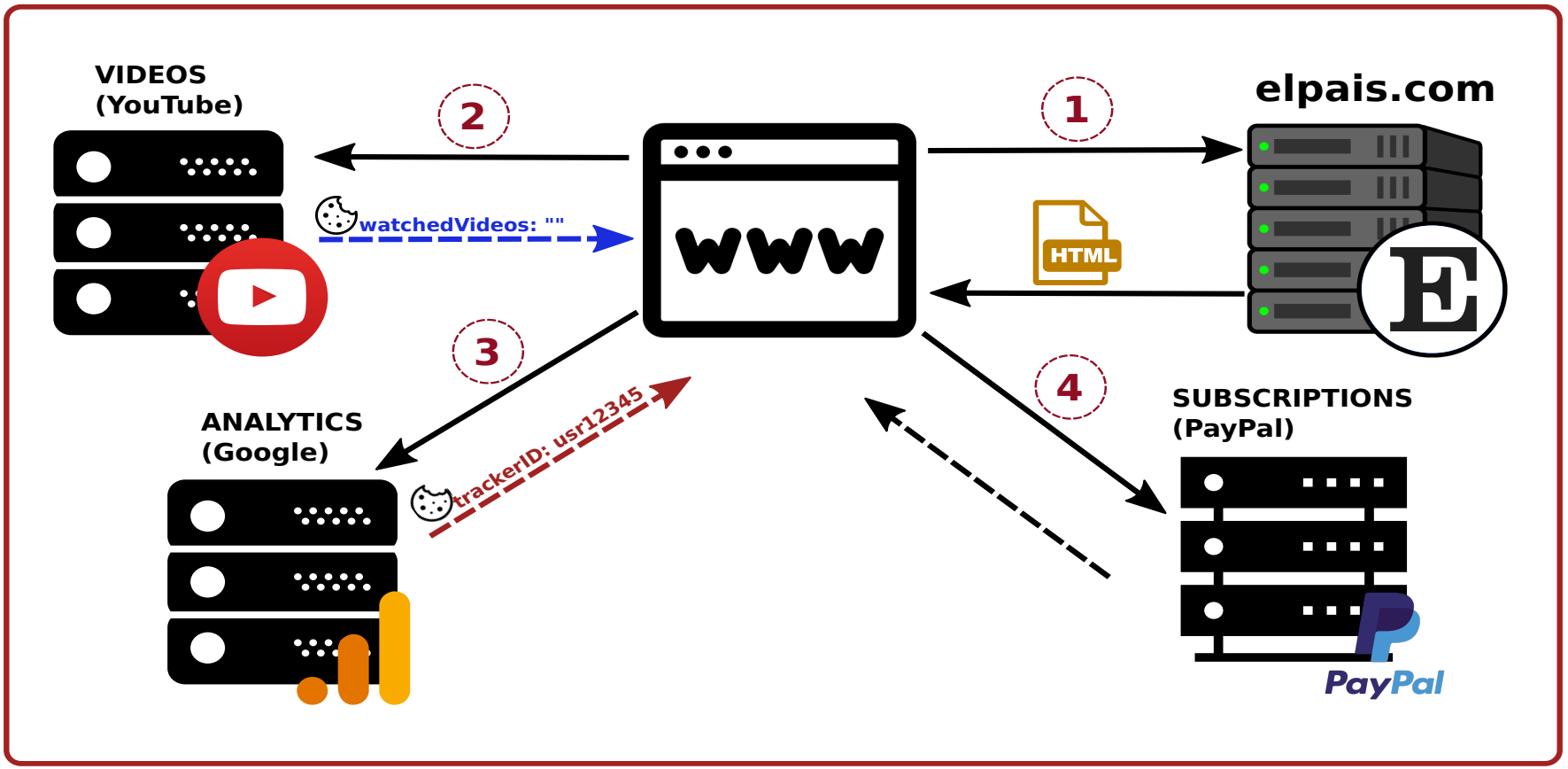
- hyperlinks
- images
- scripts



HTTP Cookies in Real-world

HTML contains:

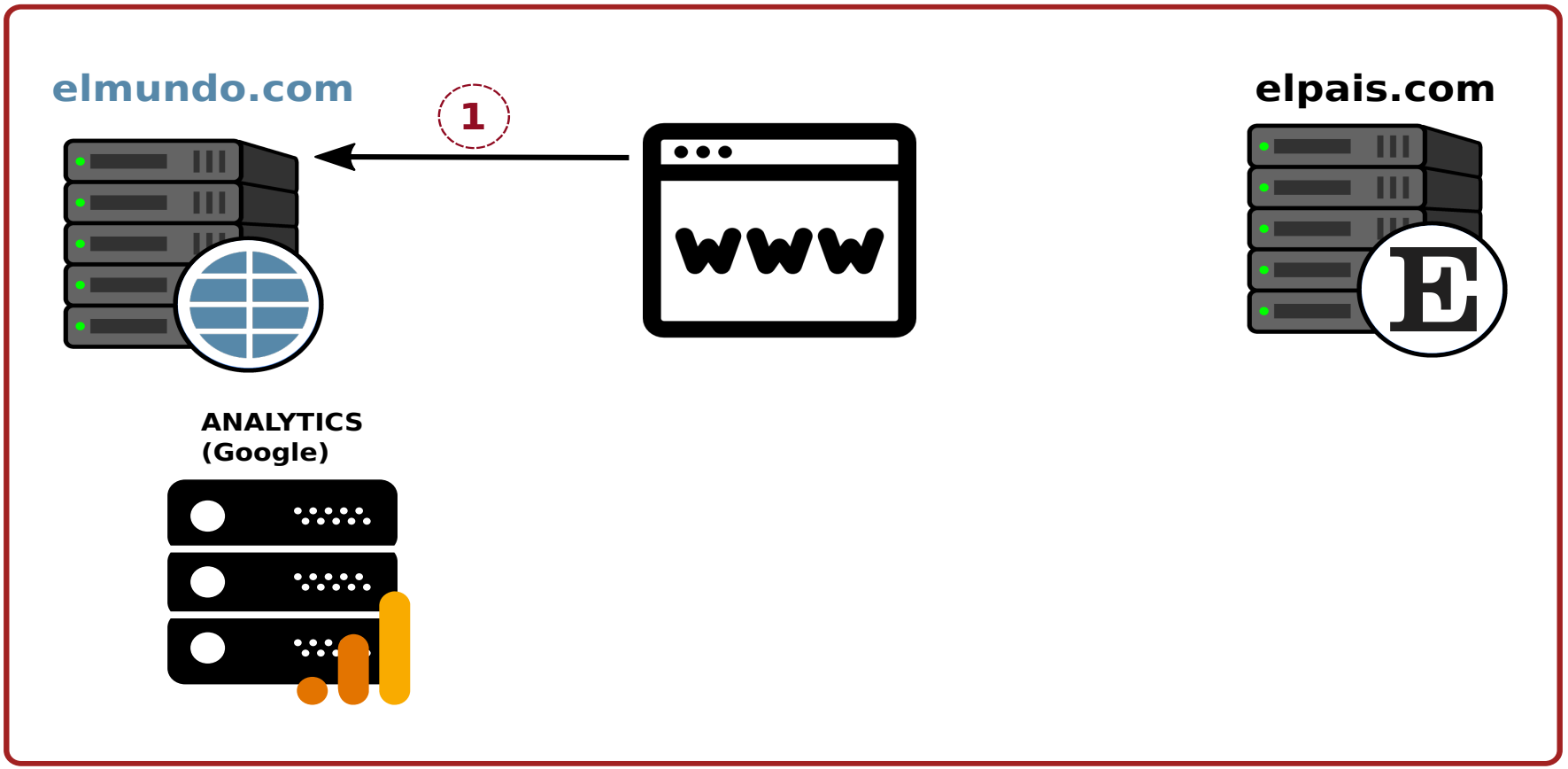
- hyperlinks
- images
- scripts



HTTP Cookies in Real-world

HTML contains:

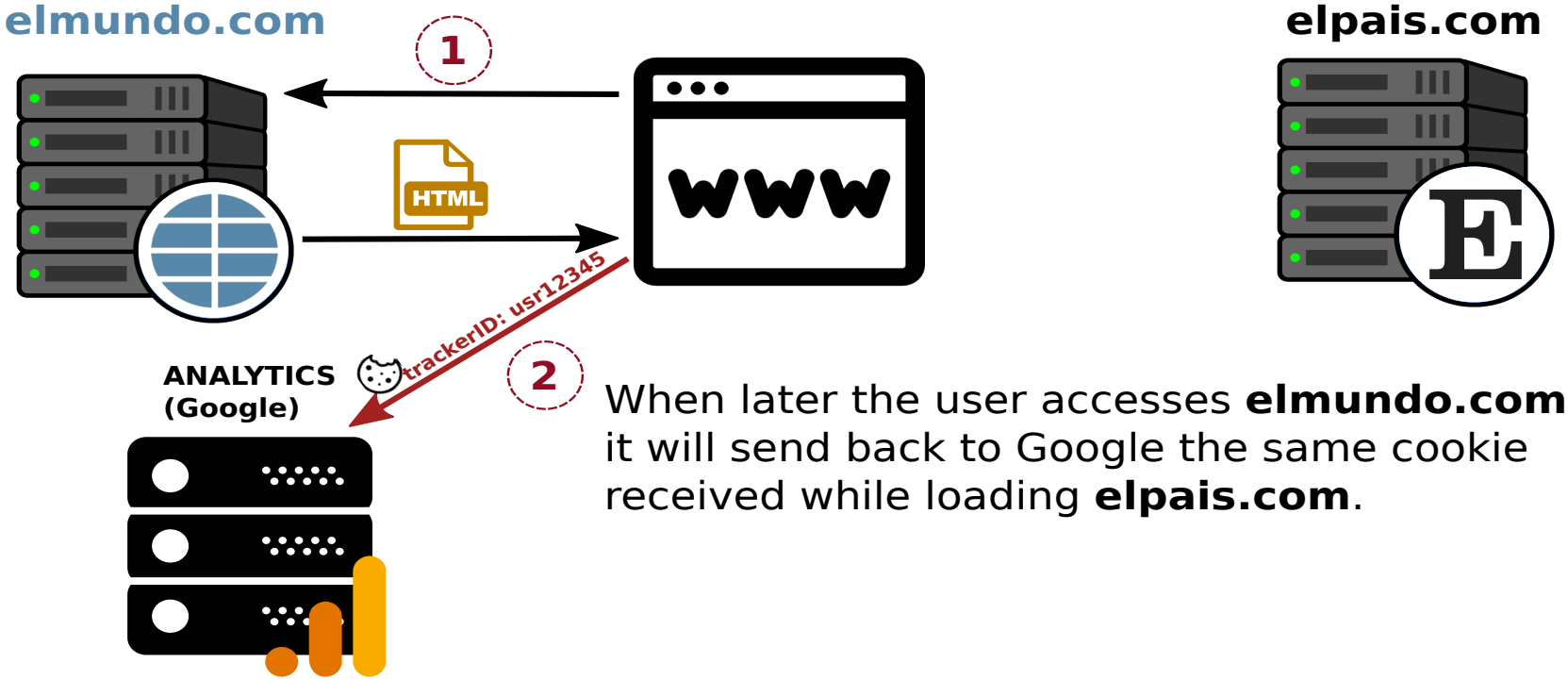
- hyperlinks
- images
- scripts



HTTP Cookies in Real-world

HTML contains:

- hyperlinks
- images
- scripts

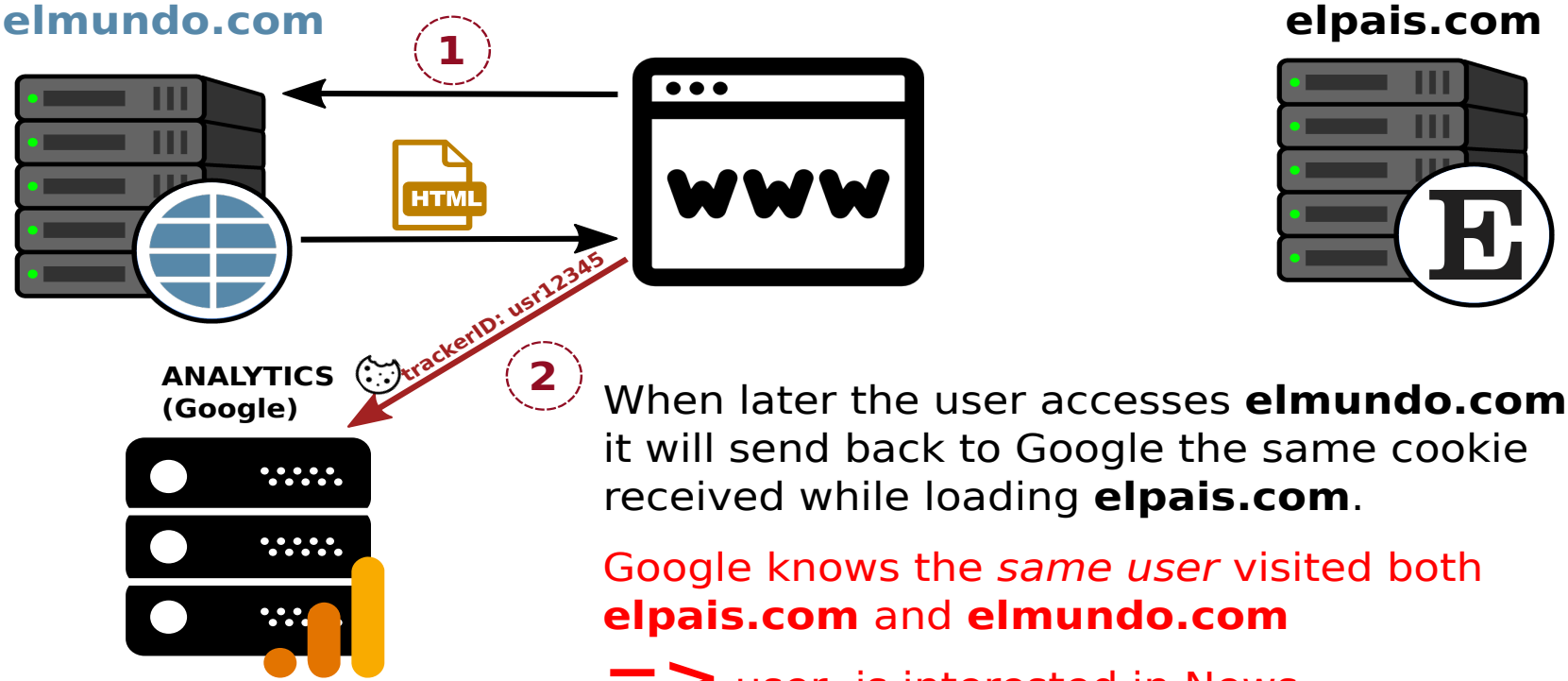


When later the user accesses **elmundo.com** it will send back to Google the same cookie received while loading **elpais.com**.

HTTP Cookies in Real-world

HTML contains:

- hyperlinks
- images
- scripts

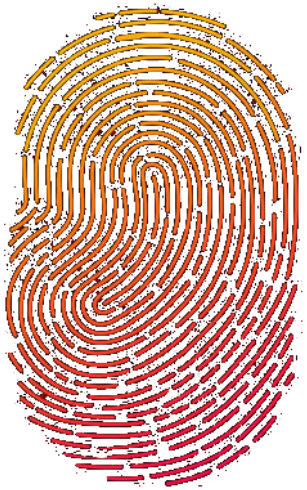


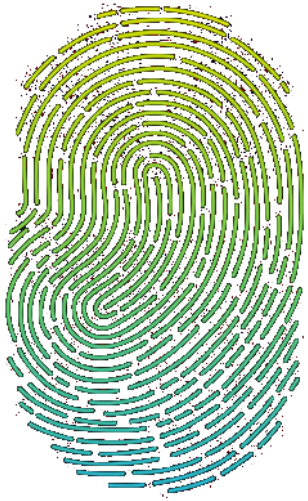
When later the user accesses **elmundo.com** it will send back to Google the same cookie received while loading **elpais.com**.

Google knows the *same user* visited both **elpais.com** and **elmundo.com**

=> user is interested in News

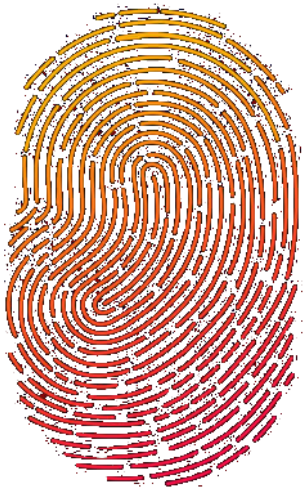
Digital Profiles



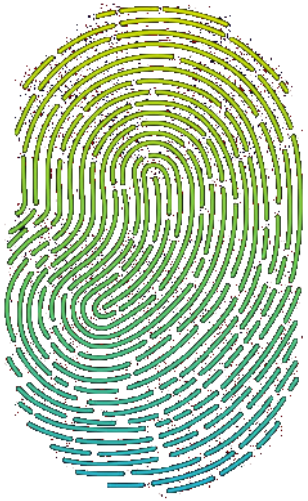


Digital Profiles

idealista



idealista



Autocasión

Digital Profiles

"*Women less likely to be shown ads for high-paid jobs on Google, study shows*" [1,2]

1. Guardian article: `https://tinyurl.com/omen-less-likely-ads`
2. PoPETS paper: `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 ⇒ **explicit** and **informed** consent
3. GDPR: **Extraterritorial** scope

Impact of GDPR (et al.)

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: Measuring the GDPR's Impact on Web Privacy **2019**

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? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework **2020**

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

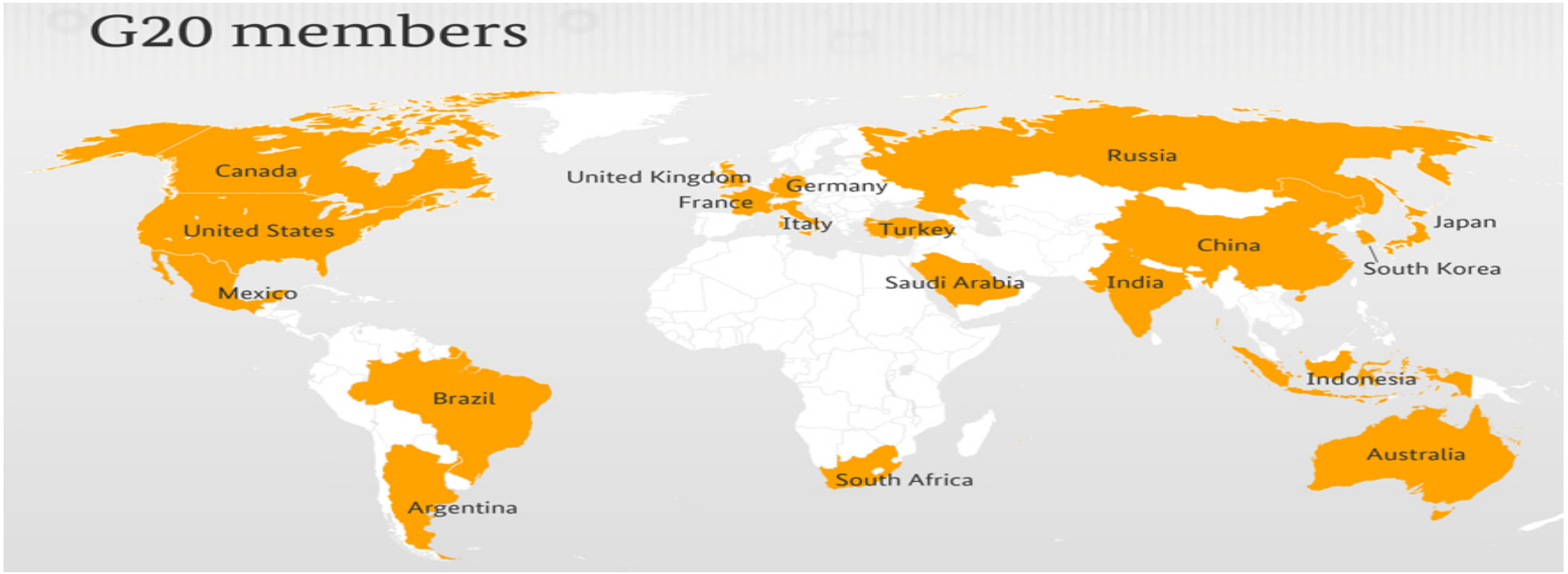
Fines for GDPR violations and on-compliance[1]: **30k** Vueling (2019), **2.2M** Carrefour (France 2020), **50M** + **50M** Google (2019, 2020)

[1] SOURCE: `https://www.cookielawinfo.com/major-gdpr-fines-2020-2021/`

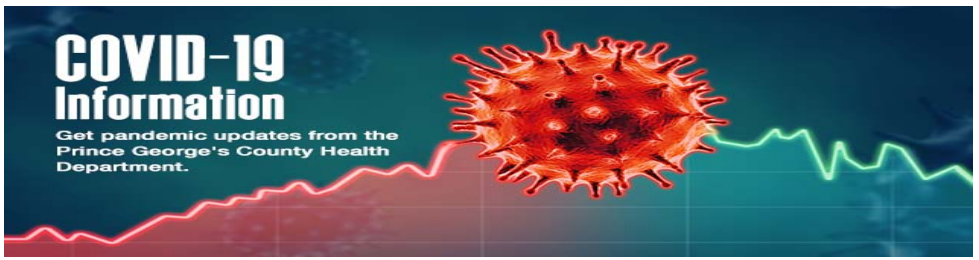
Audit Governmental and Official Websites

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

Our Study



International Organizations

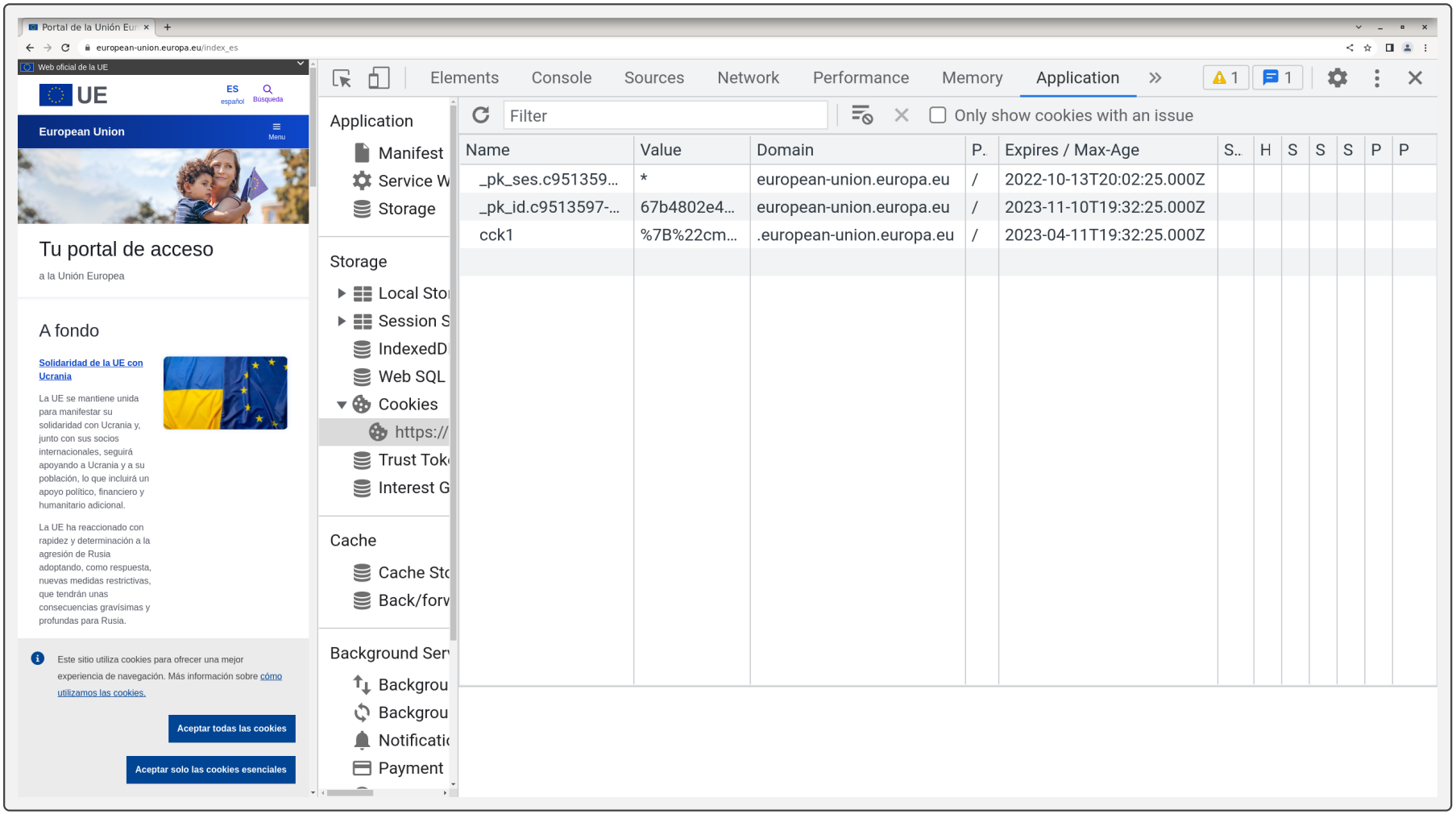


 # Websites:

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

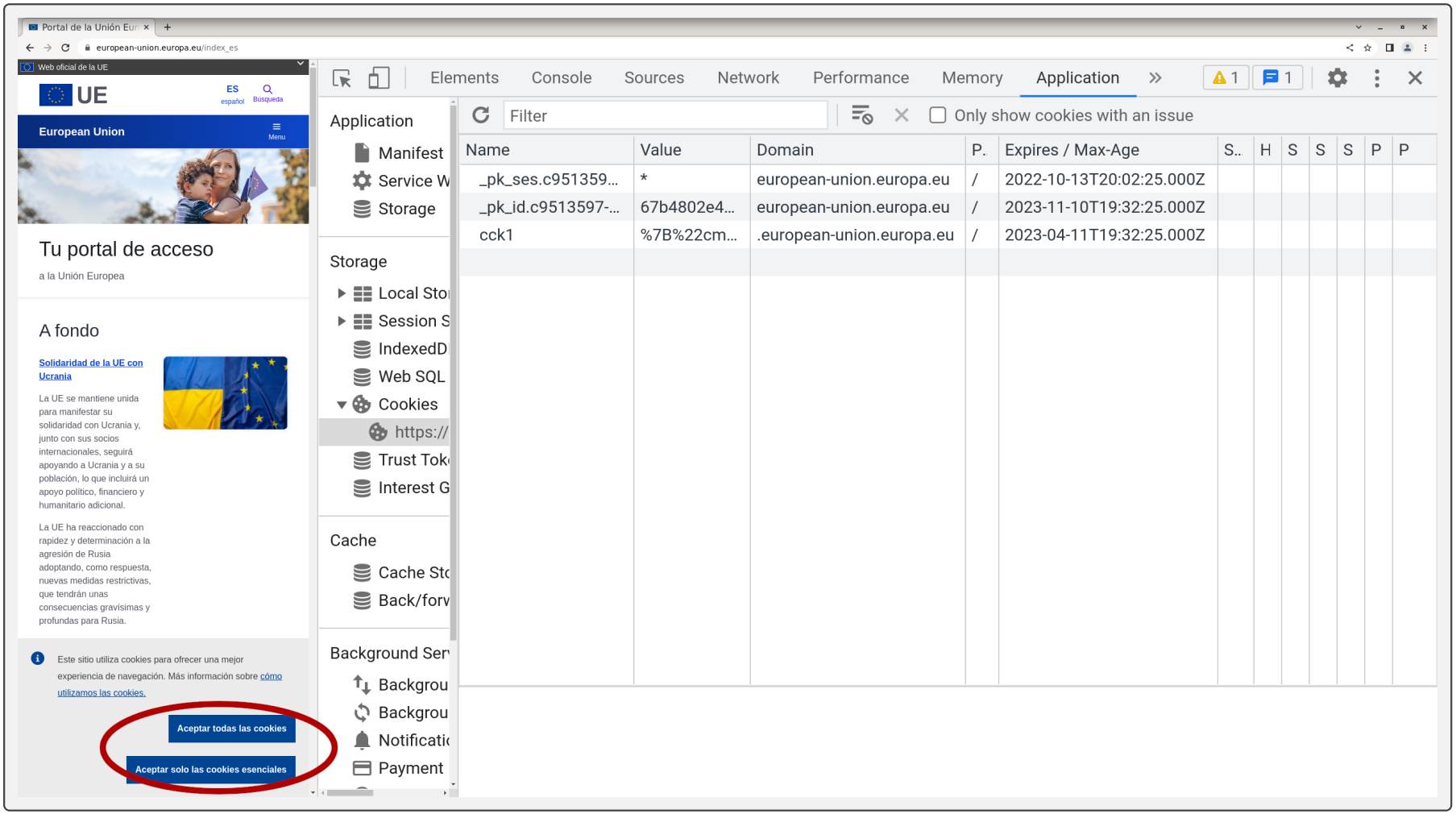
Example#1

URL: `https://european-union.europa.eu/`



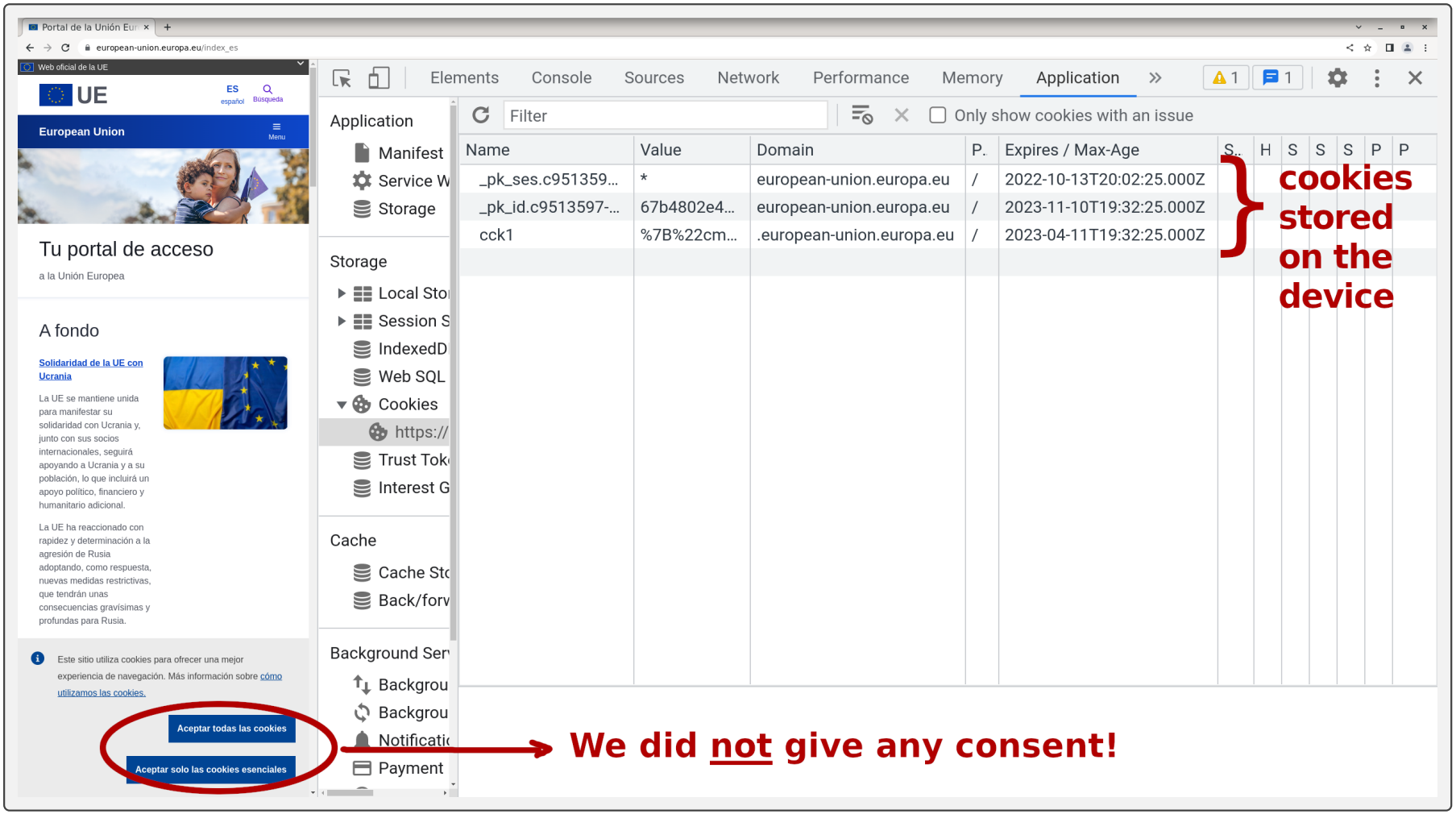
Example#1

URL: `https://european-union.europa.eu/`



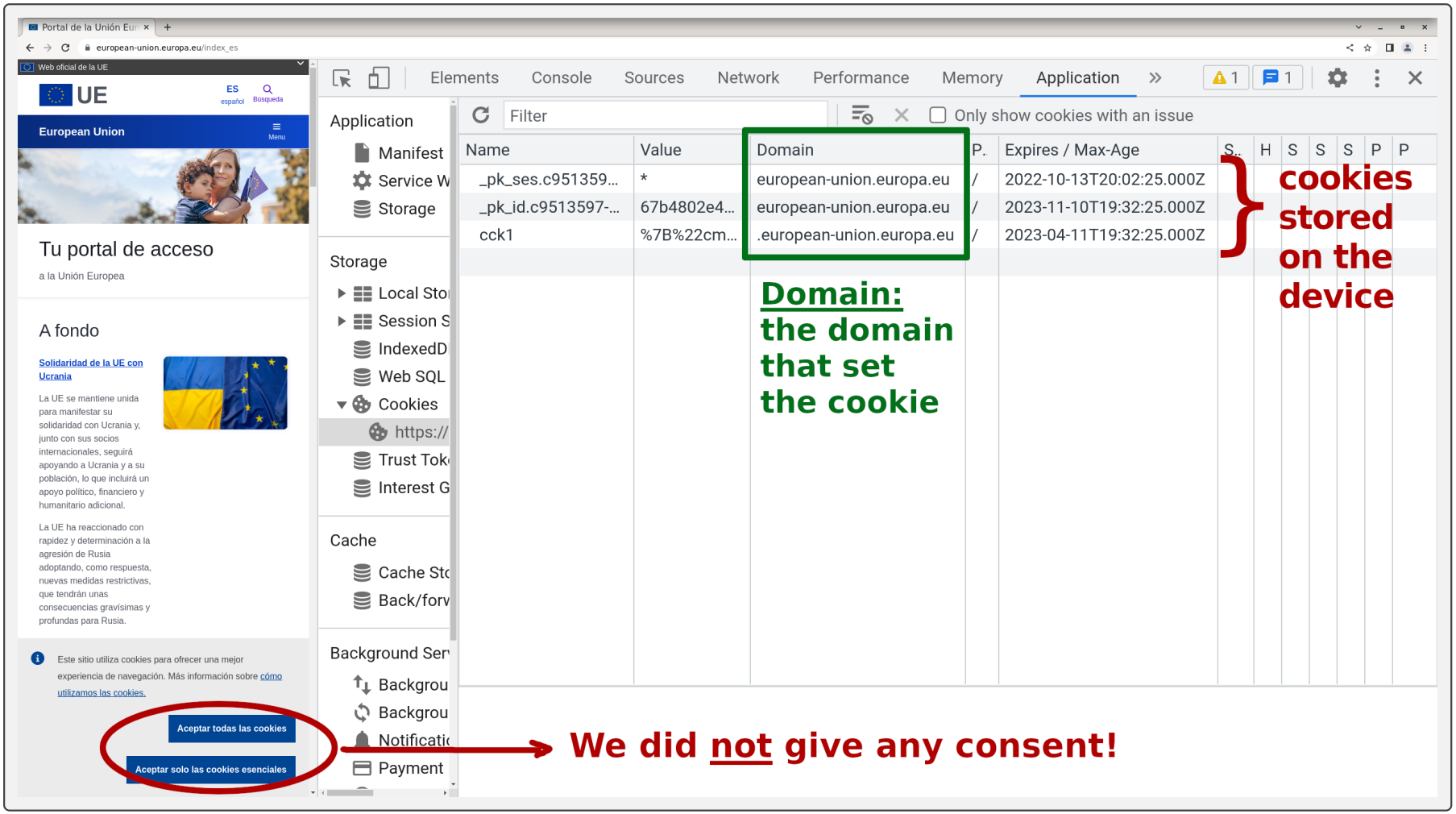
Example#1

URL: `https://european-union.europa.eu/`



Example#1

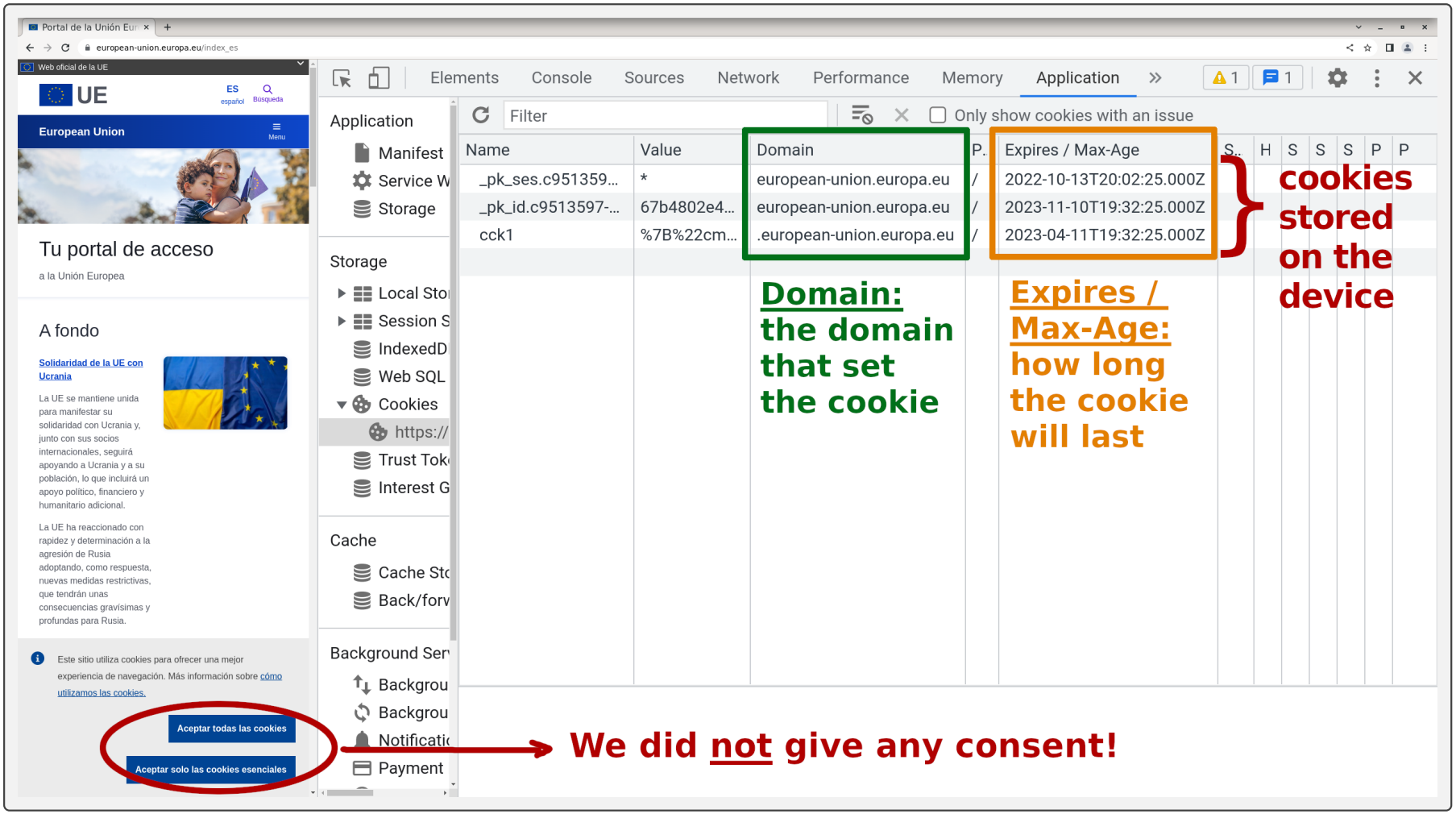
URL: `https://european-union.europa.eu/`



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

Example#1

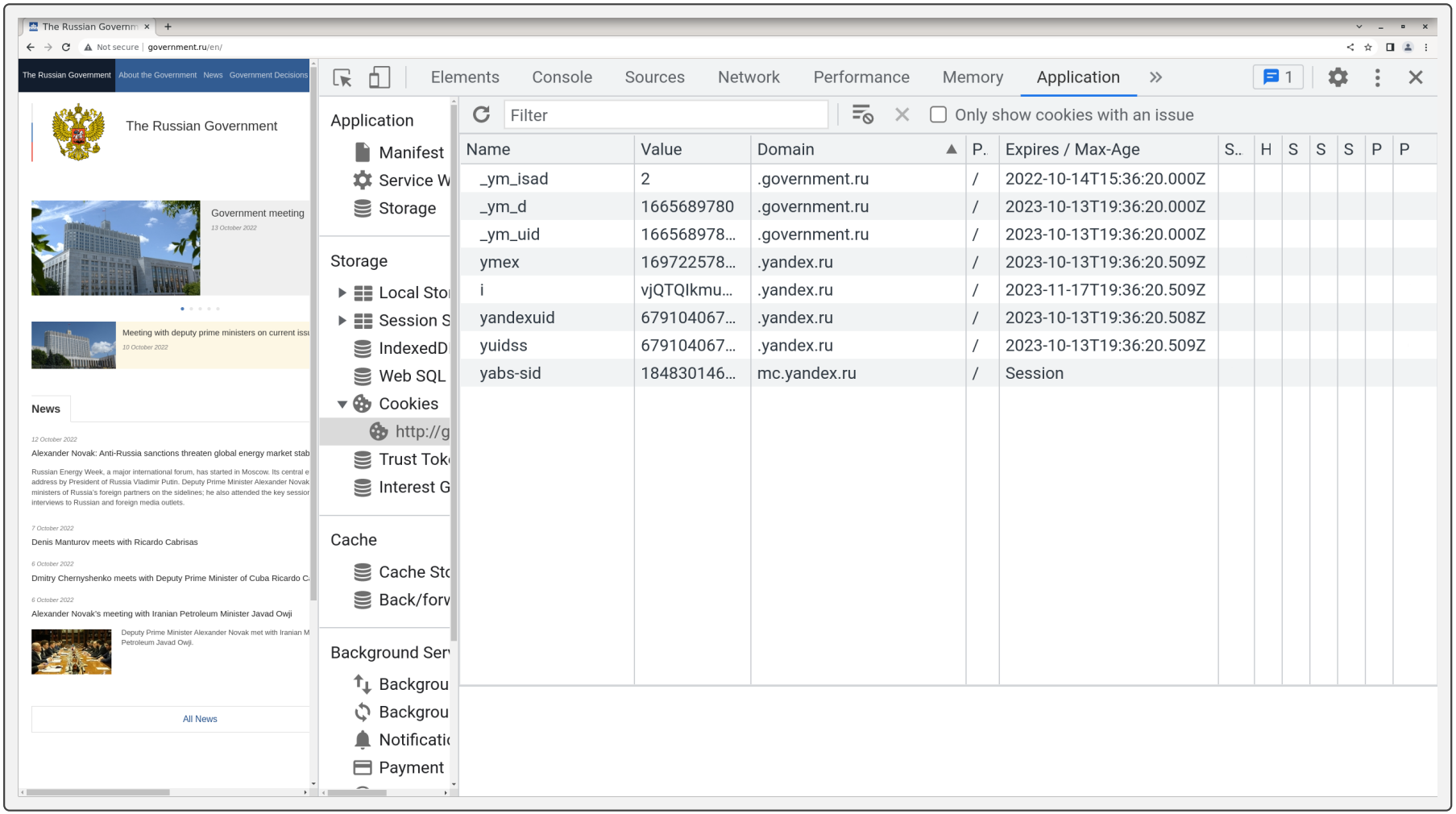
URL: `https://european-union.europa.eu/`



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

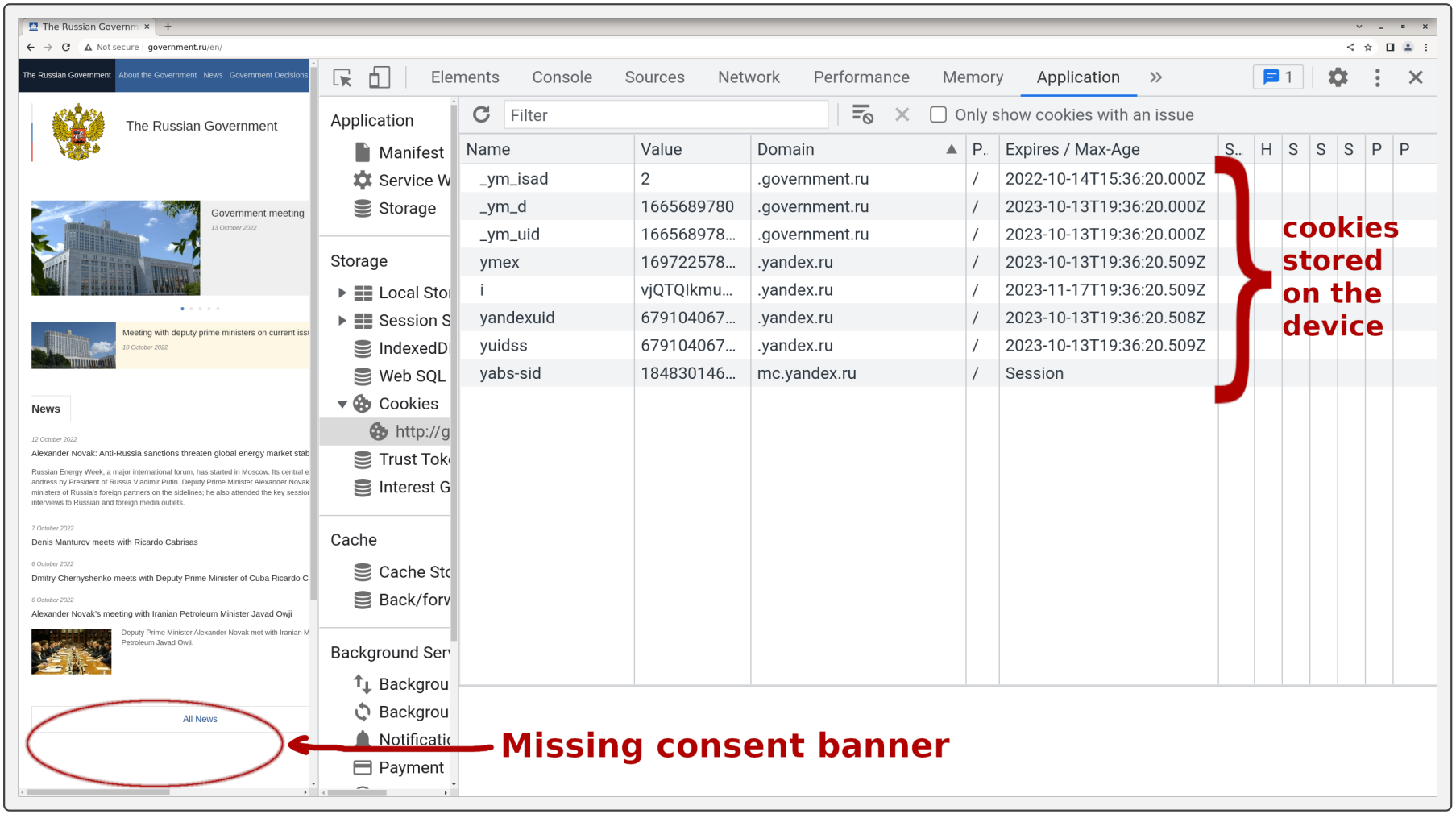
Example#2

URL: `https://government.ru/`



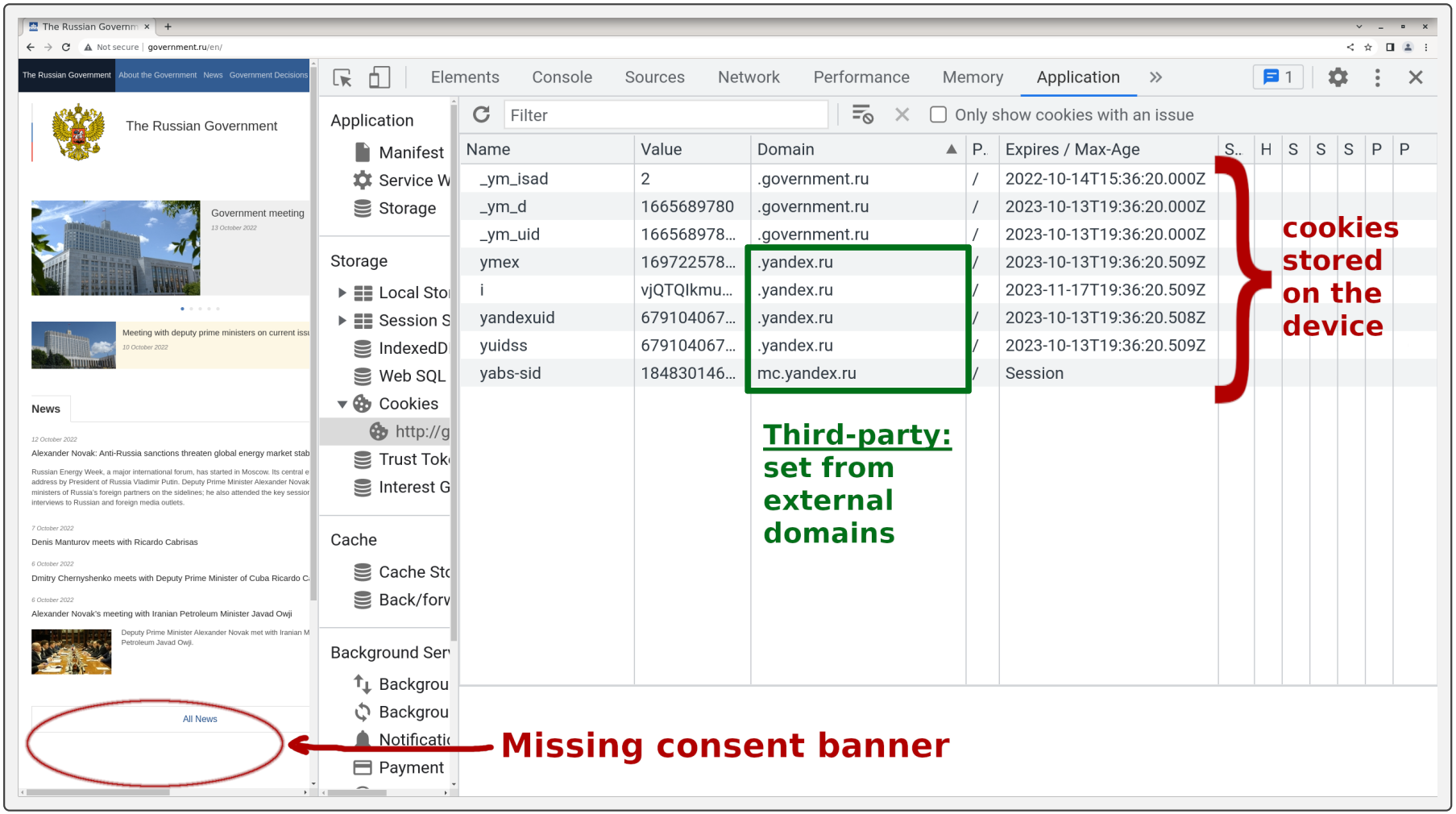
Example#2

URL: `https://government.ru/`



Example#2

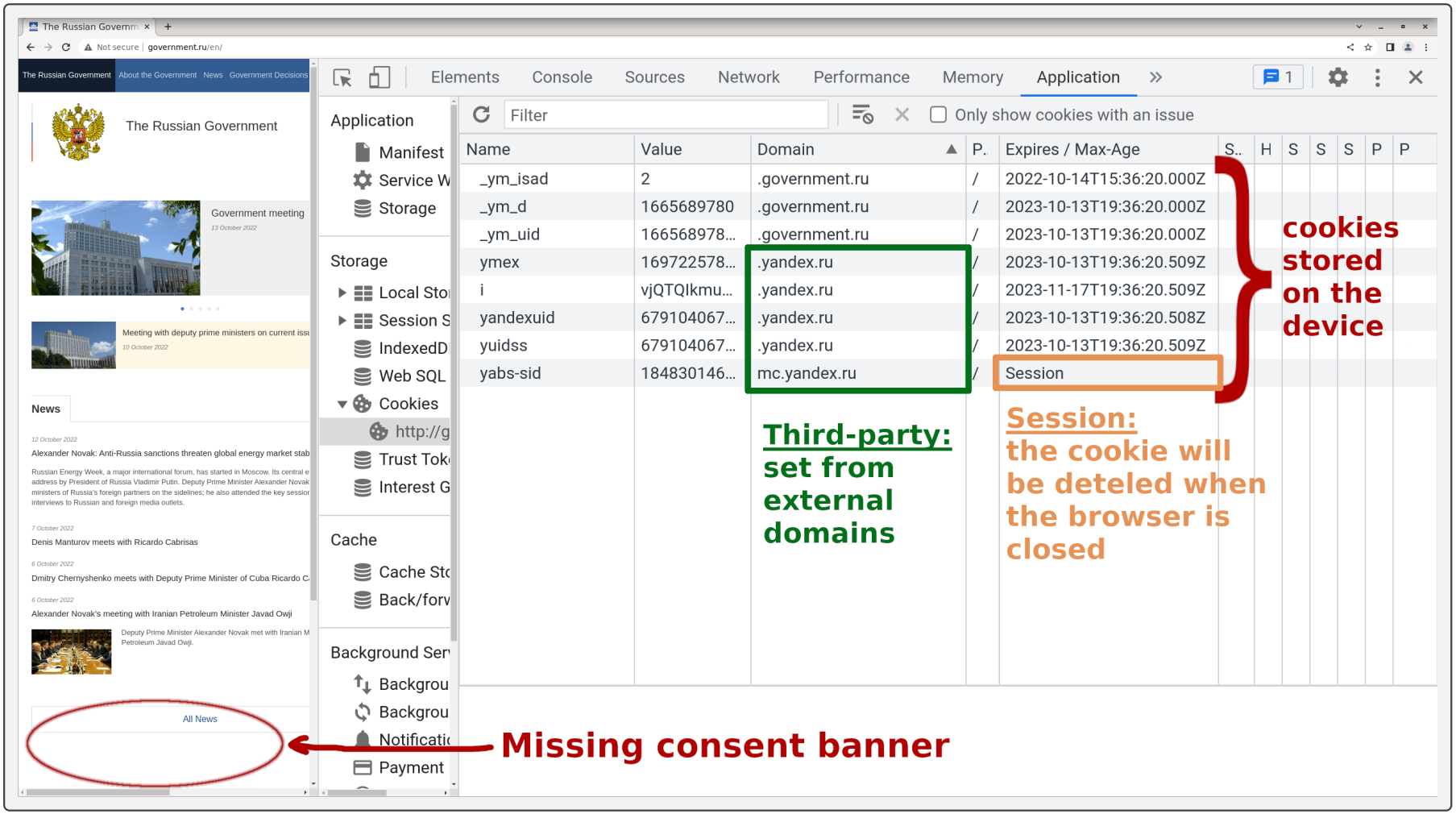
URL: `https://government.ru/`



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

Example#2

URL: `https://government.ru/`



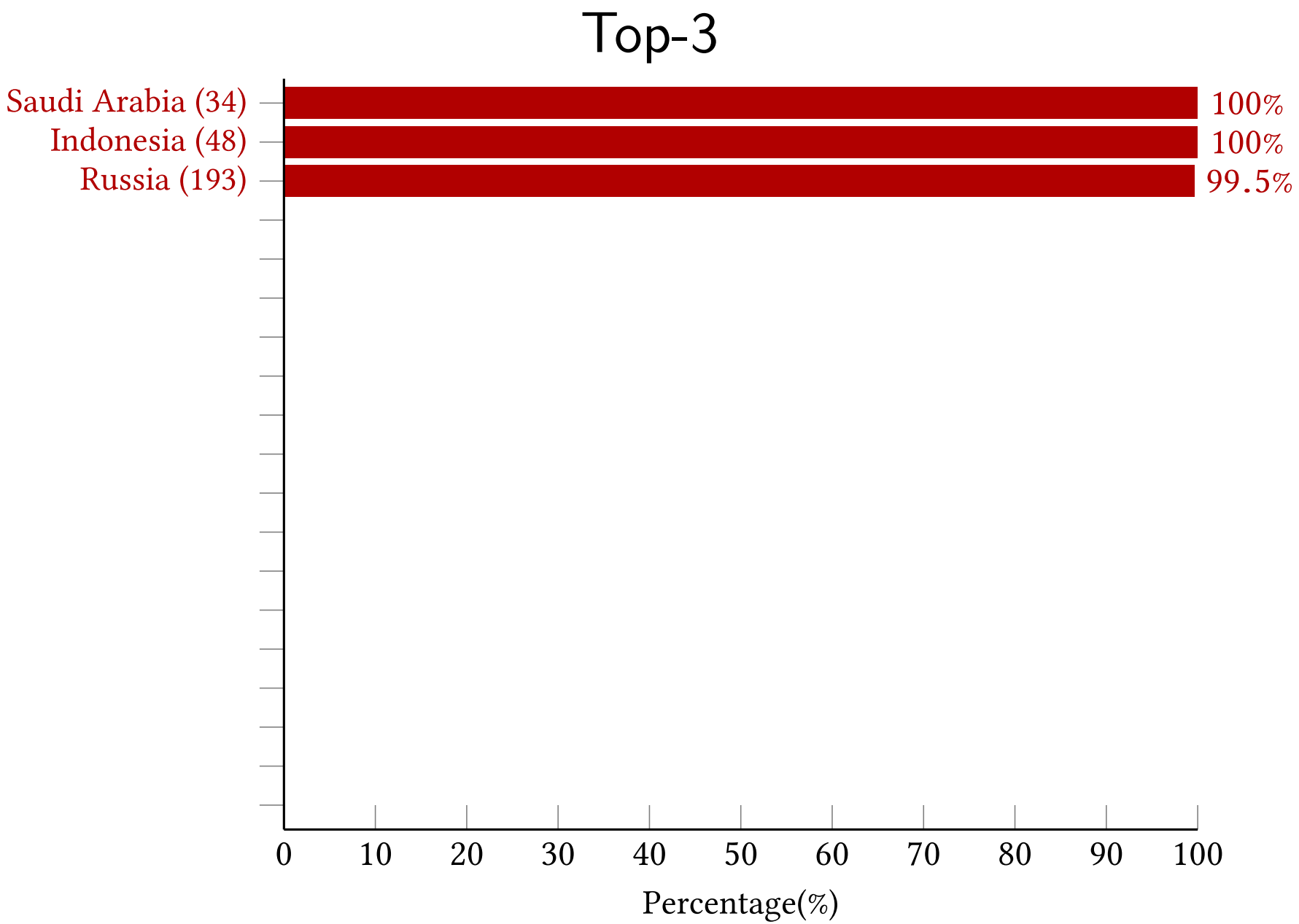
cookie origin $\neq$ visited domain $\Longrightarrow$ **Third-Party (TP)** cookie

Approach

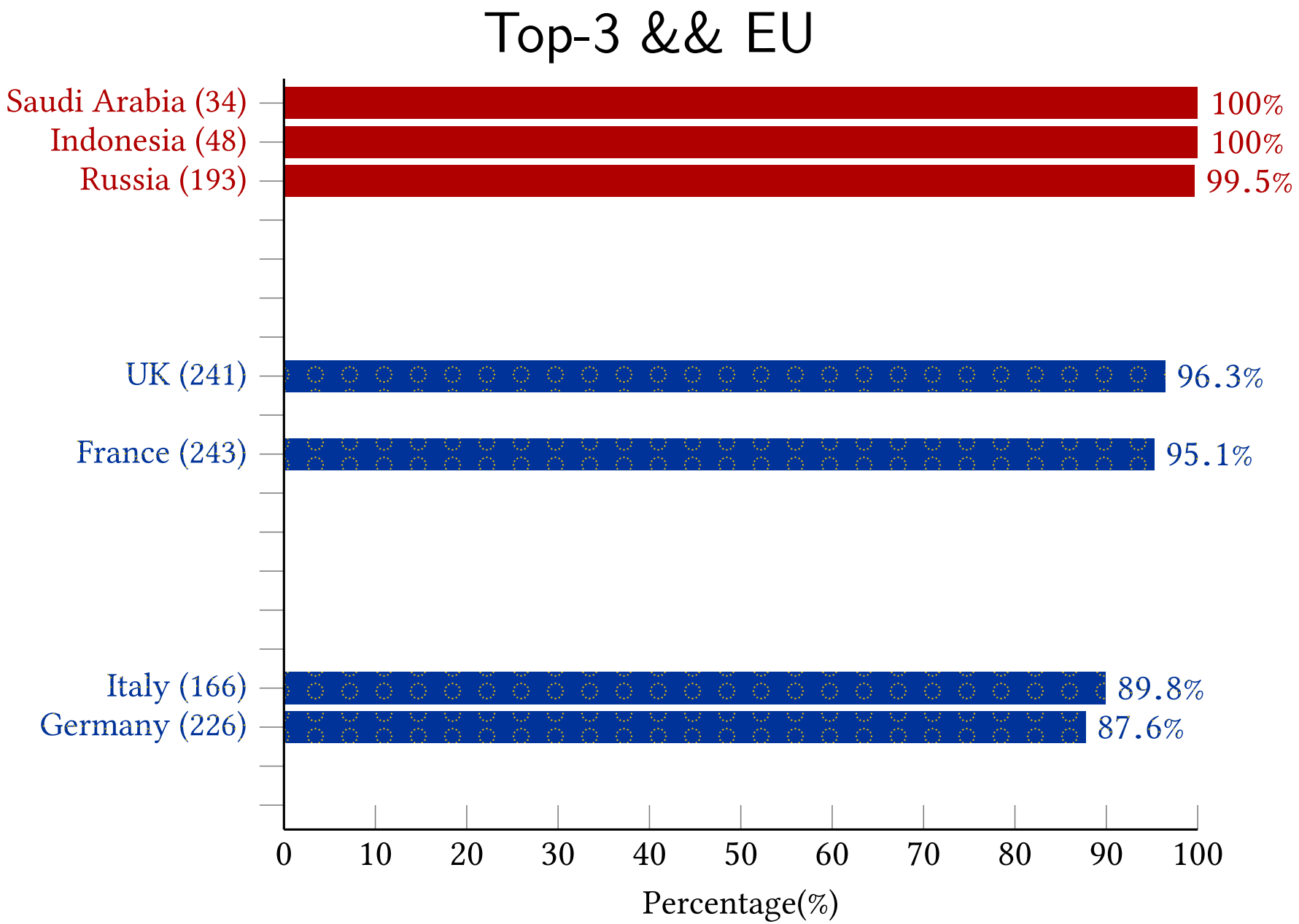
1. **Visit websites**
   - collect cookies with an instrumented web browser
   - 10 URLs per website
     - only HTML content and clickable hyperlinks
   - NO consent given (or any action performed)

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

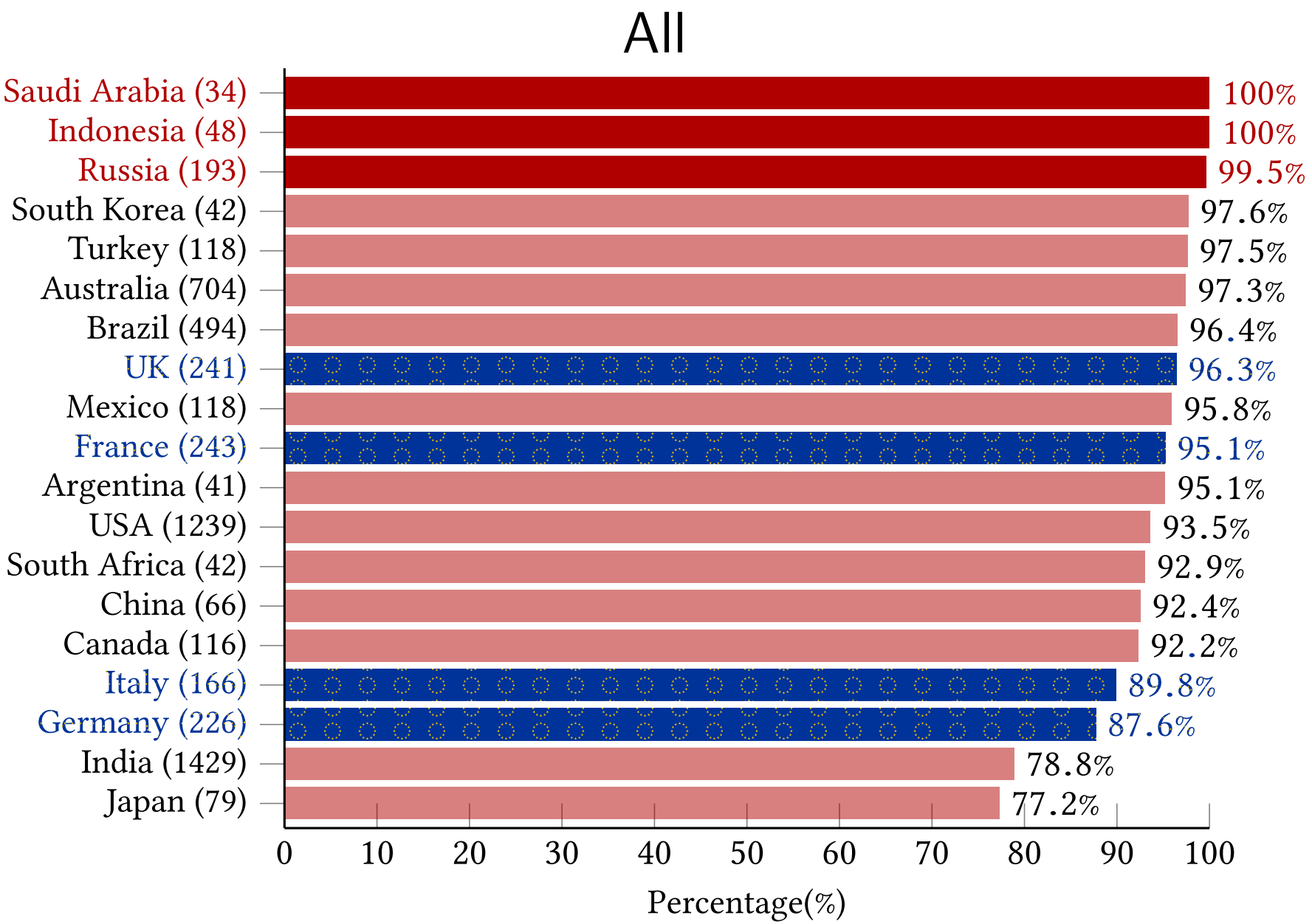
Do G20 Websites Set Cookies?



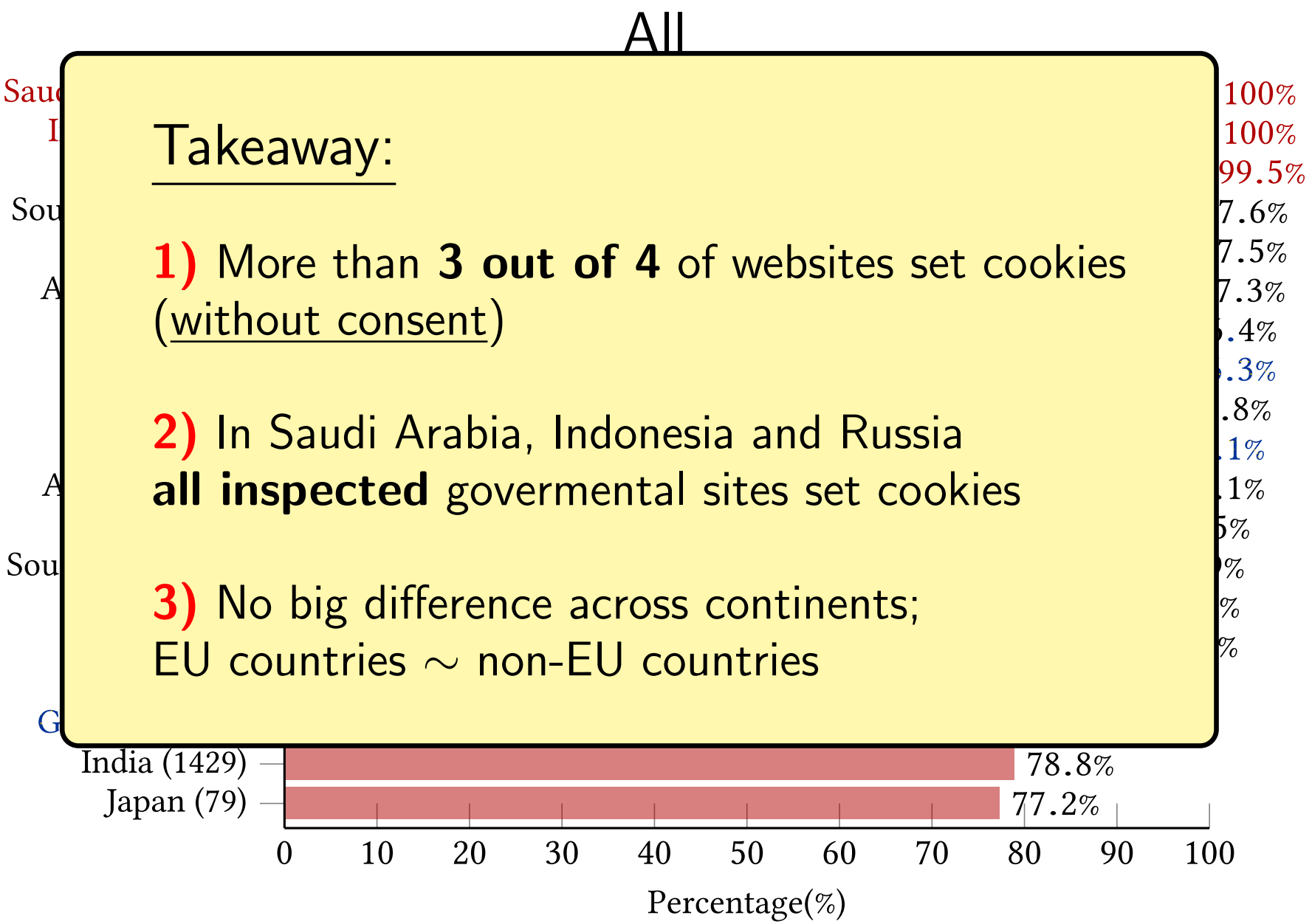
Do G20 Websites Set Cookies?



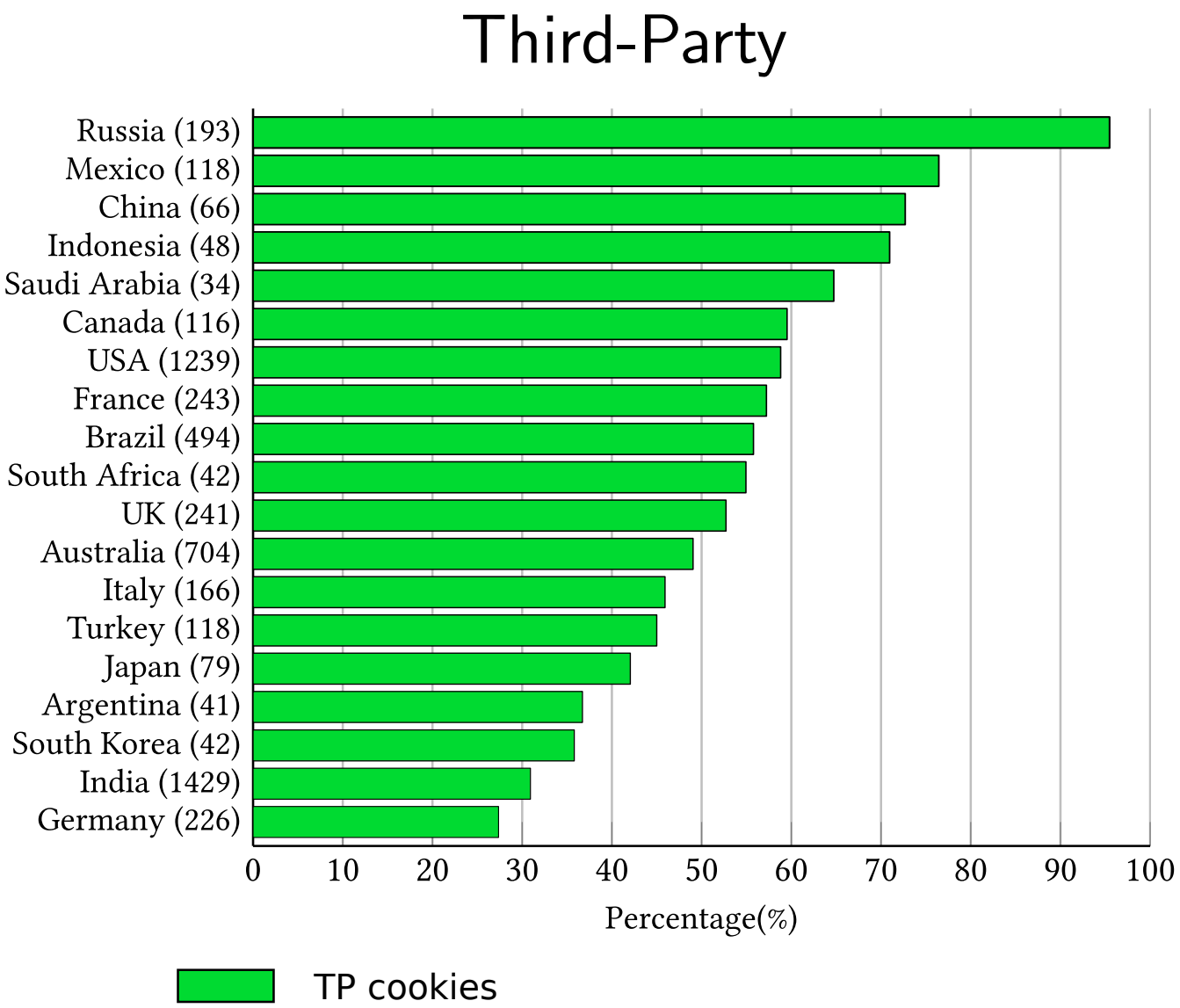
Do G20 Websites Set Cookies?



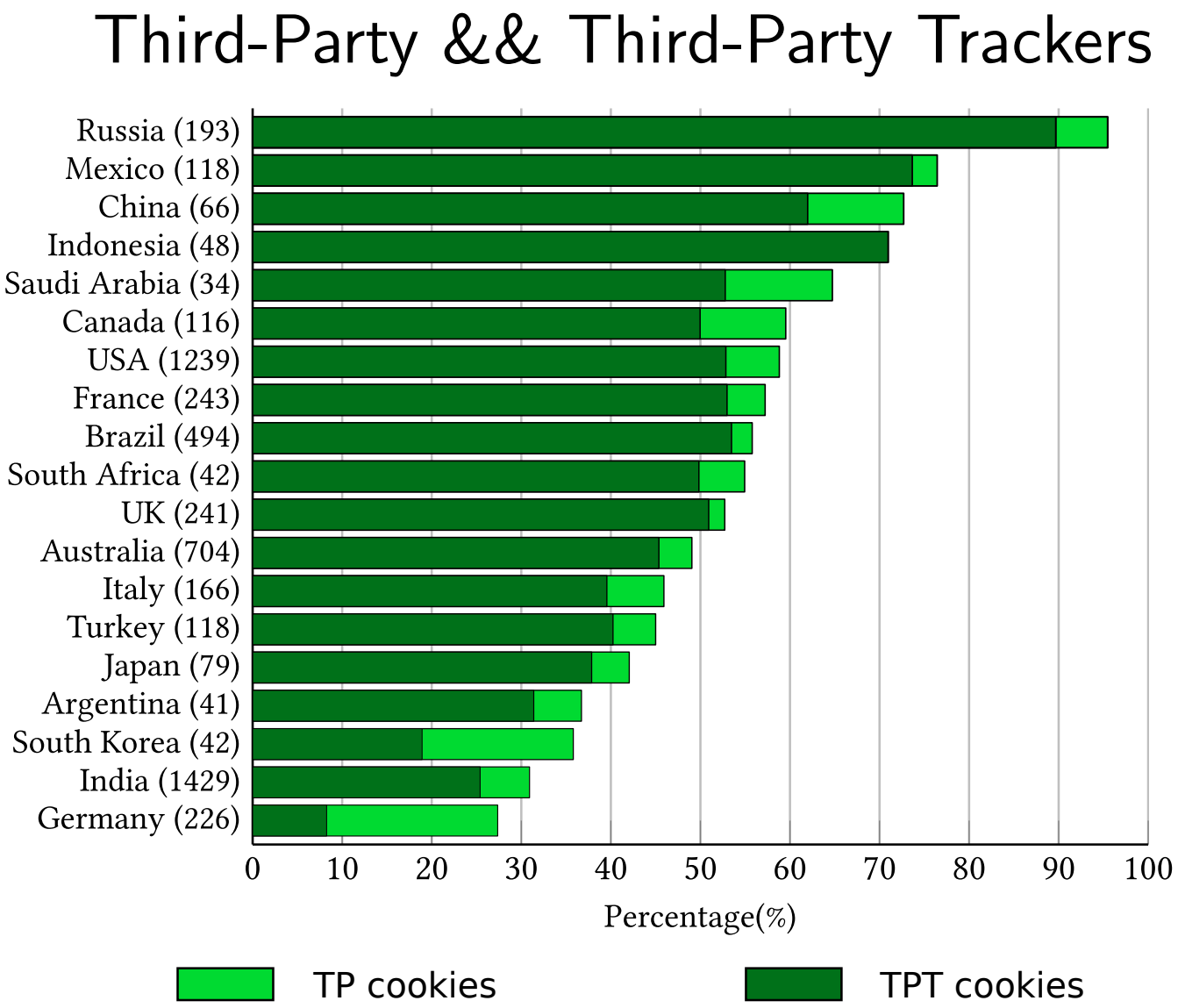
Do G20 Websites Set Cookies?



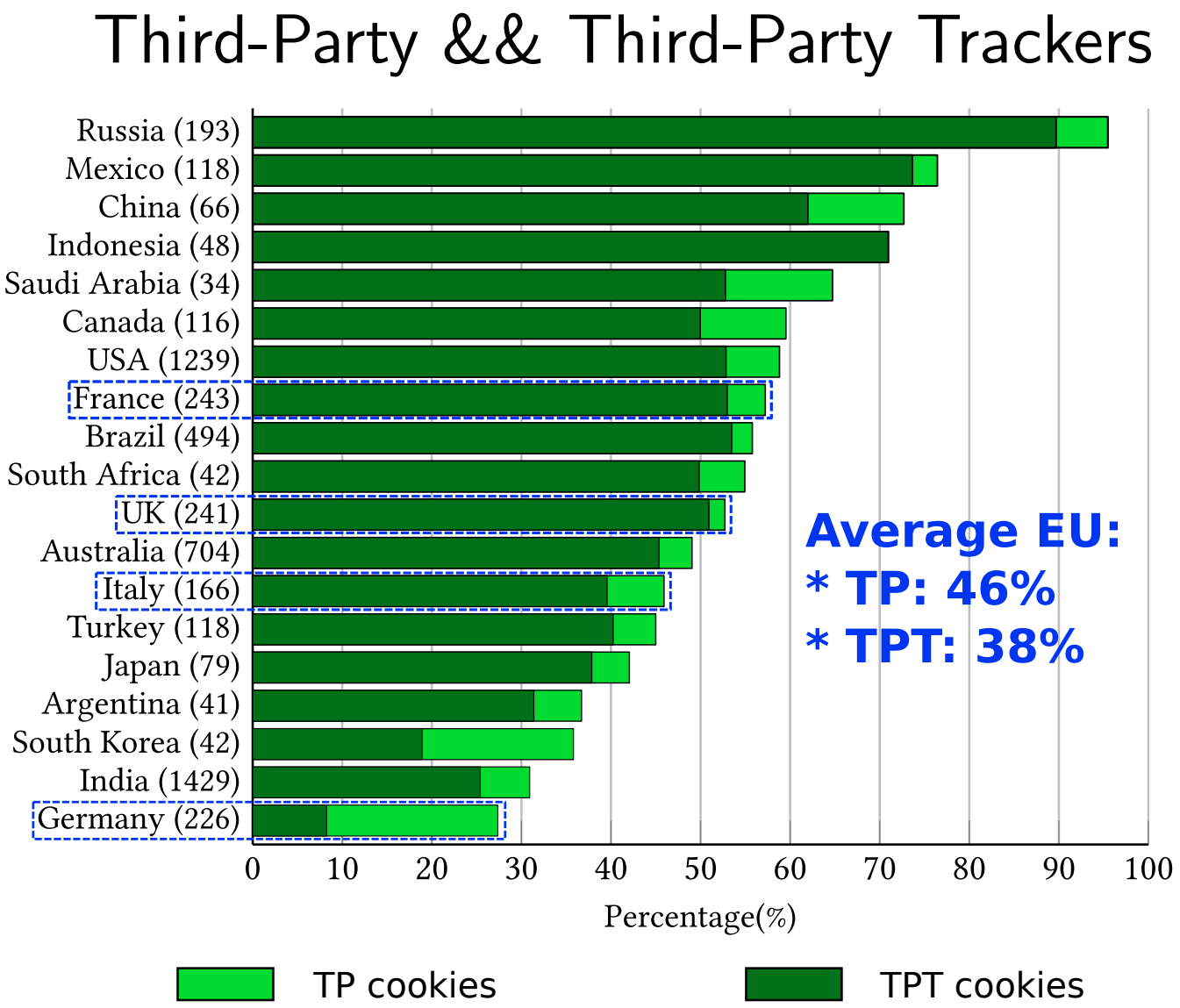
Do G20 Websites Set Third-Party Cookies?



Do G20 Websites Set Third-Party Cookies?



Do G20 Websites Set Third-Party Cookies?



Do G20 Websites Set Third-Party Cookies?

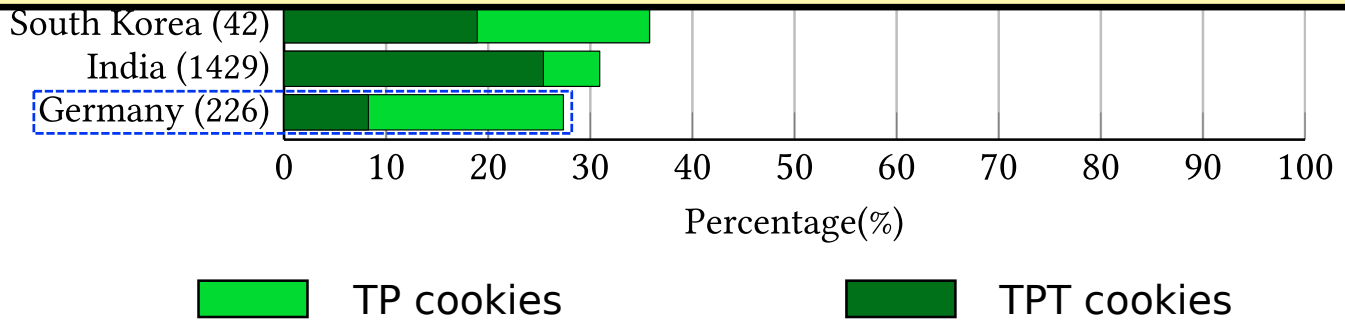
Third-Party && Third-Party Trackers

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 $\Longrightarrow$ they are from **known trackers**

3) Germany as the most privacy-oriented (EU) country

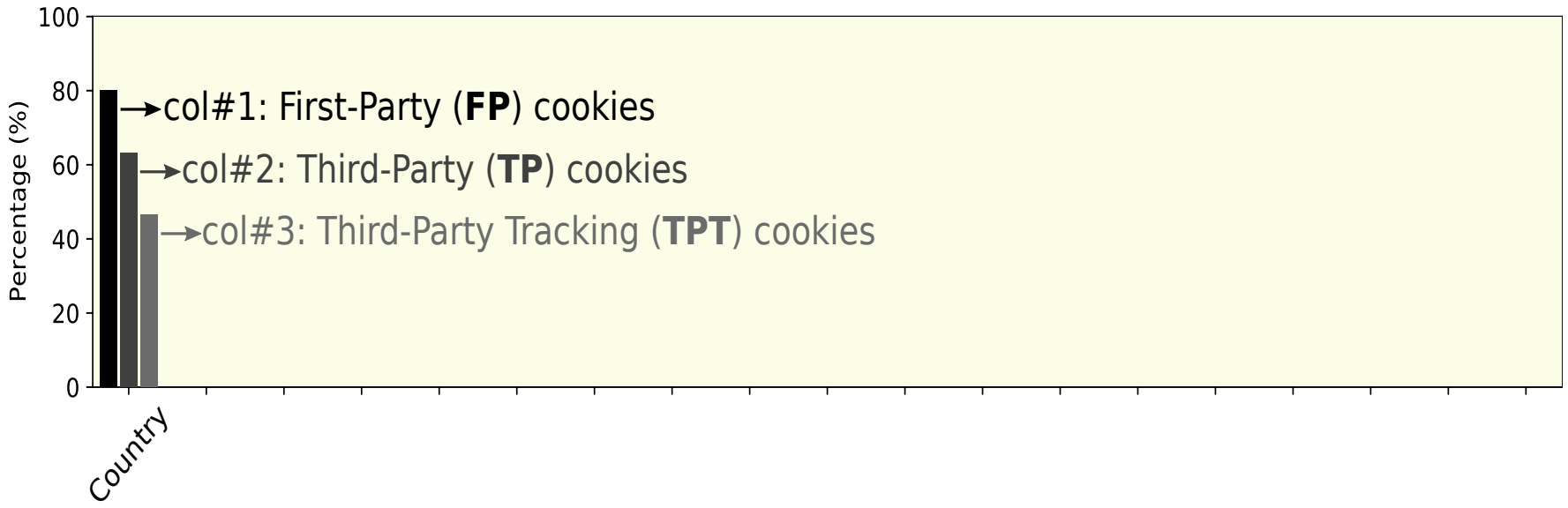


How Long do Cookies Last?

Cookie origin:

FP = First-Party

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



How Long do Cookies Last?

Cookie origin:

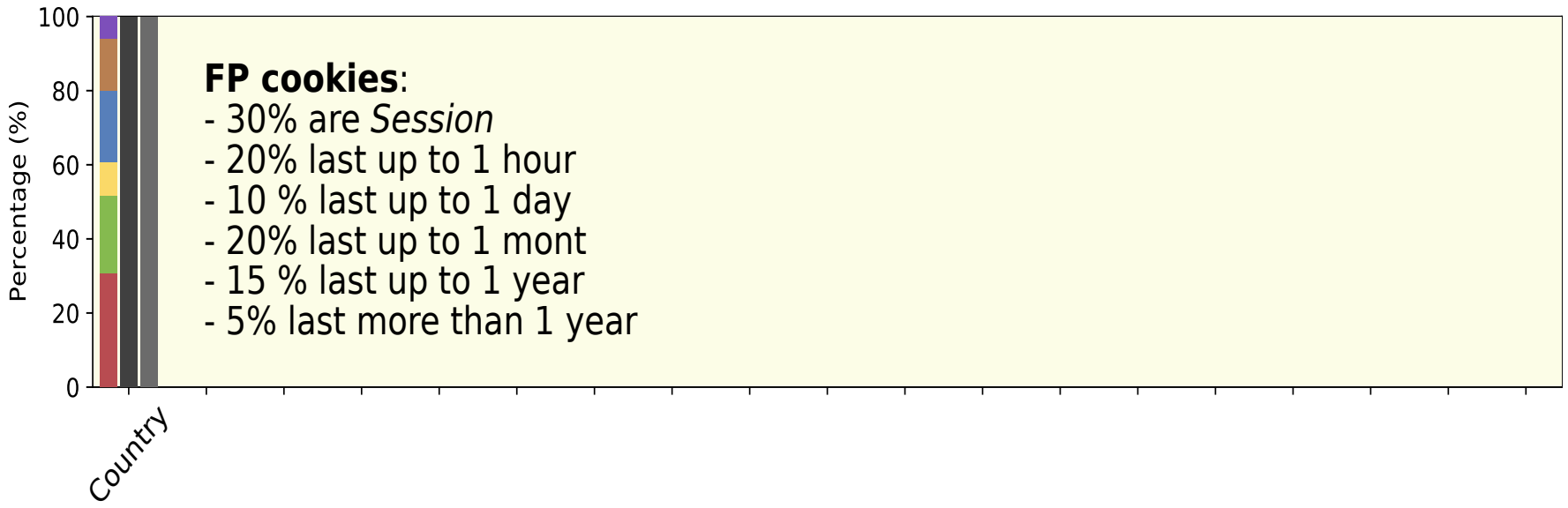
FP = First-Party

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

Cookie duration:

Session 1 hour 1 day

1 month 1 year > 1 year



FP cookies:

- 30% are *Session*
- 20% last up to 1 hour
- 10 % last up to 1 day
- 20% last up to 1 mont
- 15 % last up to 1 year
- 5% last more than 1 year

How Long do Cookies Last?

Cookie origin:

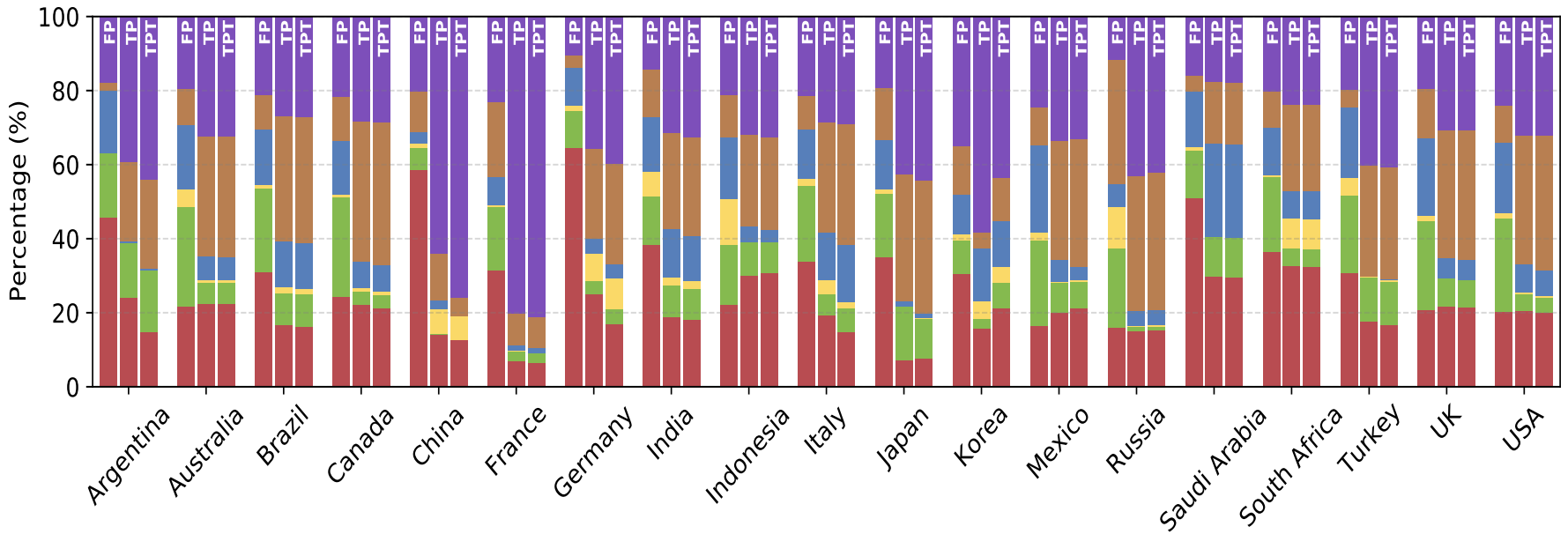
FP = First-Party

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:

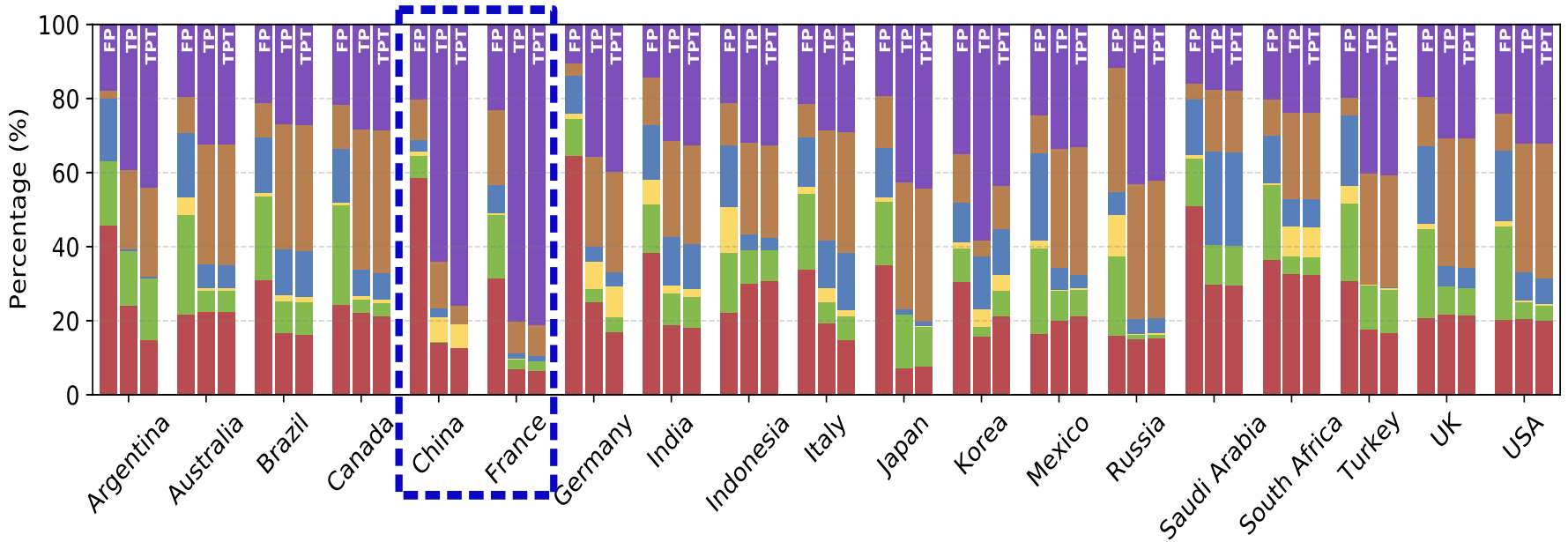
FP = First-Party

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:

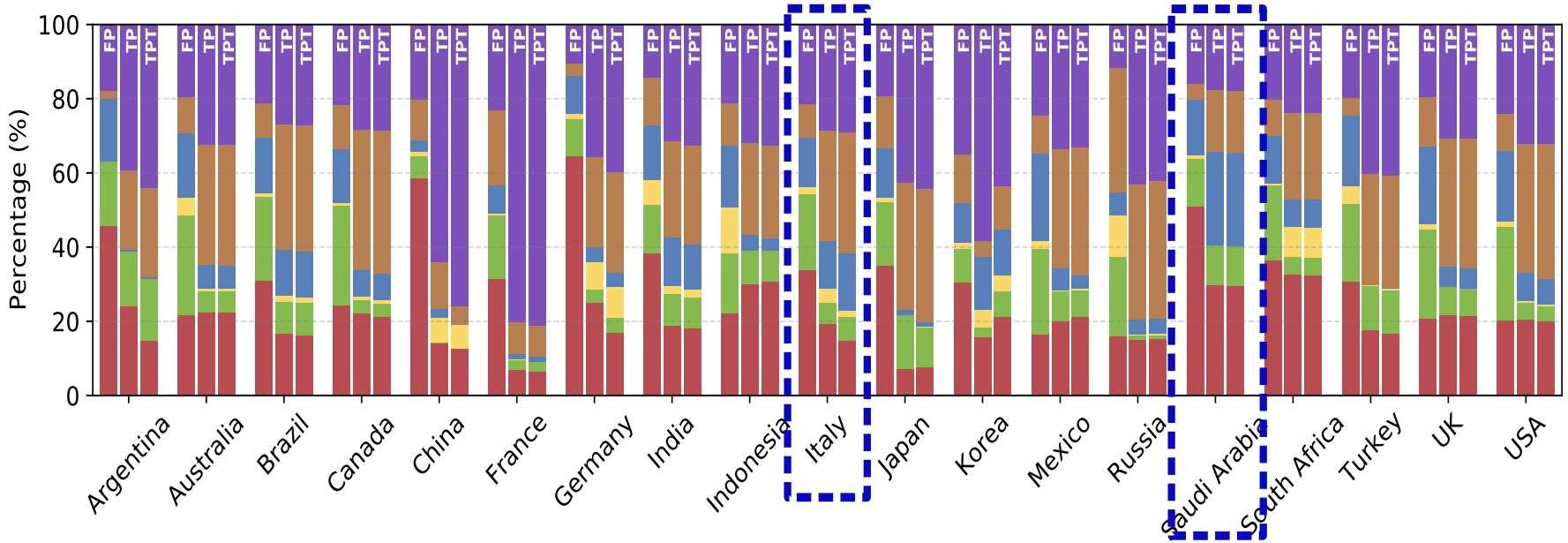
FP = First-Party

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

Cookie duration:

Session 1 hour 1 day

1 month 1 year > 1 year



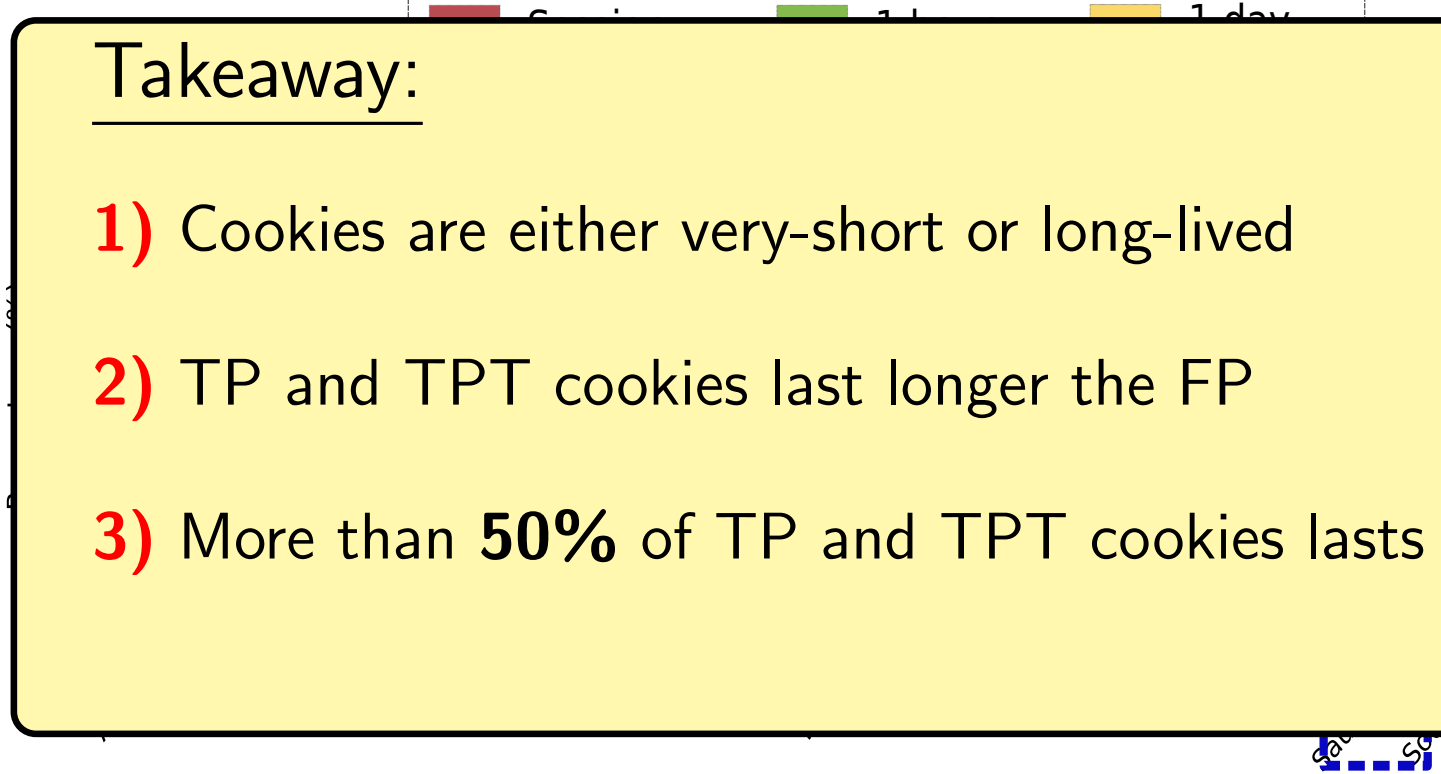
How Long do Cookies Last?

Cookie origin:

FP = First-Party

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

Cookie duration:



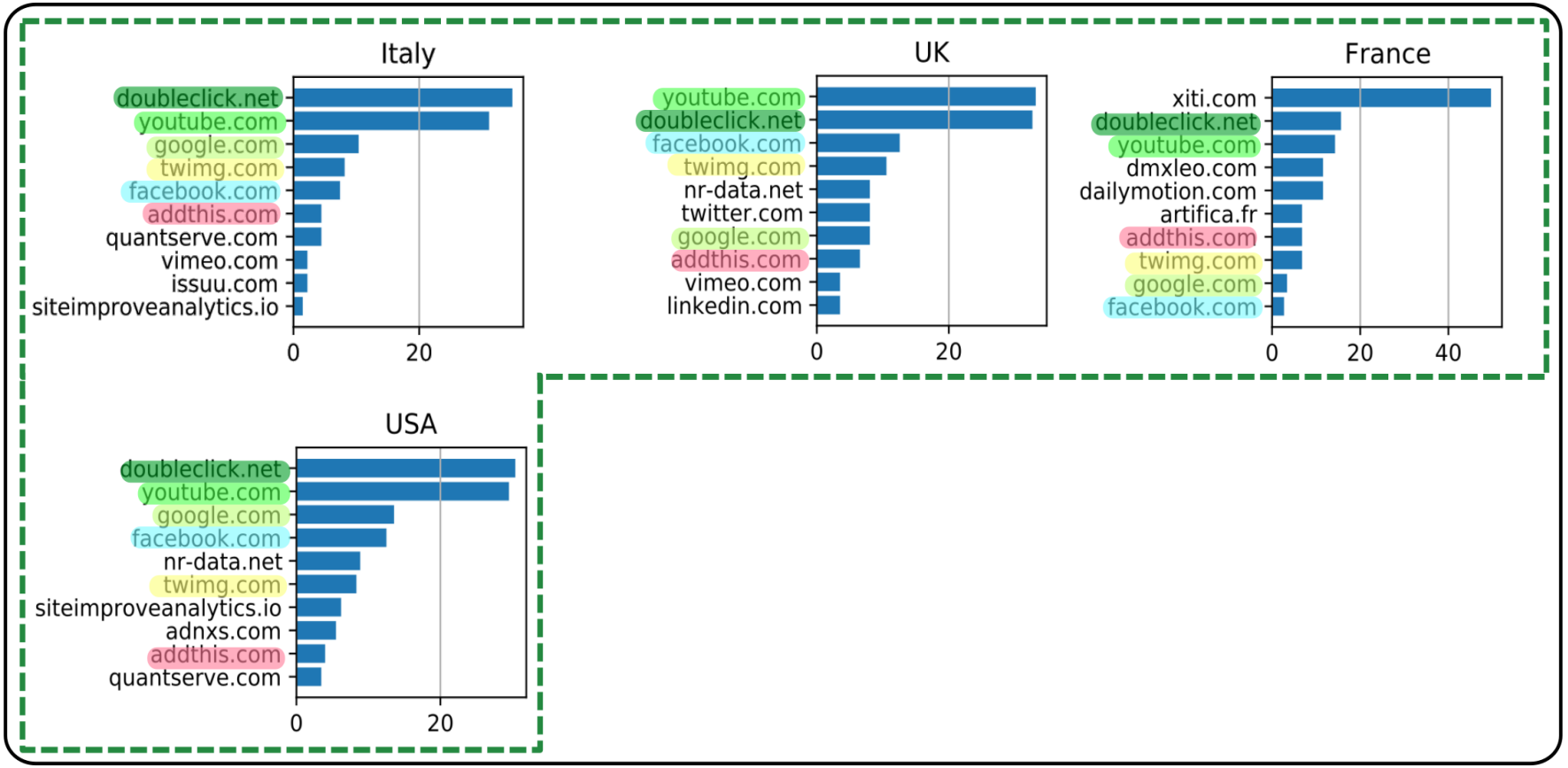
Takeaway:

1) Cookies are either very-short or long-lived

2) TP and TPT cookies last longer the FP

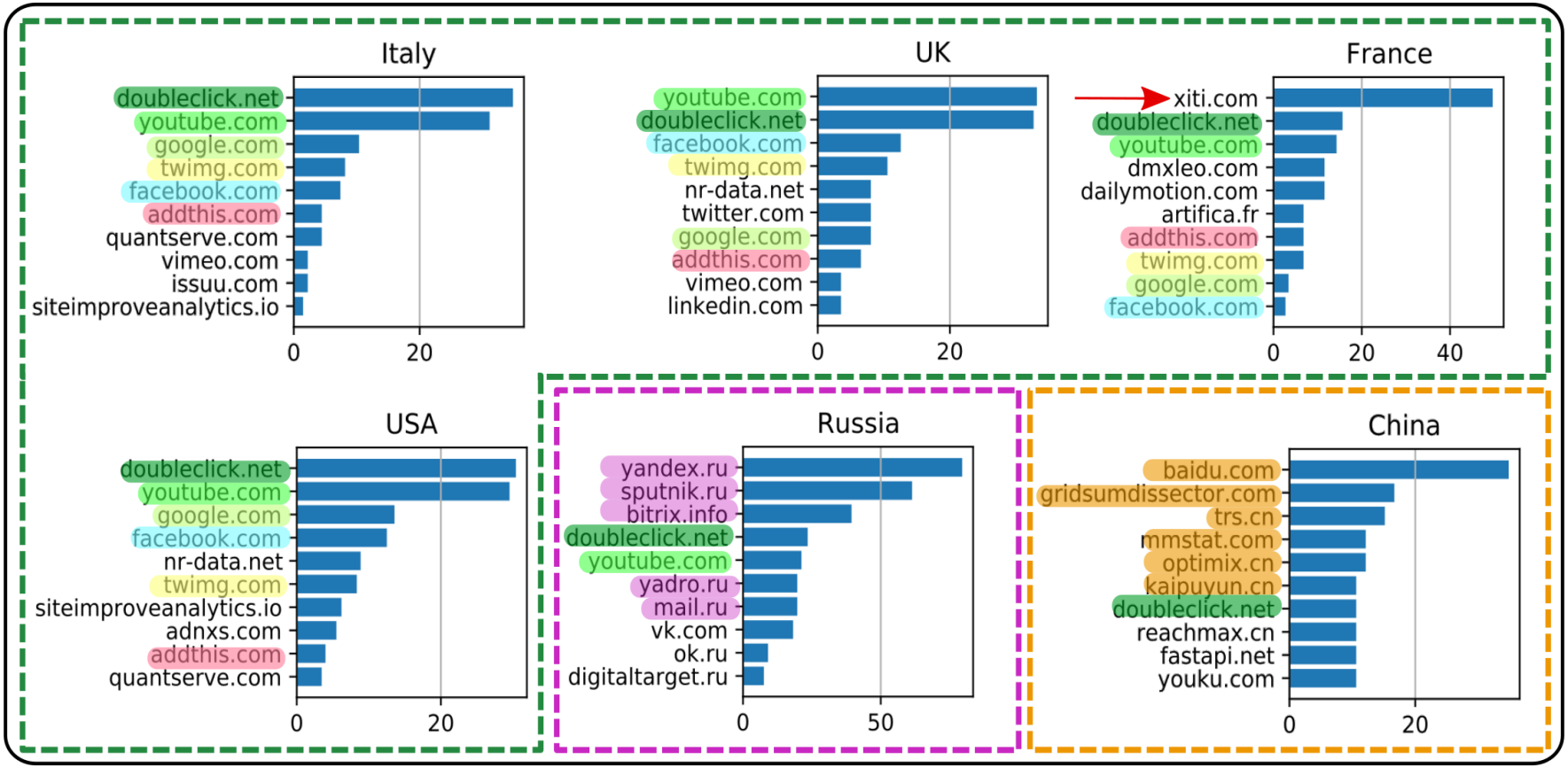
3) More than **50%** of TP and TPT cookies lasts 1+ year(s)

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:

textitThis 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).

BACKUP SLIDES

Source Selection

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`)
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

Identying Trackers

Use of **filter lists**:

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

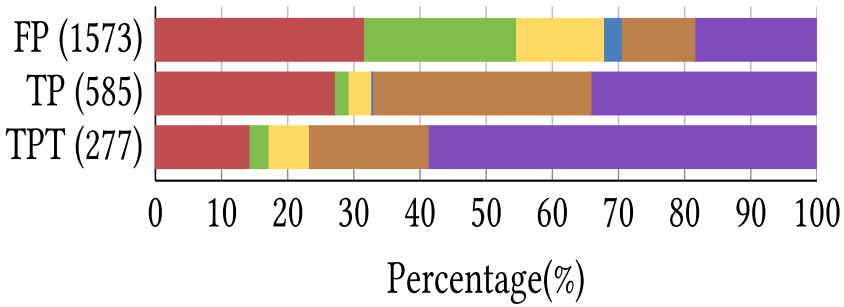
Results for Int. Orgs. and Covid-19 Websites

Cookie origin:

FP = First-Party
TP = Third-Party
TPT = Third-Party Tracker

Cookie duration:

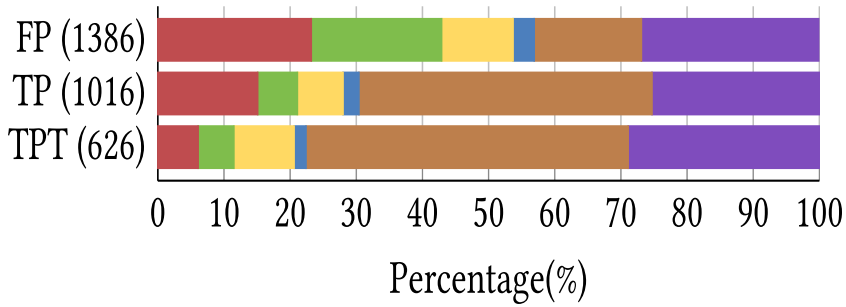
Session, 1 hour, 1 day, 1 month, 1 year, > 1 year



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

Top-5 trackers

International Organizations



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

Covid-19 websites