Lecture 15: Online Tracking

CMSC 25910

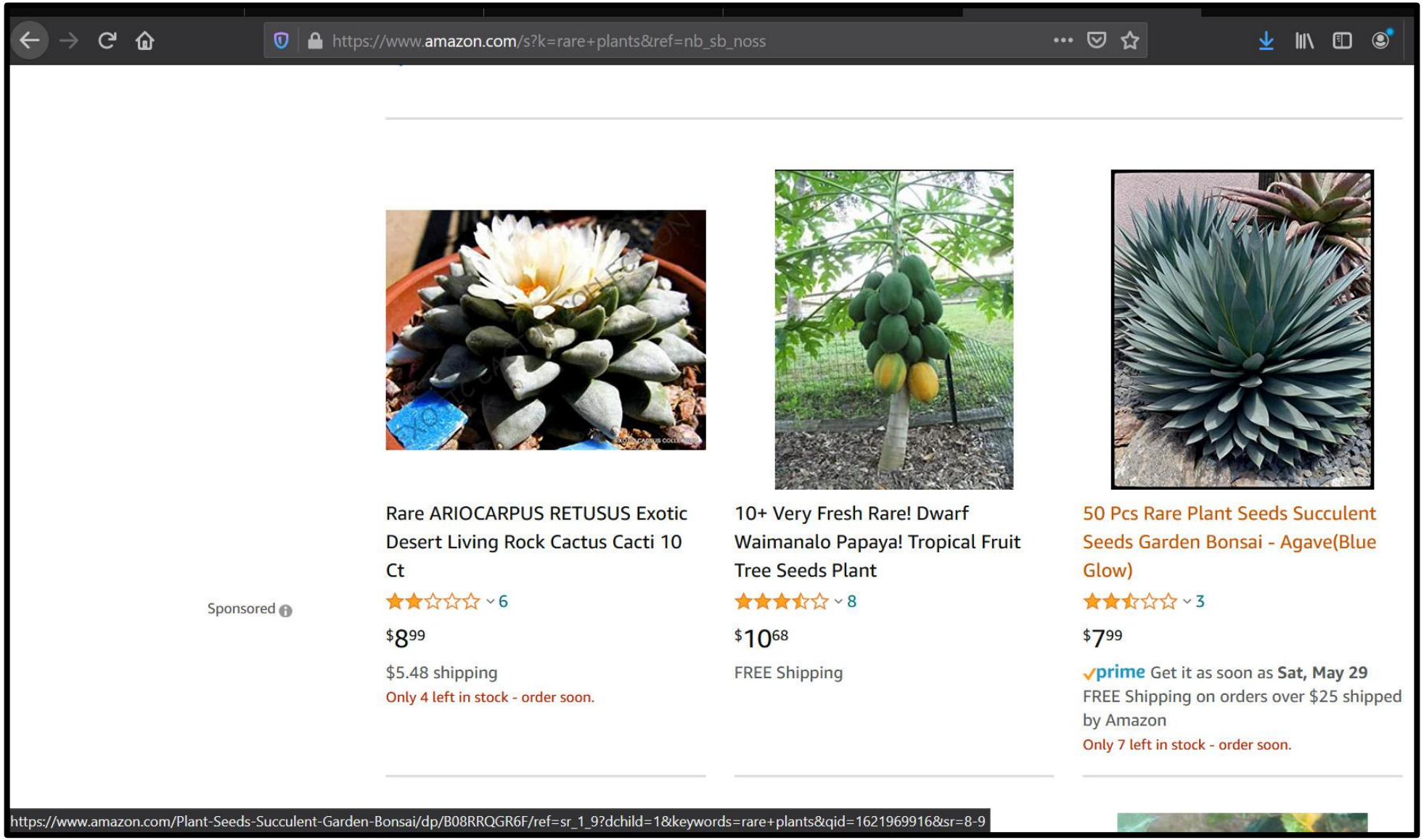
Spring 2022

The University of Chicago

THE UNIVERSITY OF CHICAGO

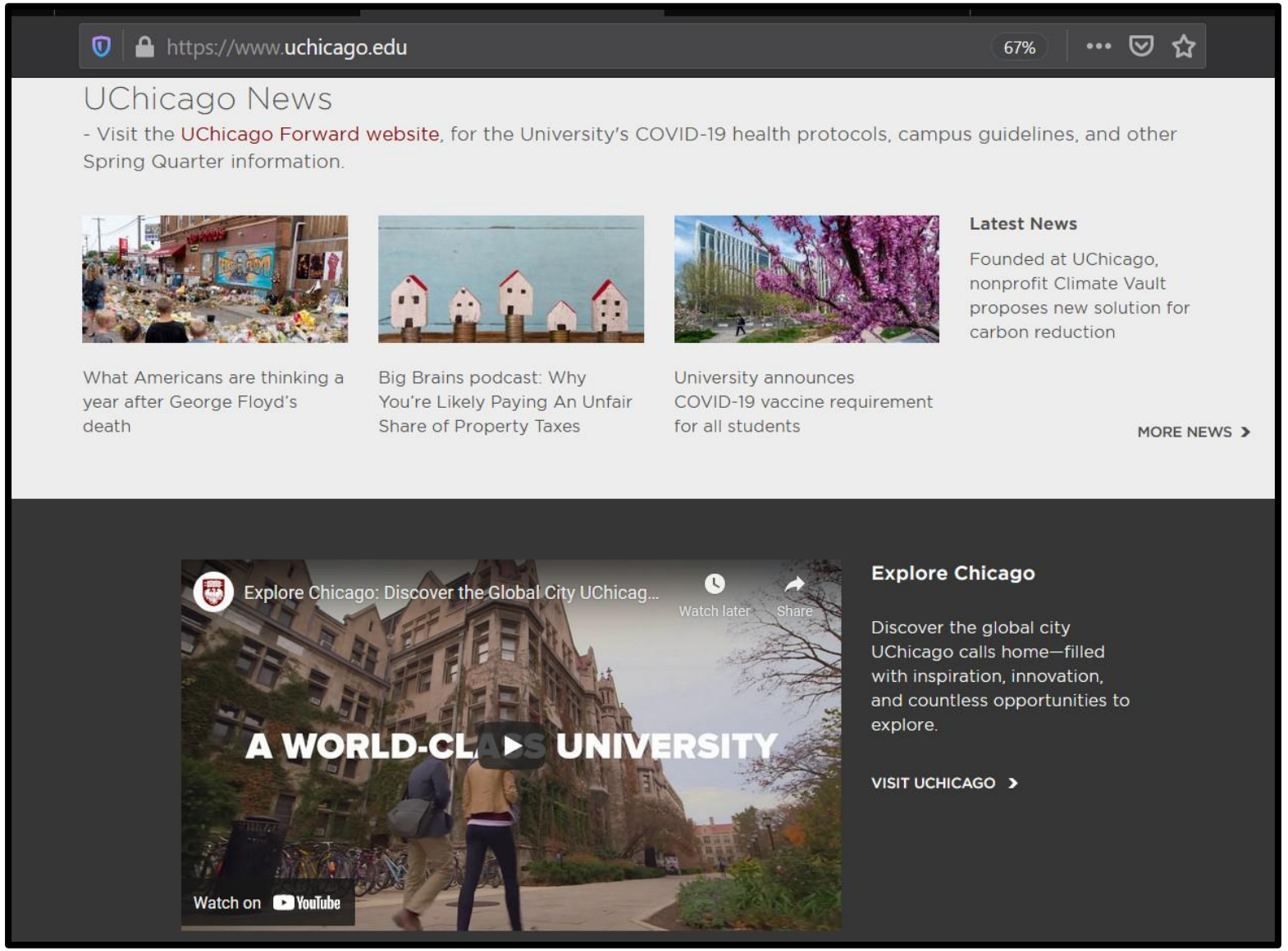
Tracking Website Visits

First-Party Tracking (The Website You Are Visiting)

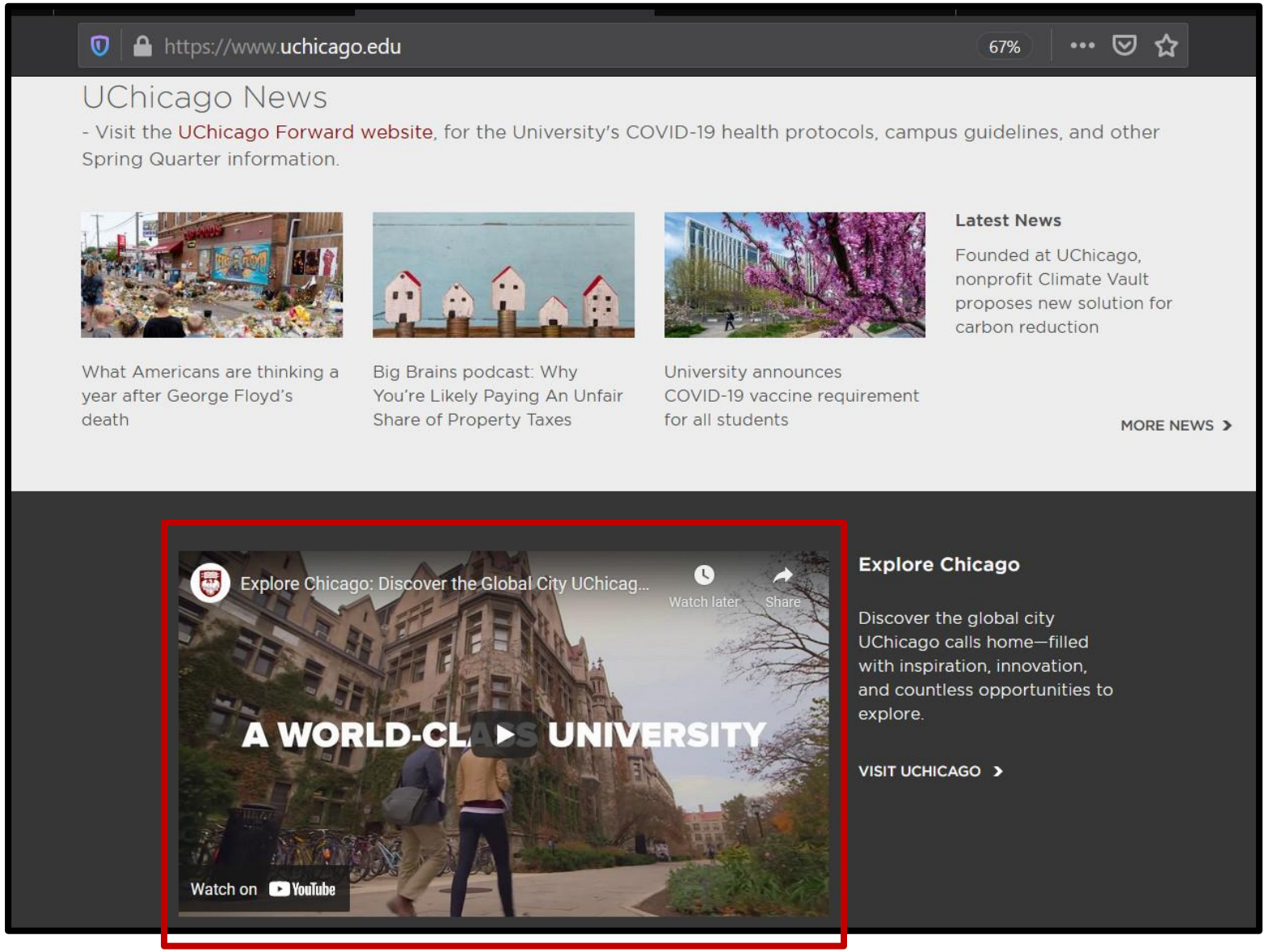


First-Party Tracking (The Website You Are Visiting)

Third-Party Tracking (***Not*** The Website You Are Visiting)

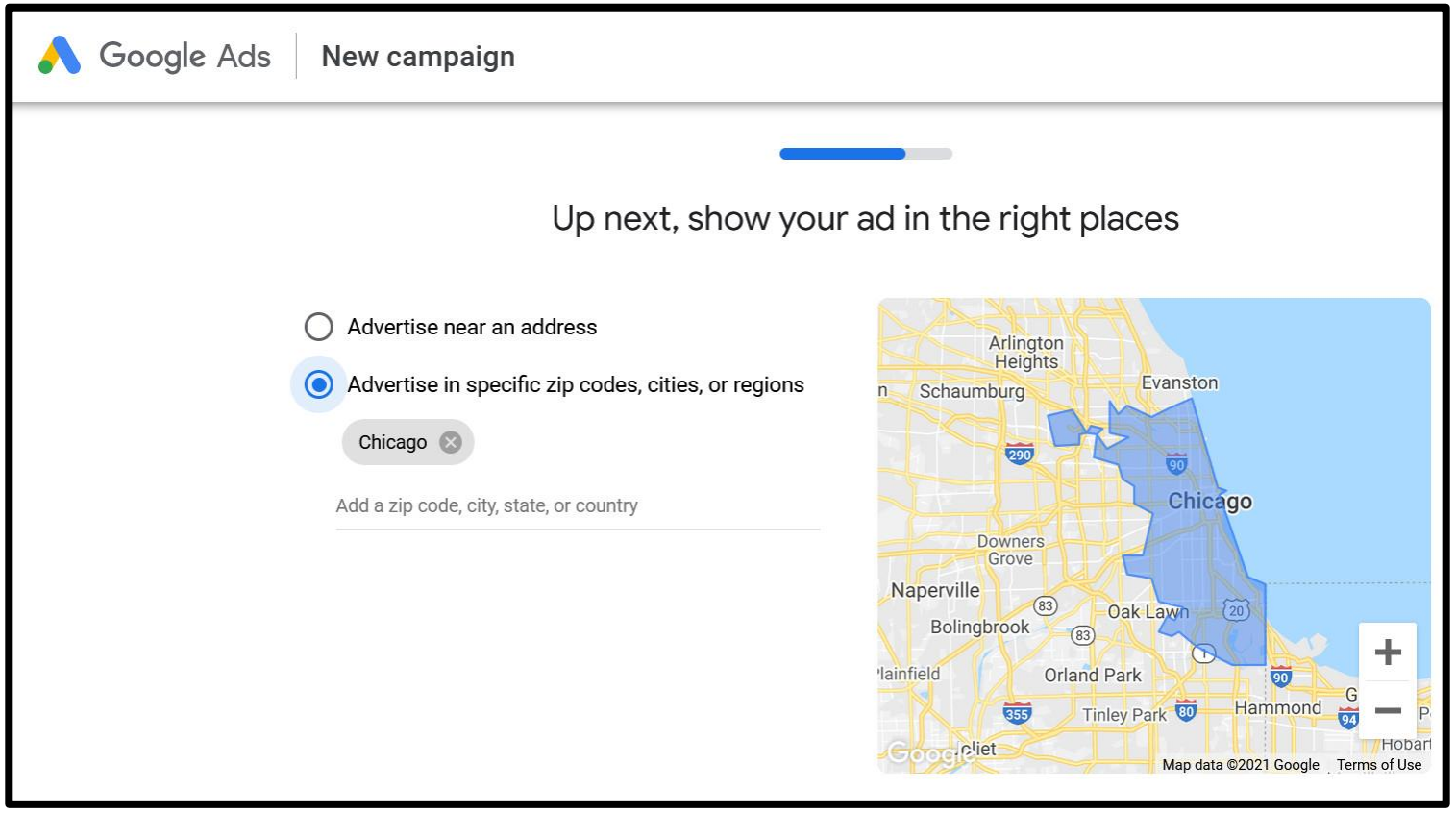


Third-Party Tracking (***Not*** The Website You Are Visiting)



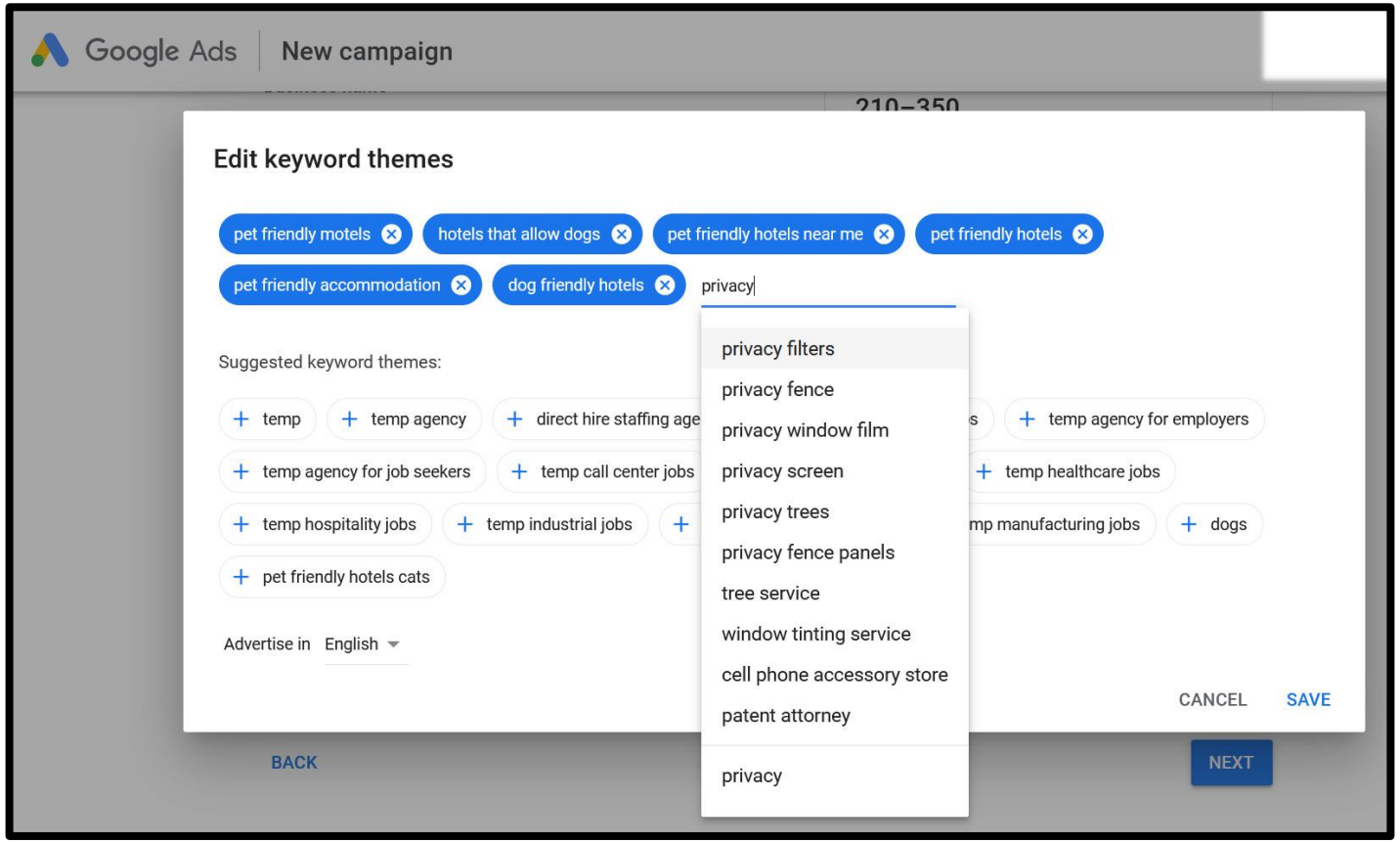
Ad Targeting

- Goal: Target advertisements to particular users
  - Based on interests, demographics, past interactions (online/offline)



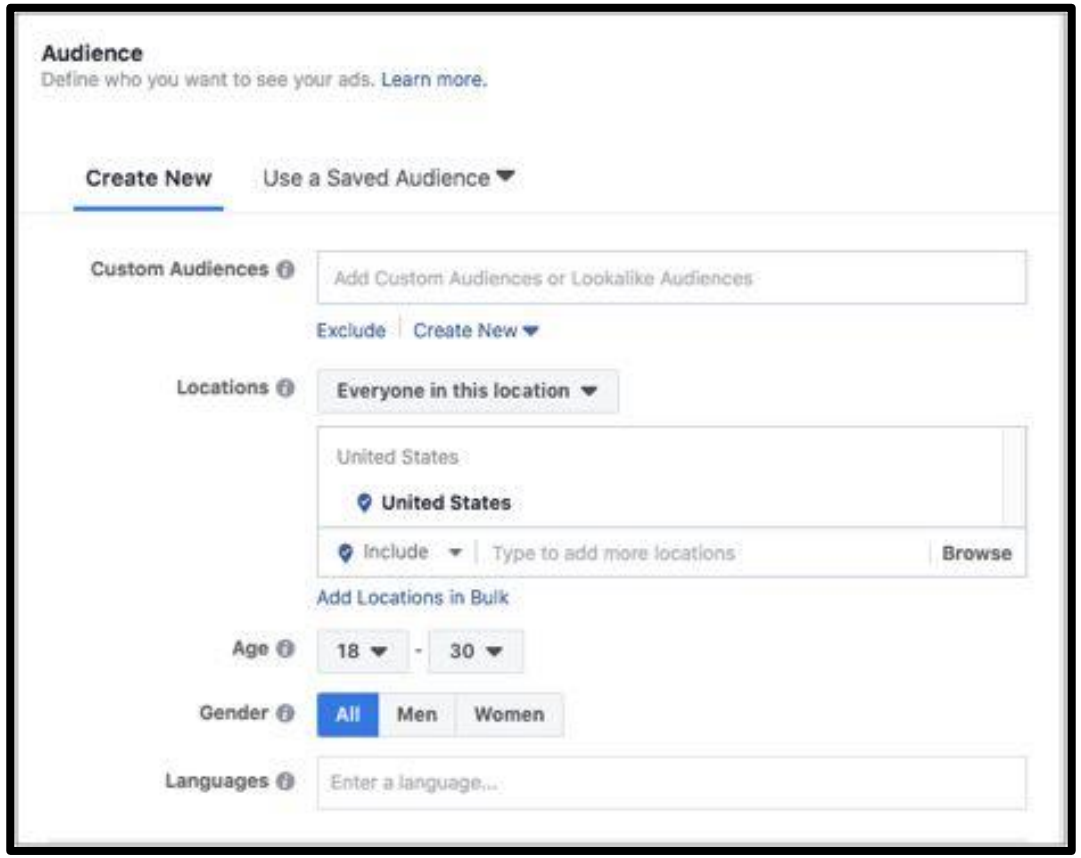
Ad Targeting

- Goal: Target advertisements to particular users
  - Based on interests, demographics, past interactions (online/offline)

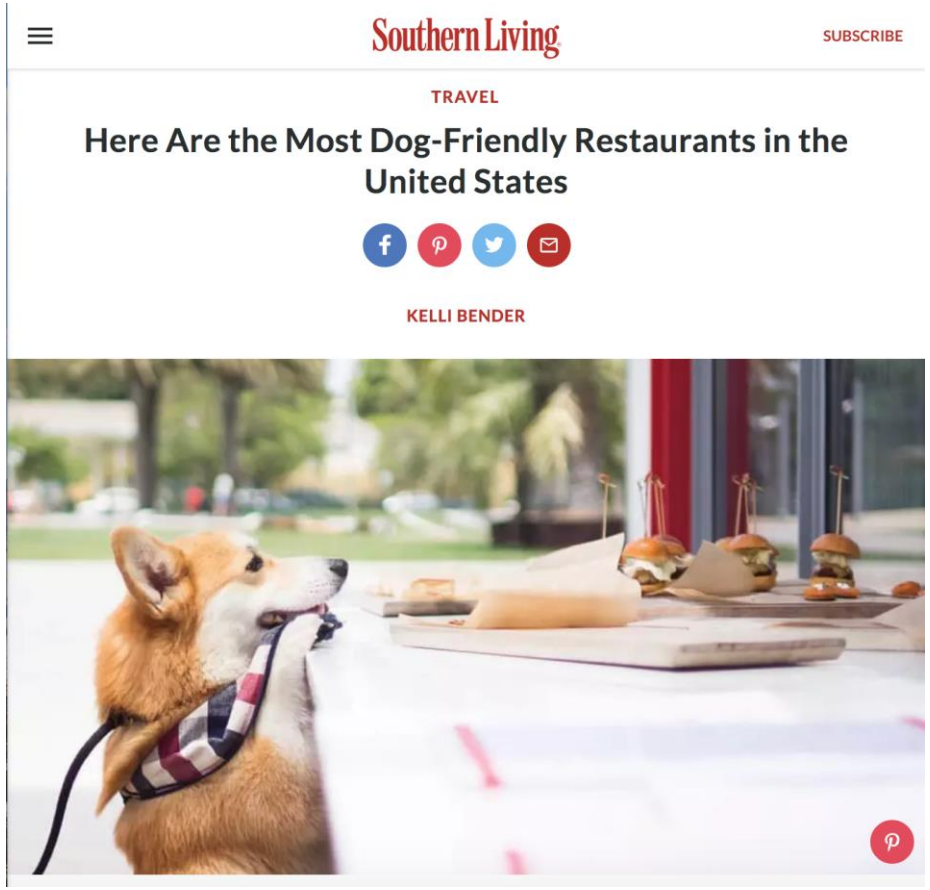


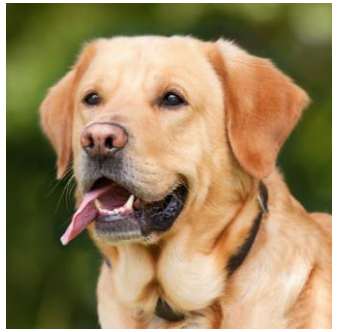
Ad Targeting

- Goal: Target advertisements to particular users
  - Based on interests, demographics, past interactions (online/offline)



Data-Driven Inferences





You might like

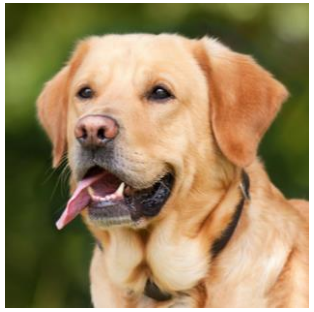
dogs!

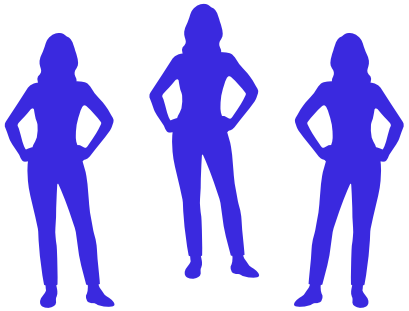
Targeted Advertising

Interests & remarketing

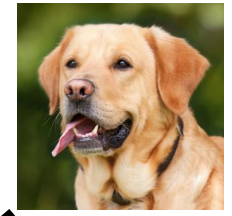
Affinity audiences

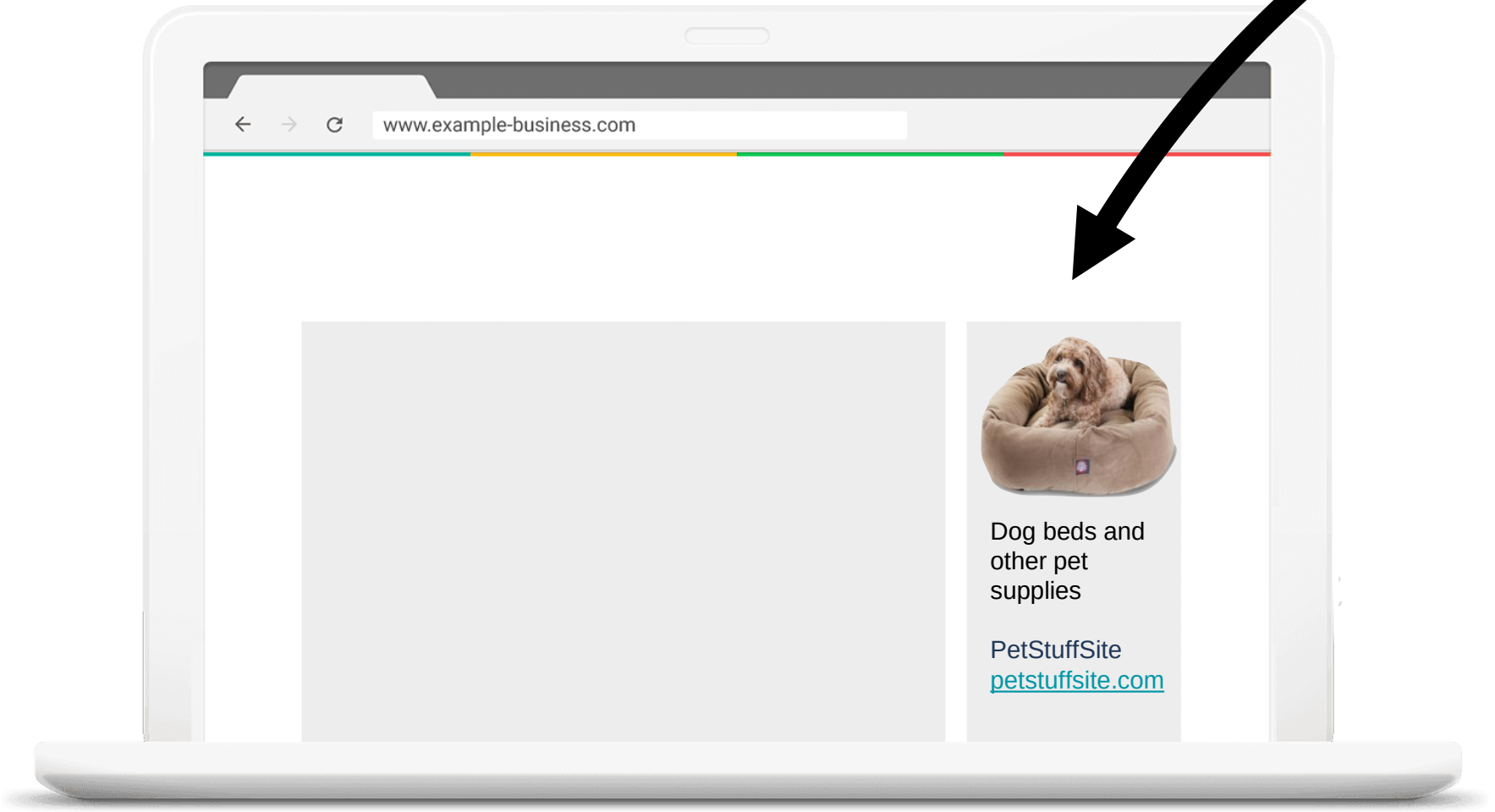
- Music Lovers »
- News Junkies »
- Nightlife Enthusiasts »
- Outdoor Enthusiasts »
- Pet Lovers »



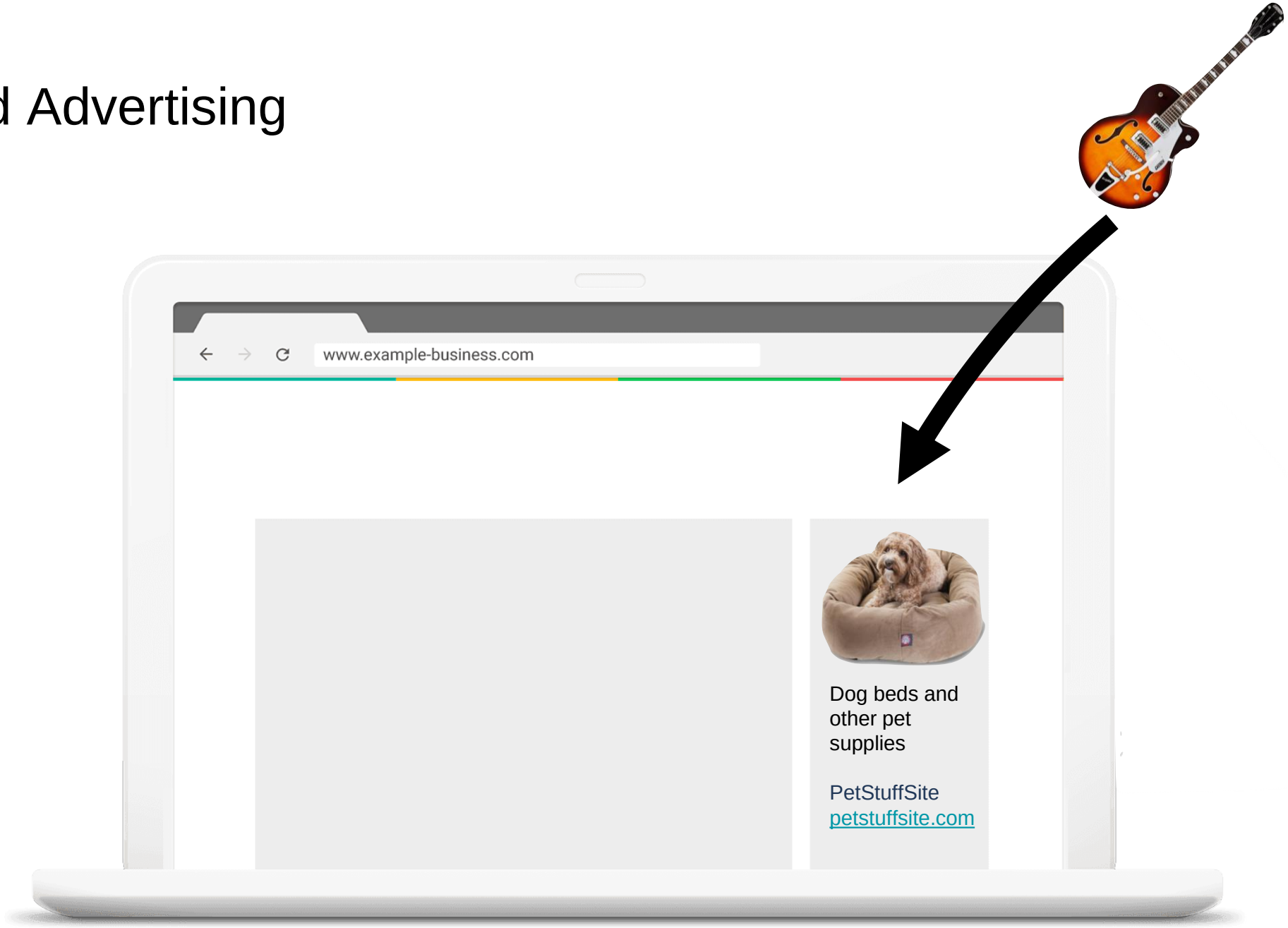


Targeted Advertising





Targeted Advertising



Mechanics of Tracking (Cookie Version)

HTTP Basics

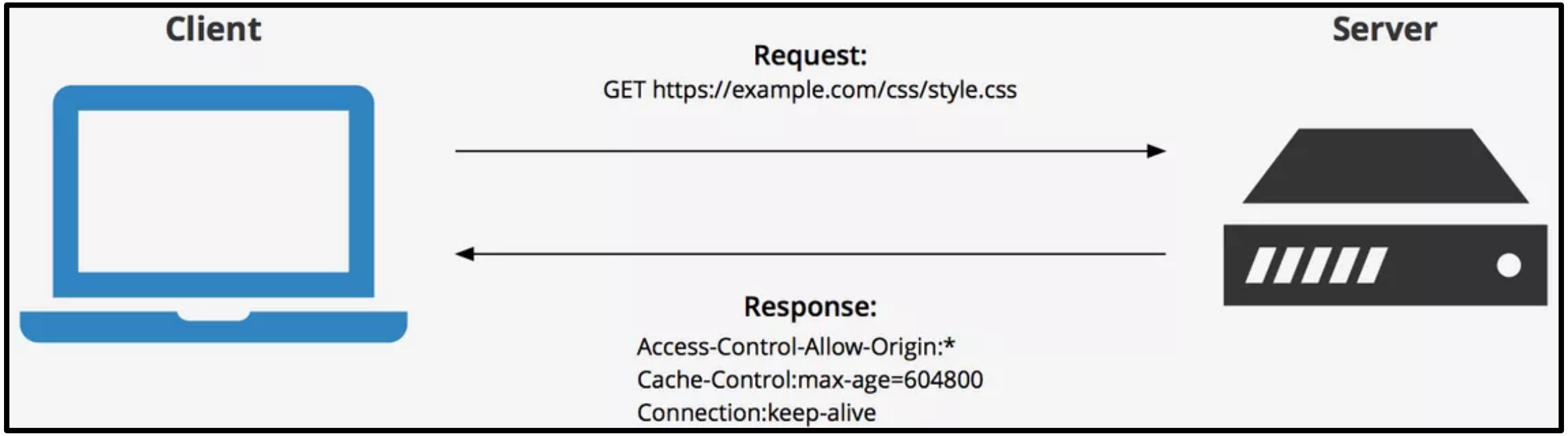
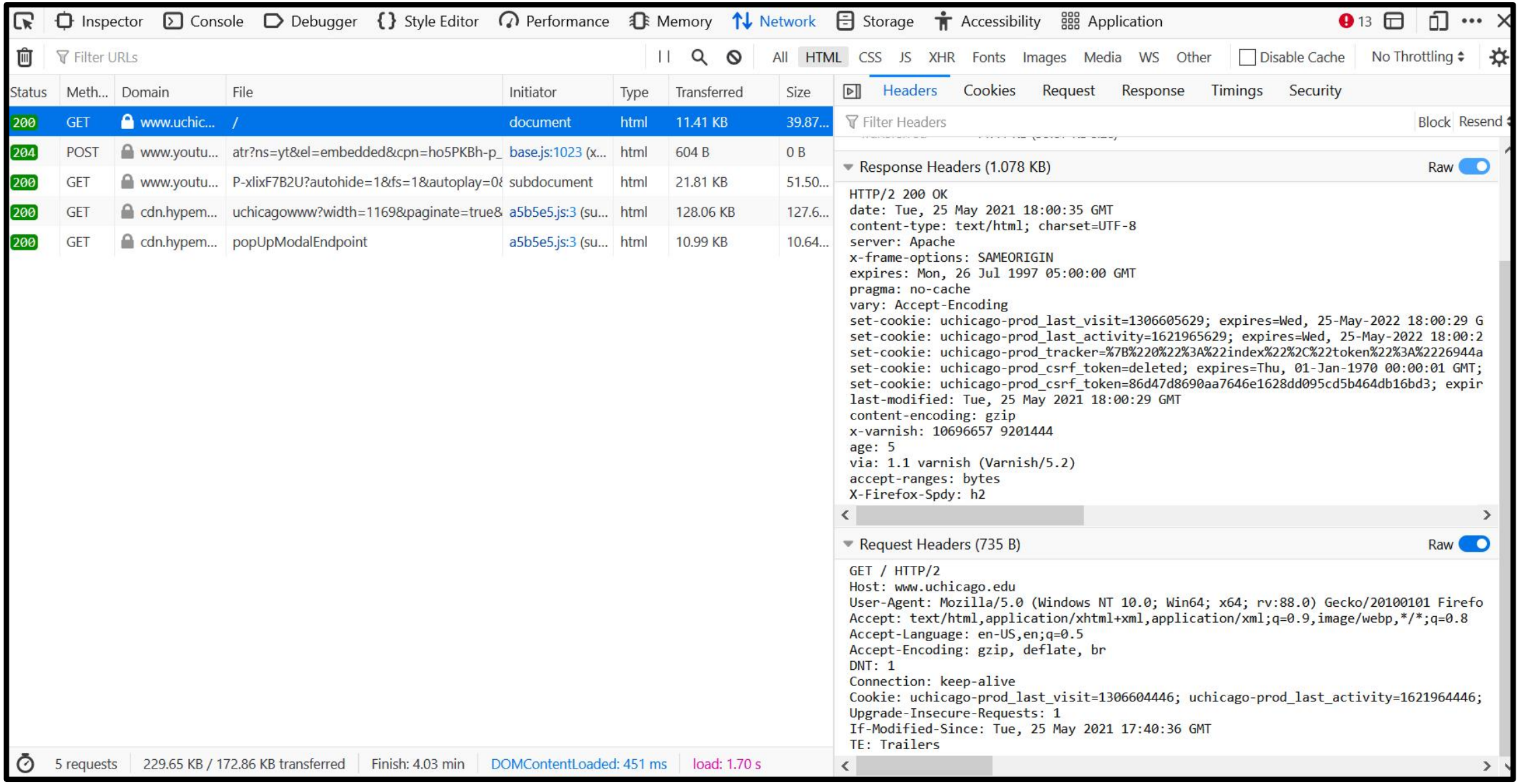
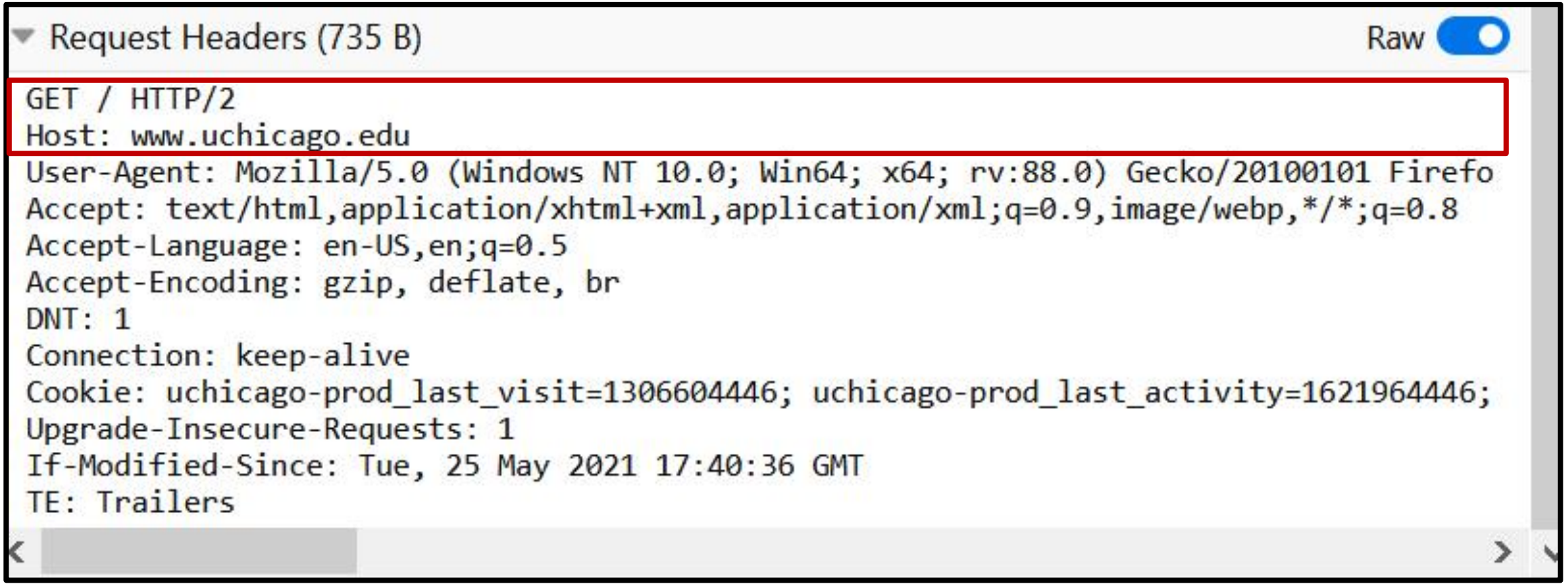


Image taken from https://www.keycdn.com/support/custom-http-headers

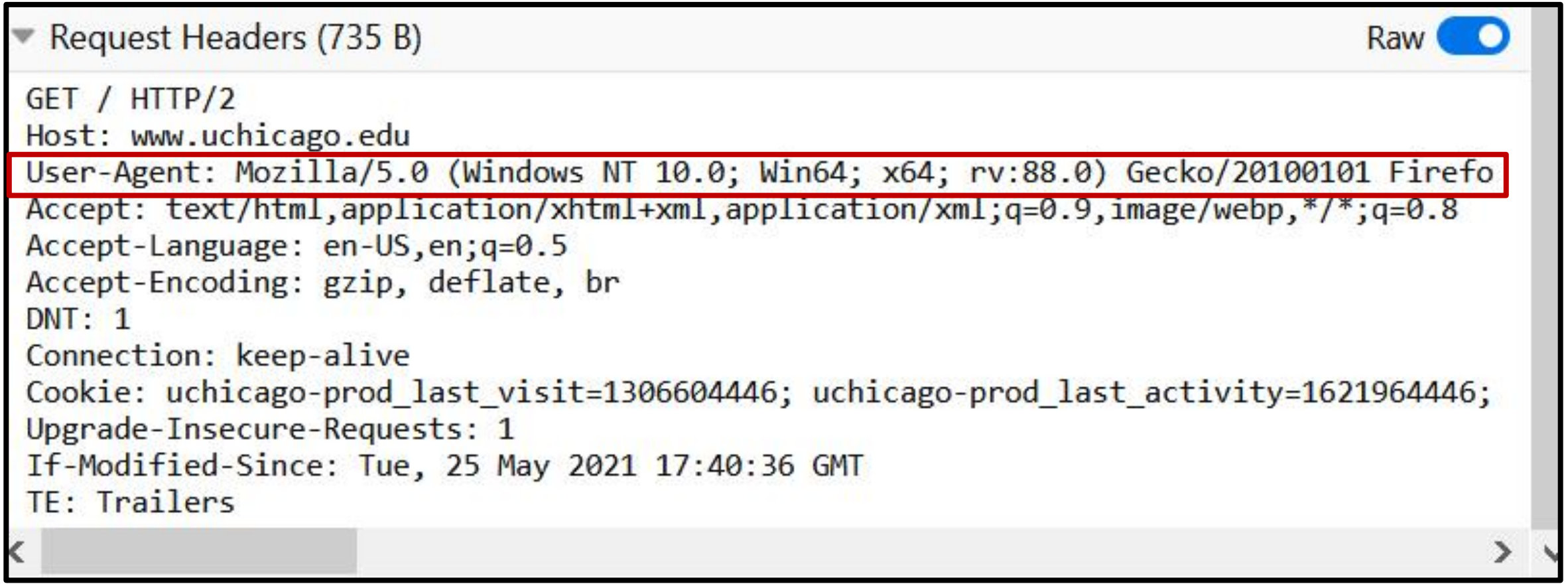
HTTP Headers (uchicago.edu)



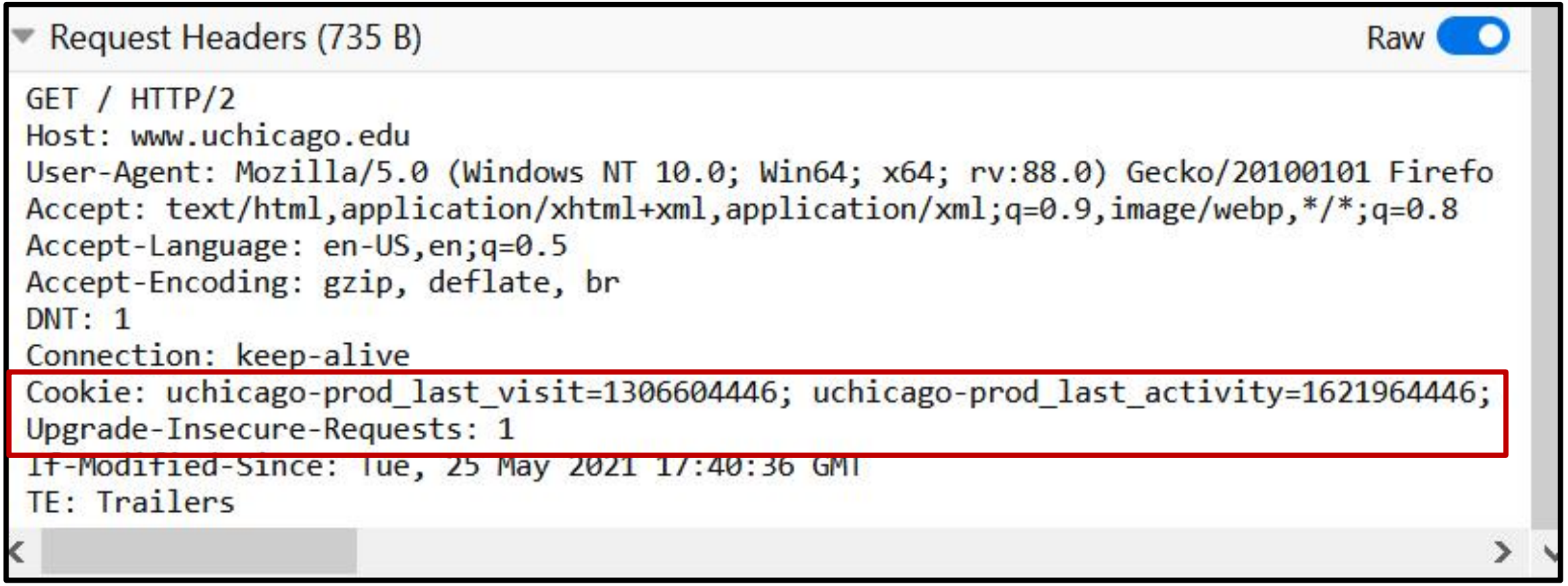
HTTP Headers (uchicago.edu): **Request**



HTTP Headers (uchicago.edu): **Request**



HTTP Headers (uchicago.edu): **Request**



Detail: HTTP Cookies

Image taken from https://networkencyclopedia.com/http-cookie/

HTTP Headers (uchicago.edu): **Response**

Response Headers (1.078 KB) Raw

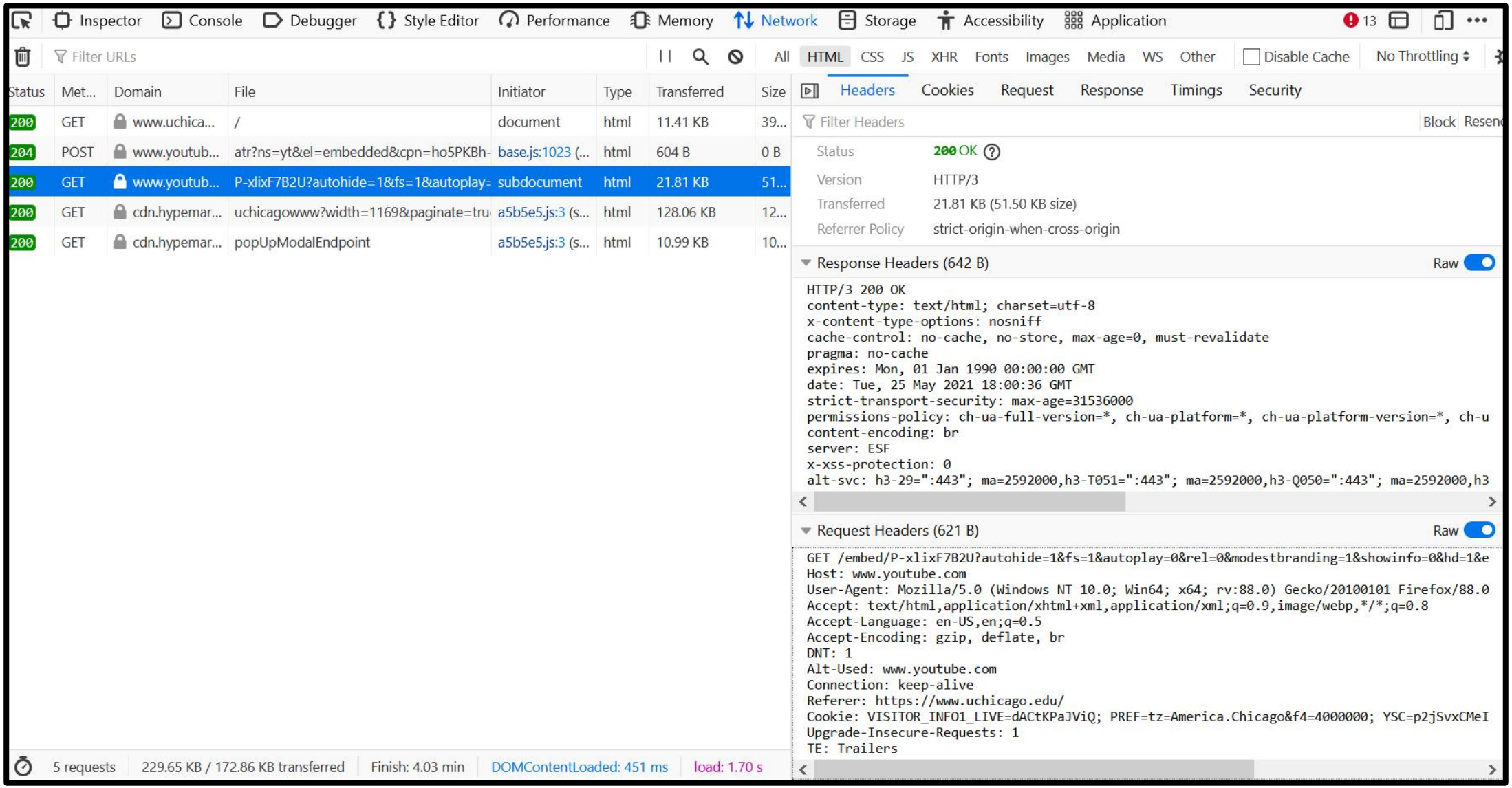
```
HTTP/2 200 OK
date: Tue, 25 May 2021 18:00:35 GMT
content-type: text/html; charset=UTF-8
server: Apache
x-frame-options: SAMEORIGIN
expires: Mon, 26 Jul 1997 05:00:00 GMT
pragma: no-cache
vary: Accept-Encoding
set-cookie: uchicago-prod_last_visit=1306605629; expires=Wed, 25-May-2022 18:00:29 G
set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-2022 18:00:2
set-cookie: uchicago-prod_tracker=%7B%220%22%3A%22index%22%2C%22token%22%3A%2226944a
set-cookie: uchicago-prod_csrf_token=deleted; expires=Thu, 01-Jan-1970 00:00:01 GMT;
set-cookie: uchicago-prod_csrf_token=86d47d8690aa7646e1628dd095cd5b464db16bd3; expir
last-modified: Tue, 25 May 2021 18:00:29 GMT
content-encoding: gzip
x-varnish: 10696657 9201444
age: 5
via: 1.1 varnish (Varnish/5.2)
accept-ranges: bytes
X-Firefox-Spdy: h2
```

HTTP Headers (uchicago.edu): **Response**

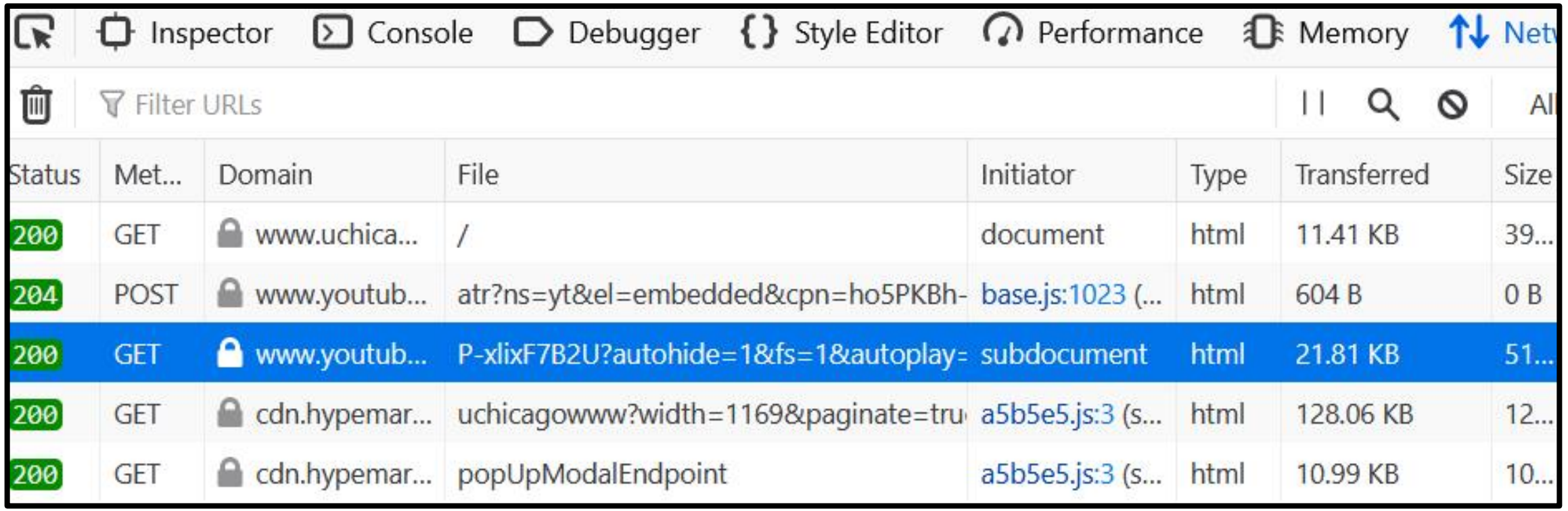
Response Headers (1.078 KB) Raw

```
HTTP/2 200 OK
date: Tue, 25 May 2021 18:00:35 GMT
content-type: text/html; charset=UTF-8
server: Apache
x-frame-options: SAMEORIGIN
expires: Mon, 26 Jul 1997 05:00:00 GMT
pragma: no-cache
vary: Accept-Encoding
set-cookie: uchicago-prod_last_visit=1306605629; expires=Wed, 25-May-2022 18:00:29 G
set-cookie: uchicago-prod_last_activity=1621965629; expires=Wed, 25-May-2022 18:00:2
set-cookie: uchicago-prod_tracker=%7B%220%22%3A%22index%22%2C%22token%22%3A%2226944a
set-cookie: uchicago-prod_csrf_token=deleted; expires=Thu, 01-Jan-1970 00:00:01 GMT;
set-cookie: uchicago-prod_csrf_token=86d47d8690aa7646e1628dd095cd5b464db16bd3; expir
last-modified: Tue, 25 May 2021 18:00:29 GMT
content-encoding: gzip
x-varnish: 10696657 9201444
age: 5
via: 1.1 varnish (Varnish/5.2)
accept-ranges: bytes
X-Firefox-Spdy: h2
```

HTTP Headers (uchicago.edu → youtube.com)

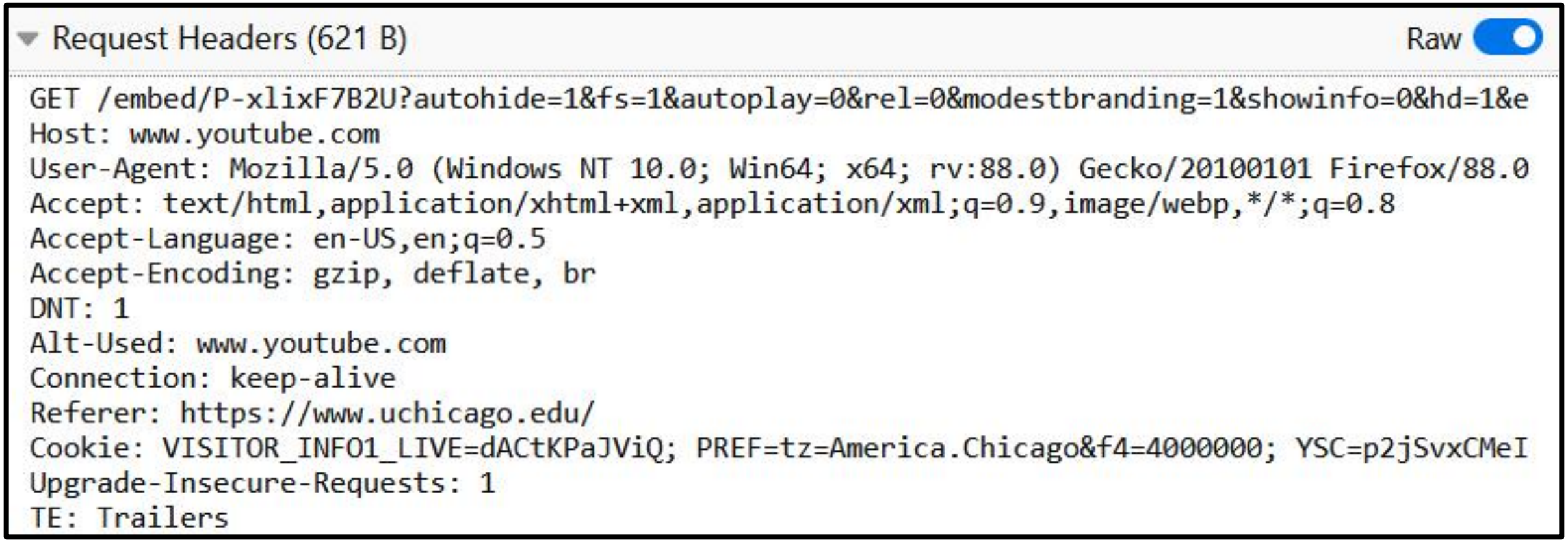


HTTP Headers (uchicago.edu → youtube.com)



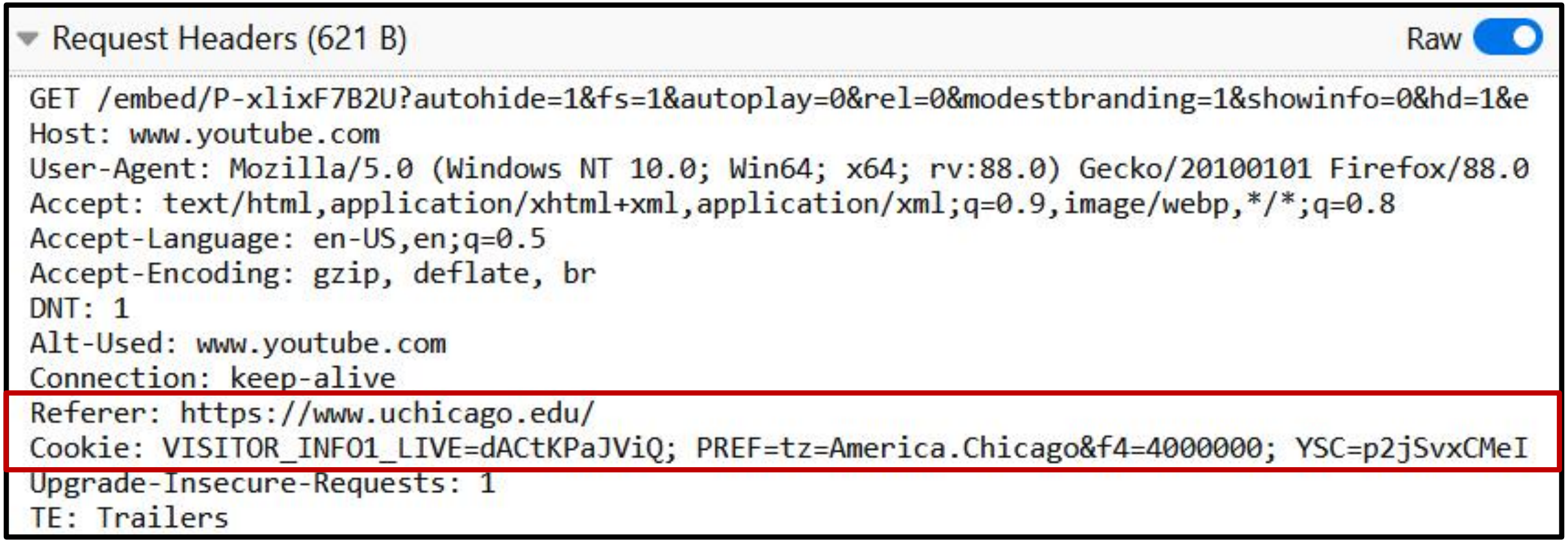
Status	Met...	Domain	File	Initiator	Type	Transferred	Size
200	GET	www.uchica...	/	document	html	11.41 KB	39...
204	POST	www.youtub...	atr?ns=yt&el=embedded&cpn=ho5PKBh-	base.js:1023 (...	html	604 B	0 B
200	GET	www.youtub...	P-xlixF7B2U?autohide=1&fs=1&autoplay=	subdocument	html	21.81 KB	51...
200	GET	cdn.hypemar...	uchicagowww?width=1169&paginate=tru	a5b5e5.js:3 (s...	html	128.06 KB	12...
200	GET	cdn.hypemar...	popUpModalEndpoint	a5b5e5.js:3 (s...	html	10.99 KB	10...

HTTP Headers (uchicago.edu → youtube.com)



HTTP Headers (uchicago.edu → youtube.com)

HTTP Headers (uchicago.edu → youtube.com)



Cookie Syncing

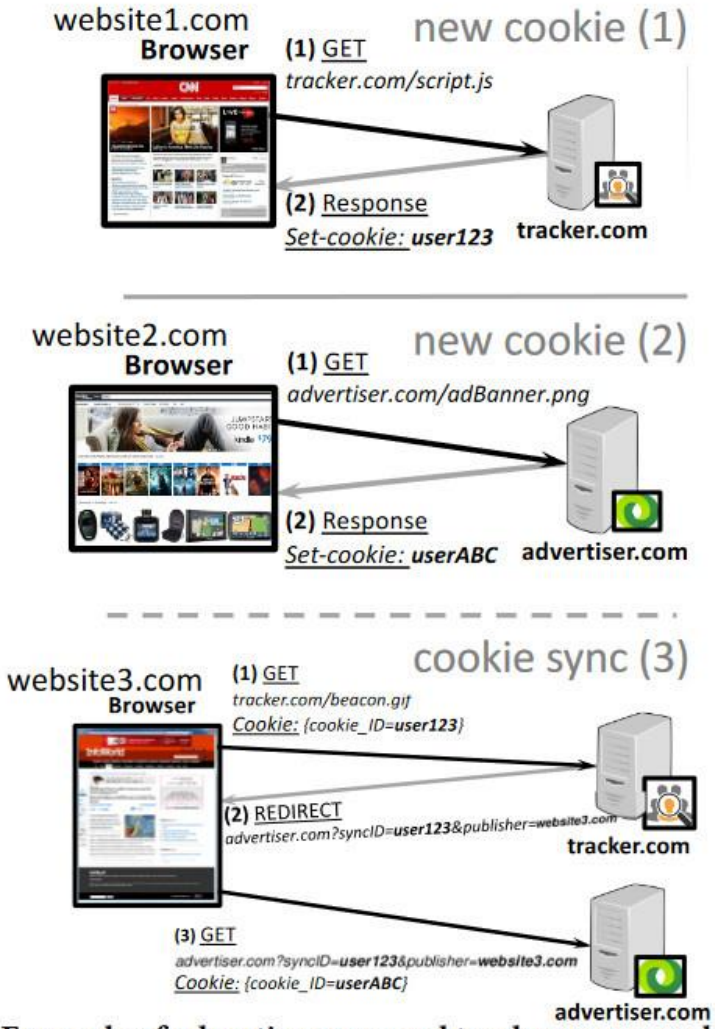


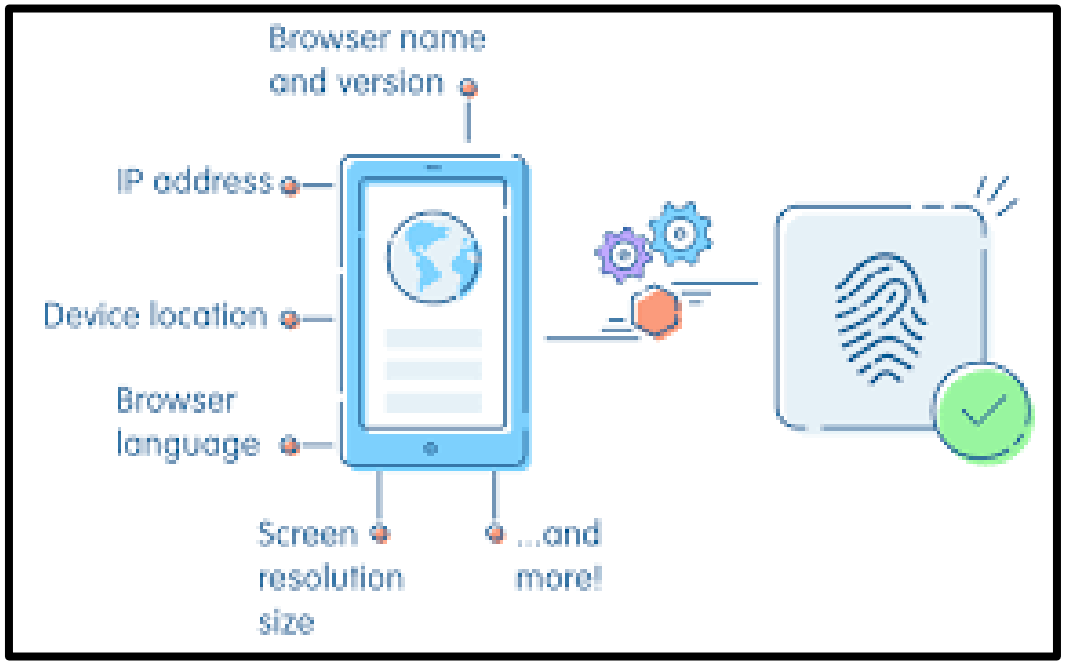
Figure 1: Example of advertiser.com and tracker.com synchronizing their cookieIDs. Interestingly, and without having any code in website3, advertiser.com learns that: (i) cookieIDs userABC==user123 and (ii) userABC has just visited the given website. Finally, both domains can conduct server-to-server user data merges.

From Papadopoulos et al. “Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask,” in Proc. WWW, 2019.

Mechanics of Tracking (Fingerprinting Version)

Browser Fingerprinting

- Use features of the browser that are relatively unique to your machine
  - Fonts
  - GPU model anti-aliasing (Canvas fingerprinting)
  - User-agent string
  - *(Often not)* IP address *(Why not?)*



See Fingerprinting Happening

The Ecosystem Underlying Targeted Ads

Ad Bidding Marketplaces

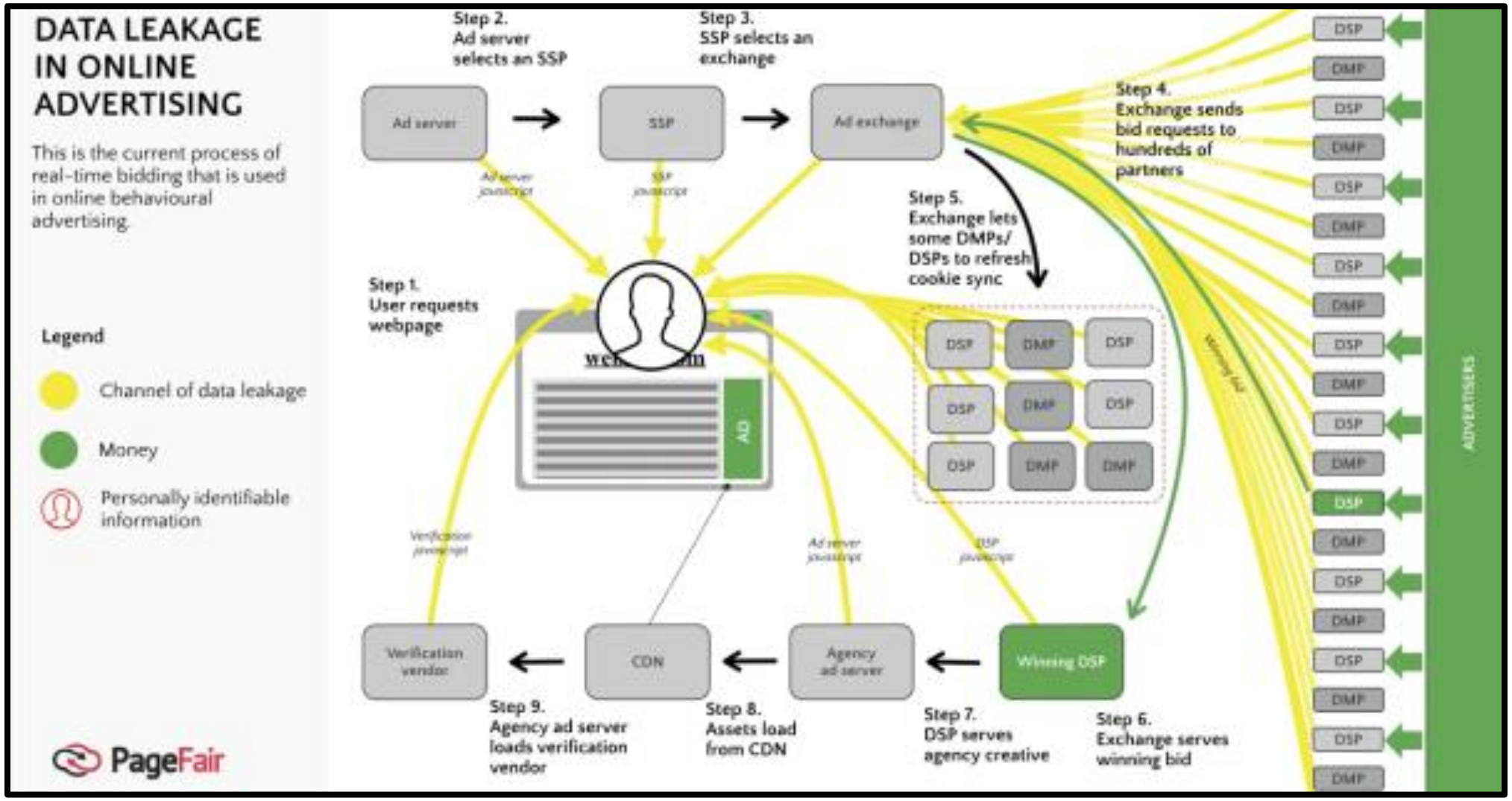
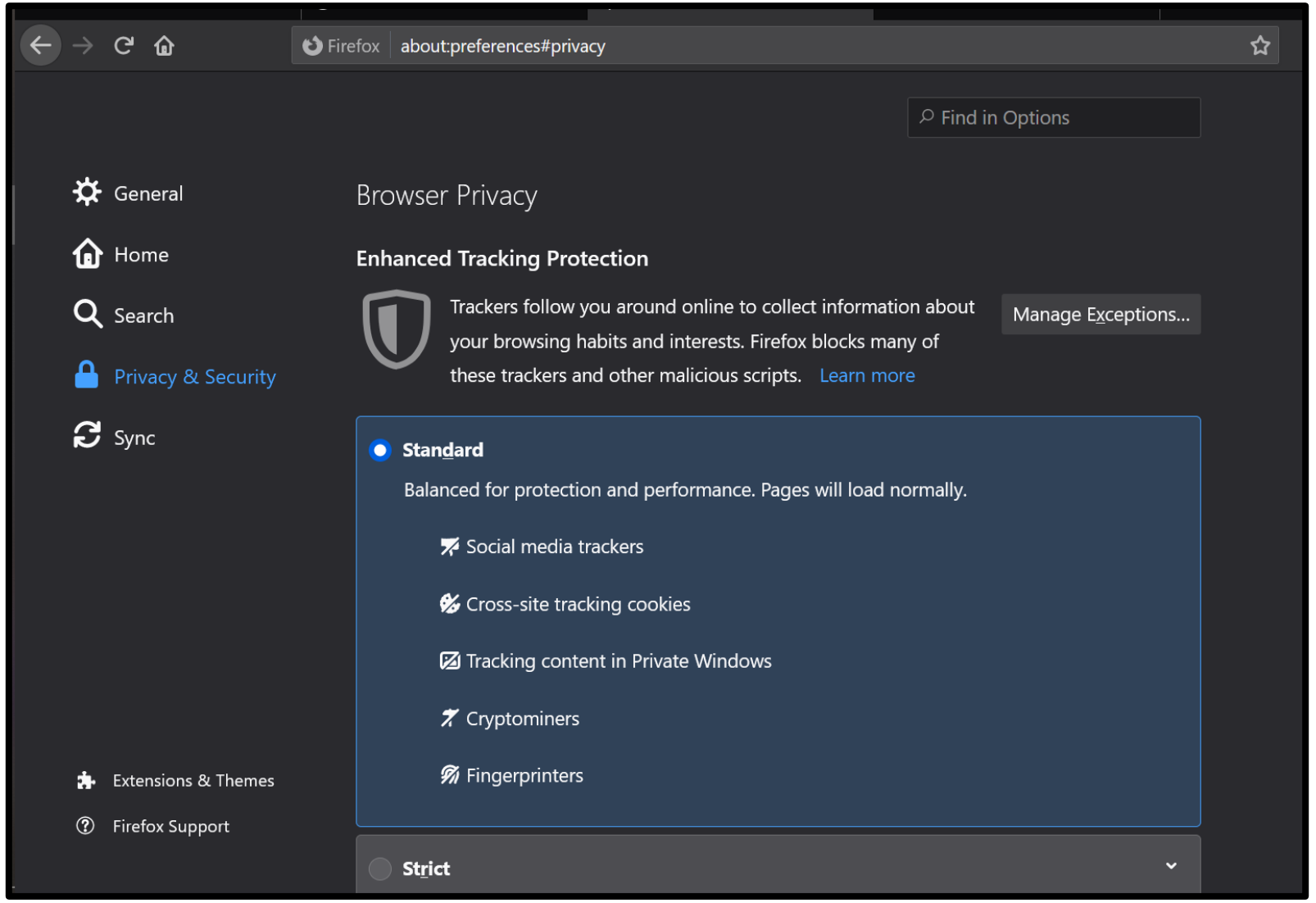


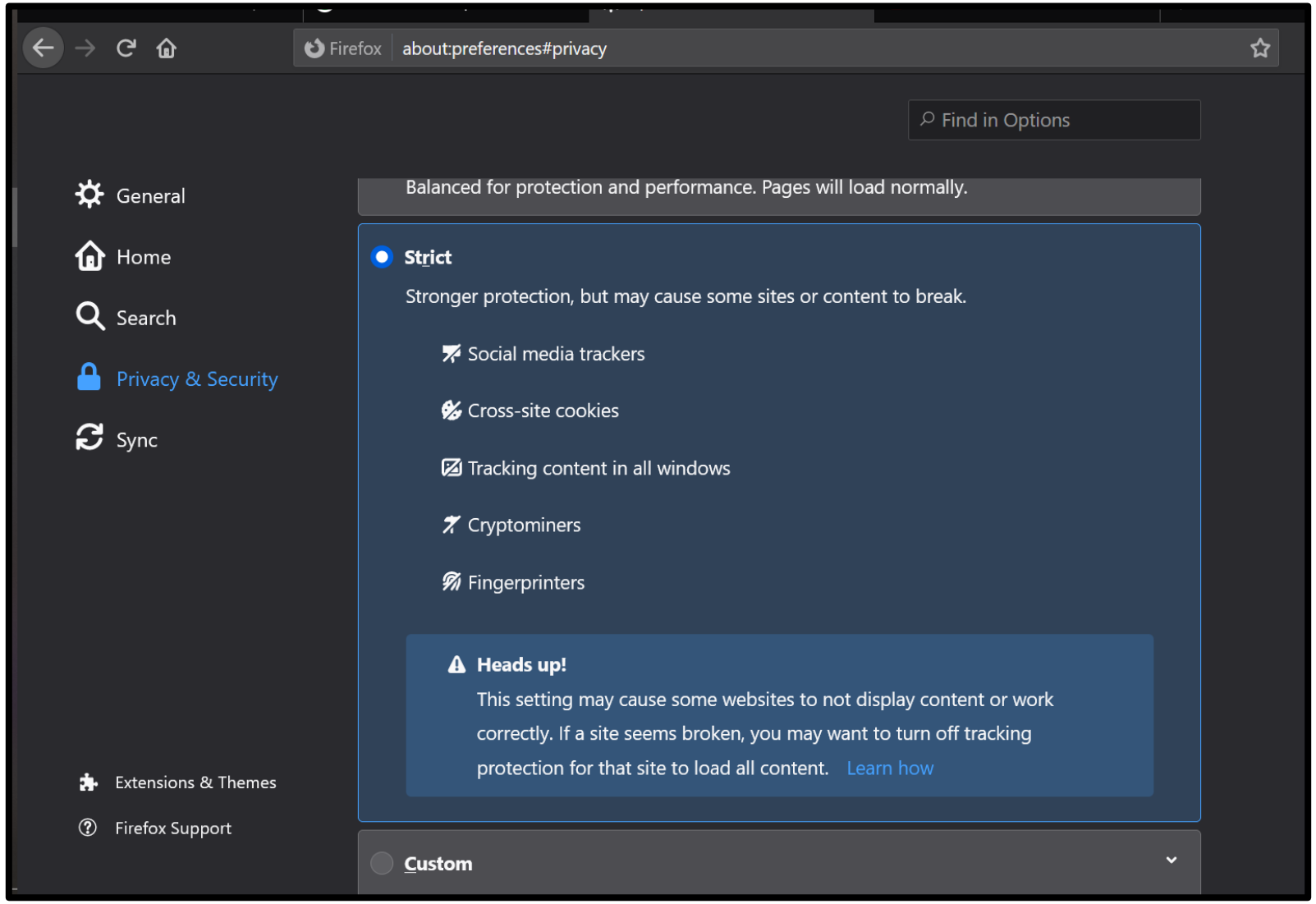
Image taken from https://vimeo.com/219820643

Anti-Tracking Mechanisms

Firefox Browser (New-ish) Default Settings

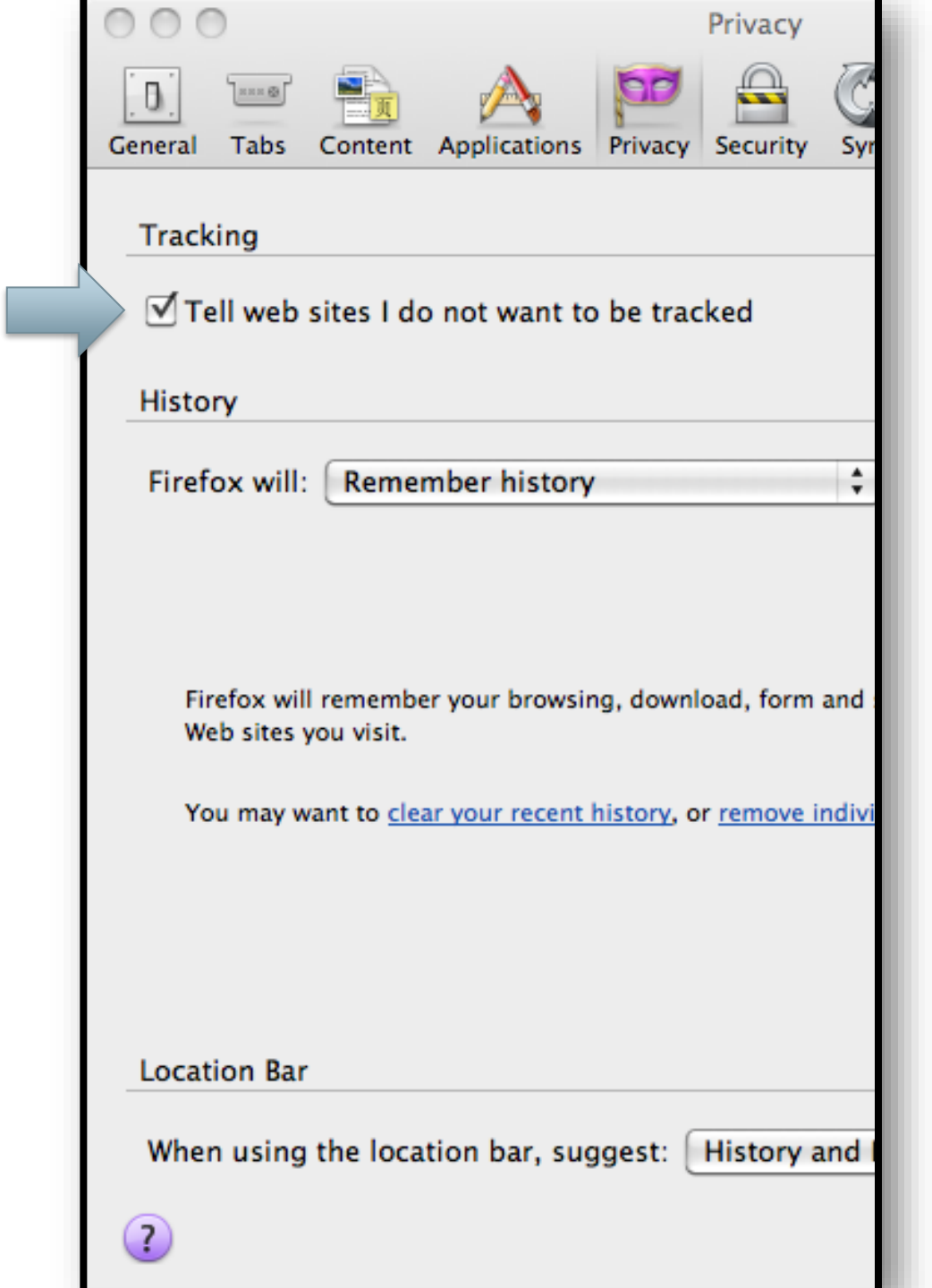


Firefox Browser Strict Settings



Do not track

- W3C standard
- User checks a box
- Browser sends a “do not track” header to website
- Website stops “tracking”
- W3C working group tried to define what “stopping tracking” meant



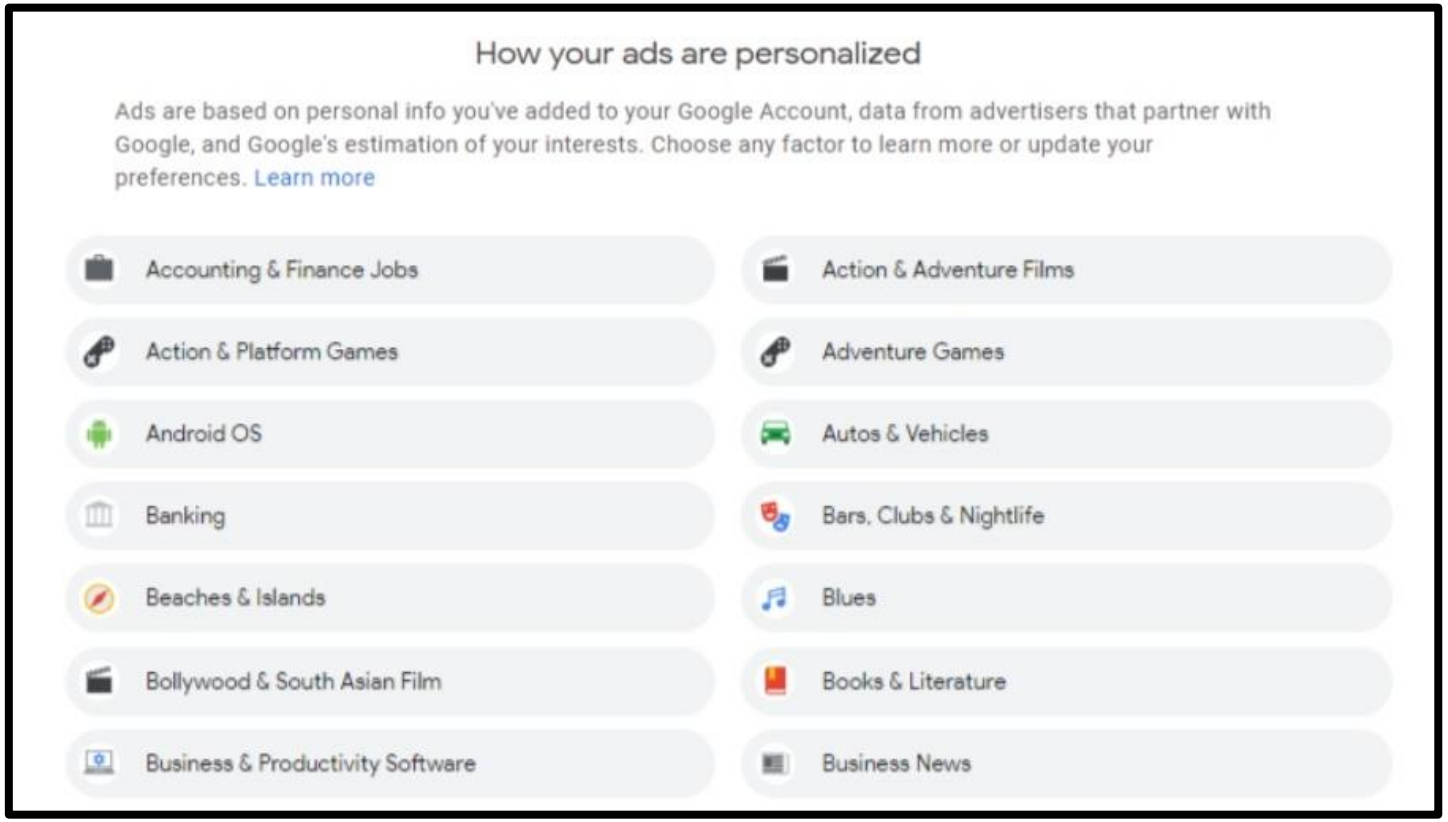
Tools to stop tracking, effective?

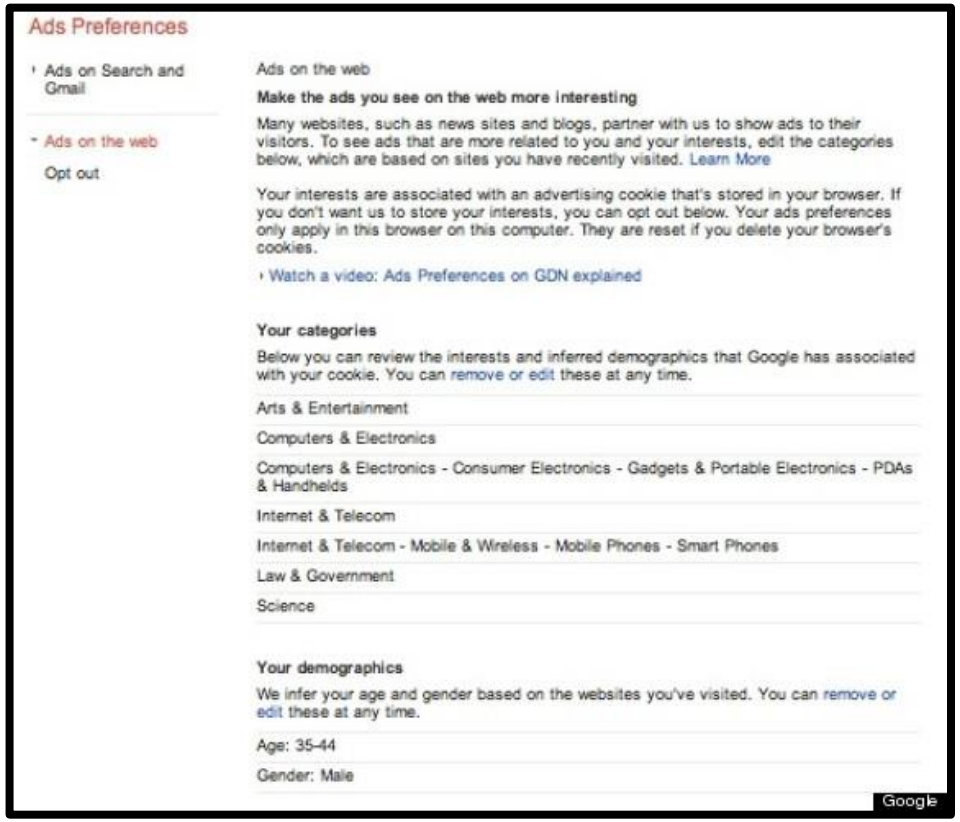
- Browser privacy settings
  - Cookie blocking
  - P3P
  - Tracking Protection Lists
  - Do Not Track
- Browser add-ons
- Opt-out cookies
- Digital Advertising Alliance (DAA) AdChoices icon and associated opt-out pages

Transparency Mechanisms

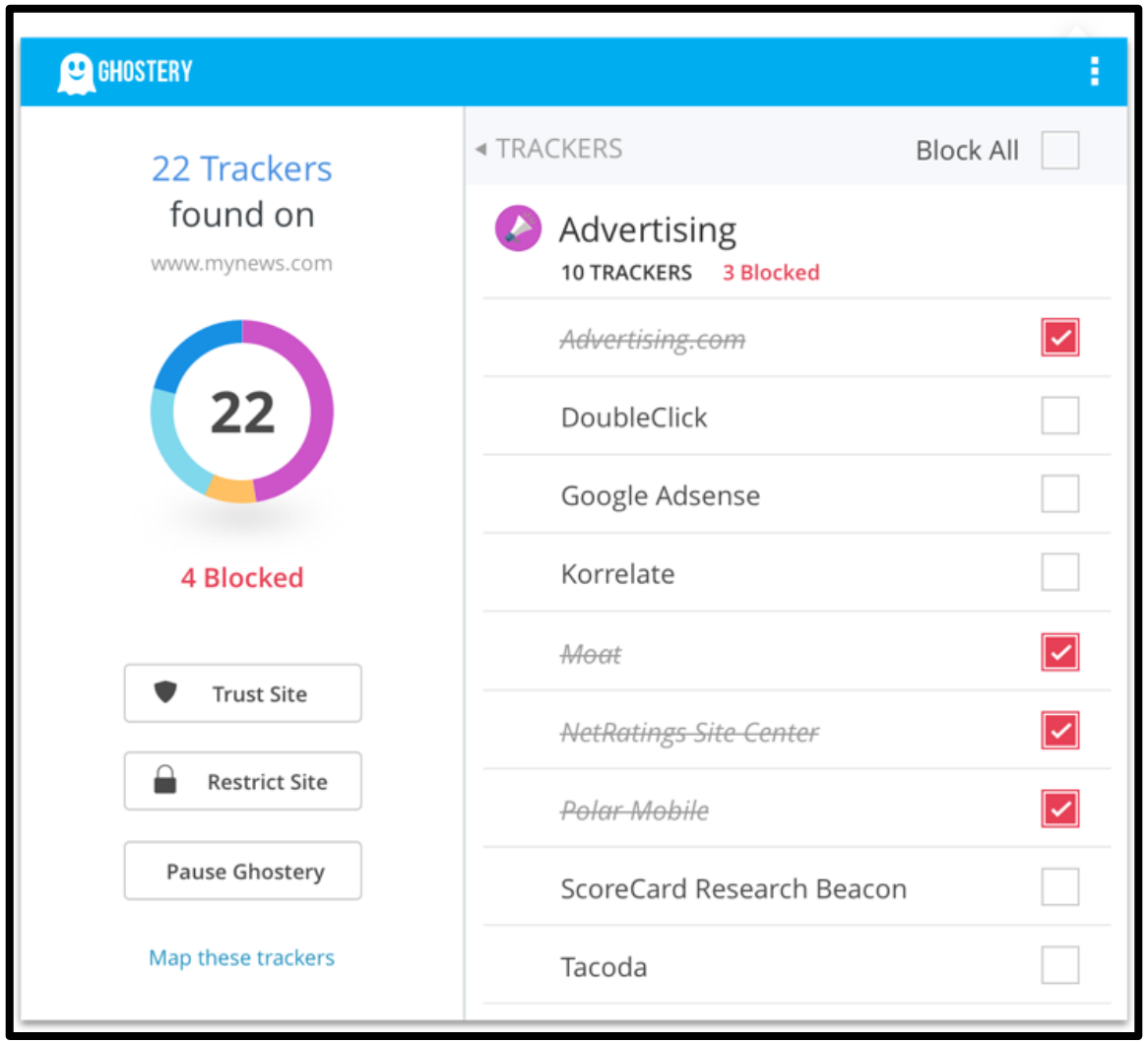
Transparency Dashboards

- Summaries of inferred preferences and demographics

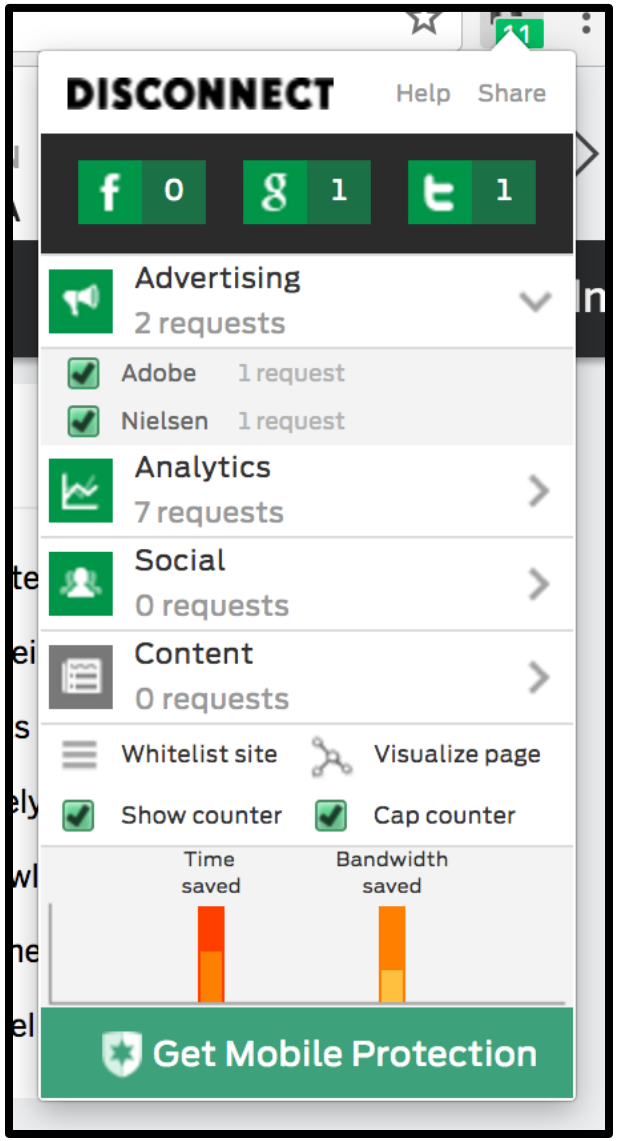


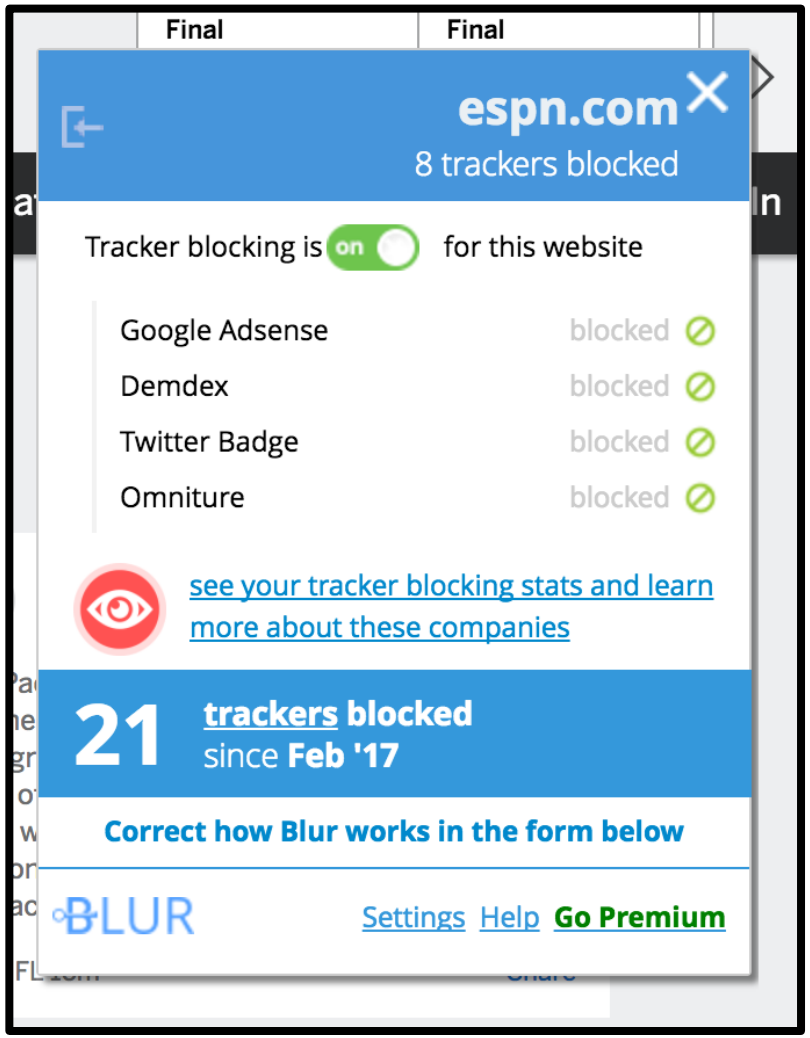


Ghostery

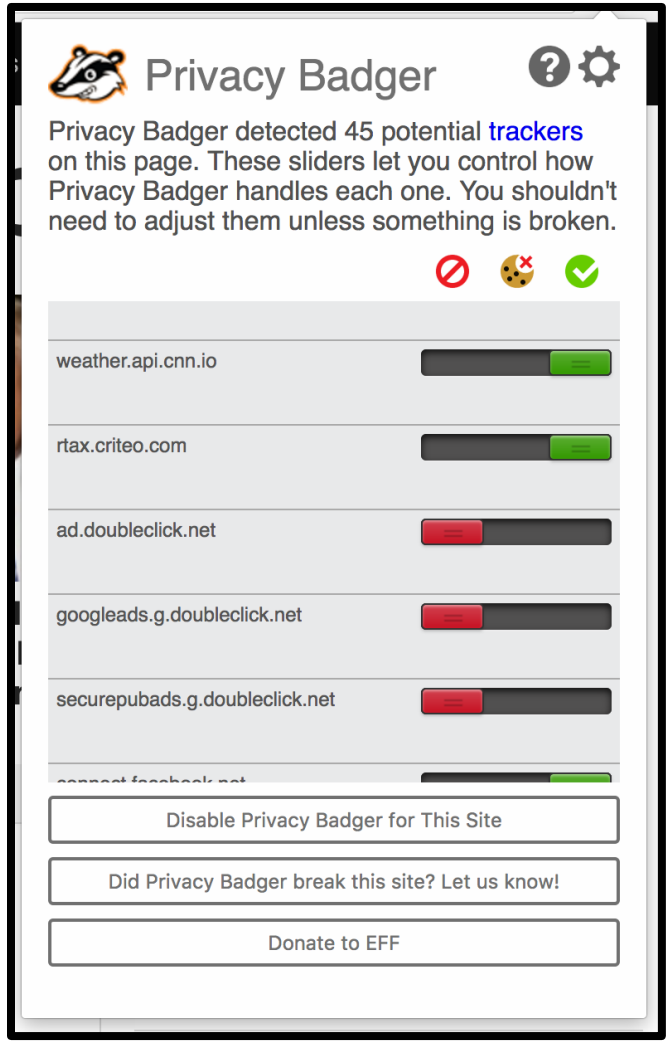

GHOSTERY
22 Trackers
found on
www.mynews.com
22
4 Blocked
Trust Site
Restrict Site
Pause Ghostery
Map these trackers
TRACKERS
Block All
Advertising
10 TRACKERS
3 Blocked
Advertising.com
DoubleClick
Google Adsense
Korrelate
Moat
NetRatings Site Center
Polar Mobile
ScoreCard Research Beacon
Tacoda

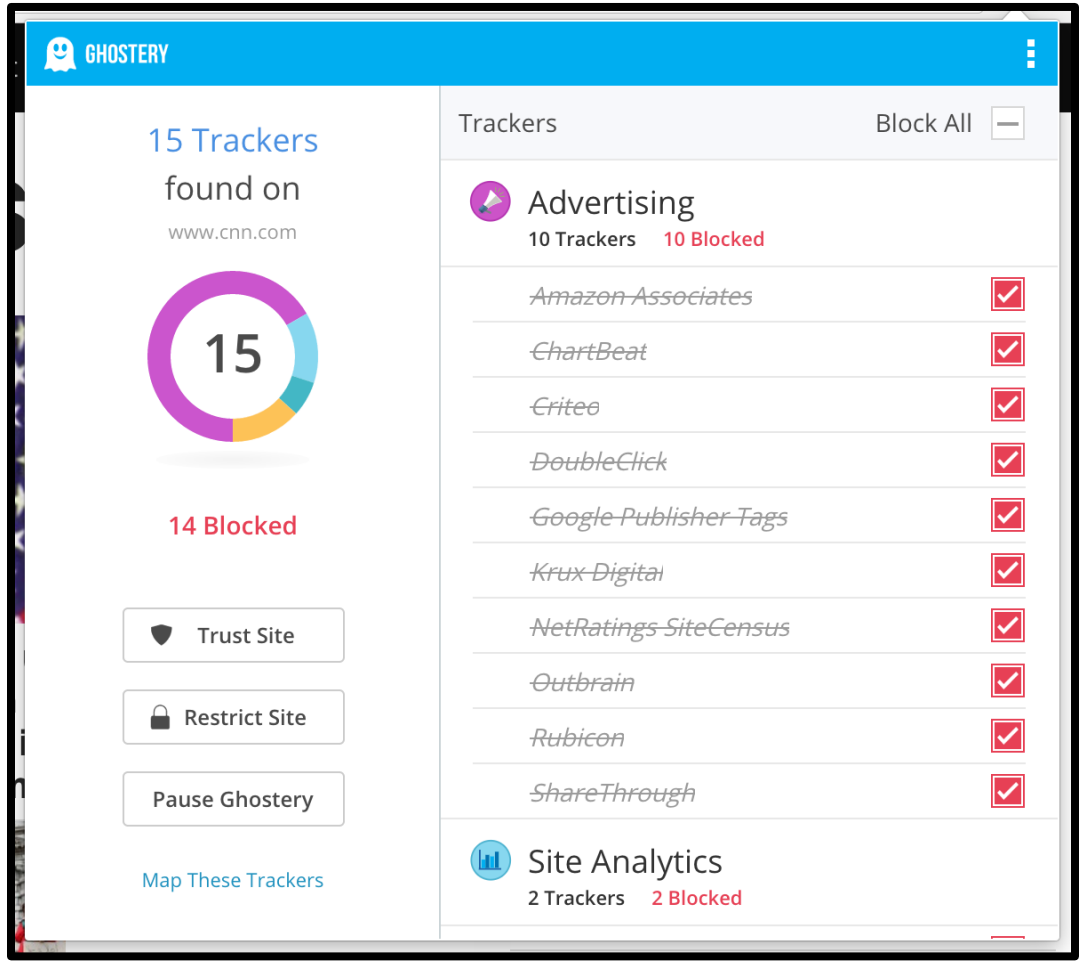
Existing Privacy Tools



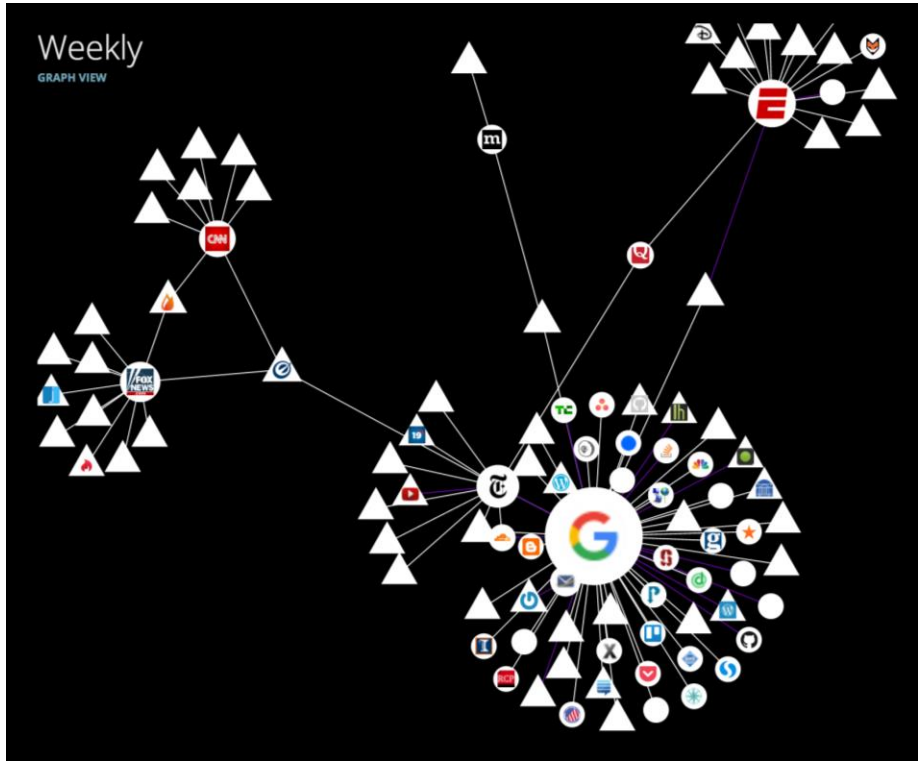


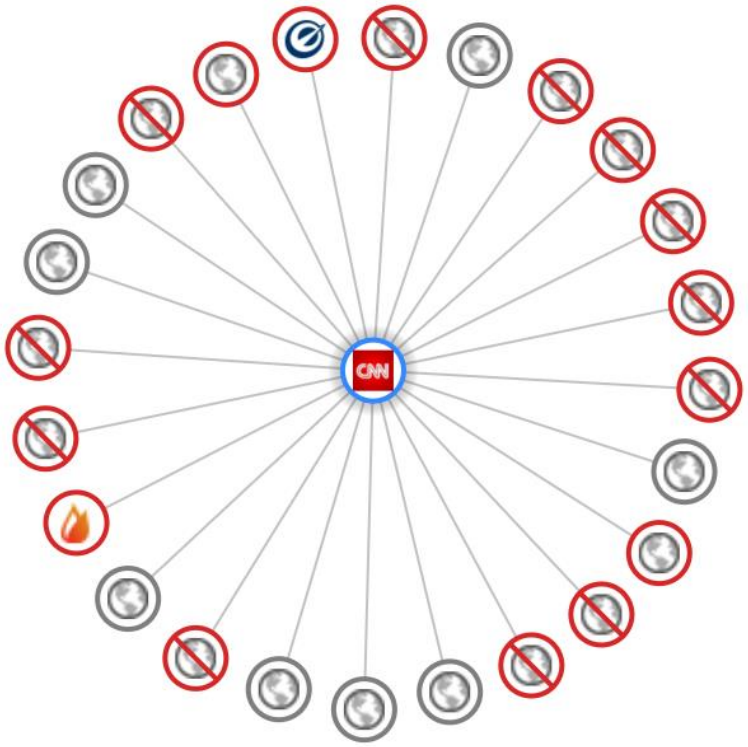
Existing Privacy Tools



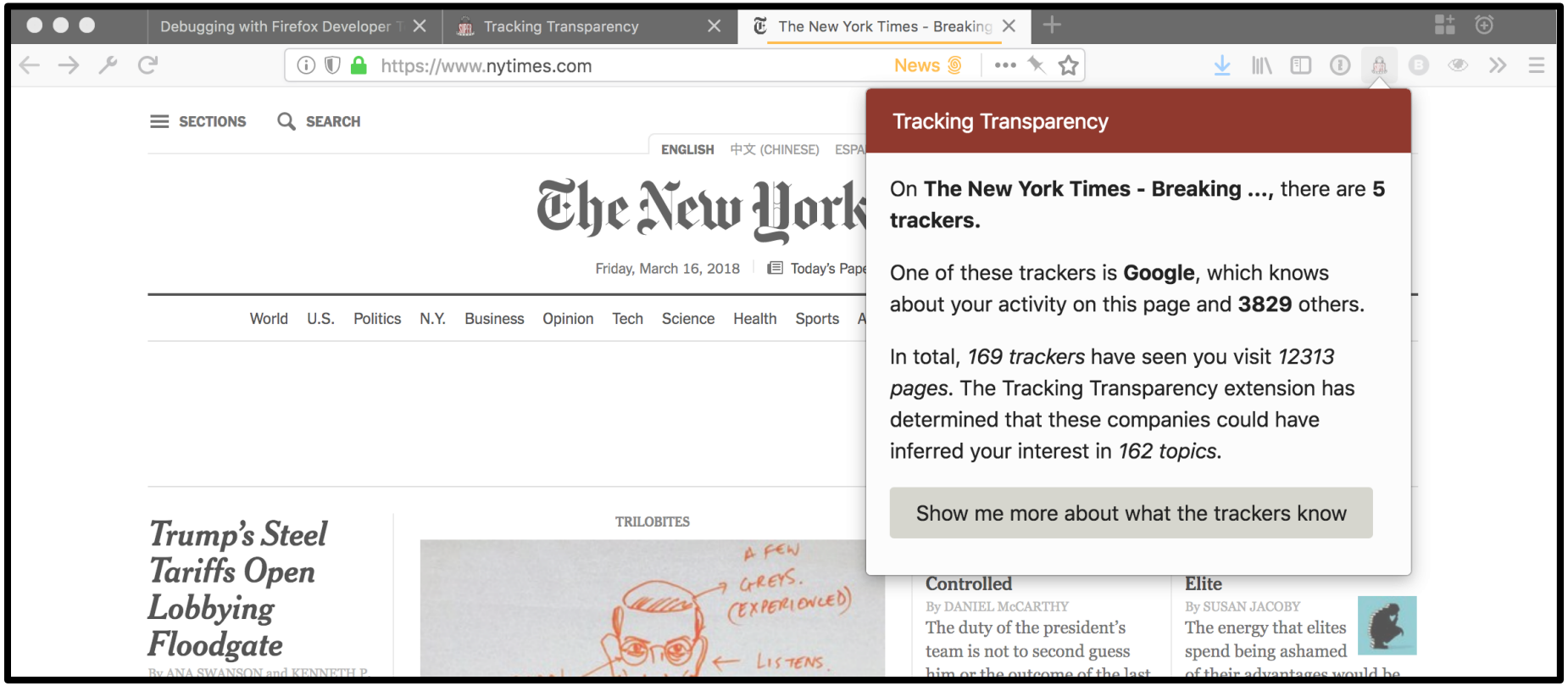


Visualization: Connection Graphs

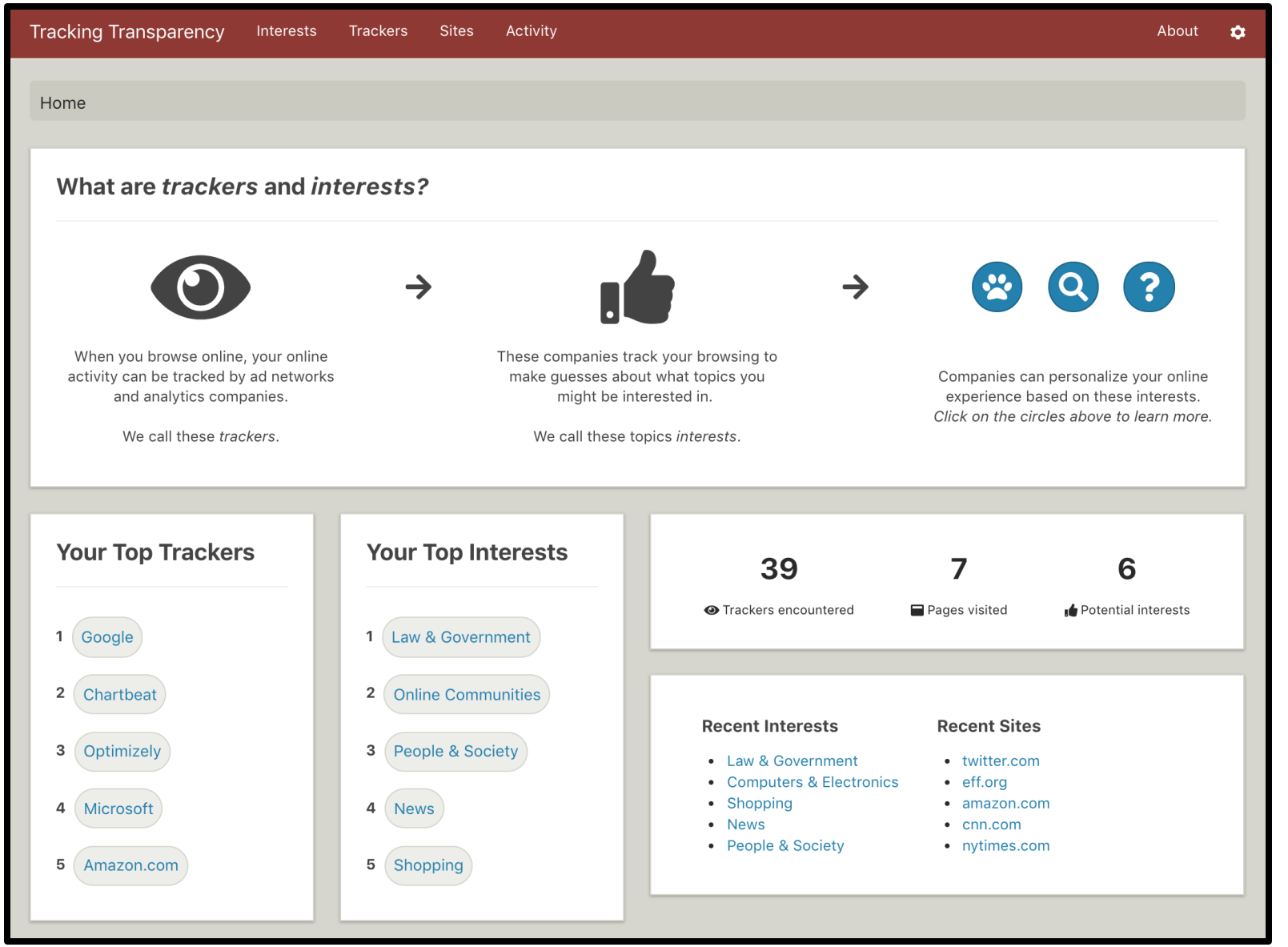




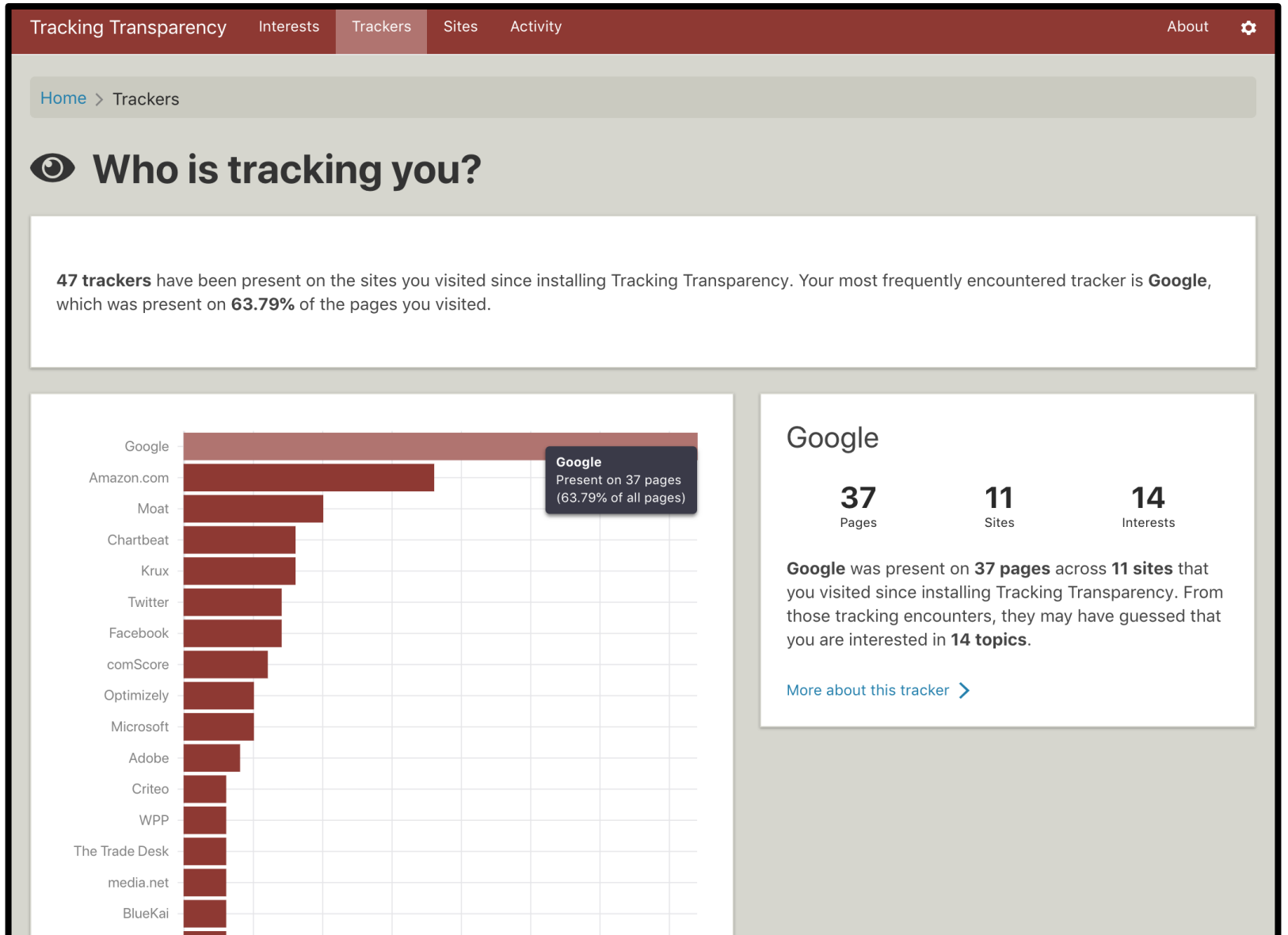
(Our Group's) Tracking Transparency



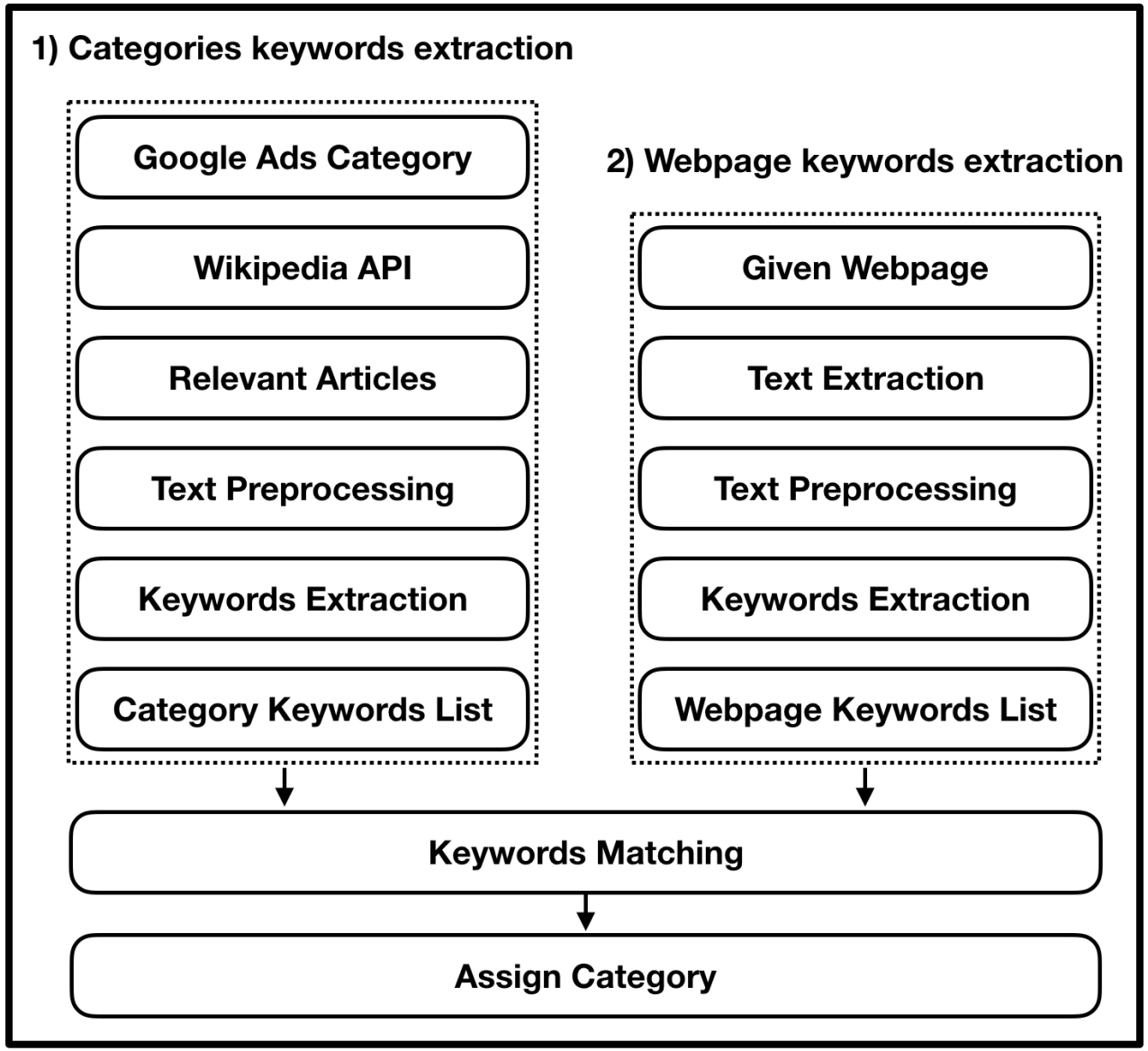
(Our Group's) Tracking Transparency



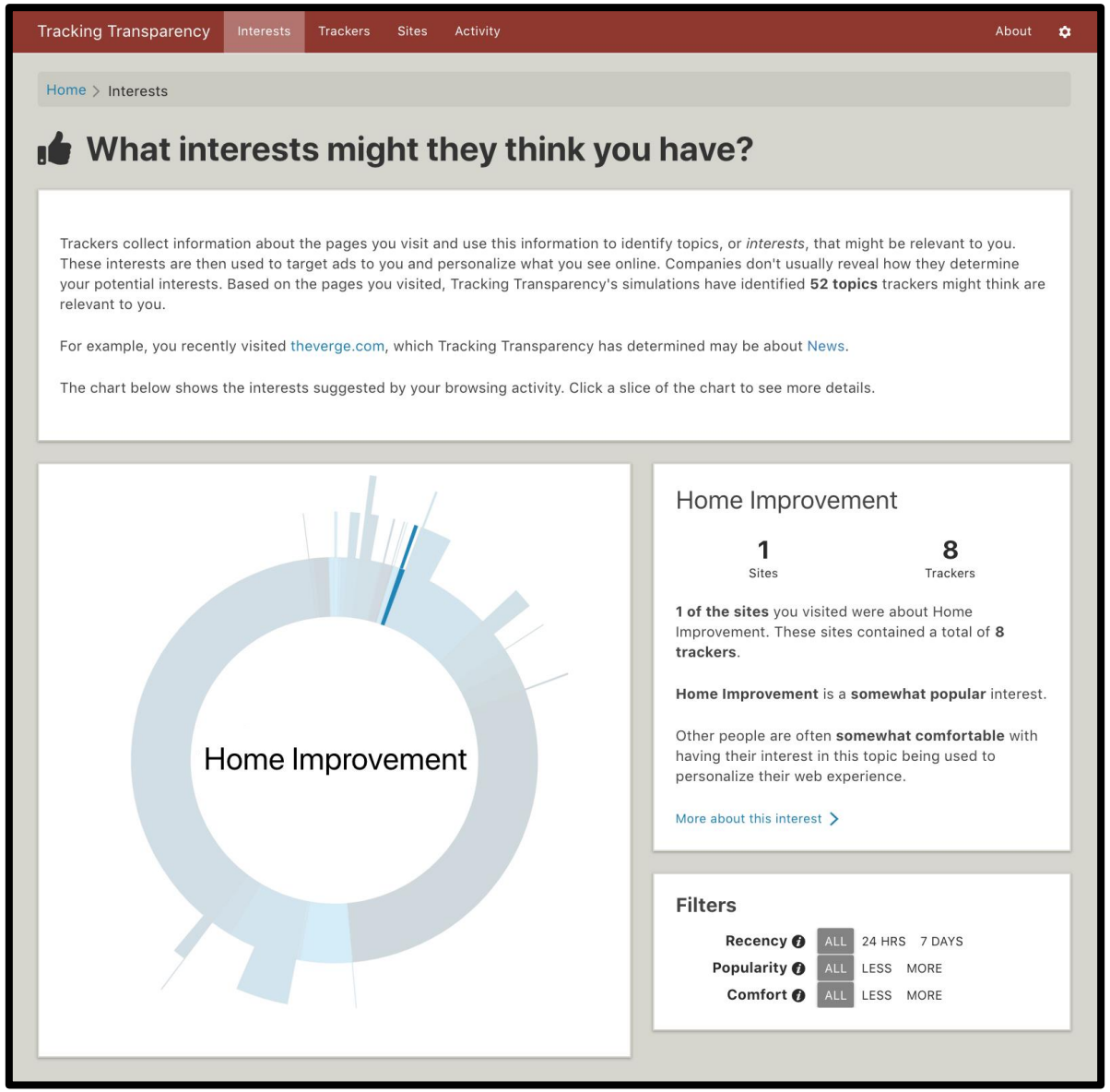
(Our Group's) Tracking Transparency



(Our Group's) Tracking Transparency



(Our Group's) Tracking Transparency



Mozilla Firefox Protections Dashboard

The Politics of Transparency

https://adobservatory.org/

The Politics of Transparency

How it works

Ad Observer is a tool you add to your Web browser. It copies the ads you see on Facebook and YouTube, so anyone can see them in our public database. (See documentation here.) If you want, you can enter basic demographic information about yourself in the tool to help improve our understanding of why advertisers targeted you. However, we'll never ask for information that could identify you.

It doesn't collect your personal information. We take your privacy very seriously. You can view the extension code on Github.

See Mozilla privacy review.

What we collect

- The advertiser's name and disclosure string.
- The ad's text, image, and link.
- The information Facebook provides about how the ad was targeted.
- When the ad was shown to you.
- Your browser language.

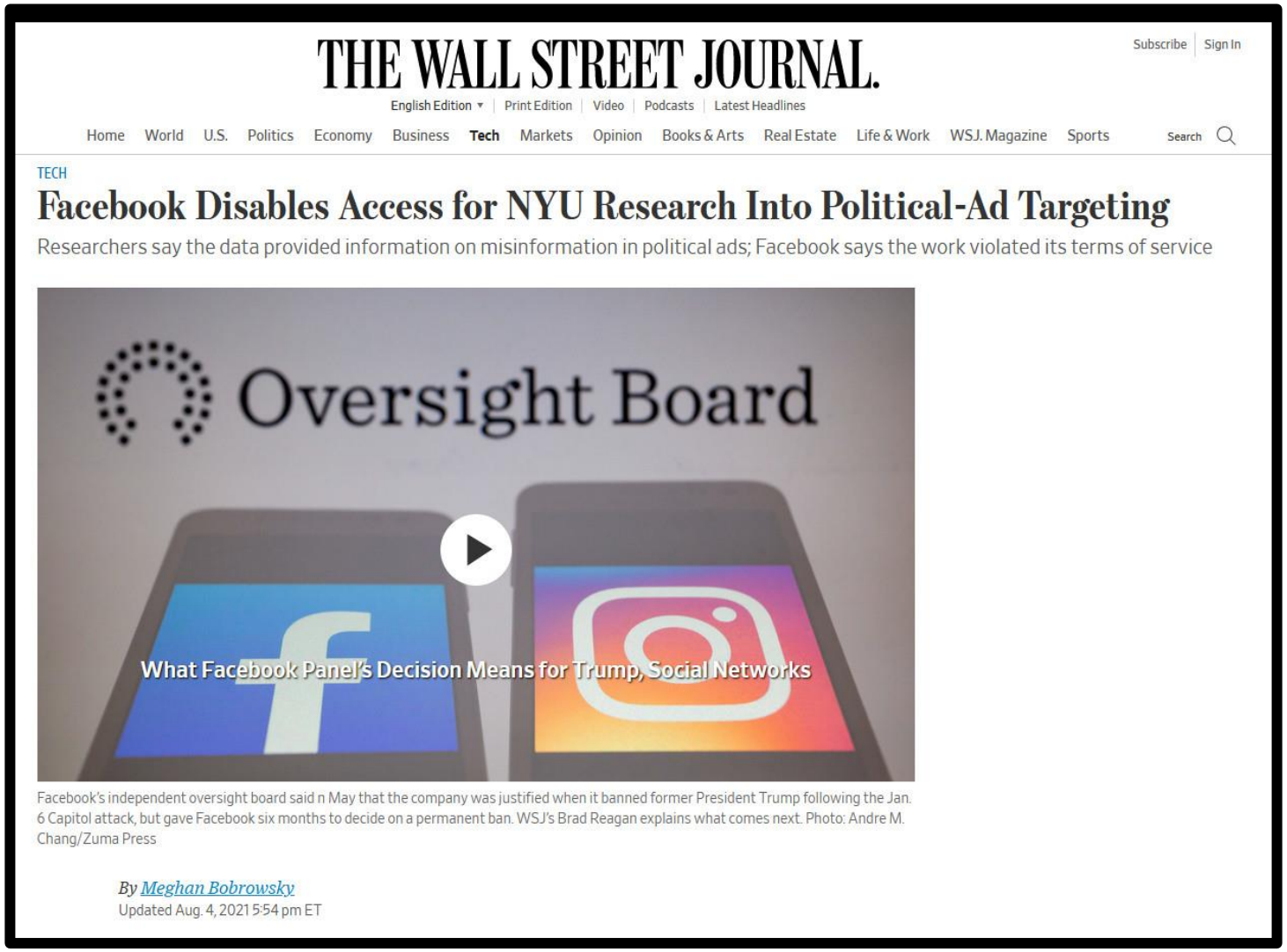
What we don't collect

Anything personally identifying, including:

- Your Facebook ID number.
- Your name, birthday, friend list, etc.
- How you interacted with ads or anything else.

https://adobservatory.org/

The Politics of Transparency



THE WALL STREET JOURNAL.

Subscribe | Sign In

English Edition | Print Edition | Video | Podcasts | Latest Headlines

Home World U.S. Politics Economy Business **Tech** Markets Opinion Books & Arts Real Estate Life & Work WSJ. Magazine Sports Search

TECH

Facebook Disables Access for NYU Research Into Political-Ad Targeting

Researchers say the data provided information on misinformation in political ads; Facebook says the work violated its terms of service

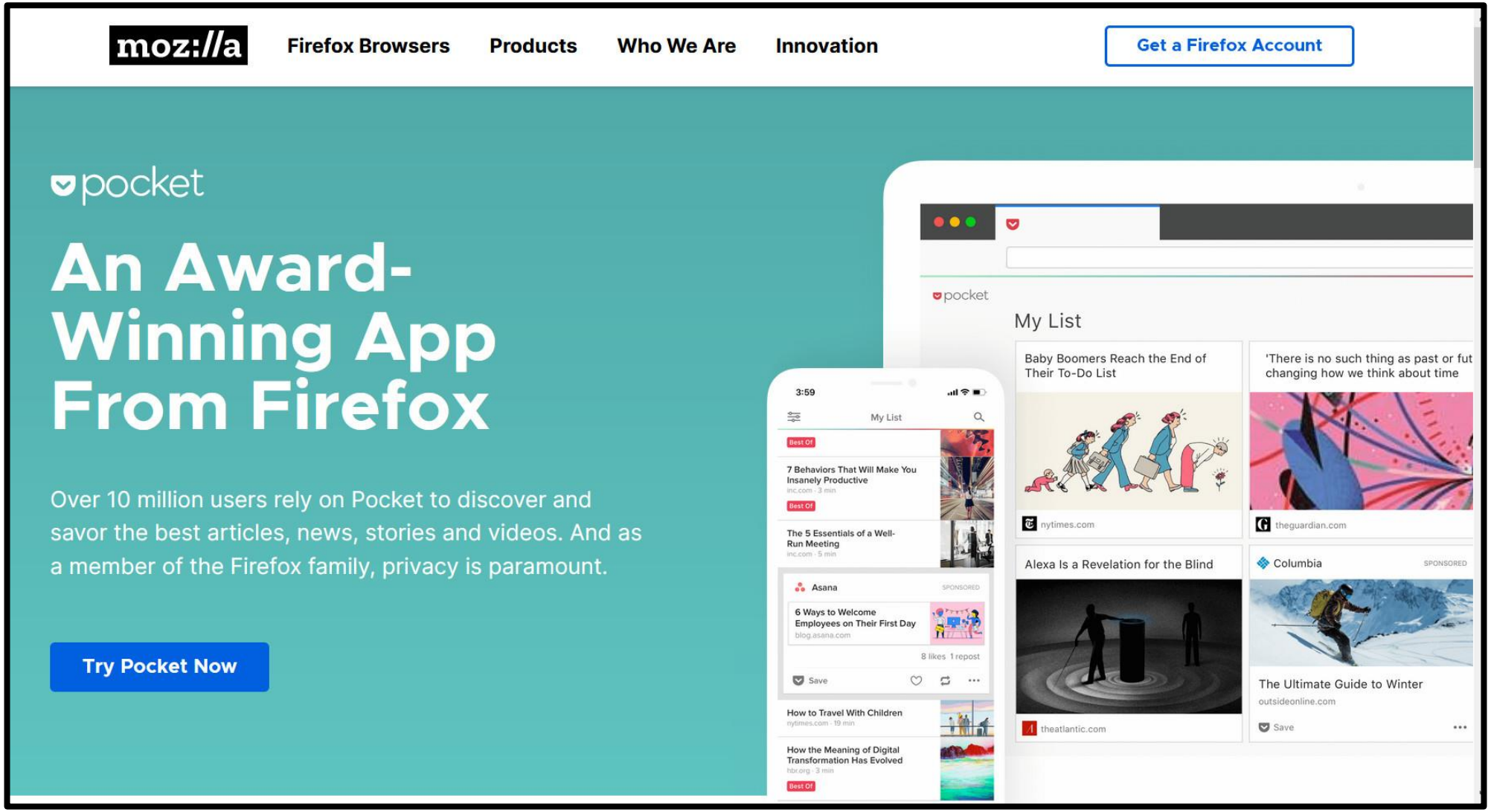
Facebook's independent oversight board said n May that the company was justified when it banned former President Trump following the Jan. 6 Capitol attack, but gave Facebook six months to decide on a permanent ban. WSJ's Brad Reagan explains what comes next. Photo: Andre M. Chang/Zuma Press

By Meghan Bobrowsky
Updated Aug. 4, 2021 5:54 pm ET

https://www.wsj.com/articles/facebook-cuts-off-access-for-nyu-research-into-political-ad-targeting-11628052204

Examining Two Technical Alternatives for Targeting

Mozilla Pocket: Customization With Only Local Tracking



Mozilla Pocket: Customization With Only Local Tracking

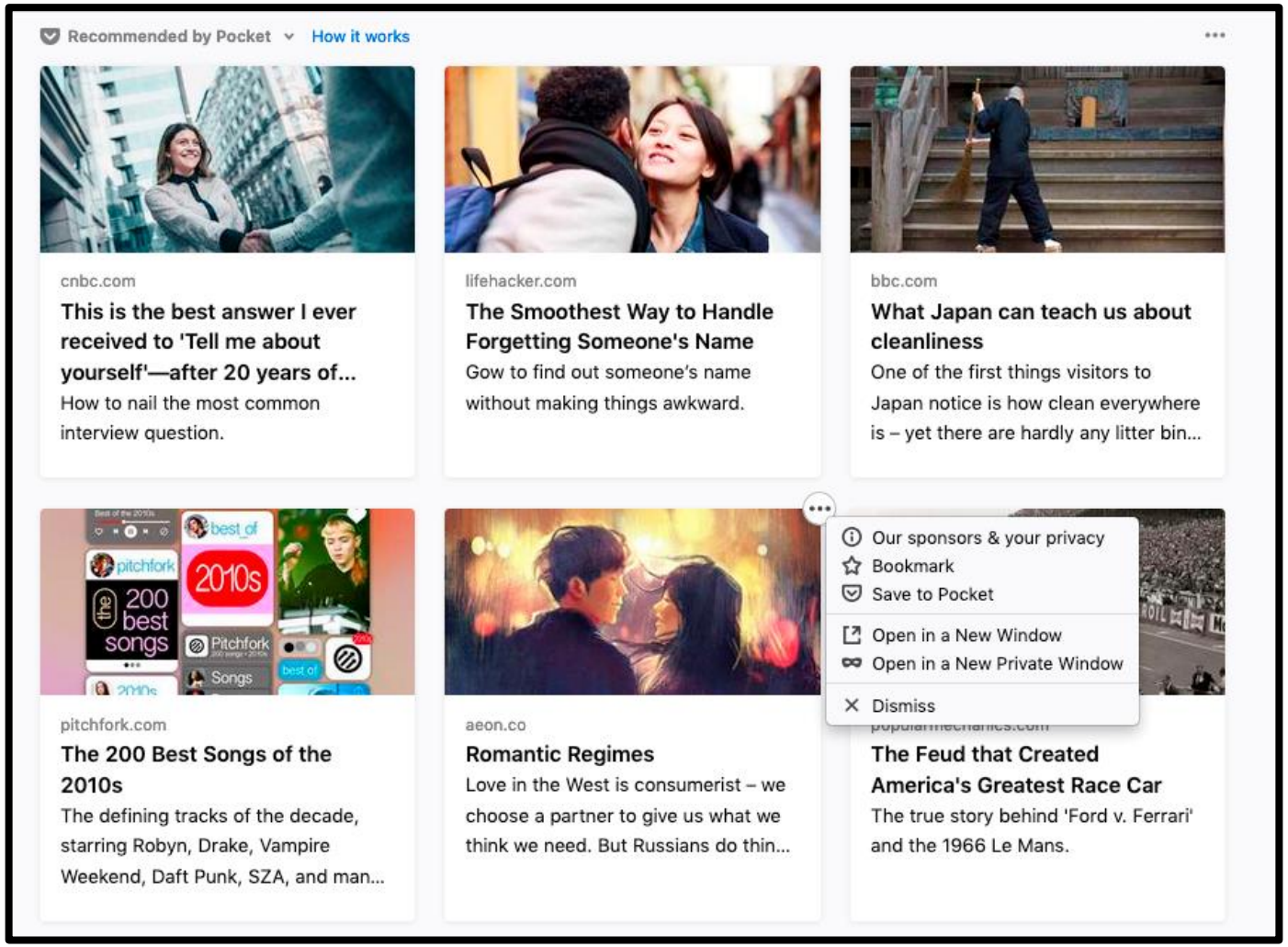


Image taken from https://blog.mozilla.org/en/products/firefox/pocket-in-firefox/

Mozilla Pocket: Customization With Only Local Tracking

Your data stays private. Always.

In addition to dishing up captivating stories, we also show you relevant, highly-vetted content from select sponsors. Rest assured, your browsing data never leaves your personal copy of Firefox—we don't see it, and our sponsors don't either. Want more details? Here's the full scoop on how privacy works in Pocket.

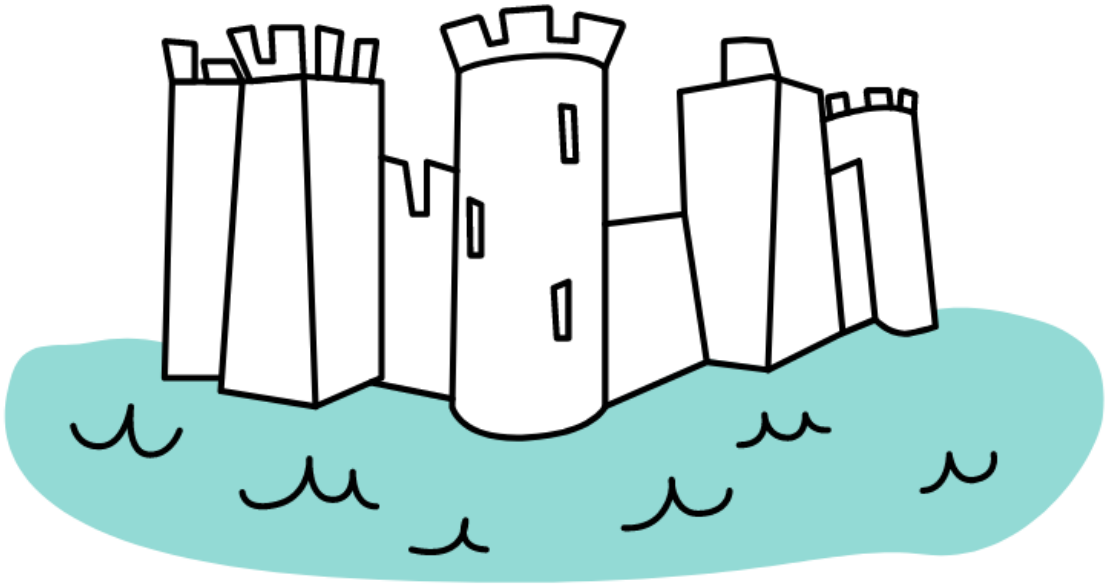


Image taken from https://getpocket.com/en/firefox/new_tab_learn_more/

Mozilla Pocket: Customization With Only Local Tracking

Are these stories in the new tab personalized to me?

For the most part, no. Most recommendations on your new tab come from a general list of the best of stories on the web. But Pocket is actively exploring ways to deliver personalized recommendations in a way that vigorously protects users' privacy. Importantly, **neither Mozilla nor Pocket ever receives a copy of your browser history.** When personalization does occur, recommendations rely on a process of story sorting and filtering that happens locally in your personal copy of Firefox.

Is information collected when I interact with these new tab recommendations? What information is collected and what controls do I have?

By default, when recommendations from Pocket are displayed on your new tab, we collect information about how many times they appear and how many times they are clicked. However, this information is not associated with any of the technical and interaction information about you or your copy of Firefox. You can learn more about the data we collect and you can opt out of data collection.

Image taken from https://help.getpocket.com/article/1142-firefox-new-tab-recommendations-faq#personalized

Google FLoC (Google's Perspective)

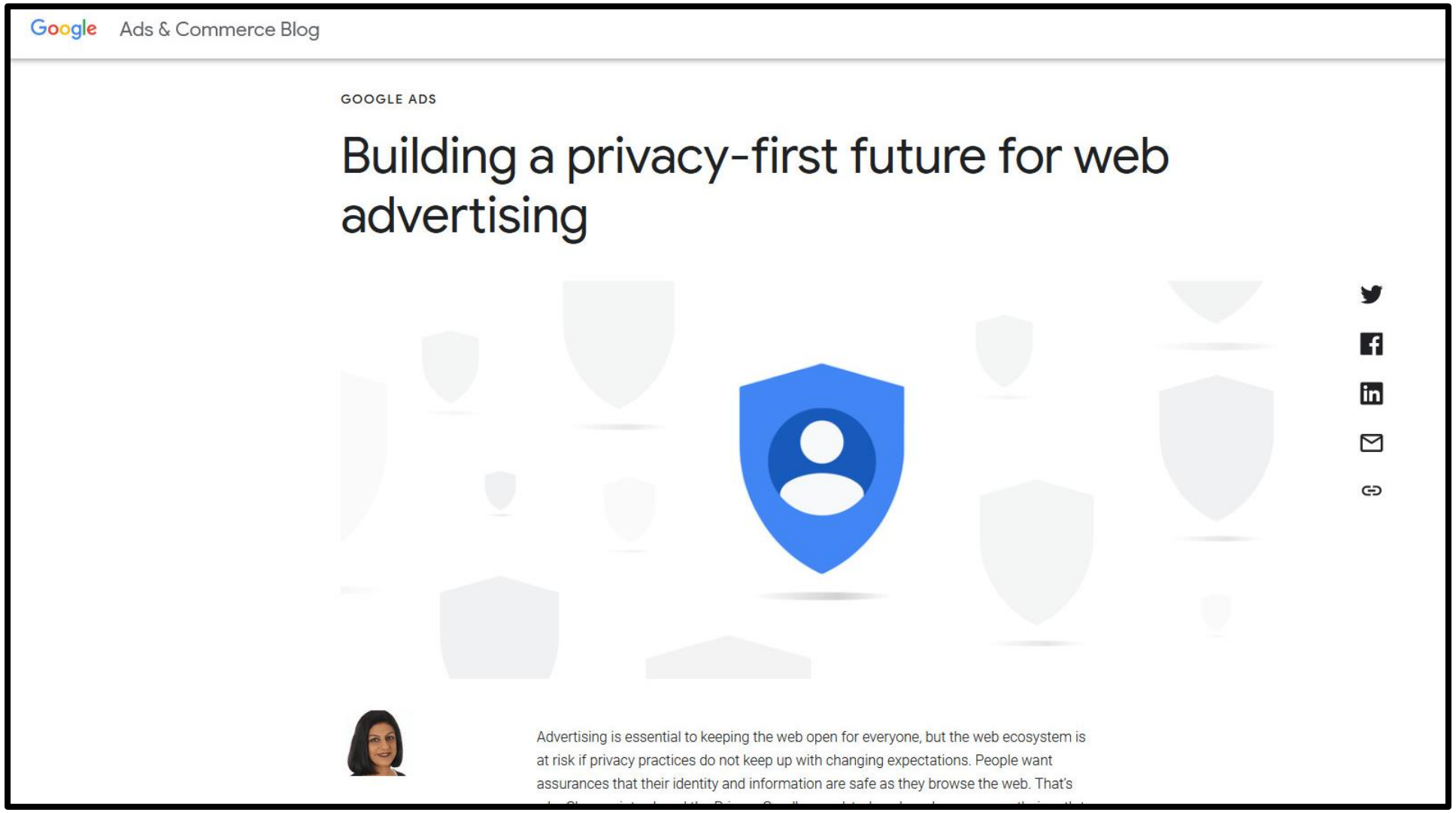


Image taken from https://blog.google/products/ads-commerce/2021-01-privacy-sandbox/

Google FLoC (EFF's Perspective)

Image taken from https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea

Google FLoC

- Federated Learning of Cohorts
- Clusters users based on their browsing activity and assigns a cohort ID
  - Uses SimHash for clustering
  - Clusters intended to contain 1,000s of users
- Criticisms include fingerprintability, ability to tie cohort to PII, and collapse of different browsing contexts
- (Abandoned in early 2022)

Google FLoC: Architecture

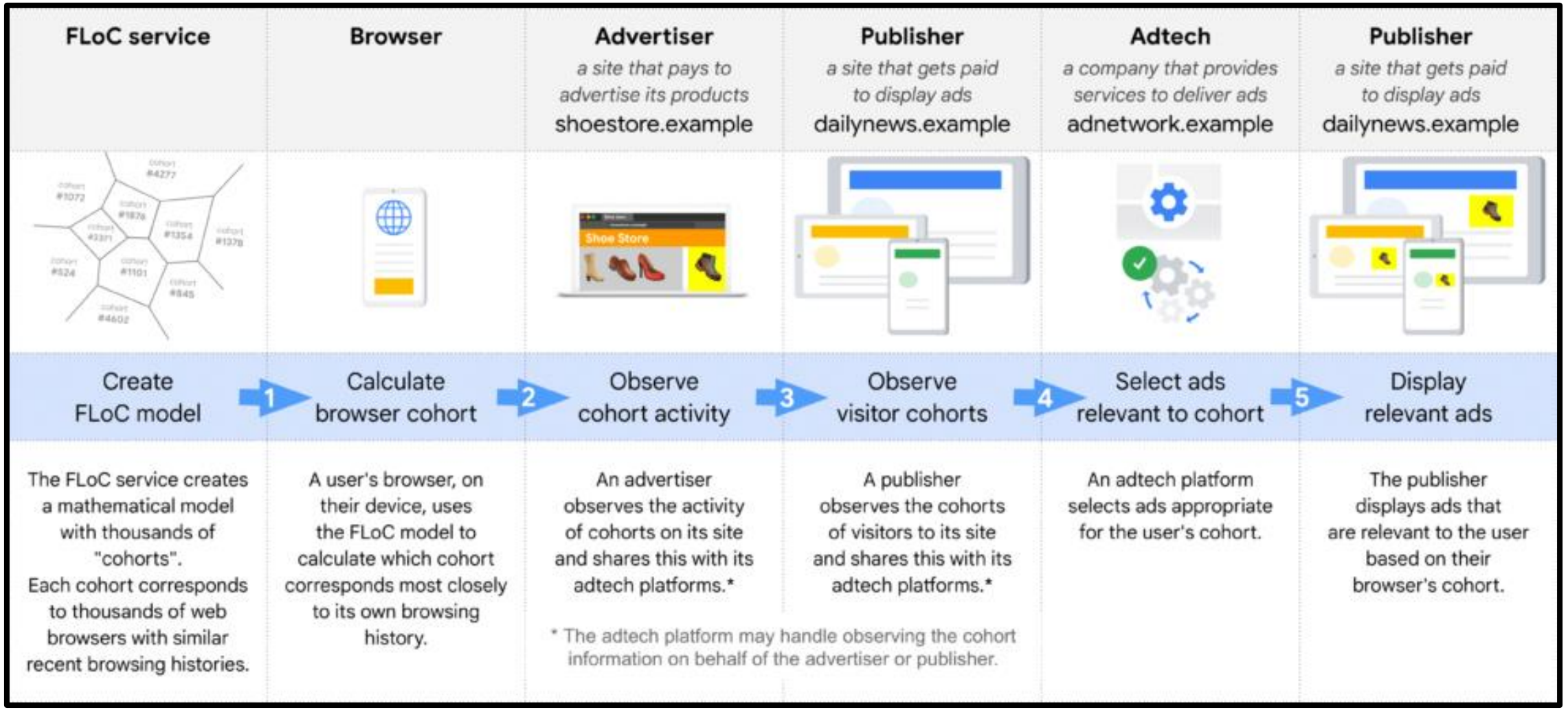
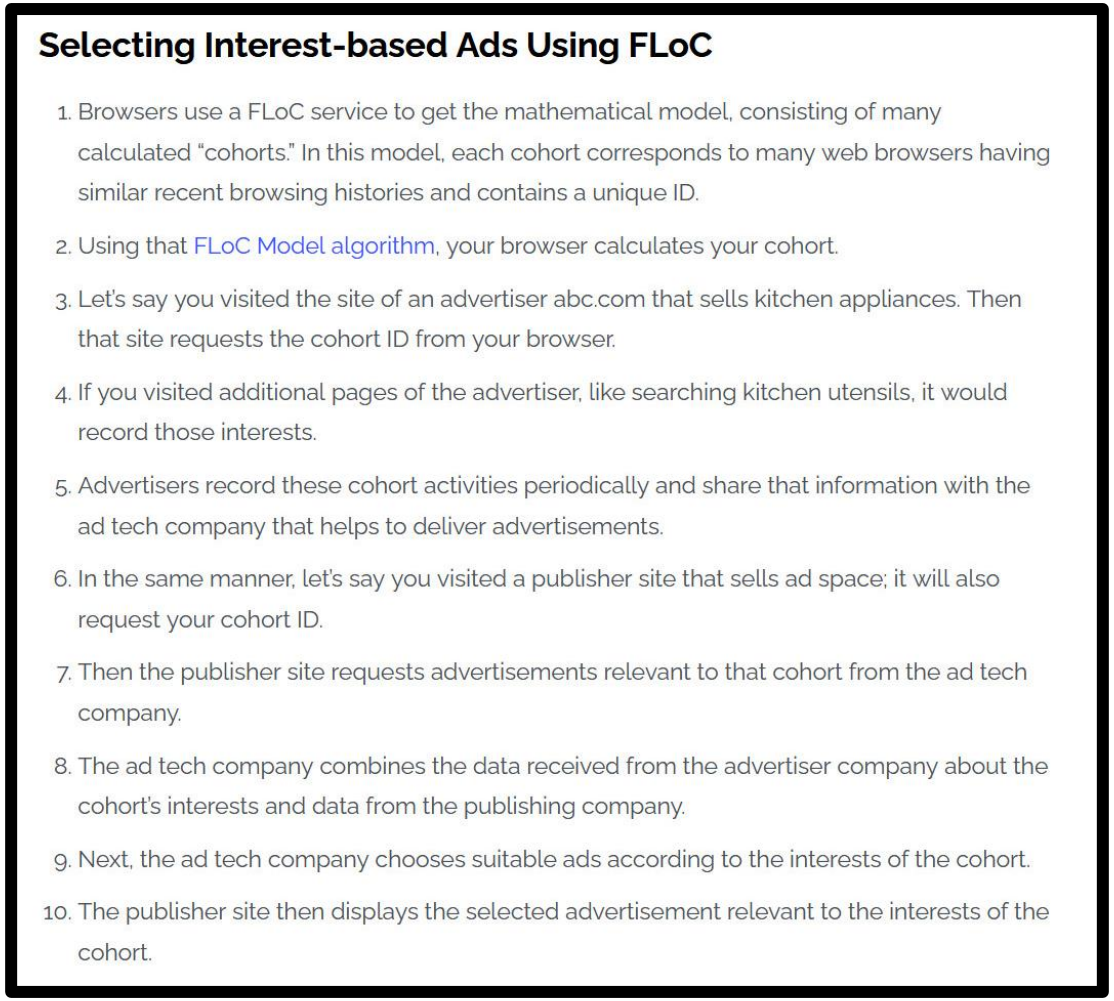


Image taken from https://arstechnica.com/gadgets/2021/04/everybody-hates-floc-googles-tracking-plan-for-chrome-ads/

Google FLoC: Architecture



Selecting Interest-based Ads Using FLoC

1. Browsers use a FLoC service to get the mathematical model, consisting of many calculated "cohorts." In this model, each cohort corresponds to many web browsers having similar recent browsing histories and contains a unique ID.
2. Using that FLoC Model algorithm, your browser calculates your cohort.
3. Let's say you visited the site of an advertiser abc.com that sells kitchen appliances. Then that site requests the cohort ID from your browser.
4. If you visited additional pages of the advertiser, like searching kitchen utensils, it would record those interests.
5. Advertisers record these cohort activities periodically and share that information with the ad tech company that helps to deliver advertisements.
6. In the same manner, let's say you visited a publisher site that sells ad space; it will also request your cohort ID.
7. Then the publisher site requests advertisements relevant to that cohort from the ad tech company.
8. The ad tech company combines the data received from the advertiser company about the cohort's interests and data from the publishing company.
9. Next, the ad tech company chooses suitable ads according to the interests of the cohort.
10. The publisher site then displays the selected advertisement relevant to the interests of the cohort.

Image taken from https://www.privacyaffairs.com/google-floc/

Google FLoC (Brave Browser's Perspective)

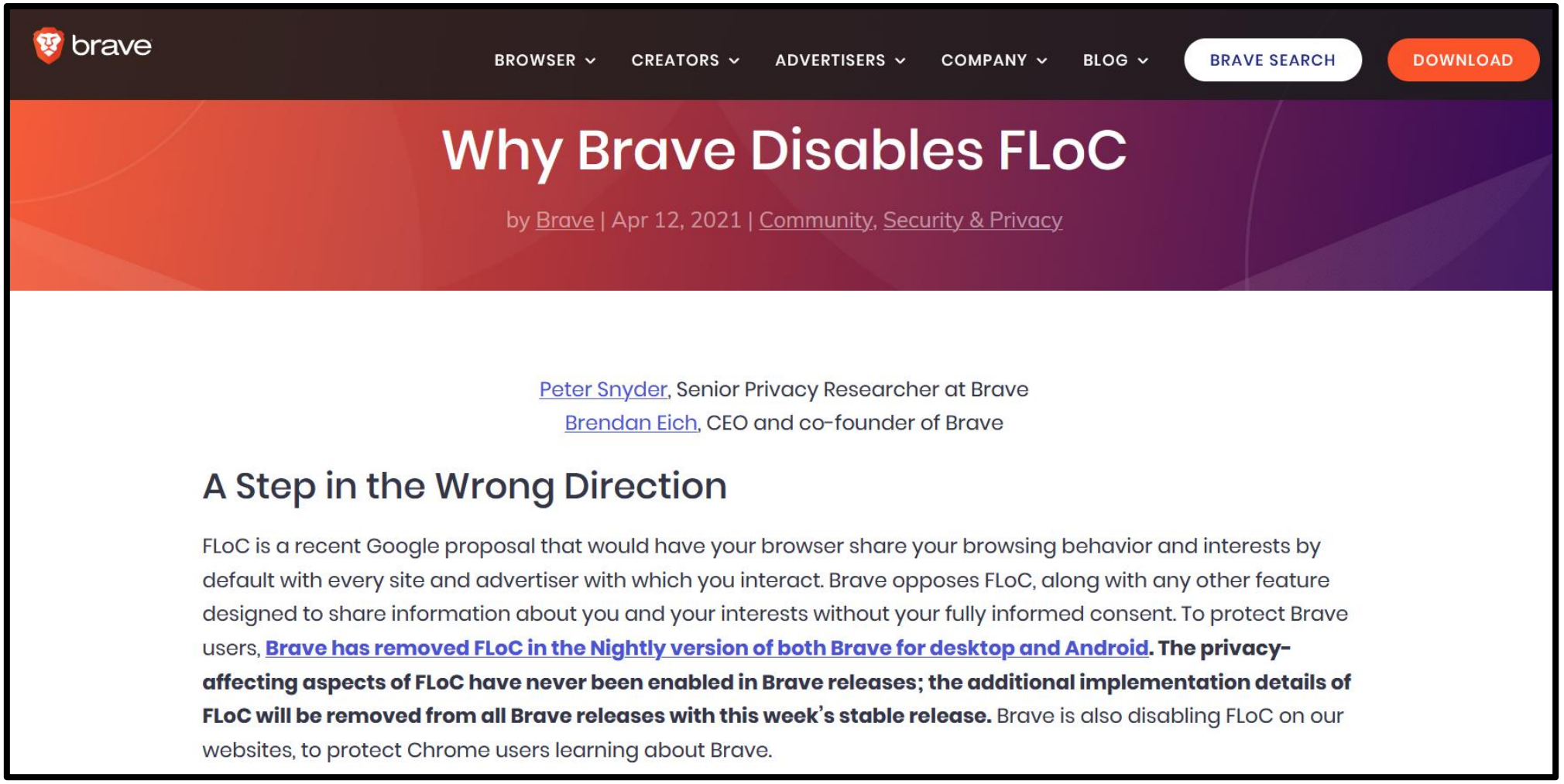


Image taken from https://brave.com/why-brave-disables-floc/

Google Topics API

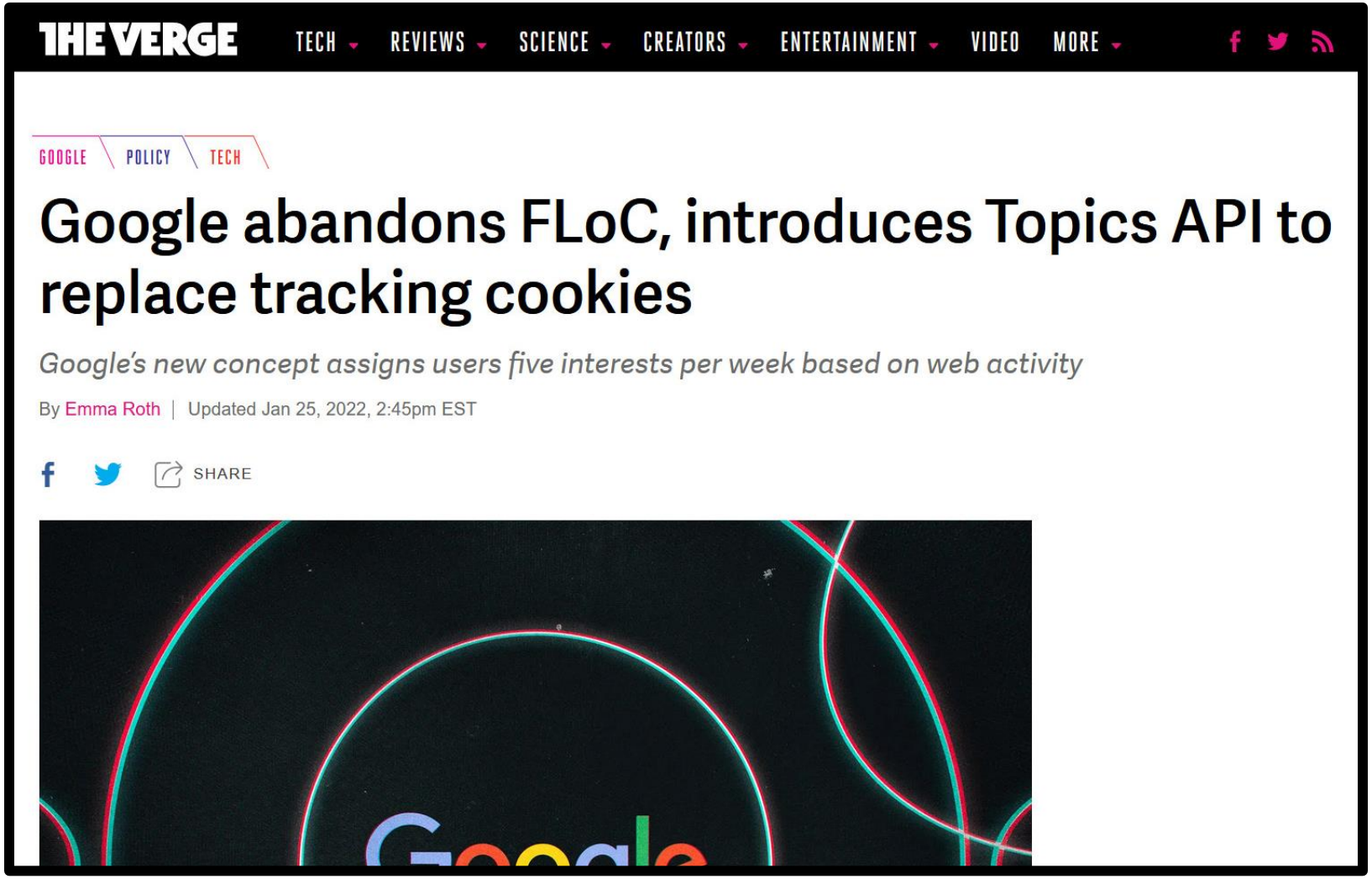


Image taken from https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking

Google Topics API

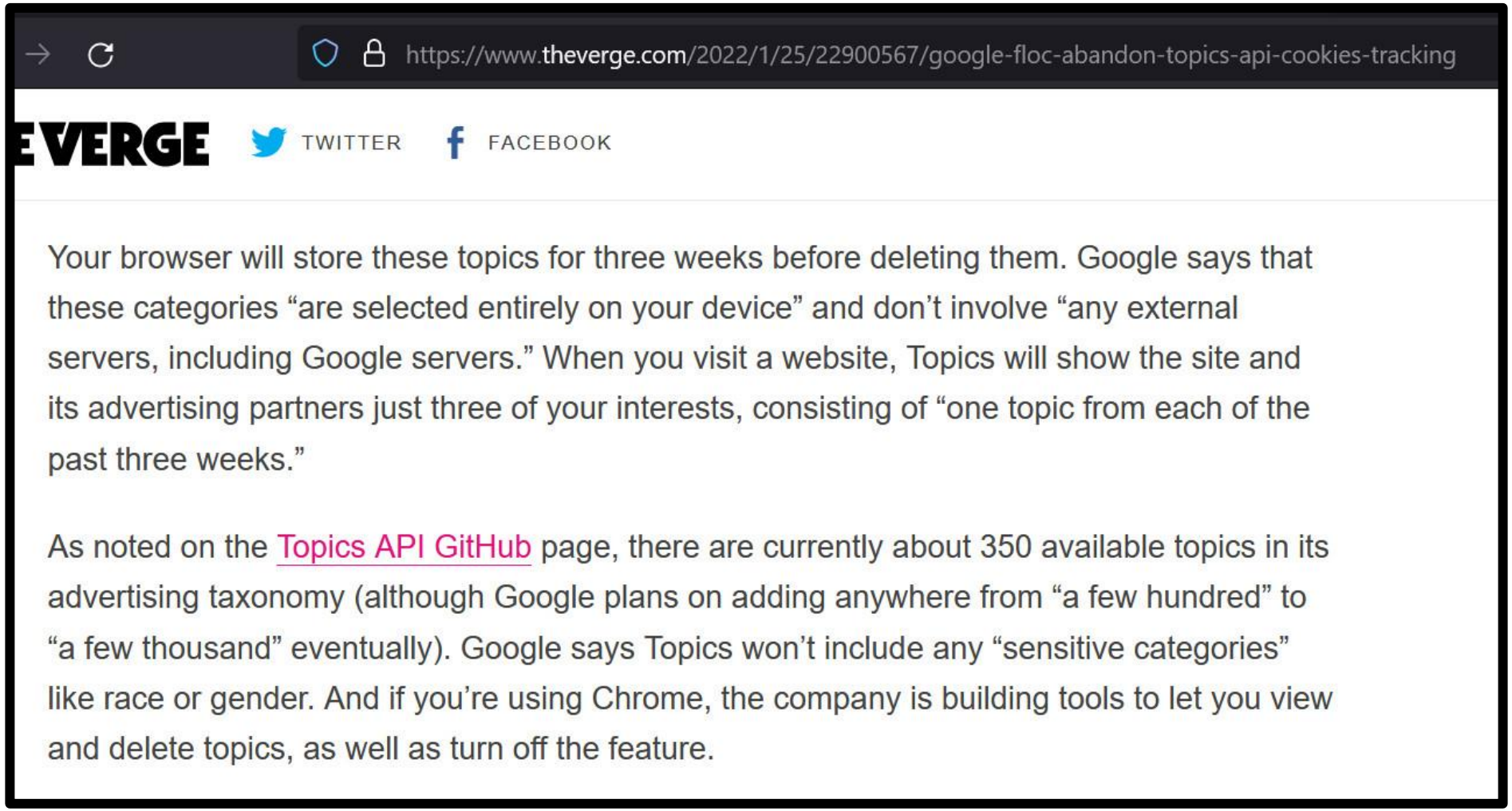


Image taken from https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking

Google Topics API

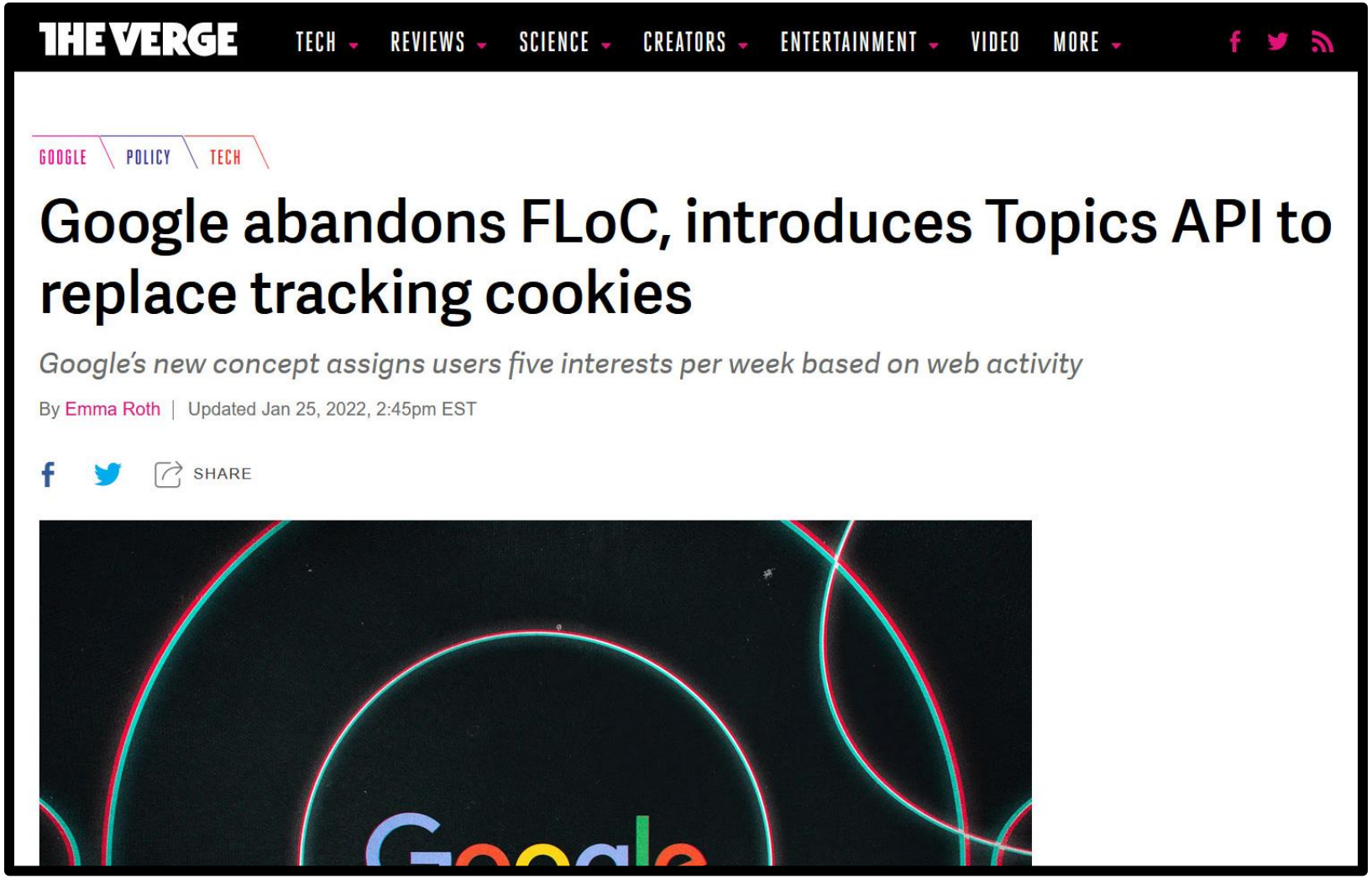


Image taken from https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking

The Impact of Companies' Decisions

iOS App Tracking Transparency

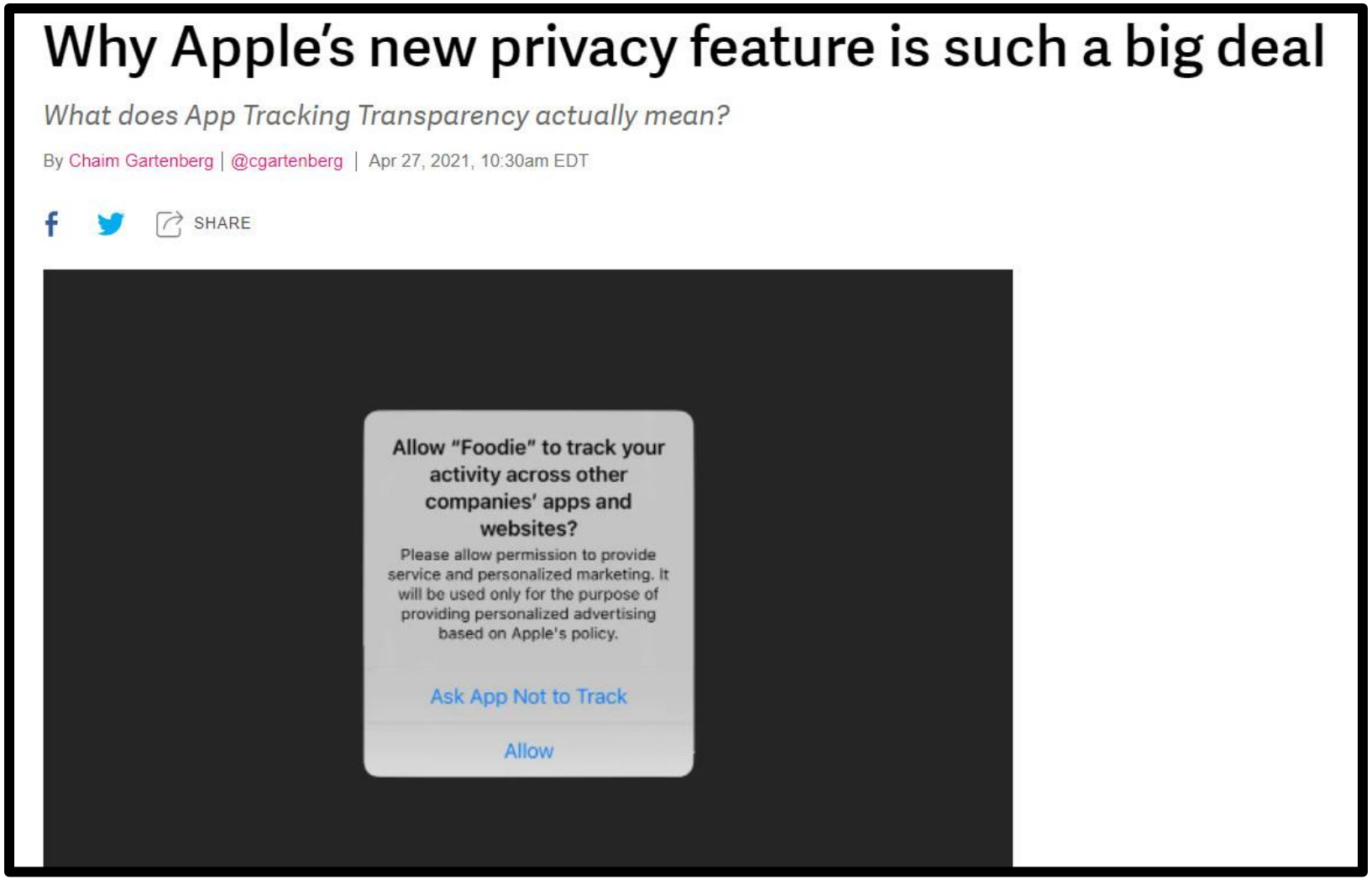
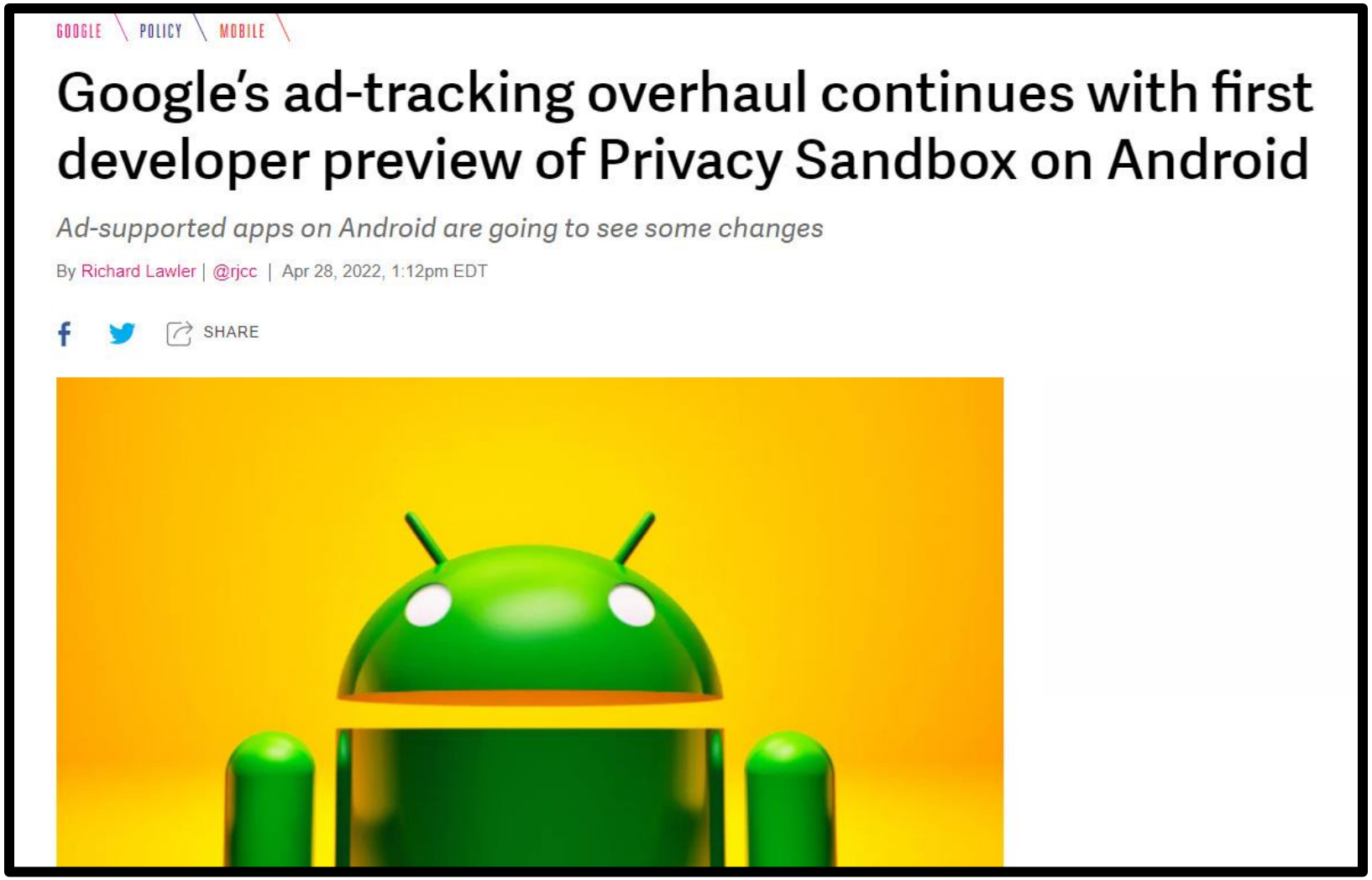


Image taken from https://www.theverge.com/2021/4/27/22405474/apple-app-tracking-transparency-ios-14-5-privacy-update-facebook-data

Android Tracking



GOOGLE \ POLICY \ MOBILE \

Google's ad-tracking overhaul continues with first developer preview of Privacy Sandbox on Android

Ad-supported apps on Android are going to see some changes

By Richard Lawler | @rjcc | Apr 28, 2022, 1:12pm EDT

SHARE

Image taken from https://www.theverge.com/2022/4/28/23046608/android-advertising-privacy-sandbox-google-preview