



Institut Mines-Télécom

# MARCHÉ DES DONNÉES PERSONNELLES ET RÉGULATION

VINCENT LEFRERE



# RESEARCH STRATEGY

## MAIN RESEARCH QUESTIONS:

- How personal data are used by stakeholders?
- What is the effect of the Privacy regulation on economics welfare?
- To determine which stakeholders are reaping OFF the gains of the data economy
- Can policy constrained by geographic boundaries shape digital be

**METHODOLOGIES:** For collecting, analyzing and interpreting the data

I used **web scraping** and **natural experiments** to generate original databases.

1. Simulating browsing behaviours
2. Python, sql, openwpm

# RESEARCH PAPERS

- “Economics of Free Mobile Applications: Personal Data and Thirds parties” With Cecere, G and Le Guel, F.
- “Privacy and Children: What drives digital data protection for very young children? With Cecere, G and Le Guel, F., Yin, P.L., Tucker, C.
- **The Effect of the European General Data Protection Regulation on the Online Publishing Industry.”** With Warberg, L., Cheyre, C., Marotta, V., Acquisti, A.

# ACHIEVEMENT

**Visiting** at *Carnegie Mellon University, Pittsburgh, USA*, in the lab directed by Prof. Alessandro Acquisti

**Publication** Lefrere, V. (2018) Quel est le coût des applications gratuites ? Regards croisés sur l'économie, 23, 2, pp. 94-104.

**Conference:** TSE Digital Seminar, Toulouse  
Federal Trade Commission (FTC) Privacycon

...

**Organizer:** of the Doctoral seminars on Natural Language Processing (NLP), at Paris-Saclay with 2500 euros grants

# THE IMPACT OF GDPR ON AD-SUPPORTED CONTENT PROVIDERS

V. Lefrere<sup>1</sup>, L. Warberg<sup>2</sup>, C. Cheyre<sup>3</sup>, V. Marotta<sup>4</sup> & A. Acquisti<sup>2</sup>

1 Institut Mines Telecom-Business School, 2 Carnegie Mellon University,  
3 Cornell University, 4 University of Minnesota

- The General Data Protection Regulation (GDPR) was enacted in May 2018 to increase privacy protection and reestablish individuals' control over their data
- One of GDPR's defining characteristics is the requirement that organizations obtain explicit consent (opt-in) prior to collecting and/or processing individuals' data

**RollingStone** Menu Read Next: Mötley Crüe, Def Leppard, Poison Add Dates to 2020 Tour Send Us a Tip Subscribe

stage on Thursday:

Joe Biden  
Bernie Sanders  
Elizabeth Warren  
Pete Buttigieg  
Andrew Yang  
Amy Klobuchar  
Tom Steyer

**Which can**

There are eight  
Michael Bloom  
Marianne Willi

On Saturday, n  
pressing the pa  
the lofty thresh  
qualify for Thu

"The DNC will  
two consecutive  
said in a statem  
where we are in

**We value your privacy**

We and our partners use technologies, such as cookies, and process personal data, such as IP addresses and cookie identifiers, to personalise ads and content based on your interests, measure the performance of ads and content, and derive insights about the audiences who saw ads and content. Click below to consent to the use of this technology and the processing of your personal data for these purposes. You can change your mind and change your consent choices at any time by returning to this site.

**DENY ALL** **ACCEPT AND MOVE ON**

Manage my Consents | See vendors

Powered by **Quantcast**

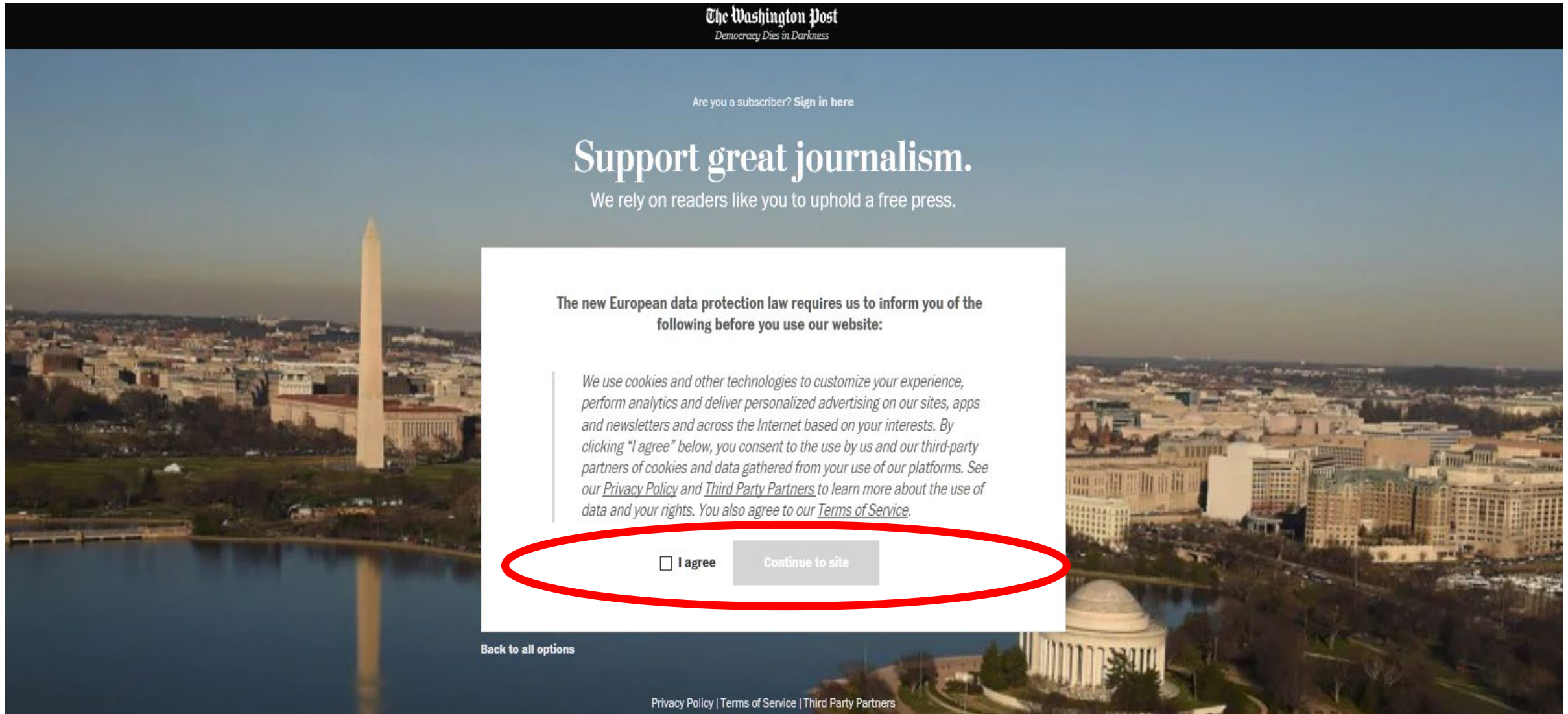
**BROWSE MORE EVENTS**  
Powered by **VIVIDSEATS**

**Sponsored Stories** by Outbrain

se einzelnst  
nachten jeden  
sucken  
in.

Hier sind die 10 tödlichsten  
Lebensmittel, die du  
regelmäßig konsumierst  
by Fit und mental.de

The next — of seven *more* debates following Thursdays — will take place January 14th in







POLITICS



## House begins day of debate ahead of vote to impeach Trump

### LATEST POLITICS >

#### POLITICS

What time is the Democratic debate in Los Angeles?

Dec. 18, 2019

#### POLITICS

Column: Let's get Rudy under oath

Dec. 18, 2019

#### POLITICS

Why is the left hating on Pete Buttigieg?

Dec. 18, 2019

#### POLITICS

Trump will address supporters in Michigan as House considers impeachment

Dec. 18, 2019

#### WORLD & NATION

My year of covering the Trump White House

Dec. 18, 2019

By JENNIFER HABERKORN, SARAH D. WIRE DEC. 18, 2019 | 3 AM

WASHINGTON — The House of Representatives began debate Wednesday leading up to a vote this evening to impeach President Trump for abusing his office and obstructing Congress, a condemnation that only two other U.S. presidents have faced in the nation's 243-year history.

By continuing to use our site, you agree to our [Terms of Service](#) and [Privacy Policy](#). You can learn more about how we use cookies by reviewing our [Privacy Policy](#). [Close](#)

Despite the historic nature of the vote to charge the president with committing high crimes and misdemeanors, the day's proceedings lack much suspense: Trump's fate

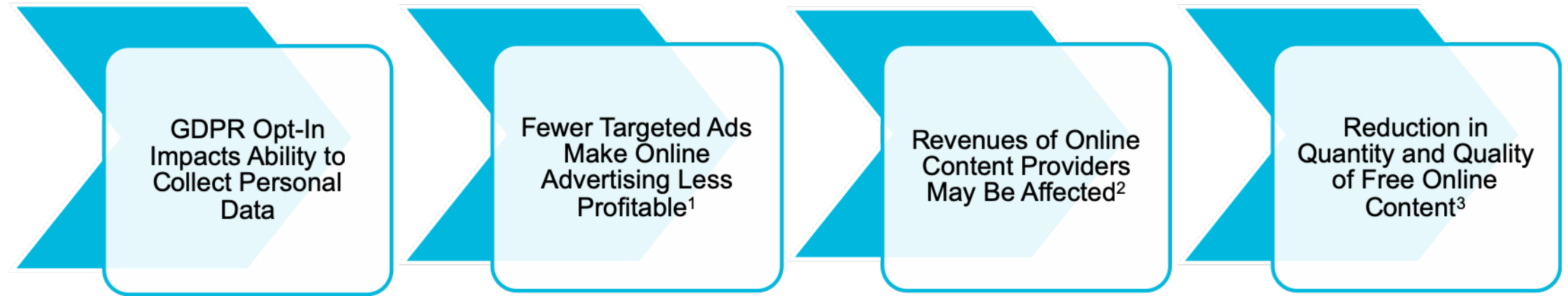
Chicago Tribune

Unfortunately, our website is currently unavailable in most European countries. We are engaged on the issue and committed to looking at options that support our full range of digital offerings to the EU market. We continue to identify technical compliance solutions that will provide all readers with our award-winning journalism.

- The General Data Protection Regulation (GDPR) was enacted in May 2018 to increase privacy protection and reestablish individuals' control over their data
- One of GDPR's defining characteristics is the requirement that organizations obtain explicit consent (opt-in) prior to collecting and/or processing individuals' data
- Many (mostly negative) claims about the economic impact of data regulation in general, and GDPR in particular...

- *“If this reduction in the use of OBA were to occur businesses would need to reallocate a large amount of marketing expenditure from OBA to other online channels. The Deloitte Business Survey suggests that, on average, the **ROI that businesses would expect to receive from other channels would decrease by 27%**” – Deloitte 2013*
- *“The evidence clearly suggests that the tradeoffs of stronger privacy laws result **in less free and low-cost content** and more spam (i.e. unwanted ads) which is not in the interests of consumers” - ITIF 2010*
- *“The **data protection text currently on the table** would essentially impose a one-size-fits-all framework to an almost infinitely broad range of data that may be considered “personal”, without adequately matching regulatory scrutiny and the onerousness of the obligations on data controllers, to the real risks arising for consumers from a particular act of processing. **This may limit digital advertising’s ability to continue to deliver a wide range of online content to users at little or no cost at the point of consumption**” - Townsend Feehan, CEO of IAB Europe, IHS Technology Report, November 2015*
- Forecast: 2.8 millions jobs lost and 1.34% EU GDP reduction following GDPR – Deloitte 2013

- The General Data Protection Regulation (GDPR) was enacted in May 2018 to increase privacy protection and reestablish individuals' control over their data
- One of GDPR's defining characteristics is the requirement that organizations obtain explicit consent (opt-in) prior to collecting and/or processing individuals' data
- Many (mostly negative) claims about the economic impact of data regulation in general, and GDPR in particular...
- Significant academic interest in IS/Marketing/Econ (Jia, Jin, Wagman 2018; Goldberg, Johnson, Shriver 2018; Godinho de Matos and Adjerd, 2019; ...)
- However, still limited evidence on GDPR effects on **downstream economic variables**
  - In particular, **unknown impact on profitability of ad-supported content providers, and their ability to keep providing (free, quality) online content**



## The impact of GDPR on content providers' ability to provide online content

### Specifically, we study:

- Whether and how websites in the EU and in the US implemented GDPR, and
- Whether and how this affected the quantity and quality of the content those websites produce

The impact of GDPR on content providers' ability to provide online content

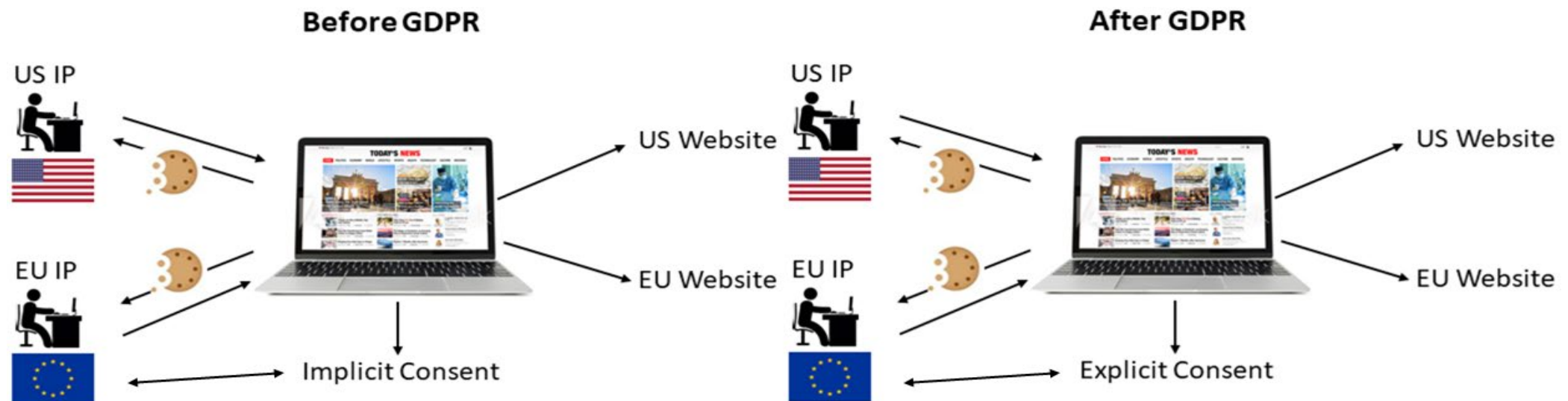
Specifically, we study:

- Whether and how websites in the EU and in the US implemented GDPR, and
- Whether and how this affected the quantity and quality of the content those websites produce

To answer this question, we used a Natural Field Experiments:

- We analyze a sample of 6,000+ websites based in the EU and the US
- Multiple categories, including those (such as news) which are more likely to be ad-supported
- We started collecting information some months before GDPR was implemented and we continue collecting data to this day
- We visit periodically each website in our sample (multiple “waves”) from both EU and US locations (i.e., IP addresses)
- We collect both “technical” and “downstream” economic variables





## ■ Mined from visiting websites directly, using OpenWPM:

- Number of cookies (1<sup>st</sup> and 3<sup>rd</sup> party)
- Size and type of cookies (1<sup>st</sup> and 3<sup>rd</sup> party)
- Number of HTTP requests (1<sup>st</sup> and 3<sup>rd</sup> party)
- Content size of HTTP requests (1<sup>st</sup> and 3<sup>rd</sup> party)
- [...]

Figure 1: 1st Cookies by EU and US Domains

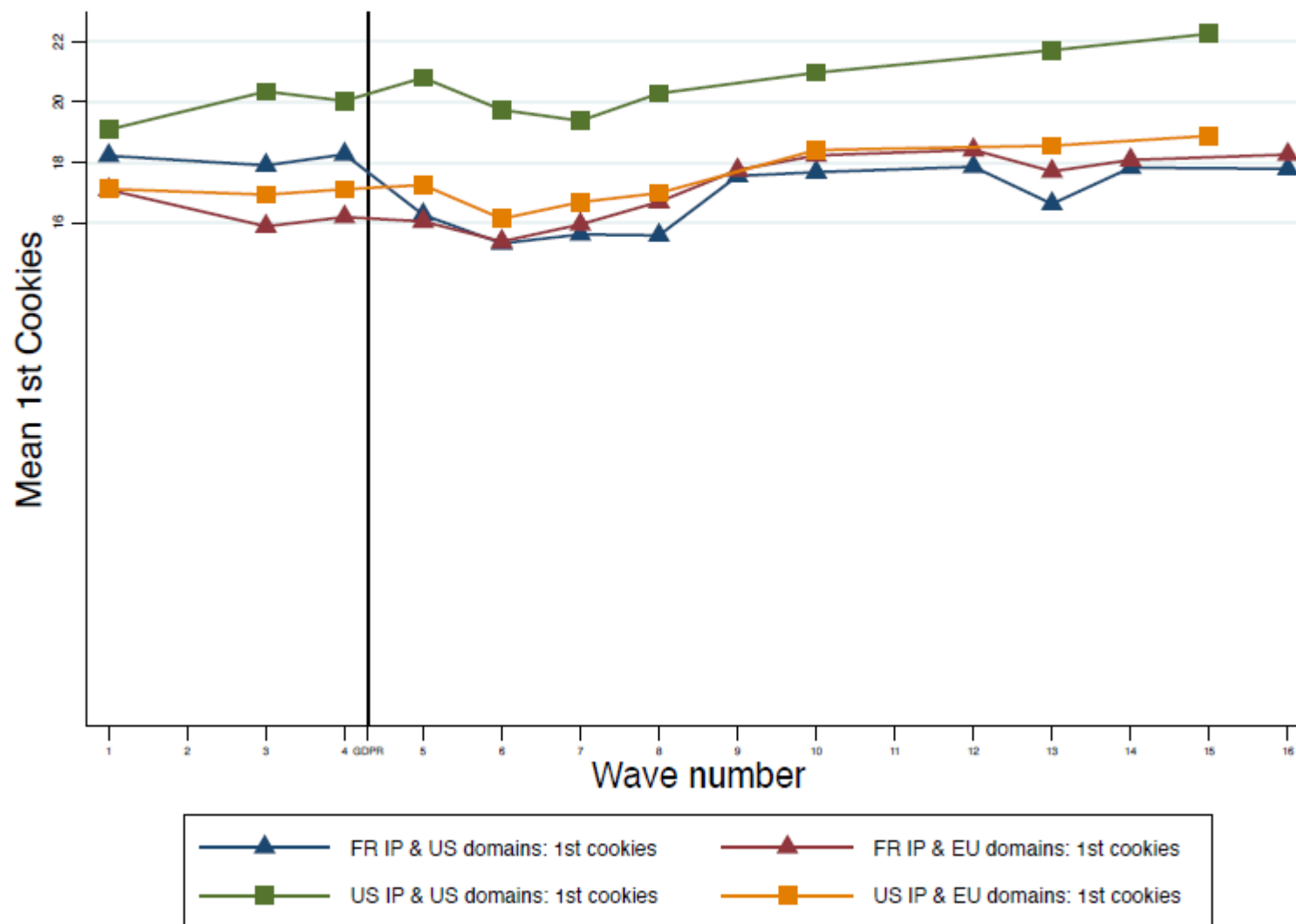


Figure 6: Consent cookies by EU and US IP and Domain

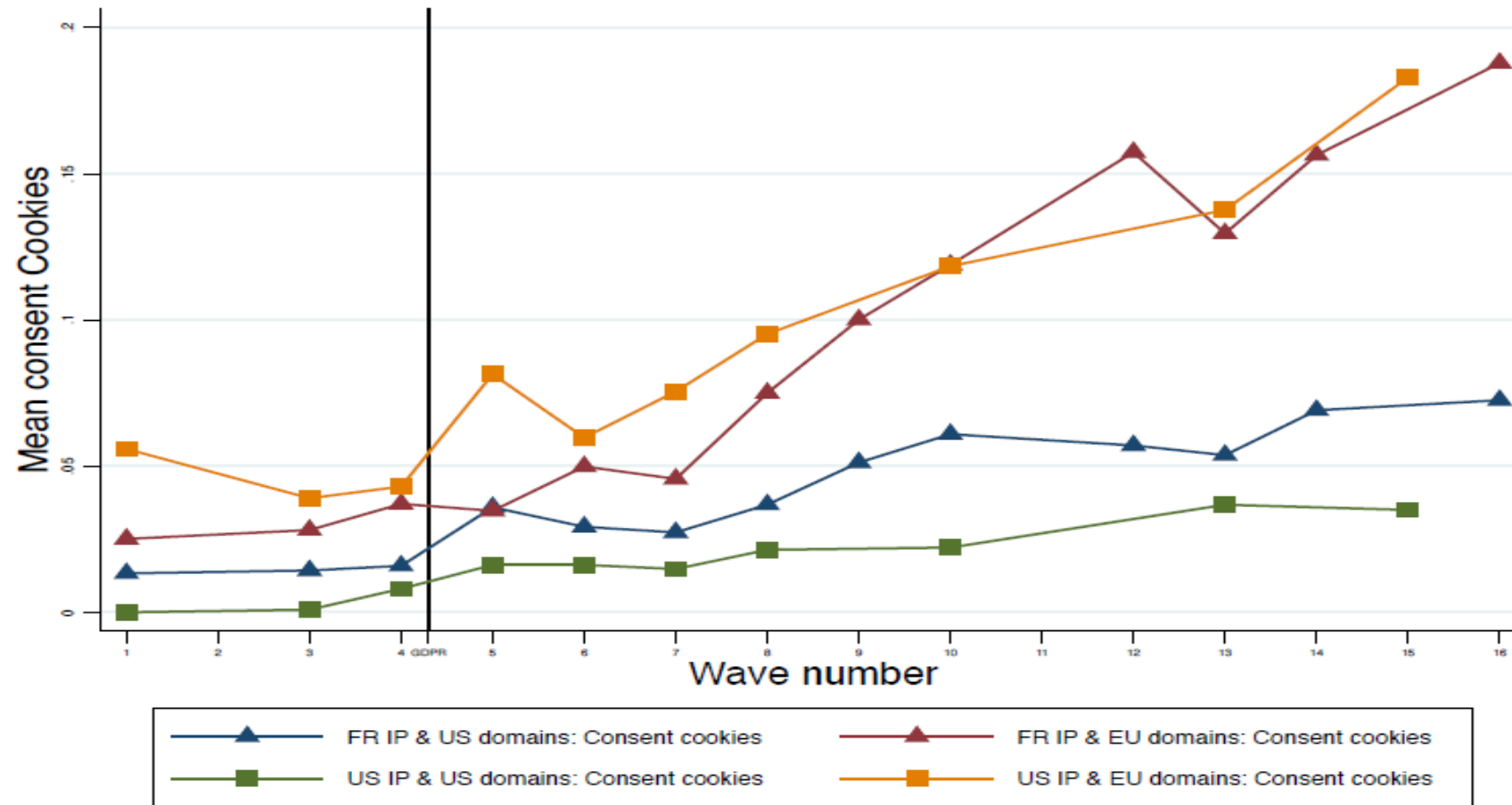


Figure 3: Tracking cookies by EU and US IP and Domain

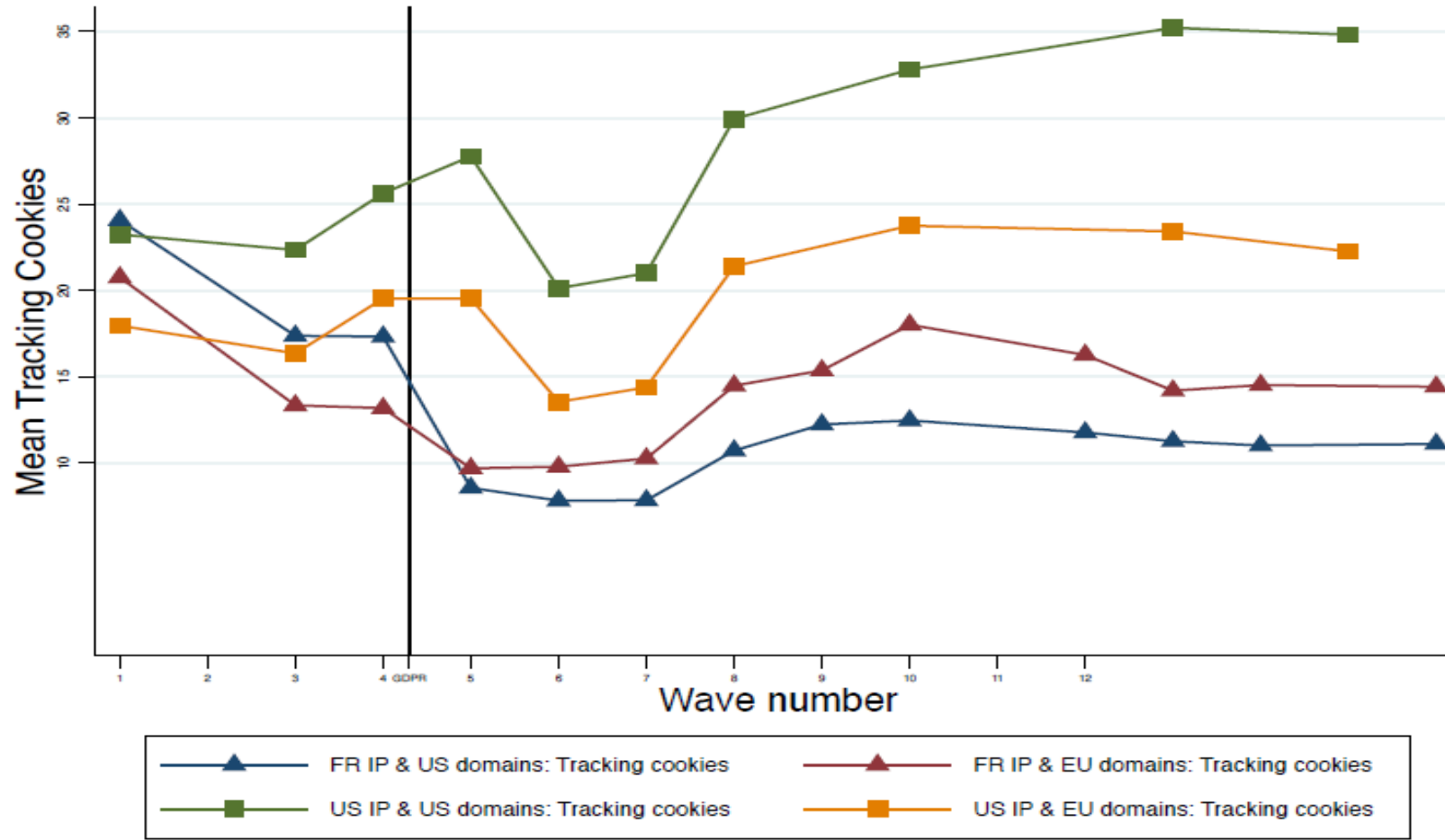
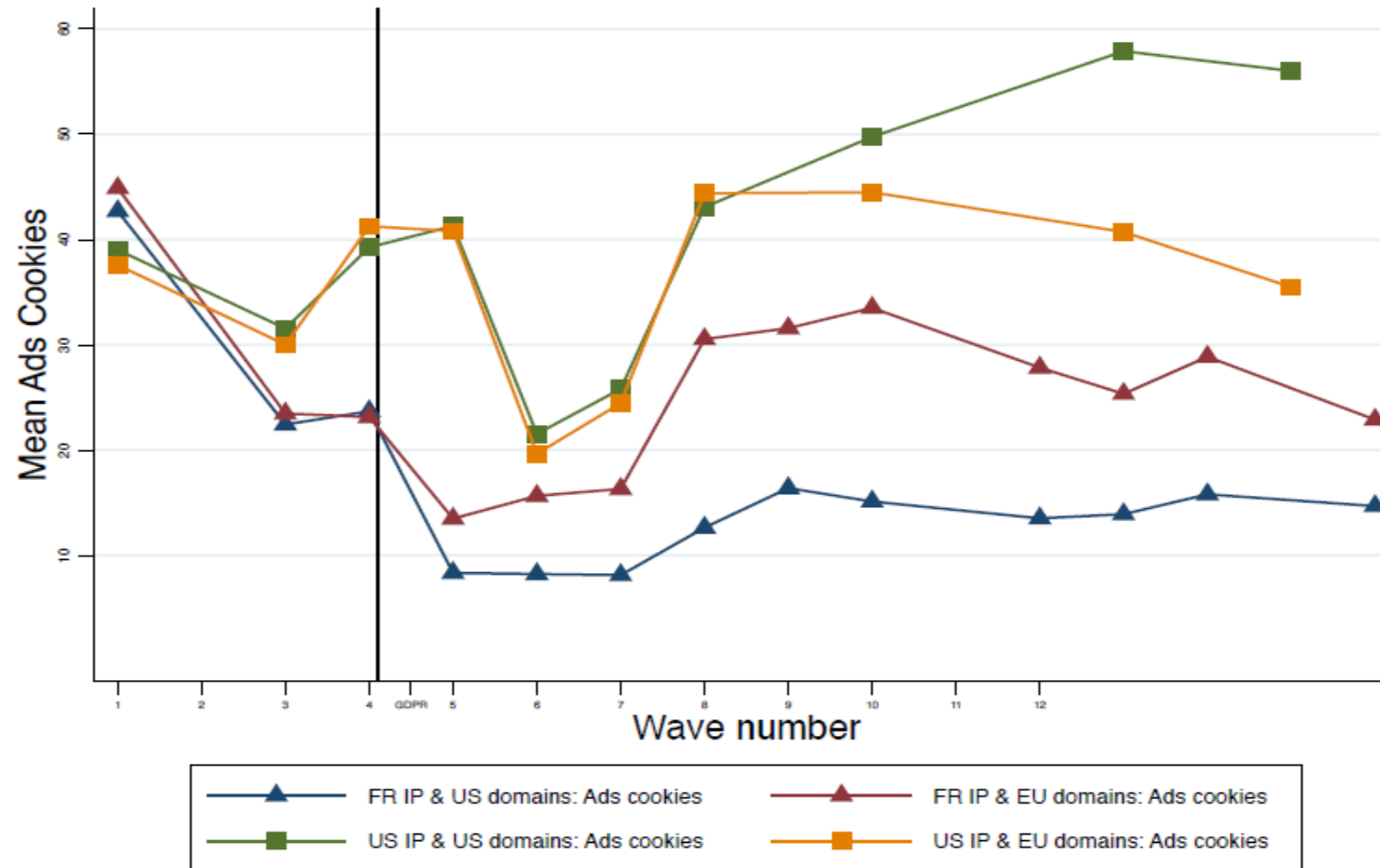


Figure 2: Ads cookies by EU and US IP and Domain



- No decrease (in fact, very slight increase) in 1<sup>st</sup> party cookies
- Significant increase over time of “consent” cookies (especially in EU domains)
- Sharp reduction in the number of 3<sup>rd</sup> party cookies (esp. ad/tracking) and http requests right around GDPR

However, sharp reduction is followed by significant recovery after some months.

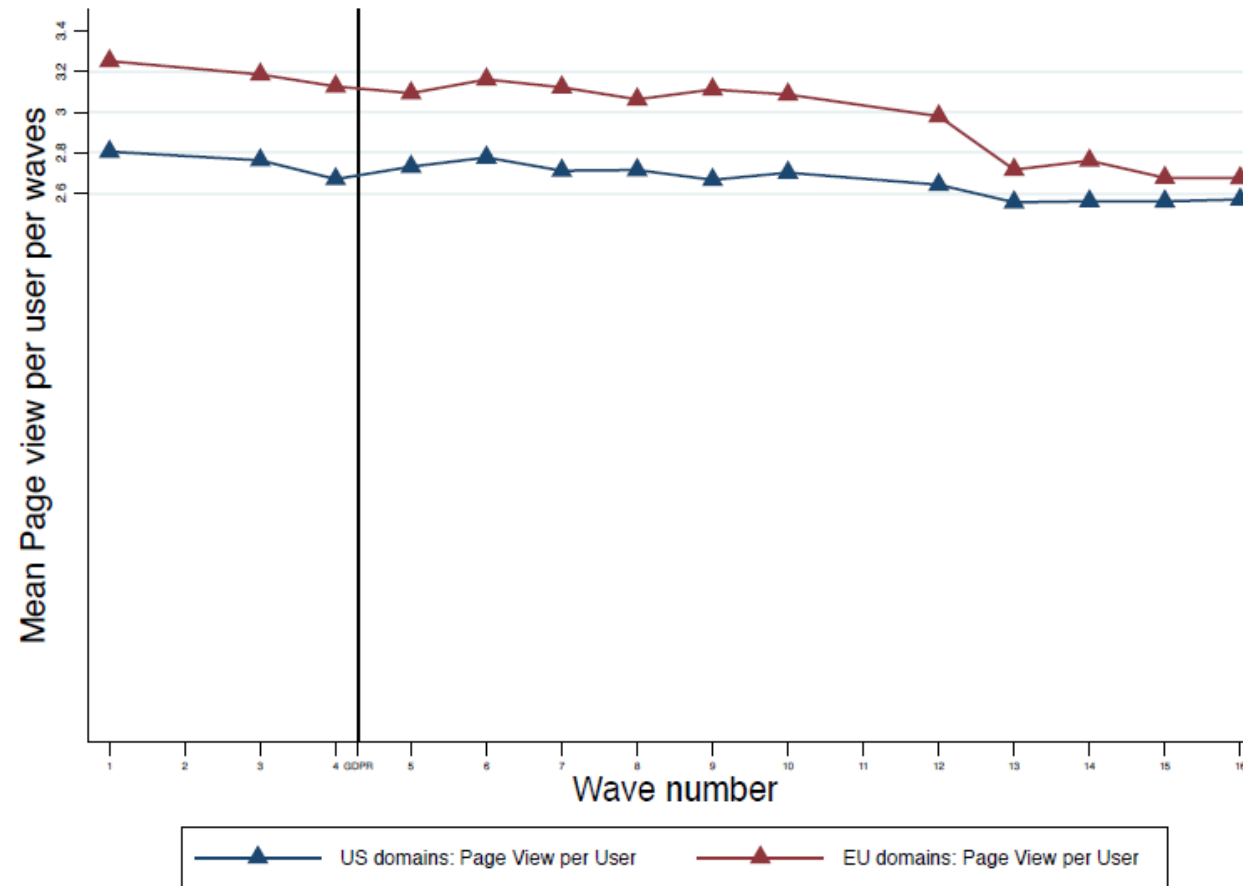
Furthermore:

- Reduction is more pronounced when visiting websites using EU IP
- Decrease is more persistent when US websites are visited from EU IPs

- A category of sites that may be particularly vulnerable to the impact of GDPR are News and Media sites
- Our sample contains 1,060 of these sites (46% in EU, 43% US, 11% other regions)
- Of US based sites, 18% are blocking EU visitors:
  - Sites that block EU: 89% of visitors from US (before GDPR)
  - Sites that don't block EU: 73% of visitors from US (before GDPR)



Figure 11: Page view per user by EU and US IP and Domain



- Mean reach (proxy for traffic) has not significantly changed for both US and EU domains
- US page views per user have increased (relative to EU sites)

Content quantity

	Panel Fixed Effects
	(1)
EU Domains × Post GDPR	0.030 (0.030)
Persistent Cookies	0.000 (0.001)
Session Cookies	0.000 (0.001)
3rd party Cookies	0.001 (0.001)
1st party Cookies	0.000 (0.001)
Cookies for Consent mechanisms	0.001 (0.008)
Advertising cookies (Disconnect)	-0.001* (0.001)
Social cookies (Disconnect)	-0.006* (0.003)
Analytics	0.008*** (0.002)
Google/Facebook/Twitter (Disconnect)	0.002 (0.002)
Nbr Cookies id by privacy	-0.001*** (0.000)
Constant	4.092*** (0.027)
Waves Fixed effects	Yes
Website fixed effects	Yes
Observations	24676
Log-likelihood	-22211.844
Std. Err	cluster

Panel fixed effects  
DV: GDelt new urls

- Mean reach (proxy for traffic) has not significantly changed for both US and EU domains
- US page views per user have increased (relative to EU sites)
- Quantity of content posted by EU sites does not seem to have decreased, relative to US sites
  - Result robust to considering only News/Media sites

We observe technical changes being implemented after GDPR:

- Initially, fewer third party cookies and requests
- Followed by a clear (but nuanced) bounce-back

~20% of US News and Media sites in our sample have restricted access to EU visitors

Some evidence that traffic has been affected

However, no evidence yet that content quantity has been affected