

# 18. Tracking on the Web



Blase Ur and David Cash  
February 21<sup>st</sup>, 2022  
CMSC 23200 / 33250



THE UNIVERSITY OF  
CHICAGO

# The Online Tracking Ecosystem

# Online Tracking

- Advertisers want to show you advertisements targeted to your interests and demographics

## Ads Preferences

† Ads on Search and Gmail

• Ads on the web

Opt out

Ads on the web

### Make the ads you see on the web more interesting

Many websites, such as news sites and blogs, partner with us to show ads to their visitors. To see ads that are more related to you and your interests, edit the categories below, which are based on sites you have recently visited. [Learn More](#)

Your interests are associated with an advertising cookie that's stored in your browser. If you don't want us to store your interests, you can opt out below. Your ads preferences only apply in this browser on this computer. They are reset if you delete your browser's cookies.

† Watch a video: [Ads Preferences on GDN explained](#)

### Your categories

Below you can review the interests and inferred demographics that Google has associated with your cookie. You can [remove](#) or [edit](#) these at any time.

Arts & Entertainment

Computers & Electronics

Computers & Electronics - Consumer Electronics - Gadgets & Portable Electronics - PDAs & Handhelds

Internet & Telecom

Internet & Telecom - Mobile & Wireless - Mobile Phones - Smart Phones

Law & Government

Science

### Your demographics

We infer your age and gender based on the websites you've visited. You can [remove](#) or [edit](#) these at any time.

Age: 35-44

Gender: Male

## How your ads are personalized

Ads are based on personal info you've added to your Google Account, data from advertisers that partner with Google, and Google's estimation of your interests. Choose any factor to learn more or update your preferences. [Learn more](#)

Accounting & Finance Jobs

Action & Adventure Films

Action & Platform Games

Adventure Games

Android OS

Autos & Vehicles

Banking

Bars, Clubs & Nightlife

Beaches & Islands

Blues

Bollywood & South Asian Film

Books & Literature

Business & Productivity Software

Business News

# Online Tracking

- First party = the site you are visiting (whose address is in the URL bar)
- Third party = other sites contacted as a result of your visit to that site
- First-party tracking (e.g., for search)
  - Consider DuckDuckGo and alternatives

# Data-Driven Inferences

☰

Southern Living

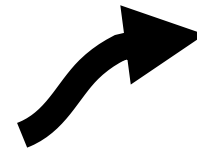
SUBSCRIBE

TRAVEL

Here Are the Most Dog-Friendly Restaurants in the United States

[f](#) [p](#) [t](#) [e](#)

KELLI BENDER

A photograph of a dog sitting at a restaurant table. The dog is looking towards the camera. On the table in front of it are several small burgers on a wooden board. The background shows a restaurant interior with large windows and greenery outside. A small red pin icon is visible in the bottom right corner of the image.

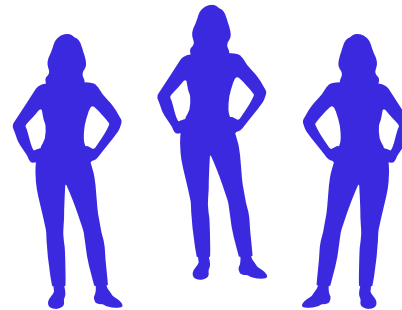
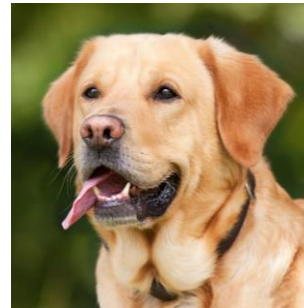
You might like dogs!

# Targeted Advertising

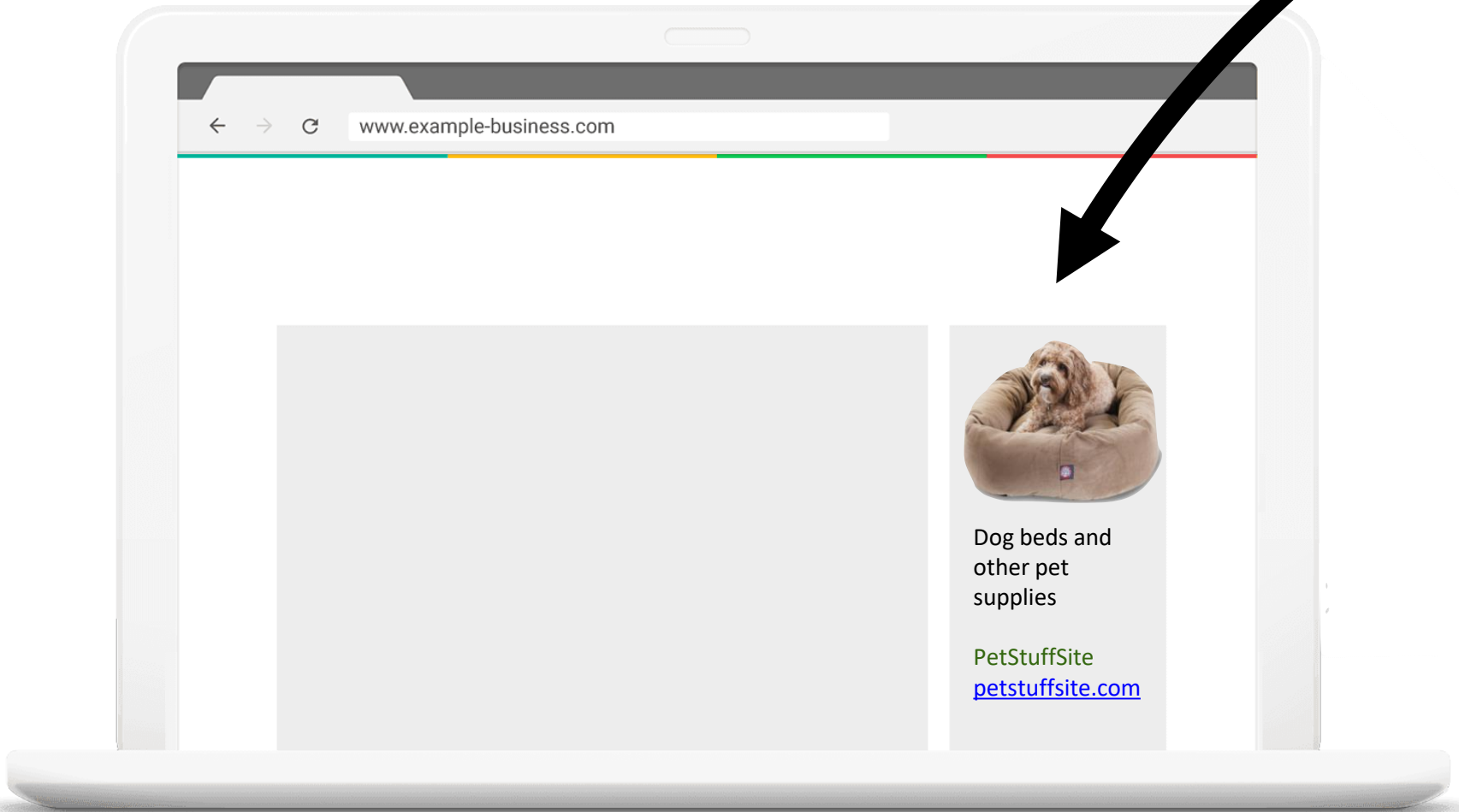
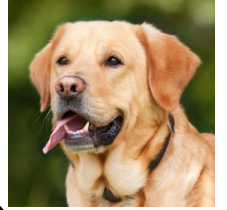
Interests & remarketing

**Affinity audiences**

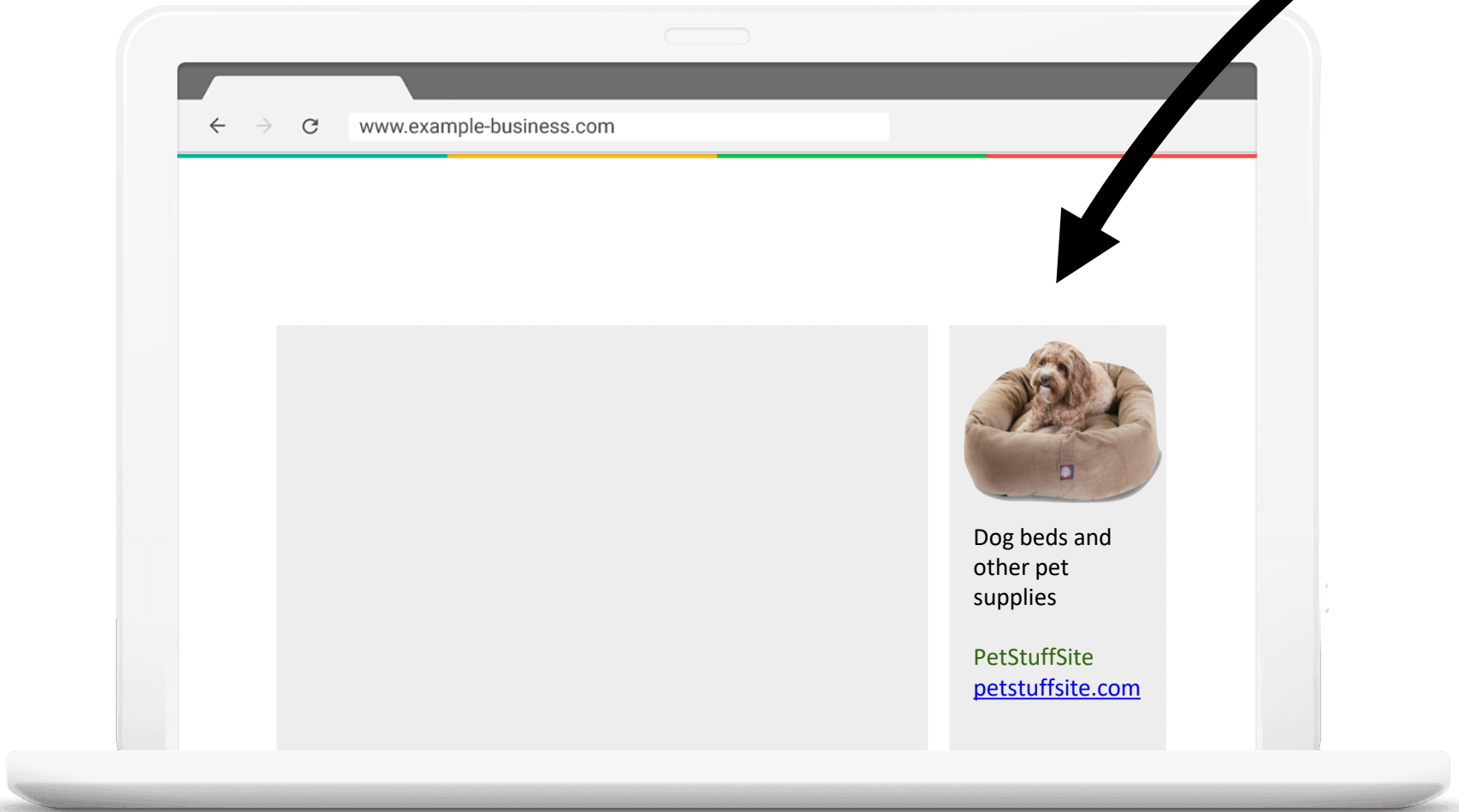
- Music Lovers >>
- News Junkies >>
- Nightlife Enthusiasts >>
- Outdoor Enthusiasts >>
- Pet Lovers >>



# Targeted Advertising



# Targeted Advertising





# Mechanics of Tracking

- Canonically, tracking is accomplished via HTTP cookies
  - Third-party cookies

# Mechanics of Online Tracking

- JavaScript / images from advertising networks loaded as part of your page
  - In iframes
  - Or sometimes not
  - Why does this matter?
- Let's discuss: what can an advertising network learn, and how?

# Mechanics of Cookie Syncing

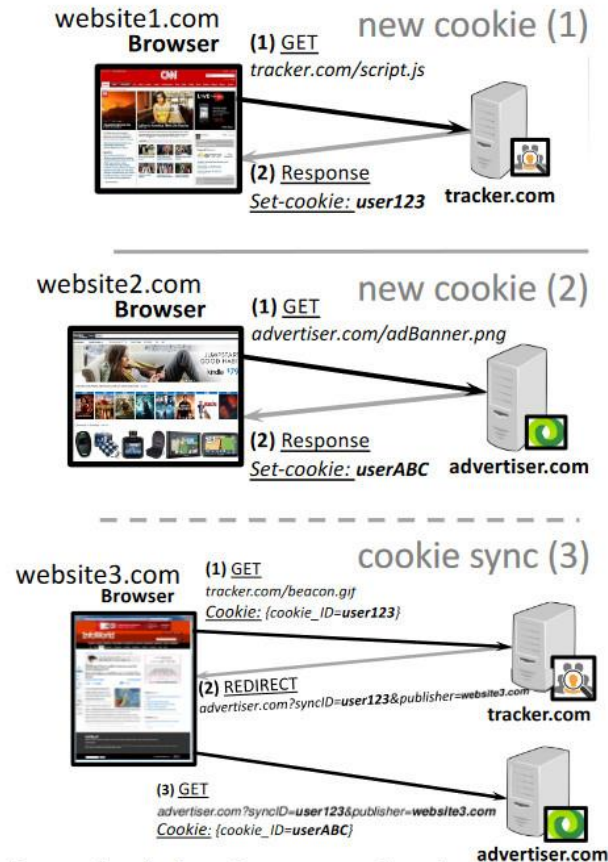


Figure 1: Example of advertiser.com and tracker.com synchronizing their cookieIDs. Interestingly, and without having any code in website3, advertiser.com learns that: (i) cookieIDs userABC==user123 and (ii) userABC has just visited the given website. Finally, both domains can conduct server-to-server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

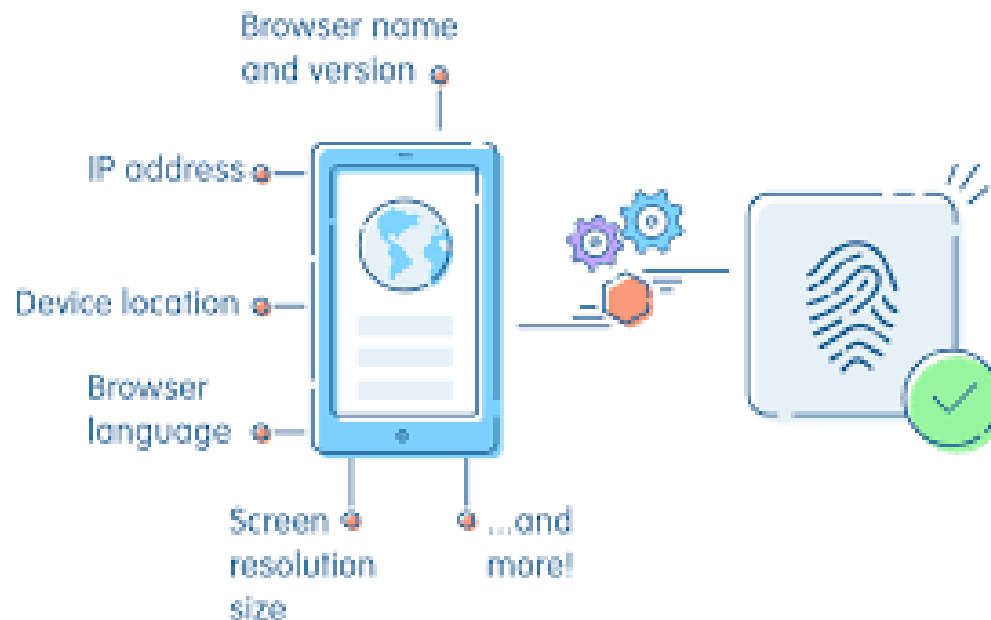
From Papadopoulos et al. "Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask," in *Proc. WWW*, 2019.

# 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?)*

# Device Fingerprinting

- Use unique(-ish) combination of device features as an identifier
- <https://panopticlick.eff.org/>



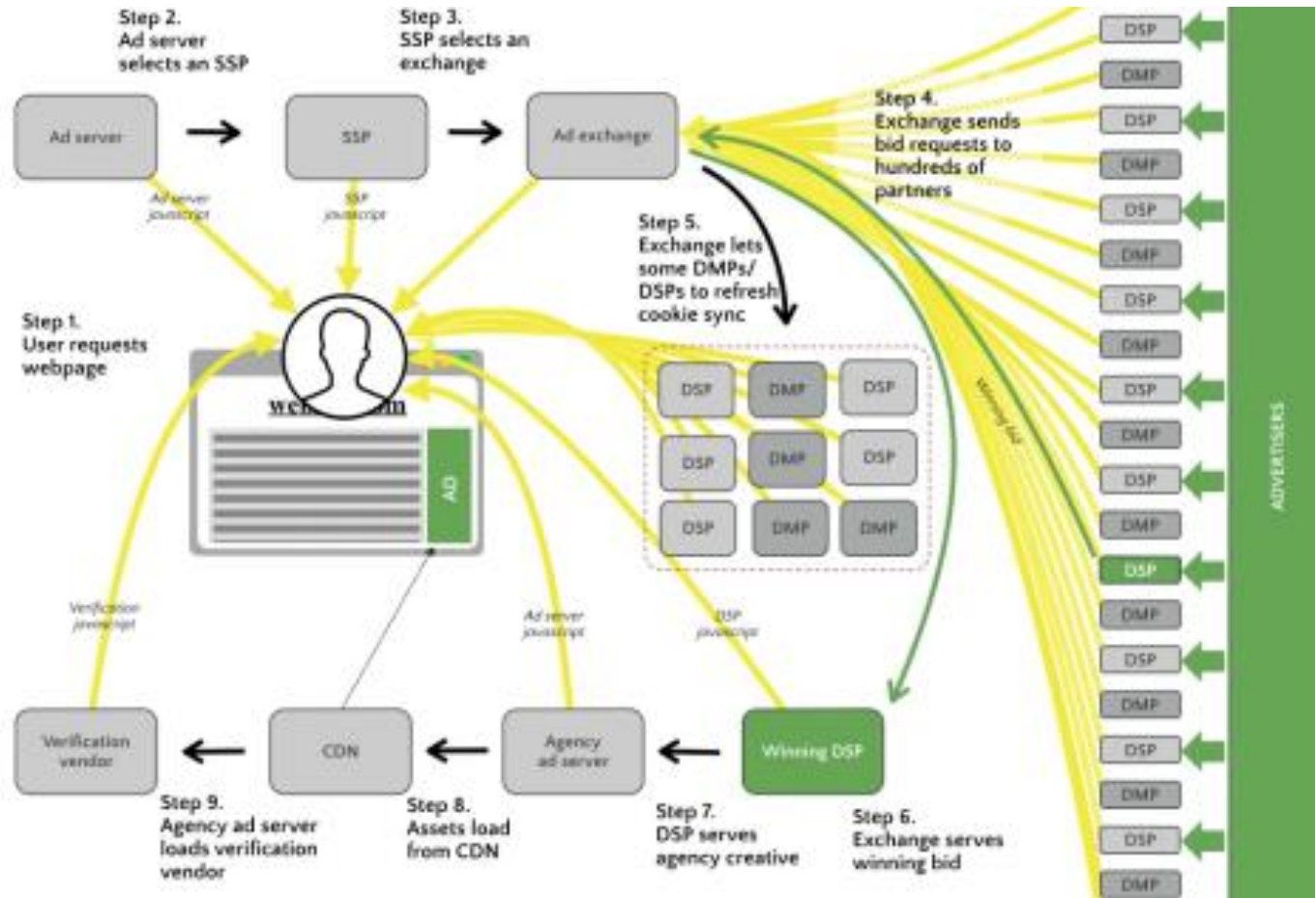
# Ad Bidding Marketplaces

## DATA LEAKAGE IN ONLINE ADVERTISING

This is the current process of real-time bidding that is used in online behavioural advertising.

### Legend

- Channel of data leakage
- Money
- ⦿ Personally identifiable information



# Transparency About Online Tracking

# Ubiquity of Online Tracking

**GHOSTERY**

22 Trackers found on [www.mynews.com](http://www.mynews.com)

**22**

4 Blocked

Trust Site

Restrict Site

Pause Ghostery

[Map these trackers](#)

TRACKERS Block All

Advertising 10 TRACKERS 3 Blocked

<i>Advertising.com</i>	<input checked="" type="checkbox"/>
DoubleClick	<input type="checkbox"/>
Google Adsense	<input type="checkbox"/>
Korrelate	<input type="checkbox"/>
<i>Moat</i>	<input checked="" type="checkbox"/>
<i>NetRatings-Site-Center</i>	<input checked="" type="checkbox"/>
<i>Polar-Mobile</i>	<input checked="" type="checkbox"/>
ScoreCard Research Beacon	<input type="checkbox"/>
Tacoda	<input type="checkbox"/>

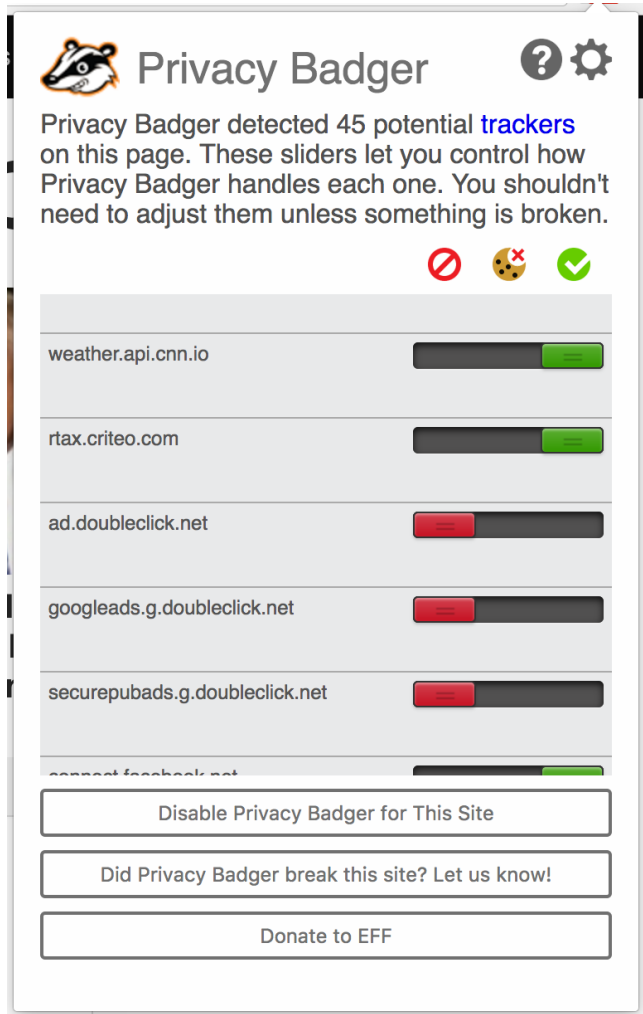


# Existing Privacy Tools

The Disconnect browser extension interface is shown. At the top, it says "DISCONNECT" with "Help" and "Share" links. Below this are social media sharing buttons for Facebook (0), Google+ (1), and Twitter (1). The main content is organized into categories: Advertising (2 requests), Analytics (7 requests), Social (0 requests), and Content (0 requests). Under Advertising, specific trackers like Adobe and Nielsen are listed with "1 request" each. At the bottom, there are utility options like "Whitelist site", "Visualize page", "Show counter", and "Cap counter". A bar chart shows "Time saved" and "Bandwidth saved". A green button at the bottom says "Get Mobile Protection".

The Blur browser extension interface is shown on a website. The top bar is blue and says "espn.com" with a close button and "8 trackers blocked". Below this, it states "Tracker blocking is on for this website" with a green toggle switch. A list of blocked trackers is shown: Google AdSense, Demdex, Twitter Badge, and Omniture, each with a "blocked" status and a green checkmark icon. A red eye icon is next to a link that says "see your tracker blocking stats and learn more about these companies". A blue bar at the bottom says "21 trackers blocked since Feb '17". Below this, it says "Correct how Blur works in the form below". At the very bottom, the "oBLUR" logo is visible along with links for "Settings", "Help", and "Go Premium".

# Existing Privacy Tools



**Privacy Badger** detected 45 potential trackers on this page. These sliders let you control how Privacy Badger handles each one. You shouldn't need to adjust them unless something is broken.

weather.api.cnn.io [Slider: Full Green]

rtax.criteo.com [Slider: Full Green]

ad.doubleclick.net [Slider: Full Red]

googleads.g.doubleclick.net [Slider: Full Red]

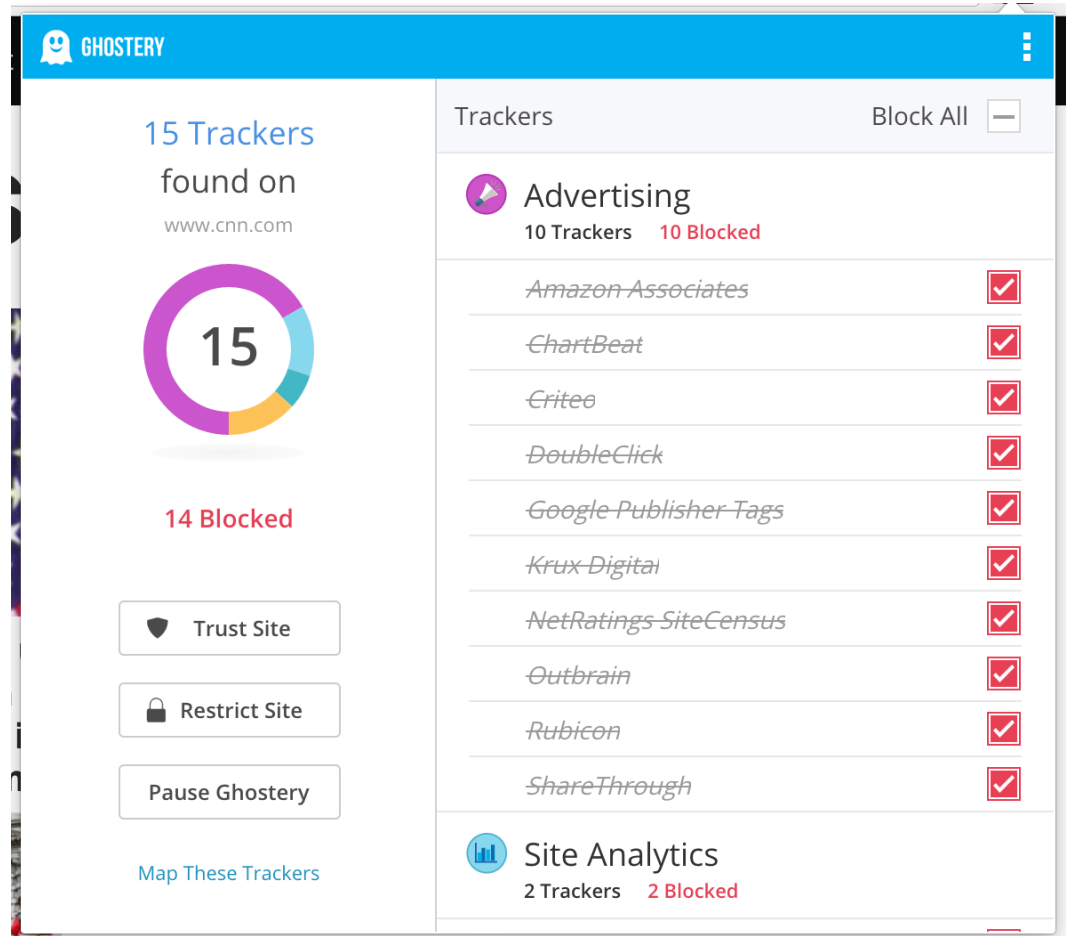
securepubads.g.doubleclick.net [Slider: Full Red]

connect.facebook.net [Slider: Full Green]

Disable Privacy Badger for This Site

Did Privacy Badger break this site? Let us know!

Donate to EFF



**GHOSTERY**

15 Trackers found on www.cnn.com

15

14 Blocked

Trust Site

Restrict Site

Pause Ghostery

Map These Trackers

Trackers Block All [Dropdown]

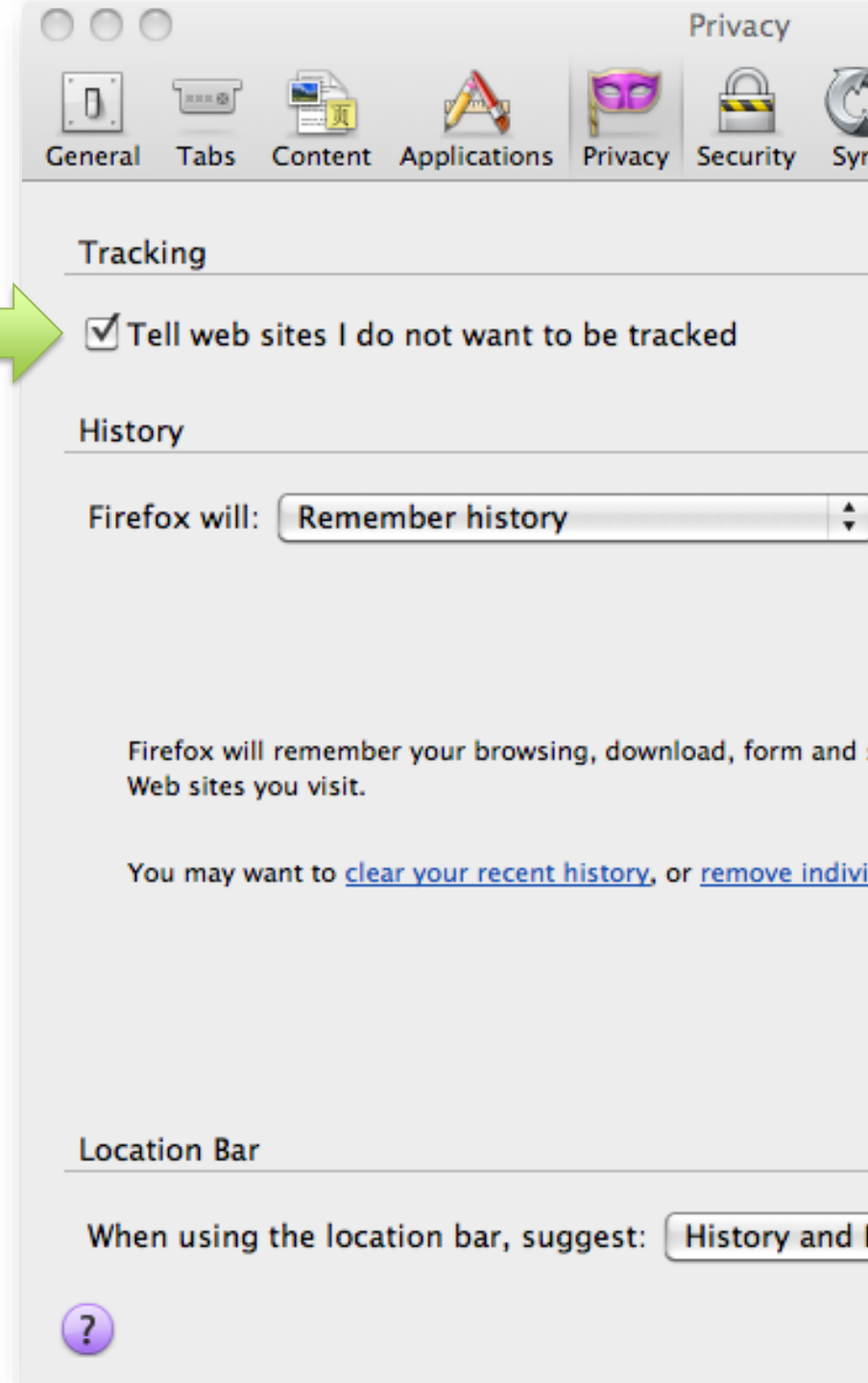
**Advertising**  
10 Trackers 10 Blocked

- Amazon Associates [Checked]
- ChartBeat [Checked]
- Criteo [Checked]
- DoubleClick [Checked]
- Google Publisher Tags [Checked]
- Krux-Digital [Checked]
- NetRatings-SiteCensus [Checked]
- Outbrain [Checked]
- Rubicon [Checked]
- ShareThrough [Checked]

**Site Analytics**  
2 Trackers 2 Blocked

# Do not track

- Proposed W3C standard
- User checks a box
- Browser sends “do not track” header to website
- Website stops “tracking”
- W3C working group tried to define what that means
- These days, mostly a no-op (it doesn't do anything)



# Tools to stop tracking, effective?

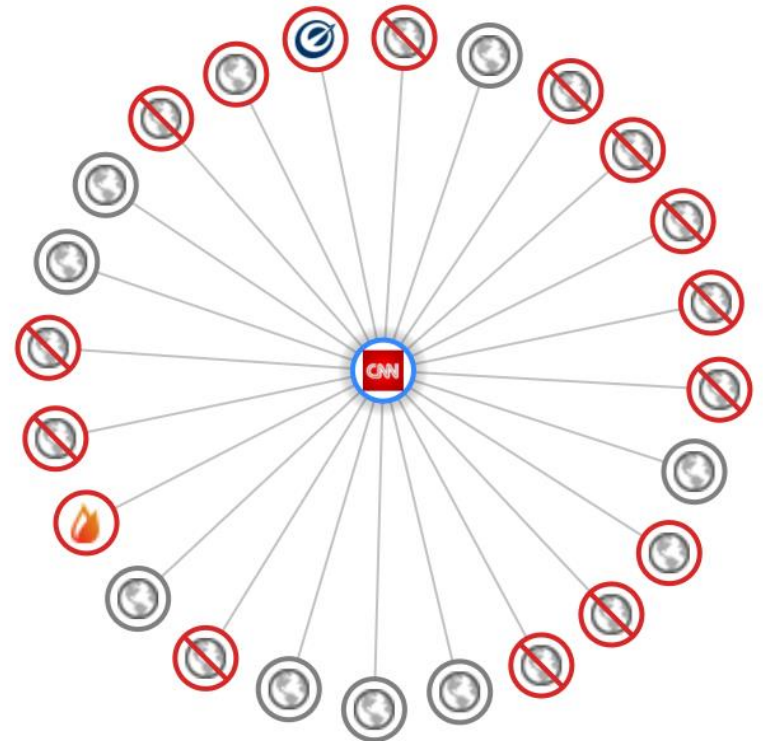
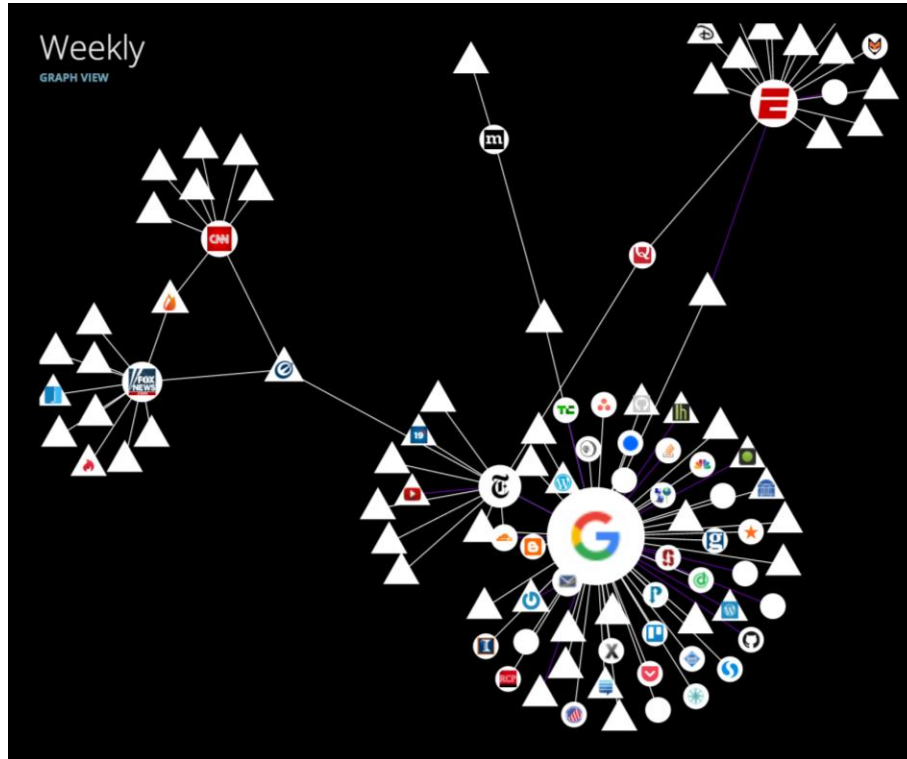
- Browser privacy settings
  - Cookie blocking
  - P3P
  - Tracking Protection Lists
  - Do Not Track
- Browser add-ons
- Opt-out cookies
- Digital Advertising Alliance (DAA) AdChoices icon and associated opt-out pages



DoNotTrackMe



# Visualization: Connection Graphs



# (My Group's) Tracking Transparency

The screenshot shows a Firefox browser window with the Tracking Transparency extension active. The browser's address bar displays <https://www.nytimes.com>. The page content is from the New York Times website, dated Friday, March 16, 2018. A notification box from the Tracking Transparency extension is overlaid on the page, providing the following information:

**Tracking Transparency**

On **The New York Times - Breaking ...**, there are **5 trackers**.

One of these trackers is **Google**, which knows about your activity on this page and **3829** others.

In total, **169 trackers** have seen you visit **12313 pages**. The Tracking Transparency extension has determined that these companies could have inferred your interest in **162 topics**.

[Show me more about what the trackers know](#)

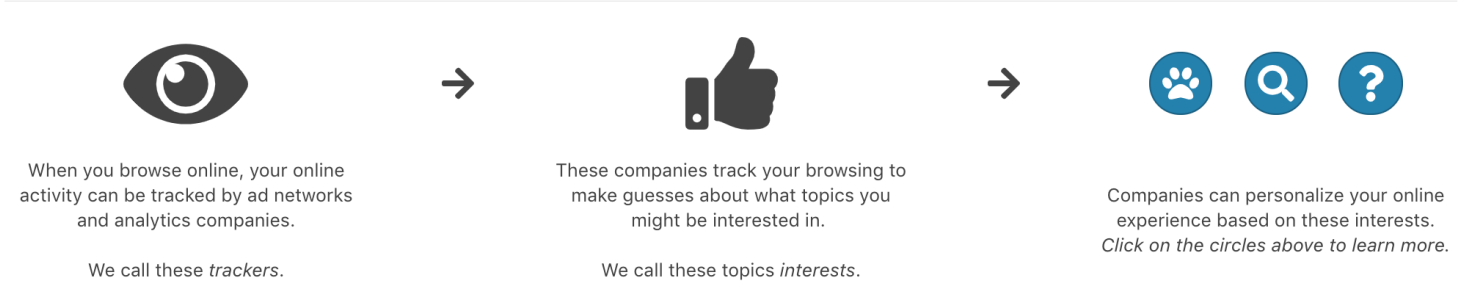
The background page shows the New York Times logo, navigation links (World, U.S., Politics, N.Y., Business, Opinion, Tech, Science, Health, Sports), and article teasers. One article is titled "Trump's Steel Tariffs Open Lobbying Floodgate" by ANA SWANSON and KENNETH P. Another article is titled "Controlled" by DANIEL McCARTHY, and another is titled "Elite" by SUSAN JACOBY. A cartoon illustration of a person's head with handwritten notes "A FEW GREYS. (EXPERIENCED)" and "LISTENS." is also visible.

# (My Group's) Tracking Transparency

Tracking Transparency   Interests   Trackers   Sites   Activity   About   ⚙️

Home

## What are *trackers* and *interests*?



The diagram illustrates the process of tracking and personalization. It starts with an eye icon representing online activity. An arrow points to a thumbs-up icon representing interests. Another arrow points to three circular icons: a paw print, a magnifying glass, and a question mark, representing personalized experiences.

When you browse online, your online activity can be tracked by ad networks and analytics companies.  
We call these *trackers*.

These companies track your browsing to make guesses about what topics you might be interested in.  
We call these topics *interests*.

Companies can personalize your online experience based on these interests.  
*Click on the circles above to learn more.*

### Your Top Trackers

- Google
- Chartbeat
- Optimizely
- Microsoft
- Amazon.com

### Your Top Interests

- Law & Government
- Online Communities
- People & Society
- News
- Shopping

**39** Trackers encountered   **7** Pages visited   **6** Potential interests

### Recent Interests

- Law & Government
- Computers & Electronics
- Shopping
- News
- People & Society

### Recent Sites

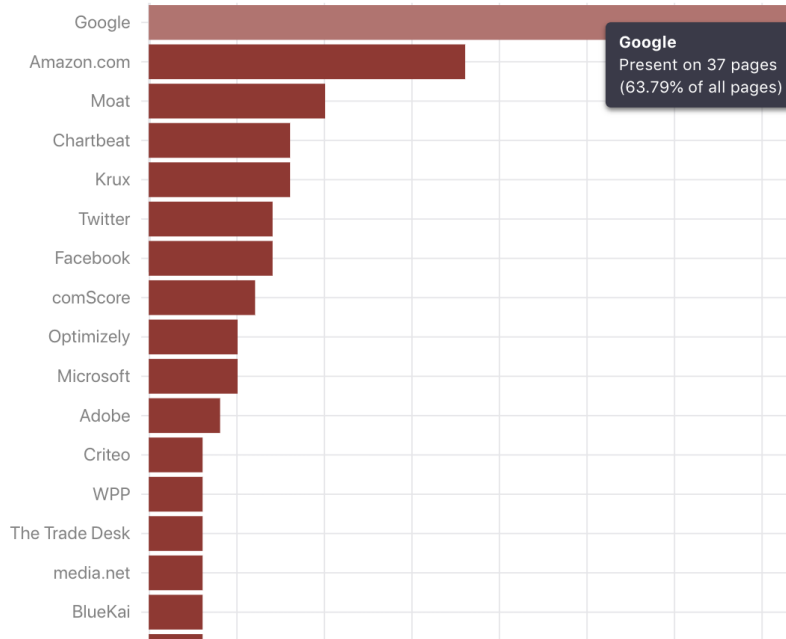
- twitter.com
- eff.org
- amazon.com
- cnn.com
- nytimes.com

# (My Group's) Tracking Transparency

[Home](#) > Trackers

## 👁️ Who is tracking you?

**47 trackers** have been present on the sites you visited since installing Tracking Transparency. Your most frequently encountered tracker is **Google**, which was present on **63.79%** of the pages you visited.



### Google

**37**  
Pages

**11**  
Sites

**14**  
Interests

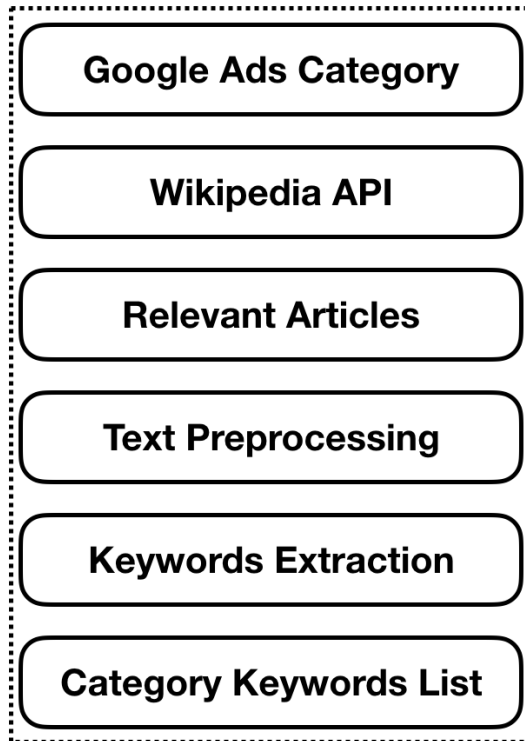
**Google** was present on **37 pages** across **11 sites** that you visited since installing Tracking Transparency. From those tracking encounters, they may have guessed that you are interested in **14 topics**.

[More about this tracker >](#)

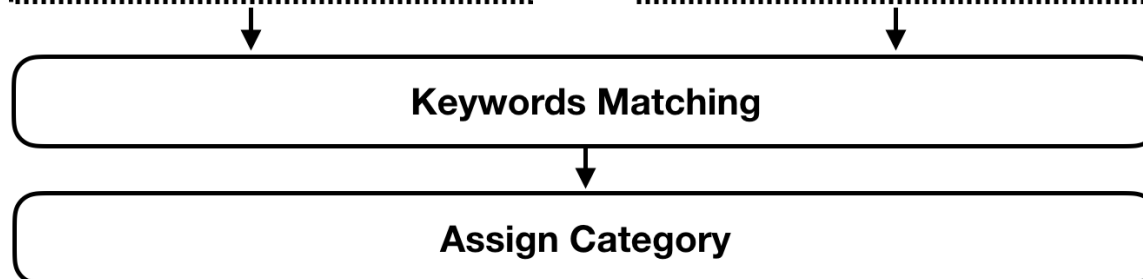
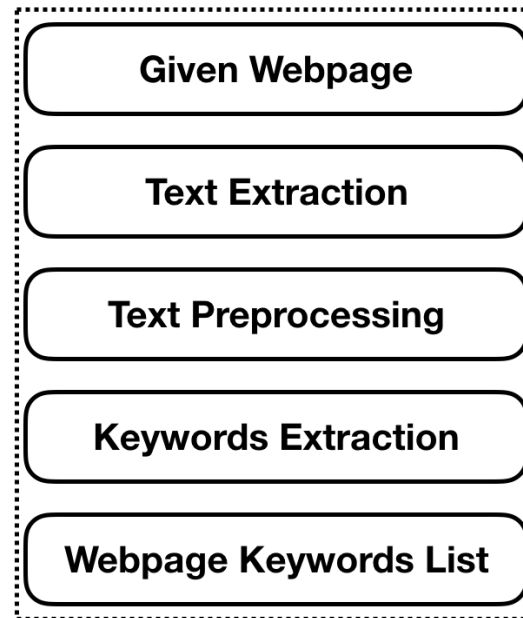


# (My Group's) Tracking Transparency

## 1) Categories keywords extraction



## 2) Webpage keywords extraction



# (My Group's) Tracking Transparency

Tracking Transparency   Interests   Trackers   Sites   Activity   About   ⚙️


Home > Interests

## 👍 What interests might they think you have?

Trackers collect information about the pages you visit and use this information to identify topics, or *interests*, that might be relevant to you. These interests are then used to target ads to you and personalize what you see online. Companies don't usually reveal how they determine your potential interests. Based on the pages you visited, Tracking Transparency's simulations have identified **52 topics** trackers might think are relevant to you.

For example, you recently visited [theverge.com](#), which Tracking Transparency has determined may be about **News**.

The chart below shows the interests suggested by your browsing activity. Click a slice of the chart to see more details.



### Home Improvement

<b>1</b>	<b>8</b>
Sites	Trackers

**1 of the sites** you visited were about Home Improvement. These sites contained a total of **8 trackers**.

**Home Improvement** is a **somewhat popular** interest.

Other people are often **somewhat comfortable** with having their interest in this topic being used to personalize their web experience.

[More about this interest >](#)

### Filters

Recency ⓘ	ALL	24 HRS	7 DAYS
Popularity ⓘ	ALL	LESS	MORE
Comfort ⓘ	ALL	LESS	MORE

# Auditing Online Tracking/Targeting

The image is a screenshot of a news article from The Wall Street Journal. At the top, the newspaper's name "THE WALL STREET JOURNAL." is displayed in a large, bold, serif font. To the right of the name are links for "Subscribe" and "Sign In". Below the name, there is a navigation bar with links for "English Edition", "Print Edition", "Video", "Podcasts", and "Latest Headlines". A secondary navigation bar includes "Home", "World", "U.S.", "Politics", "Economy", "Business", "Tech", "Markets", "Opinion", "Books & Arts", "Real Estate", "Life & Work", "WSJ Magazine", "Sports", and a search icon. The article's category is "TECH". The main headline is "Facebook Disables Access for NYU Research Into Political-Ad Targeting". Below the headline is a sub-headline: "Researchers say the data provided information on misinformation in political ads; Facebook says the work violated its terms of service". The main image shows the Oversight Board logo (a circle of dots) and the text "Oversight Board" above two smartphones. The left smartphone displays the Facebook logo, and the right one displays the Instagram logo. A play button icon is centered over the phones. Below the image is a video title: "What Facebook Panel's Decision Means for Trump, Social Networks". At the bottom of the article, there is a short paragraph of text and a byline: "By Meghan Bobrowsky" and "Updated Aug. 4, 2021 5:54 pm ET".

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

What Facebook Panel's Decision Means for Trump, Social Networks

Facebook's independent oversight board said in 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

# Alternatives to Cookies for Tracking / Profiling

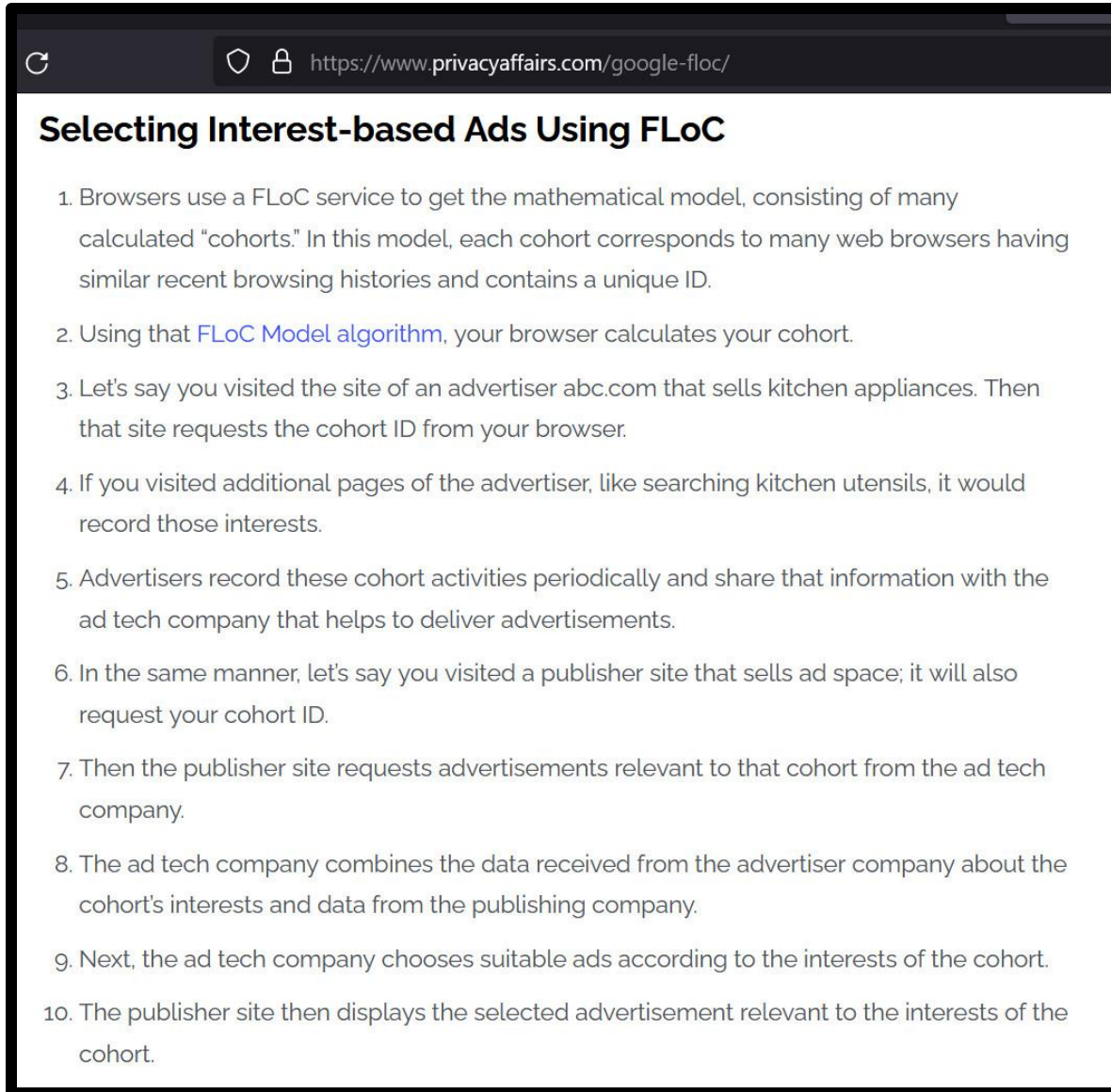
# In-browser Targeting

- The Adnostic research prototype suggested profiling users in-browser and thus choosing ads in-browser from a set
- Key issues include how to properly bill advertisers and how to prevent ad fraud

# Google's 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's FLoC

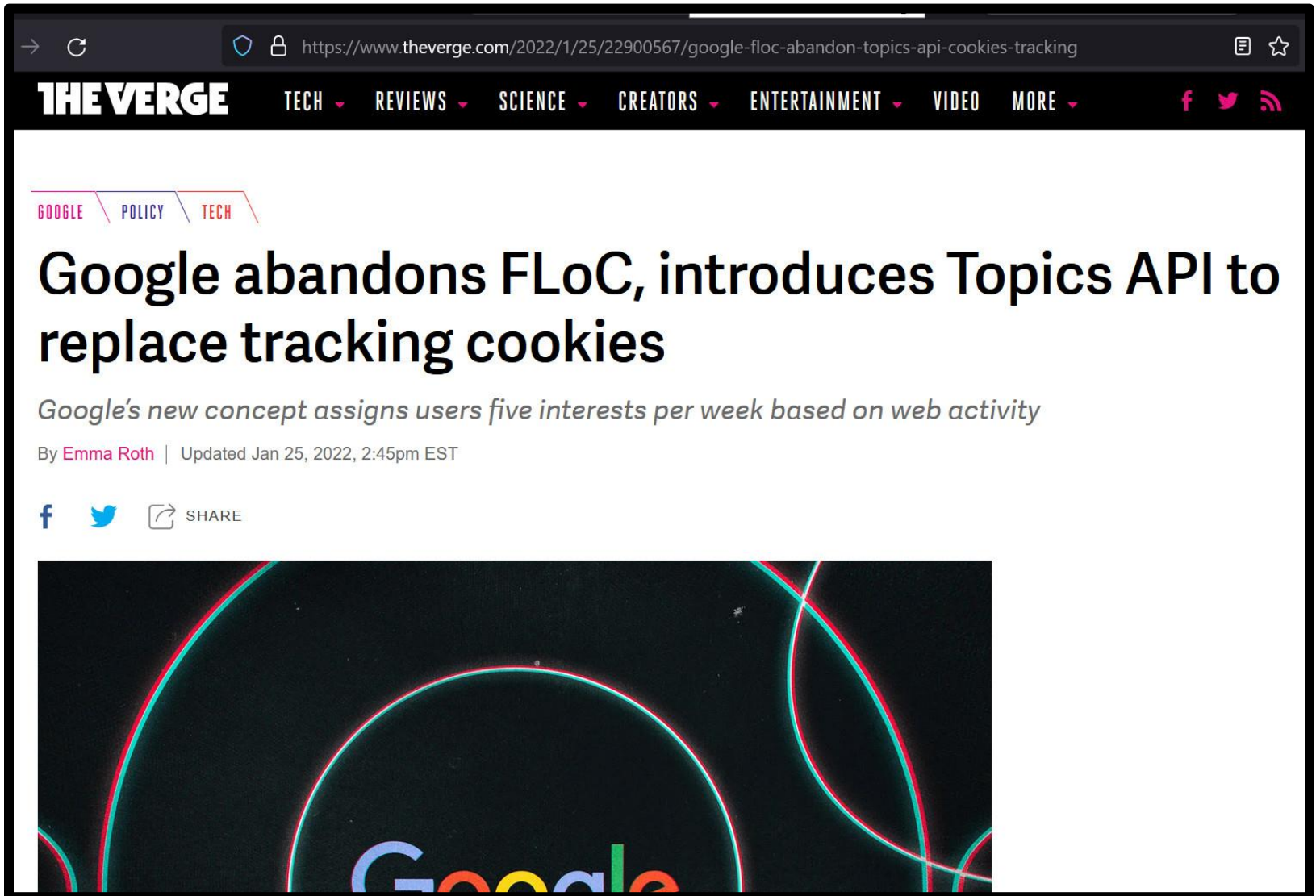


The image is a screenshot of a web browser displaying an article. The browser's address bar shows the URL <https://www.privacyaffairs.com/google-floc/>. The article title is "Selecting Interest-based Ads Using FLoC". The content consists of a numbered list of 10 steps explaining the FLoC process.

## 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.

# Google's Topics API



The screenshot shows a web browser window with the URL <https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking>. The page features the 'THE VERGE' logo and a navigation menu with categories: TECH, REVIEWS, SCIENCE, CREATORS, ENTERTAINMENT, VIDEO, and MORE. Social media icons for Facebook, Twitter, and RSS are also present. The article title is 'Google abandons FLoC, introduces Topics API to replace tracking cookies'. Below the title is a subtitle: 'Google's new concept assigns users five interests per week based on web activity'. The author is listed as 'By Emma Roth' and the update time is 'Updated Jan 25, 2022, 2:45pm EST'. There are social sharing icons for Facebook, Twitter, and a 'SHARE' button. The main image shows the Google logo at the bottom, overlaid on a dark background with glowing, overlapping circles in red and cyan.

→ ↻ <https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking> 📄 ☆

**THE VERGE** TECH ▾ REVIEWS ▾ SCIENCE ▾ CREATORS ▾ ENTERTAINMENT ▾ VIDEO ▾ MORE ▾ f 🐦 📡


GOOGLE POLICY TECH

## Google abandons FLoC, introduces Topics API to replace tracking cookies

*Google's new concept assigns users five interests per week based on web activity*

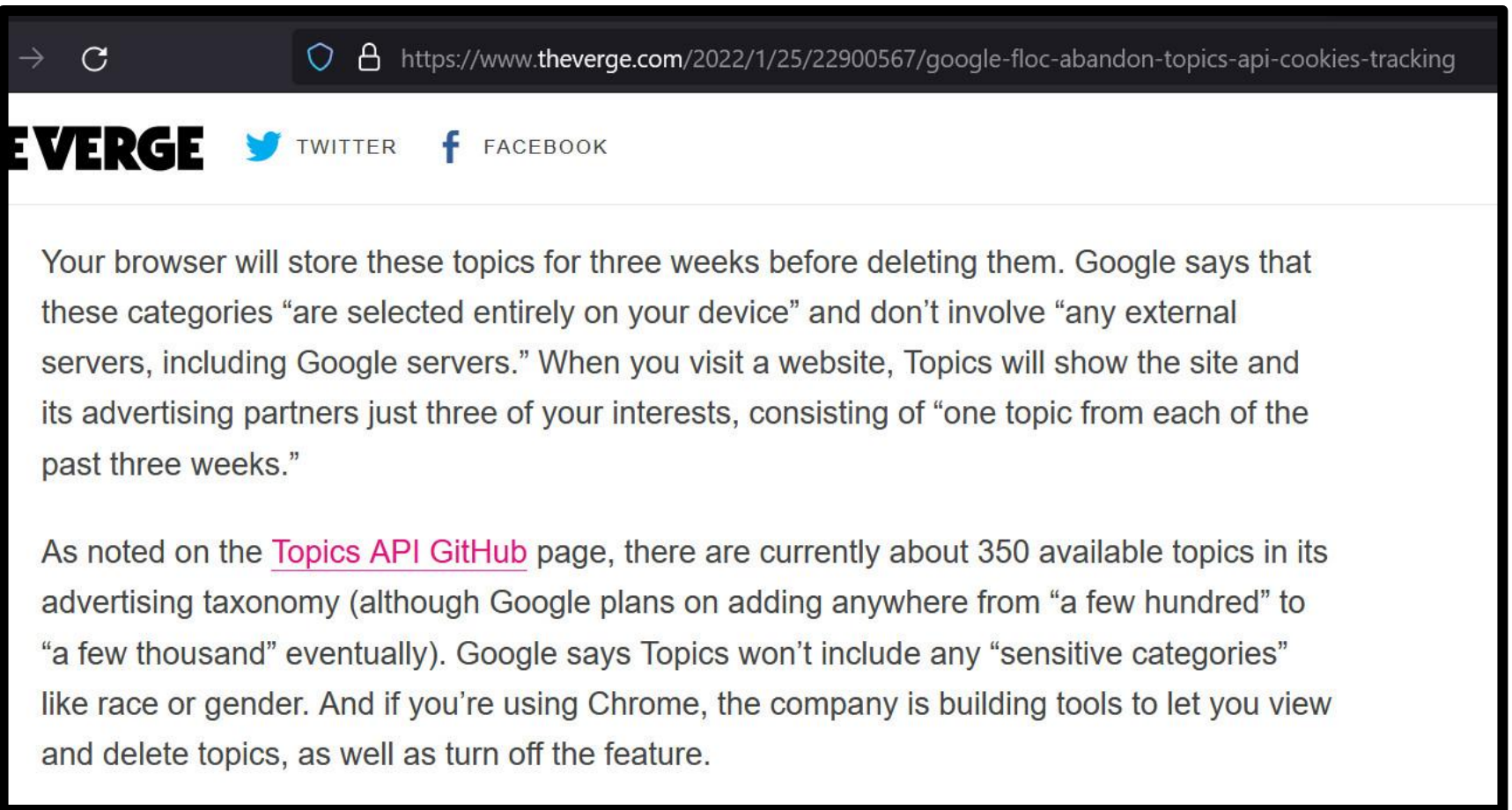
By [Emma Roth](#) | Updated Jan 25, 2022, 2:45pm EST

f 🐦 ↗ SHARE





# Google's Topics API



The screenshot shows a web browser window with the address bar displaying the URL: <https://www.theverge.com/2022/1/25/22900567/google-floc-abandon-topics-api-cookies-tracking>. The page header includes the "EVERGE" logo, a Twitter icon with the text "TWITTER", and a Facebook icon with the text "FACEBOOK".

Your browser will store these topics for three weeks before deleting them. Google says that these categories “are selected entirely on your device” and don’t involve “any external servers, including Google servers.” When you visit a website, Topics will show the site and its advertising partners just three of your interests, consisting of “one topic from each of the past three weeks.”

As noted on the [Topics API GitHub](#) page, there are currently about 350 available topics in its advertising taxonomy (although Google plans on adding anywhere from “a few hundred” to “a few thousand” eventually). Google says Topics won’t include any “sensitive categories” like race or gender. And if you’re using Chrome, the company is building tools to let you view and delete topics, as well as turn off the feature.