UNIVERSITÉ CÔTE D'AZUR

Inria INVENTEURS DU MONDE NUMÉRIQUE

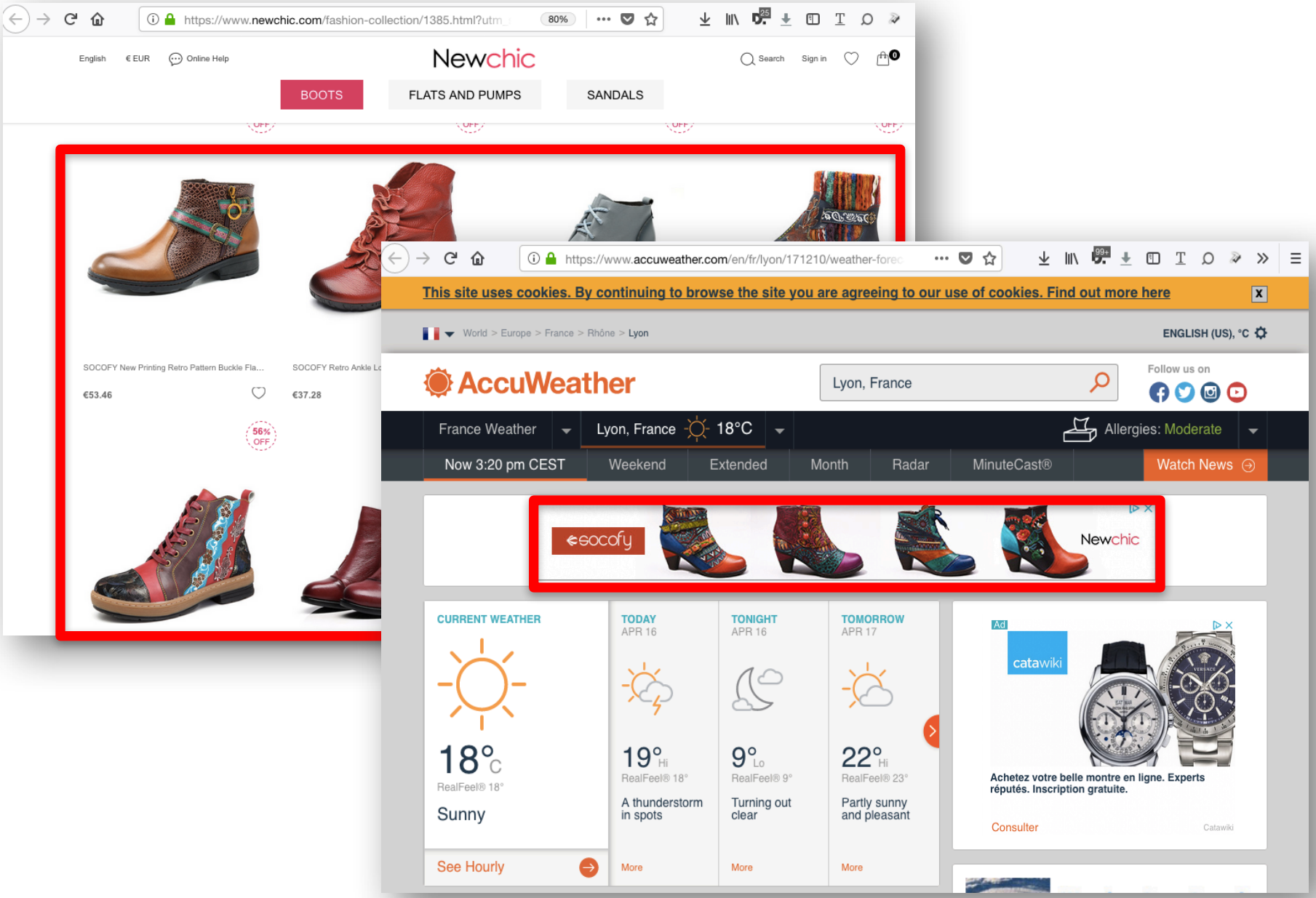
Web Tracking technologies

Nataliia Bielova

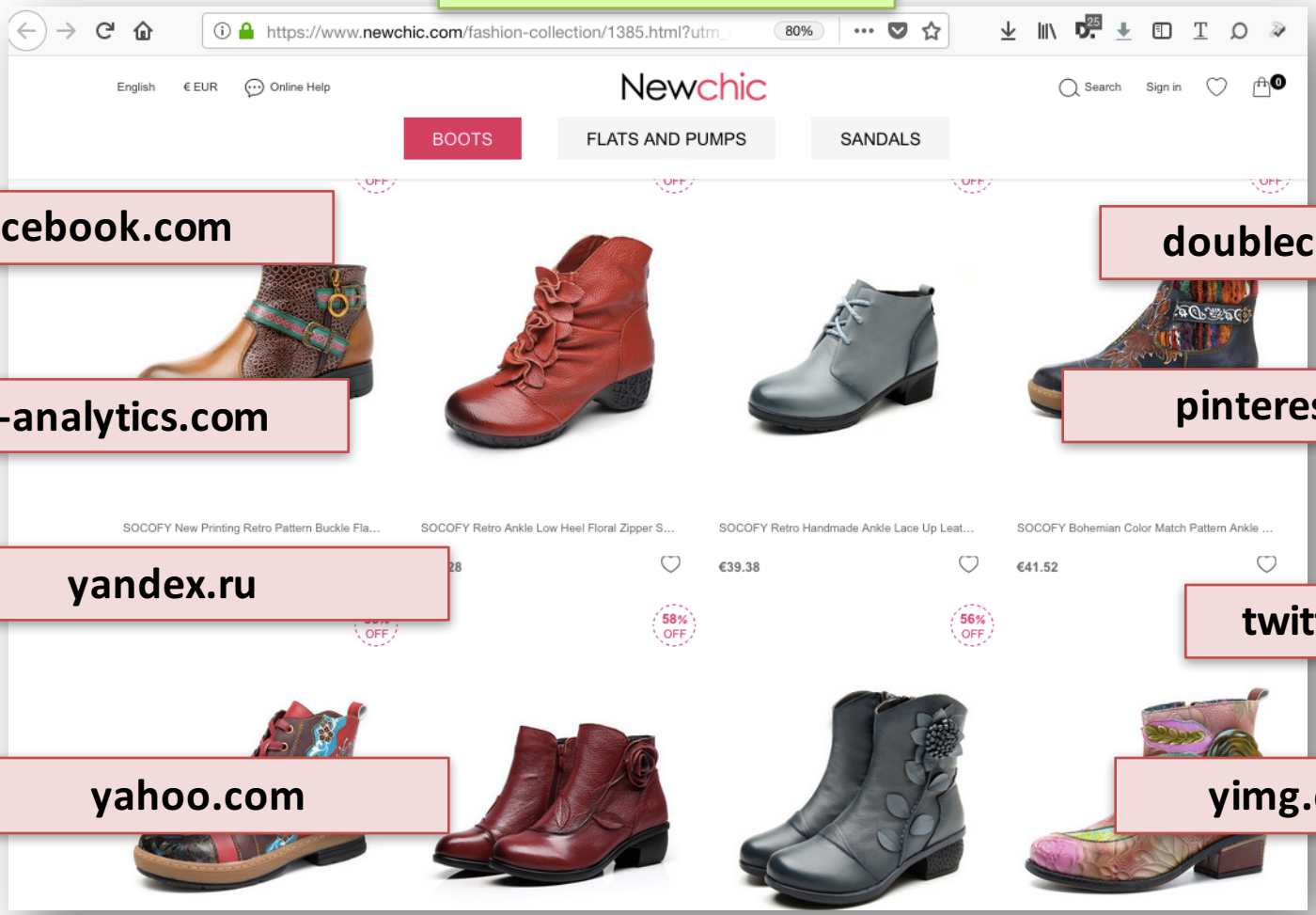
@nataliabielova

Security and ethical aspects of data

Université Cote d'Azur


https://www.newchic.com/fashion-collection/1385.html?utm
English
€ EUR
Online Help
Newchic
Search
Sign in
BOOTS
FLATS AND PUMPS
SANDALS
SOCOFY New Printing Retro Pattern Buckle Fla...
€53.46
SOCOFY Retro Ankle Lo
€37.28
56% OFF
https://www.accuweather.com/en/fr/lyon/171210/weather-forec
This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here
World > Europe > France > Rhône > Lyon
ENGLISH (US), °C
AccuWeather
Lyon, France
Follow us on
France Weather
Lyon, France 18°C
Allergies: Moderate
Now 3:20 pm CEST
Weekend
Extended
Month
Radar
MinuteCast®
Watch News
socofy
Newchic
CURRENT WEATHER
18°C
RealFeel® 18°
Sunny
See Hourly
TODAY
APR 16
19° Hi
RealFeel® 18°
A thunderstorm in spots
More
TONIGHT
APR 16
9° Lo
RealFeel® 9°
Turning out clear
More
TOMORROW
APR 17
22° Hi
RealFeel® 23°
Partly sunny and pleasant
More
Ad
catawiki
Achetez votre belle montre en ligne. Experts réputés. Inscription gratuite.
Consulter
Catawiki

newchic.com



facebook.com

doubleclick.net

google-analytics.com

pinterest.com

yandex.ru

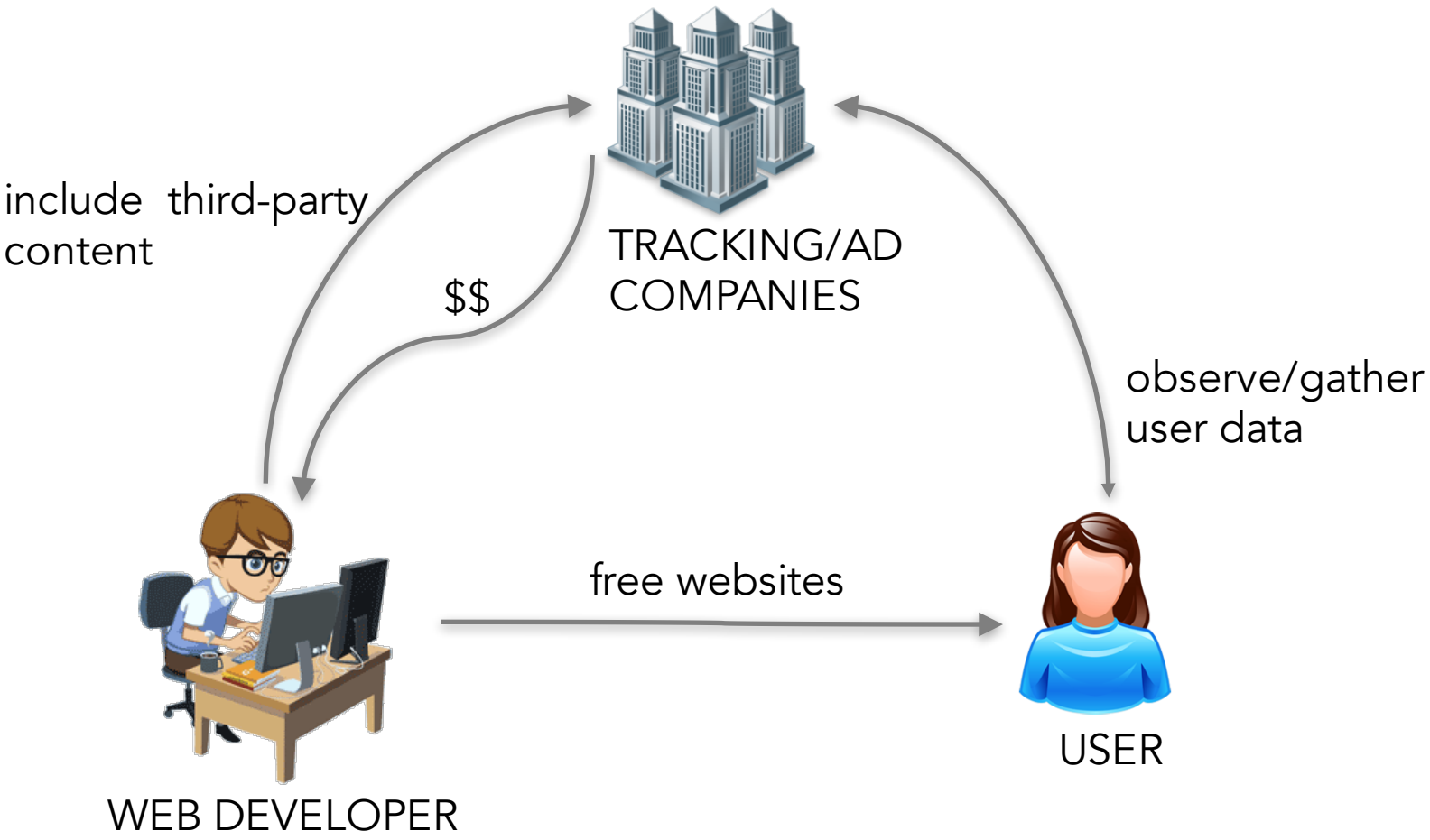
twitter.com

yahoo.com

yimg.com

What is Web Tracking?

Business model of the Web



Business model of the Web

include third-party content

TRACKING/AD

EU GDPR in force on May 25, 2018
ePrivacy Regulation under discussion

WEB DEVELOPER

free websites

USER

EU Data Protection Regulations

Applies to all people physically located in the EU

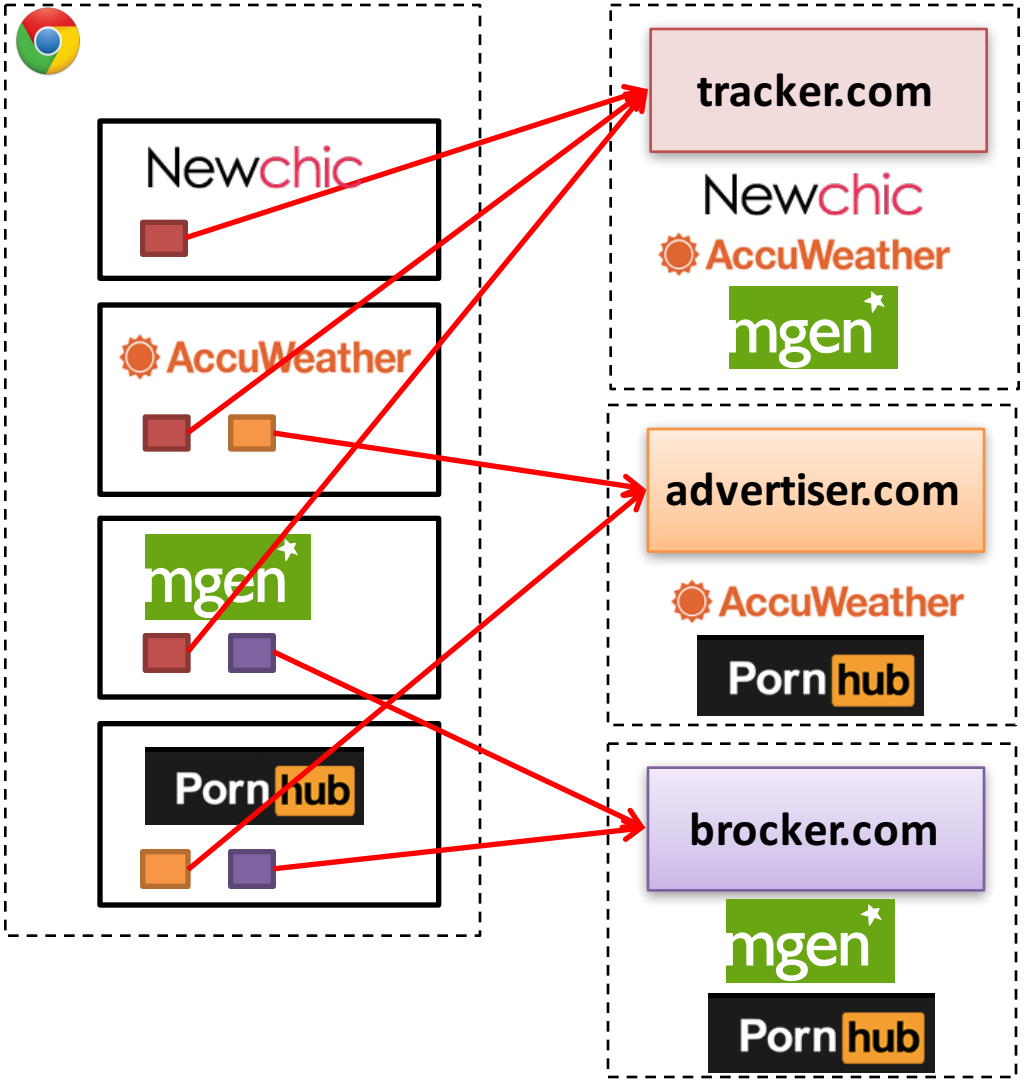
GDPR
(applies to personal data)

ePrivacy Regulation
(information stored on/retrieved from devices)

Applies to any information, not only personal data

Regulation = applies directly to all member states

Web Tracking



Tracking companies build bigger browsing profiles
= **increased value** for trackers
= **reduced privacy** for users

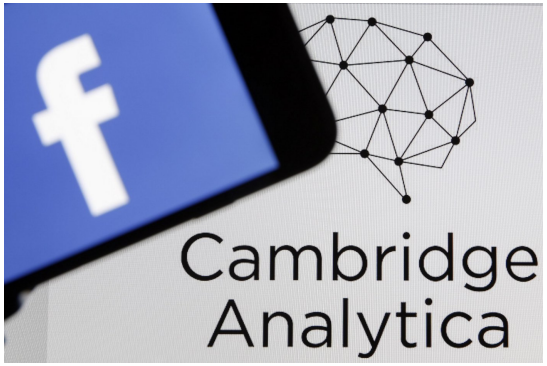
*Hypothetical relations

Based on the slide of Franzisca Roesner

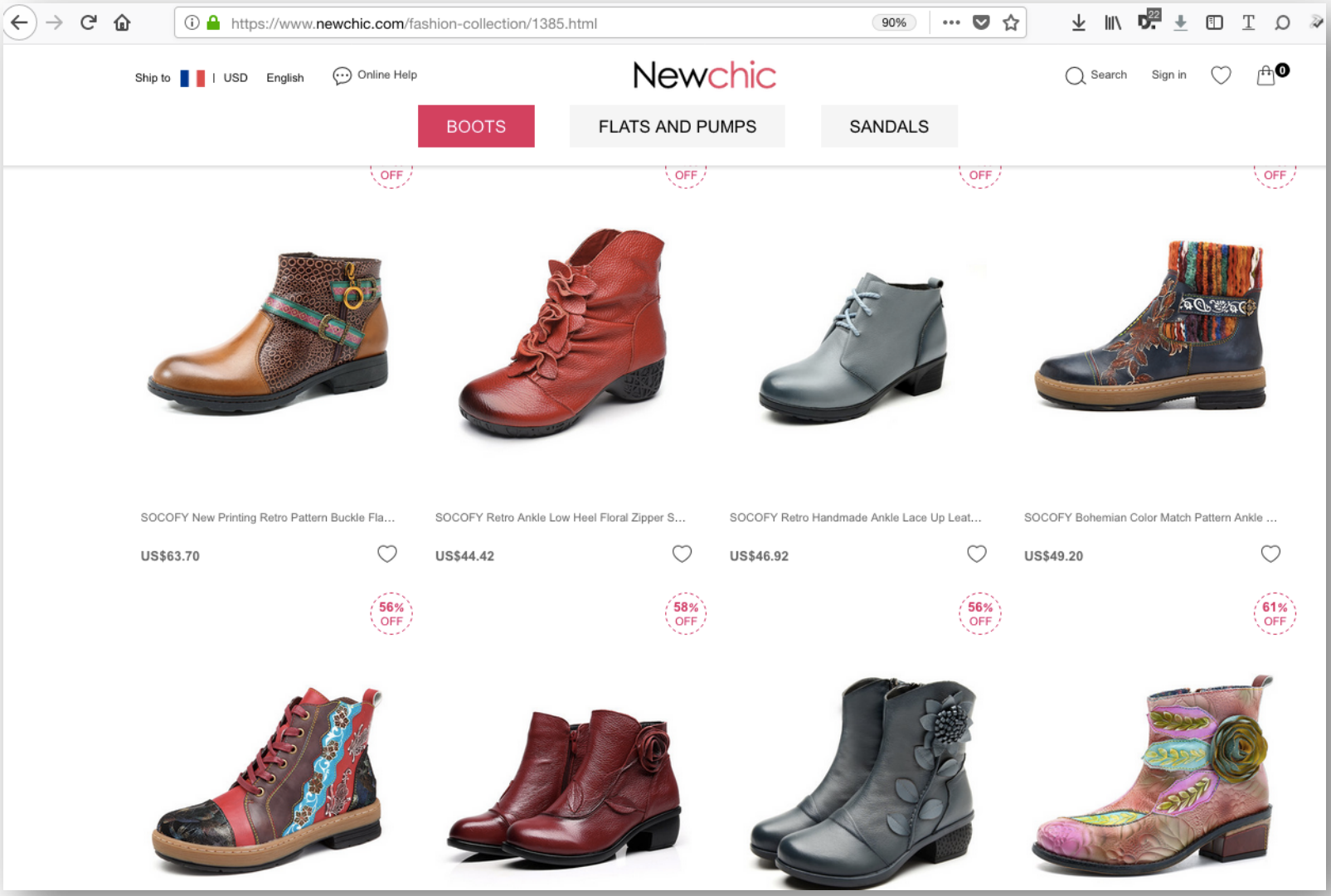
Why Web Tracking is important?

- Collection of our data without our knowledge
  - on sensitive websites
  - collection of our browsing patterns, preferences, tastes, even mood…

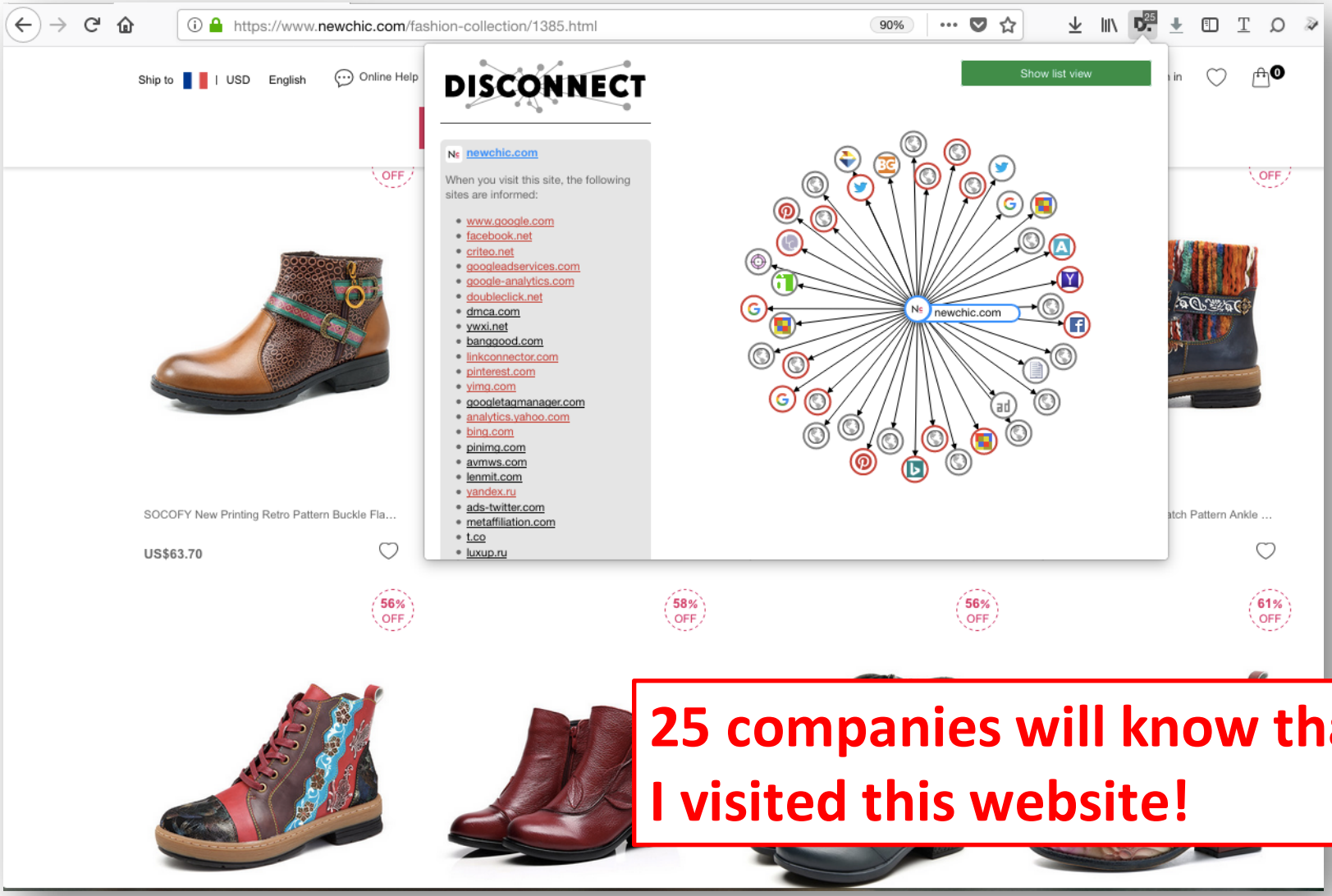
- **Usage of our data!**
  - targeted advertisement
  - **manipulation**



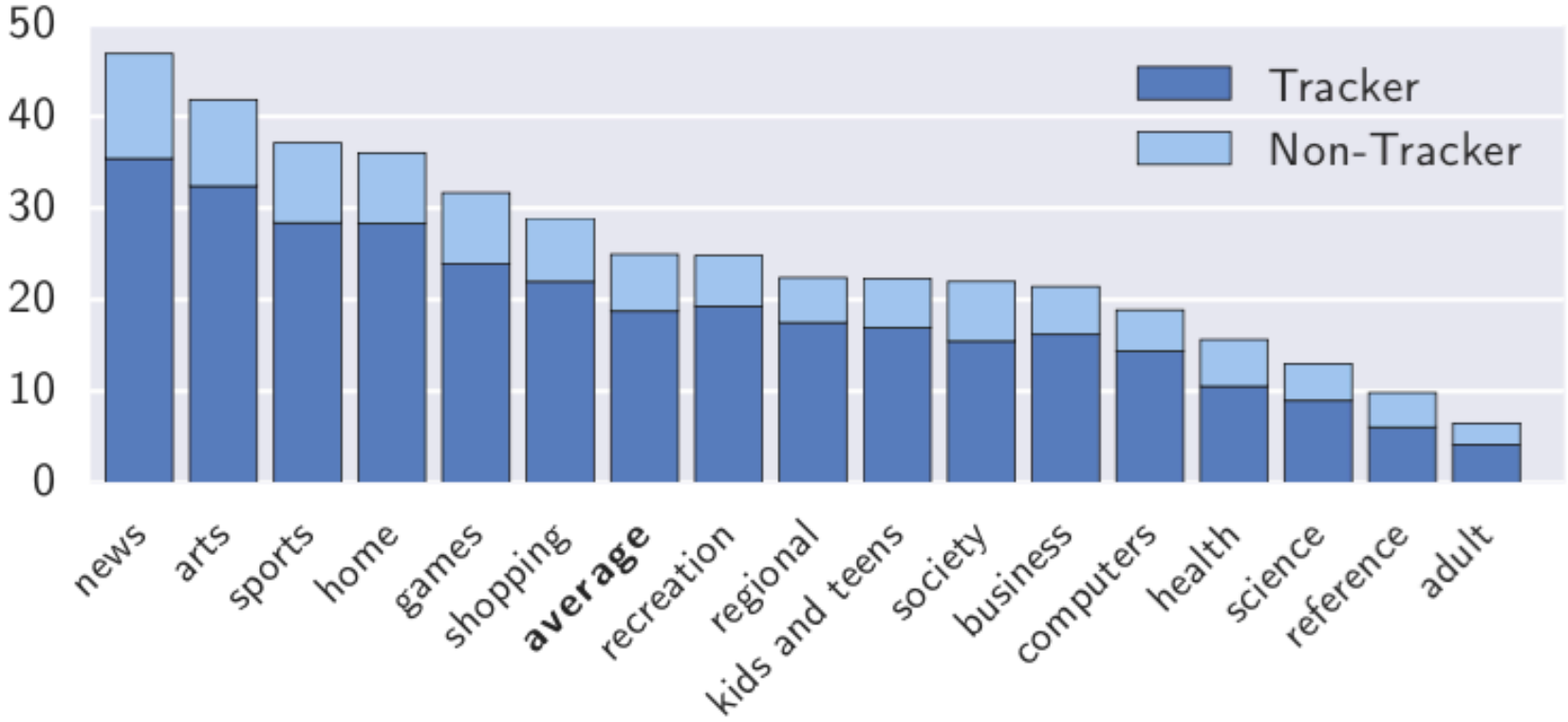
How many trackers per website?



How many trackers per website?



How many trackers per website?



Why are we tracked on news websites?

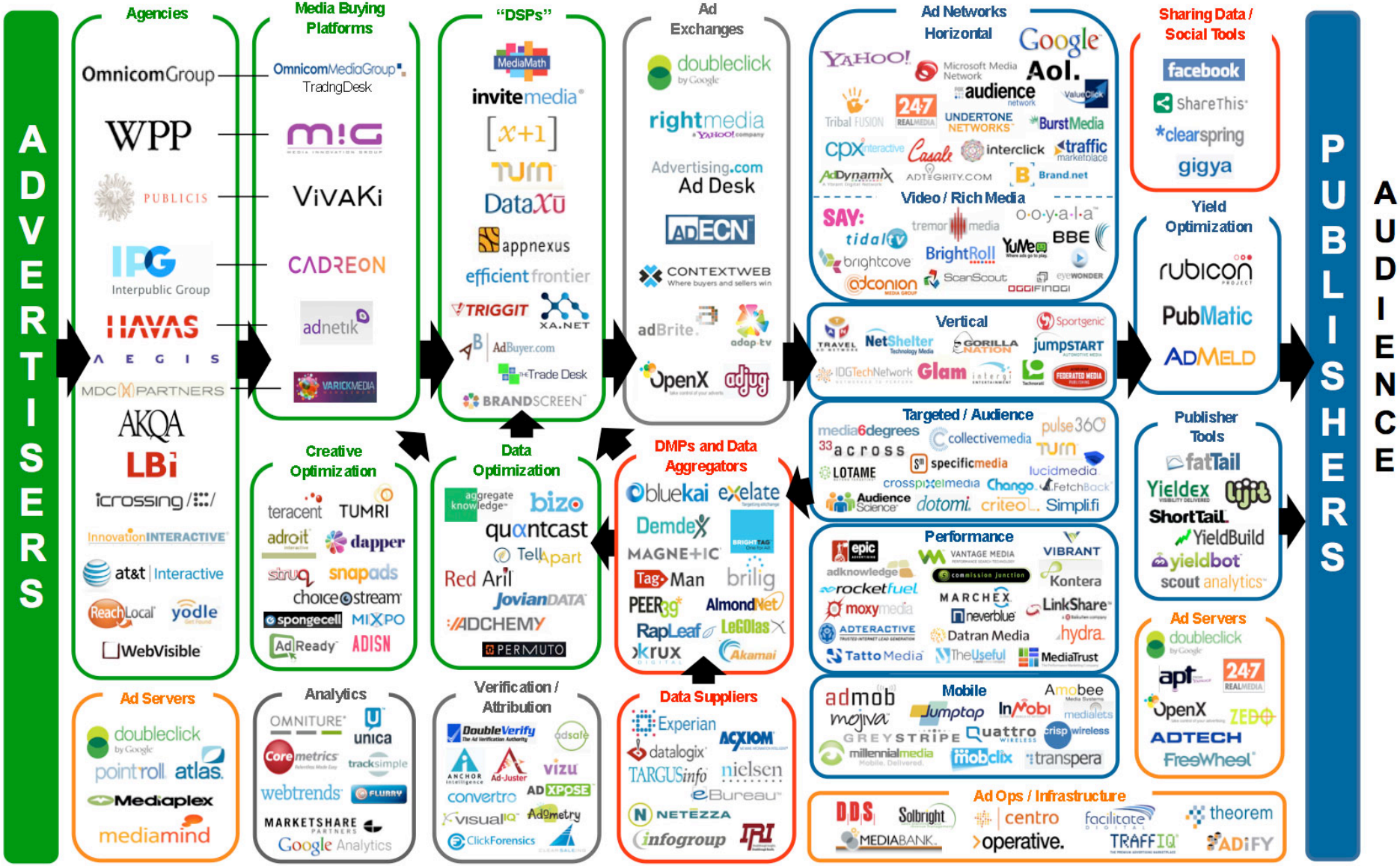
"The core business of the plaintiff is to deliver ads to its visitors. **Journalistic content is just a vehicle to get readers to view the ads.**"

Axel Springer's lawyer, 2015

Axel Springer SE is the largest digital publishing house in Europe

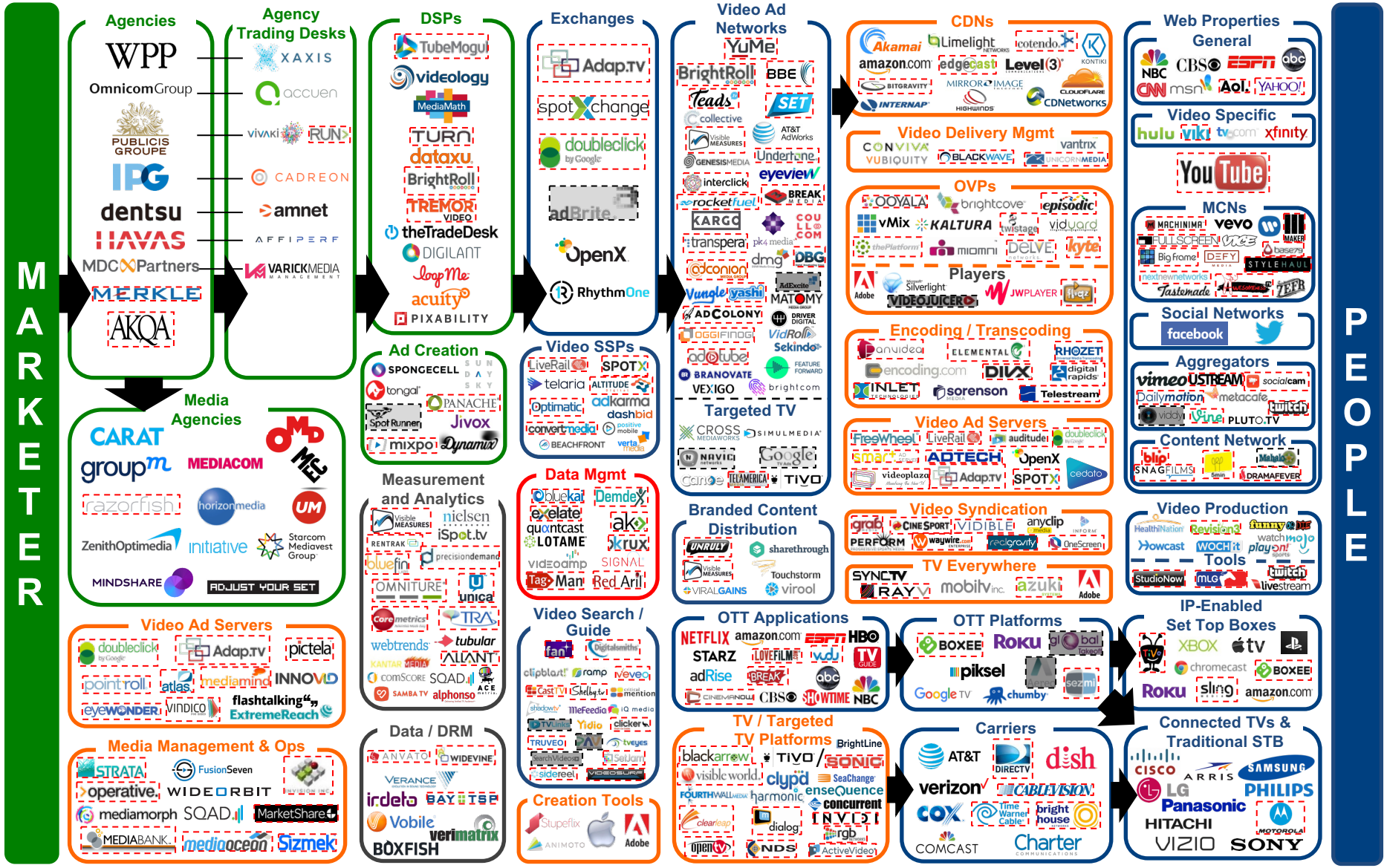
Display Advertising Technology Landscape

2010



VIDEO LUMAscape

2017

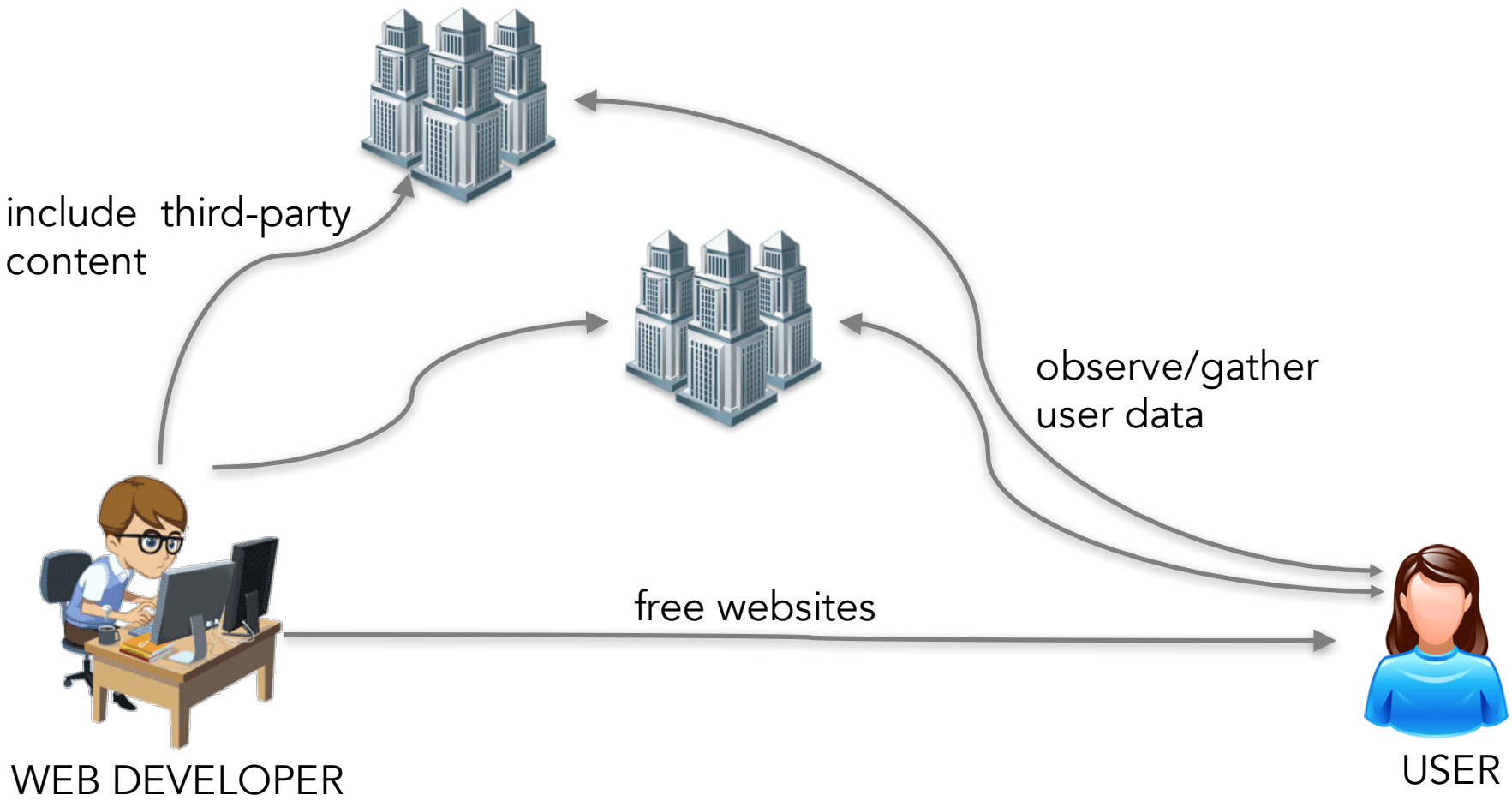


Denotes acquired company

Denotes shuttered company

© LUMA Partners LLC 20

Where is the problem?



Where is the problem?

Users have no control over their data!

Disproportion between the data collected and data actually used for functionality

free websites

WEB DEVELOPER

USER

Don’t browser extensions solve it?

AdBlockPlus: blocks scripts/requests **only from known advertising companies**

Ghostery: blocks scripts/requests **only from known tracking companies**

- They don’t protect from tracking
  - by other companies
  - by the main website
  - from cookie synchronization

How does Web Tracking work?

Cookies in HTTP header

Web browser

Cookie Database

bbc.co.uk/news:
session-id=20827872011

URL path: bbc.co.uk/news
Parameters
Method: GET
…

HTTP request

HTTP response

Status: 200 OK
Content: HTML page
Set-cookies: session-id=**20827872011** & …

…

Web server

Cookies in HTTP header

Web browser

URL path: bbc.co.uk/news…
Method: GET
Cookies: session-id=**2082787201l** & …
…

HTTP request

Web server

Cookie Database

bbc.co.uk/news:
session-id=2082787201l

Mechanisms Required By Trackers

- Ability to store/create user identity in the browser
  - HTTP cookies
  - HTTP headers
  - browser storages
  - device fingerprinting:
    - browser properties
    - OS properties
    - IP address…

- Ability to communicate user identity back to tracker
  - HTTP requests by the browser
  - JavaScript

Tracking via cookies

Cookies: first- and third-party

first-party cookie

Cookie Database

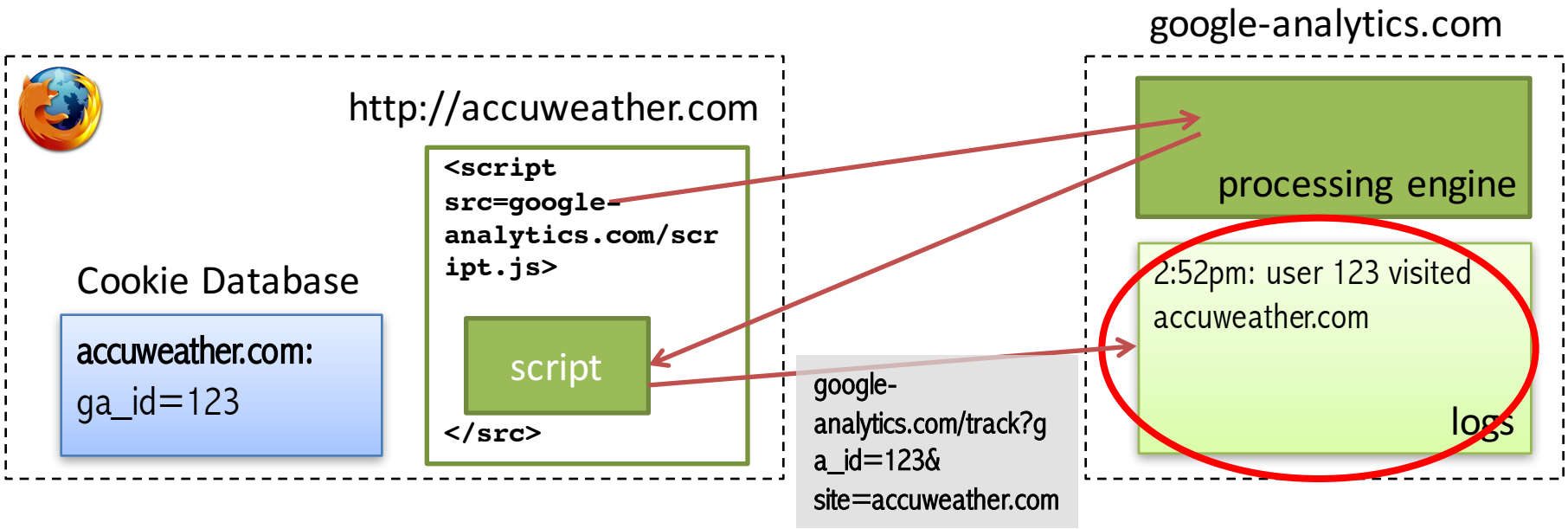
accuweather.com: id=123

doubleclick.com: id=456

third-party cookie

Within-Site Tracking

First-party cookies are used to track repeat visits to a site.



Based on the slide of Franziska Roesner

Cookies are manipulated via JavaScript

- Read/write access to cookies: **`document.cookie`**
- Script that sends cookies

```
// google-analytics.com/script.js

var url = "http://google-analytics.com/track?ga_id= "
    + encodeURI(document.cookie)
    + "&site= " + encodeURI(document.location);

document.write('<img src=' + url + '/>');
```

First-party cookies have more benefits

- Website owners can evaluate
  - website statistics
  - popularity of certain pages
  - popularity of links
  - selected and copied phrases

Cross-site tracking via Cookies

Cookies: first- and third-party

first-party cookie

Cookie Database

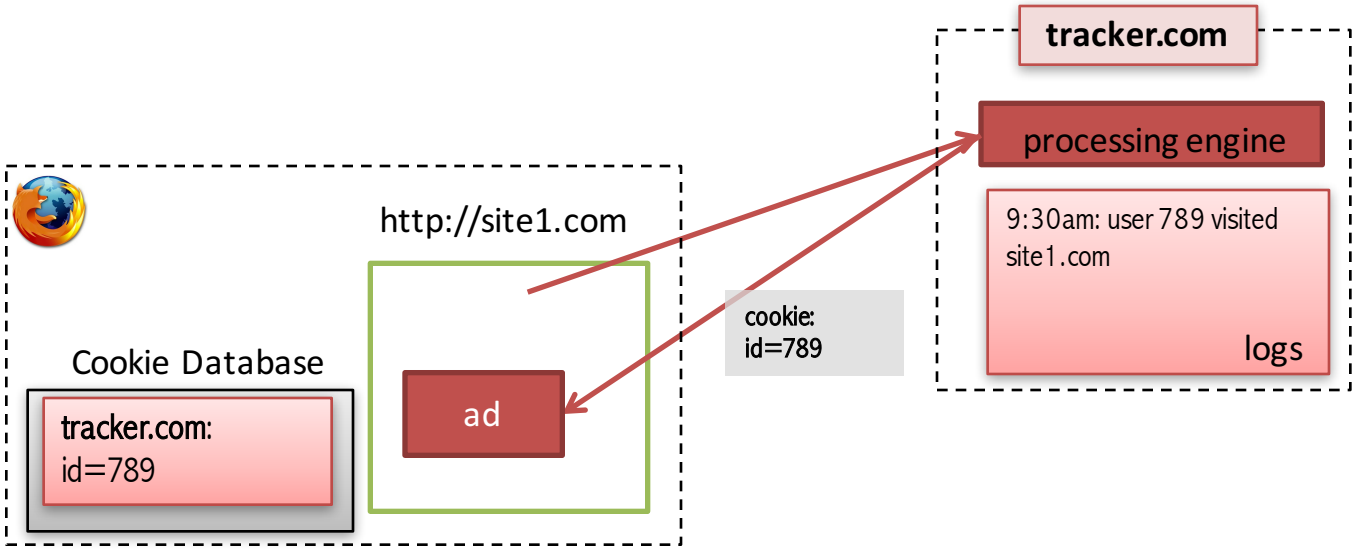
accuweather.com: id=123

doubleclick.com: id=456

third-party cookie

Cross-site tracking

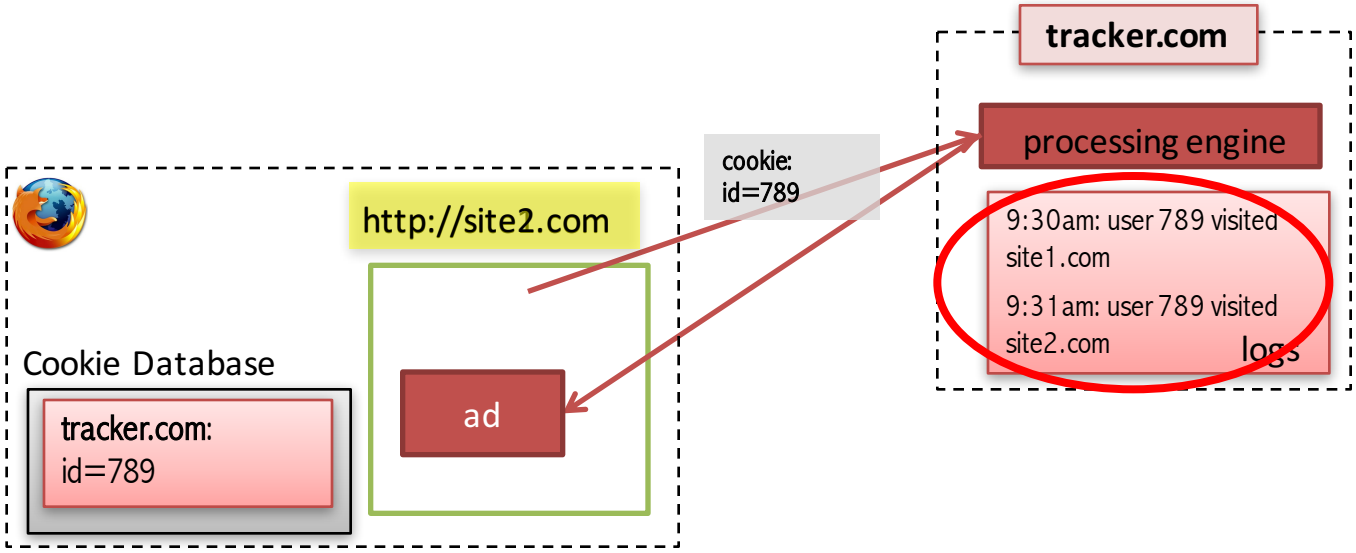
- **Third-party cookies** are used to track users across-sites and collect their browsing history.



Based on the slide of Franziska Roesner

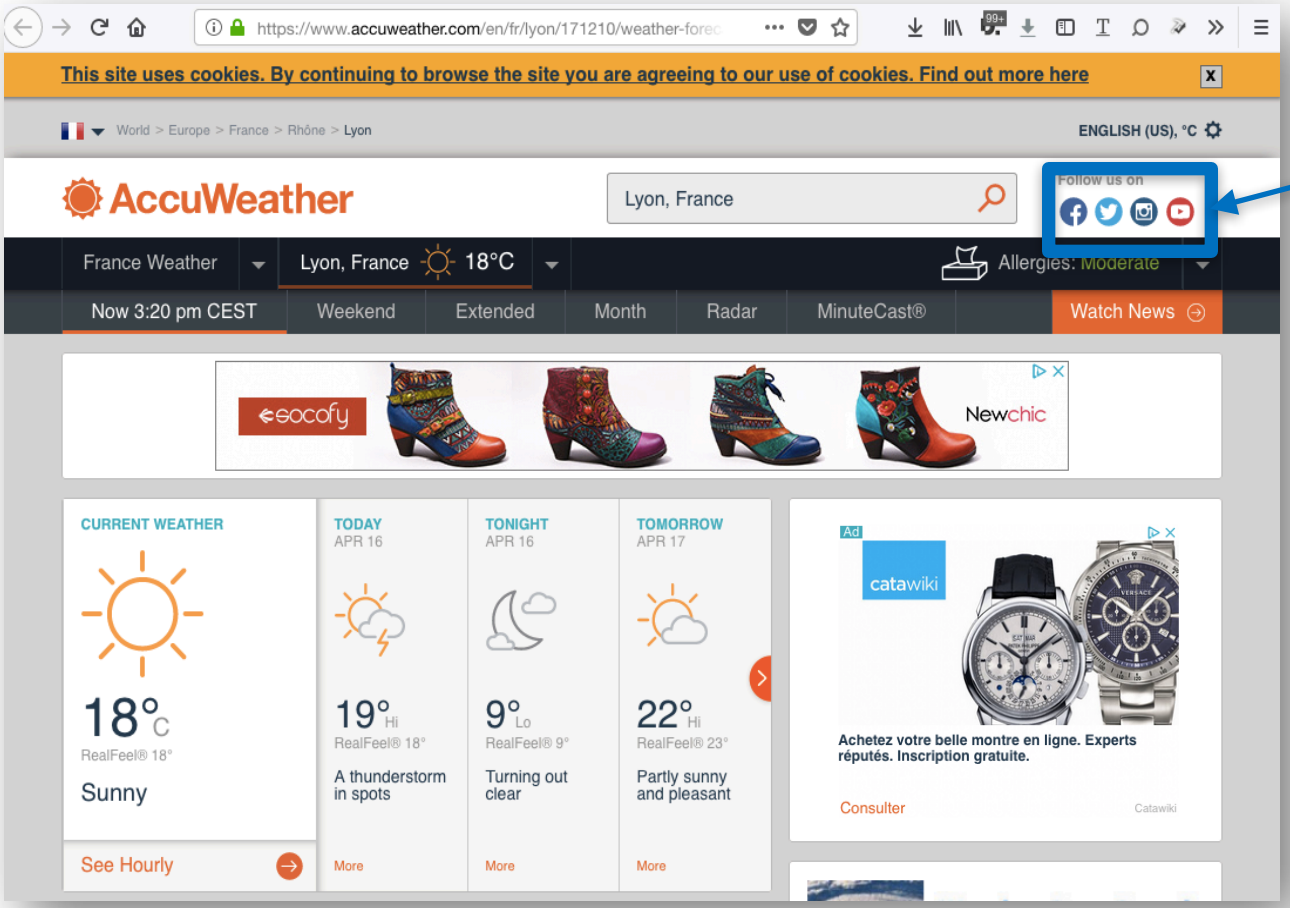
Cross-site tracking

- **Third-party cookies** are used to track users across-sites and collect their browsing history.



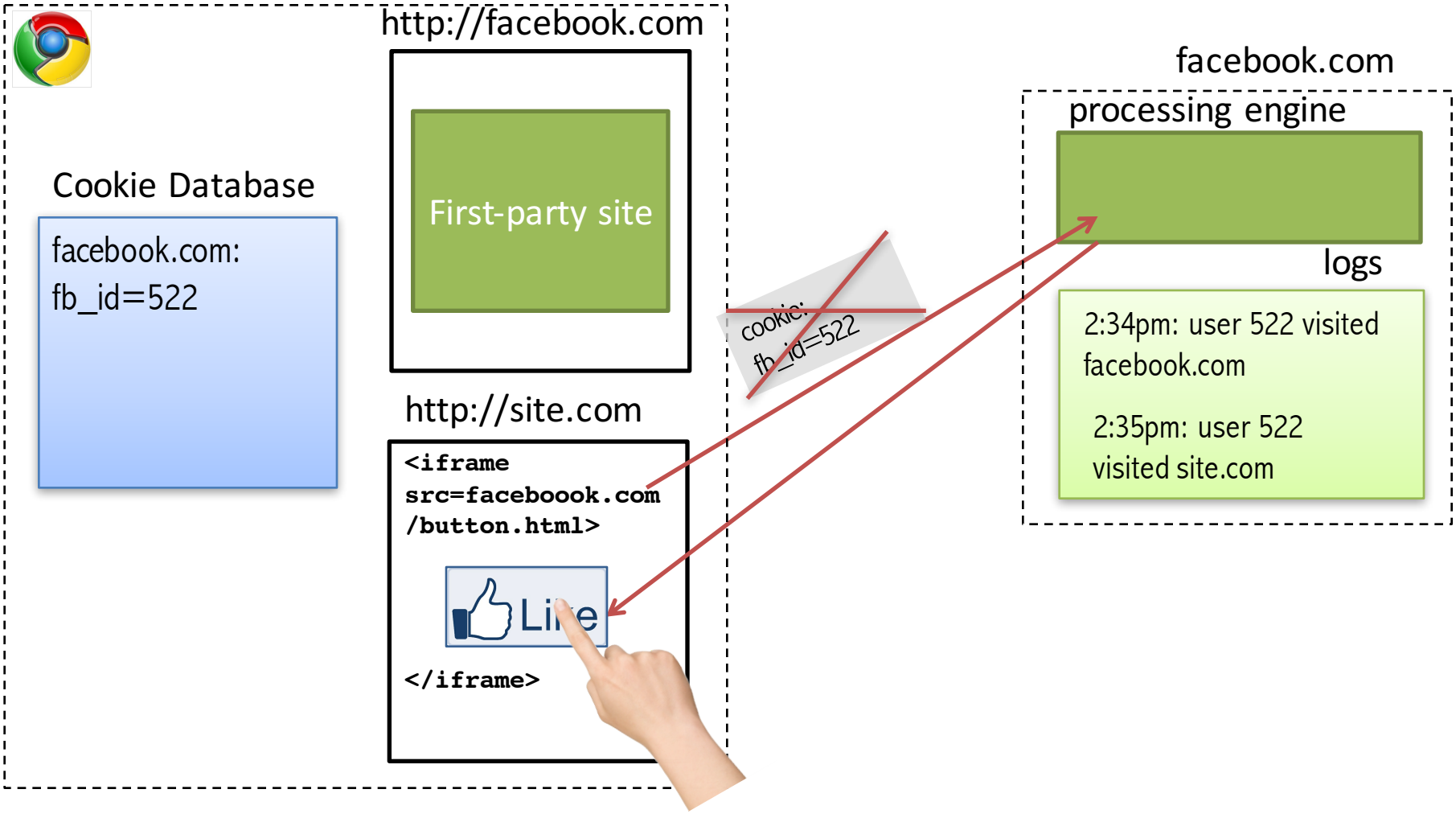
Based on the slide of Franziska Roesner

Social buttons



Can track cross-sites and collect browsing history!

ShareMeNot

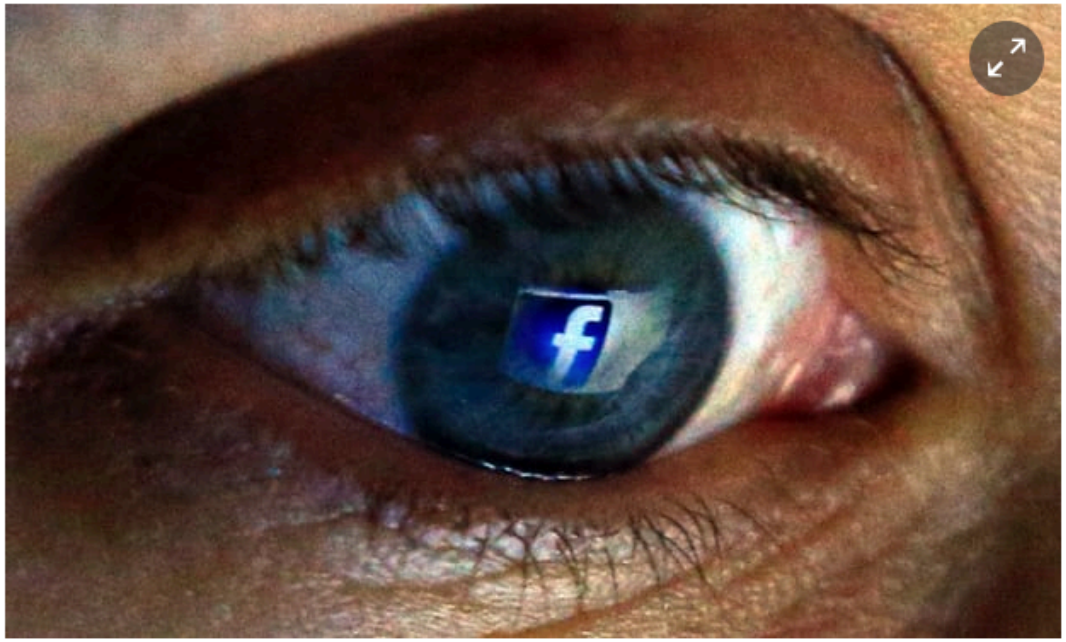


Now is a part of **Privacy Badger** https://www.eff.org/privacybadger

Facebook

Belgian court orders Facebook to stop tracking non-members

Judge threatens social network with fines of €250,000 a day over 'datr' cookie, which records visits to many websites whether or not user has Facebook account



Facebook has been ordered to stop tracking non-members in Belgium. Photograph: Jonathan Brady/PA

This article is 1 year old

Agence France-Presse in Brussels

Tuesday 10 November 2015 00.38 GMT

A Belgian court on Monday gave Facebook 48 hours to stop tracking internet users who do not have accounts with the social network or risk fines of up to €250,000 a day.

appeal against the order, which followed a case lodged by
ndog in June saying the US company indiscriminately
nen they visit pages on the site or click "like" or "share",
mbers.

This works because of third-party cookies!

CYBER RISK FEBRUARY 16, 2018 / 4:00 PM / 2 MONTHS AGO

Facebook loses Belgian privacy case, faces fine of up to $125 million

Reuters Staff

2 MIN READ

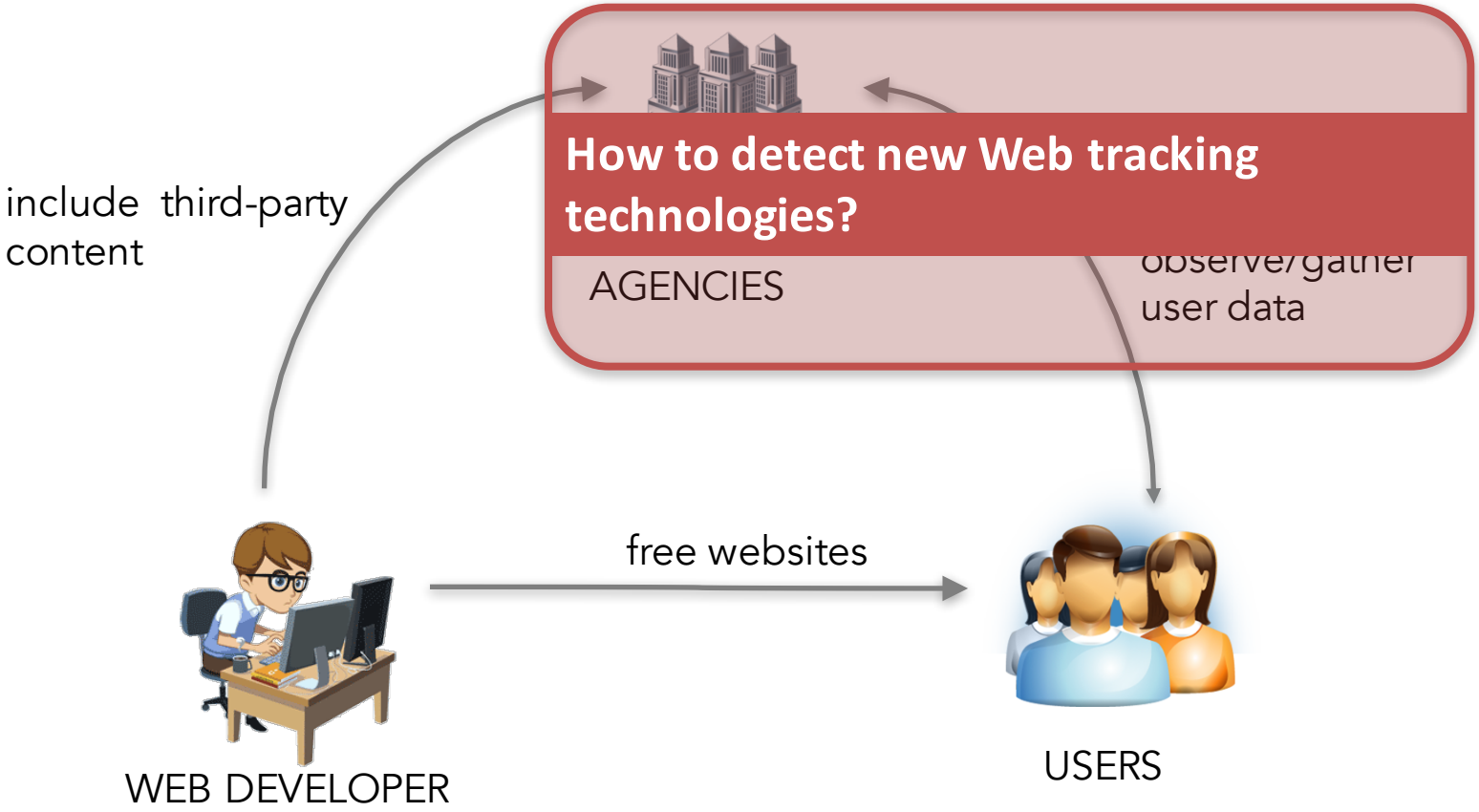
BRUSSELS (Reuters) - A Belgian court threatened Facebook (FB.O) on Friday with a fine of up to 100 million euros ($125 million) if it continued to break privacy laws by tracking people on third-party websites.

Exercise

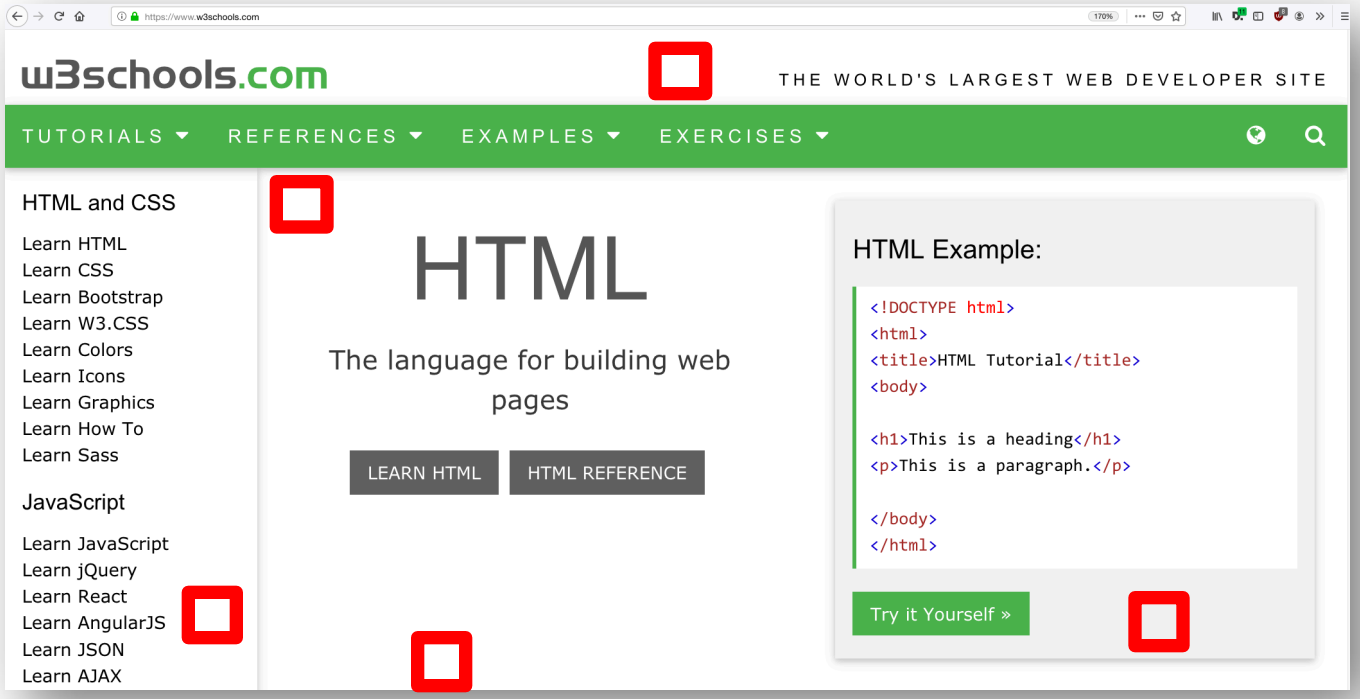
1. Try to find how to visualize all cookies in your preferred browser
2. Open a private/incognito mode
3. Visit your favorite news website
4. What cookies are stored by the website itself?
5. What cookies are stored by third-parties?
6. For how long third-party cookies will remain in your browser if you don’t delete them?

How to detect cookie-based tracking?

Research Challenges

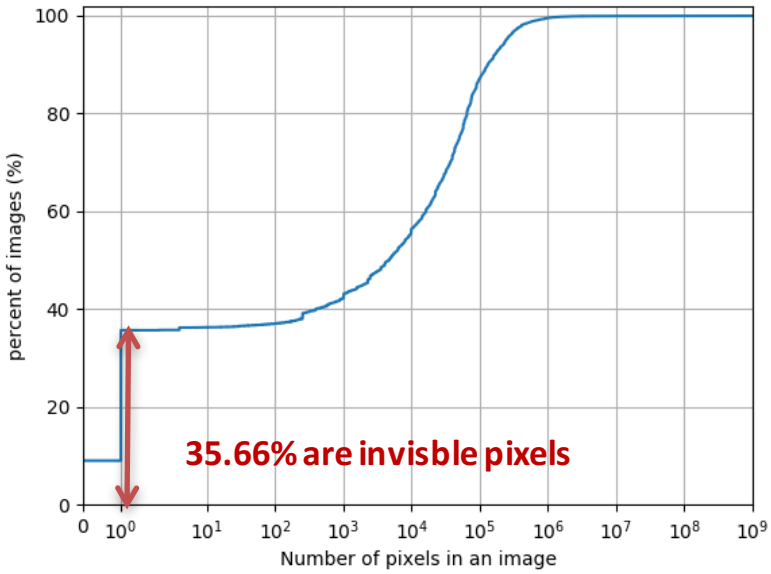


Invisible pixels



Data collection with OpenWPM

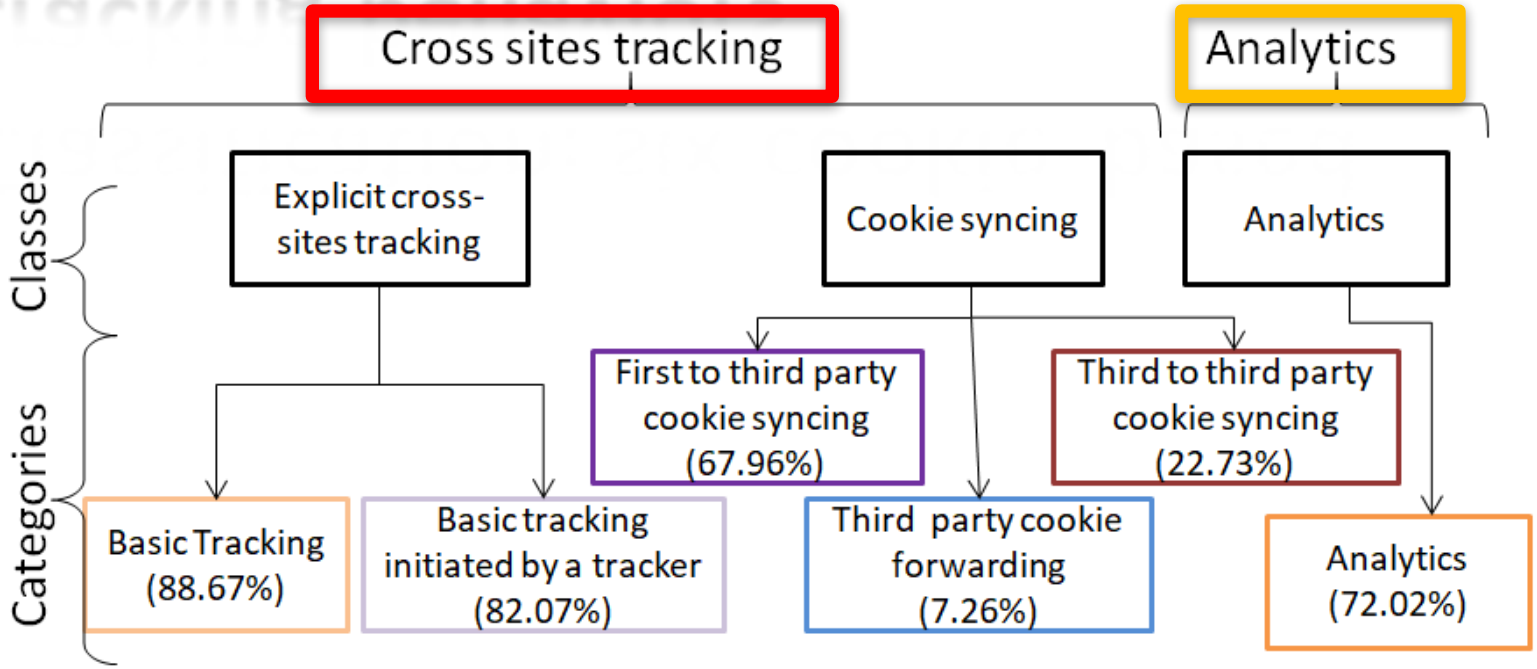
- Crawl Top 10,000 Alexa domains in February 2019
- For each domain we visit
  - Homepage + 10 first links
- Successfully crawled:
  - 8,744 domains, 84,658 pages
- Results:
  - 2,297,716 images <100KB collected
  - **35.66% images are invisible**
  - **95% domains** contain at least one invisible image



Invisible pixels are perfect suspect for tracking and widely present on the Web

Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels. Imane Fouad, Nataliia Bielova, Arnaud Legout, Natasa Sarafijanovic-Djukic.To appear in Privacy Enhancing Technologies Symposium (PETS 2020).

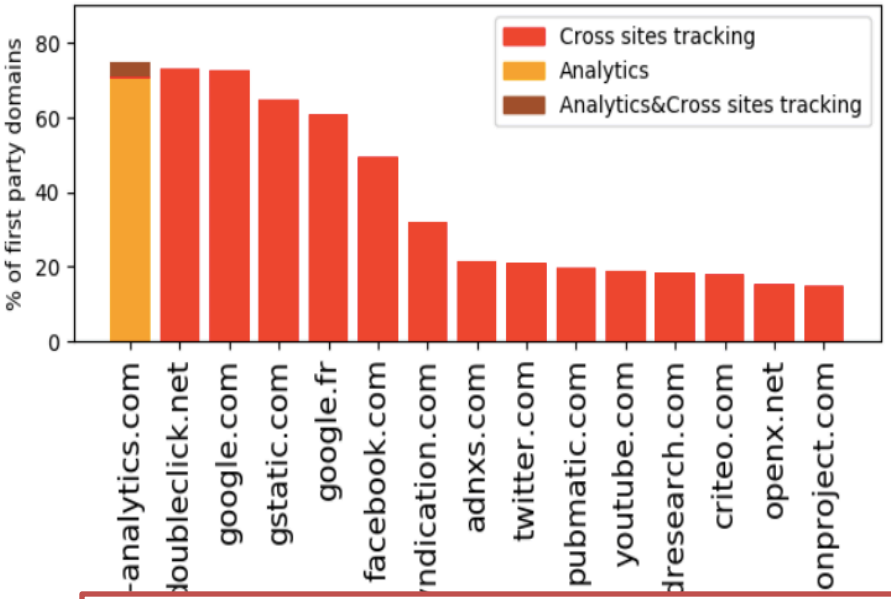
Classification: six cookie-based tracking behaviors

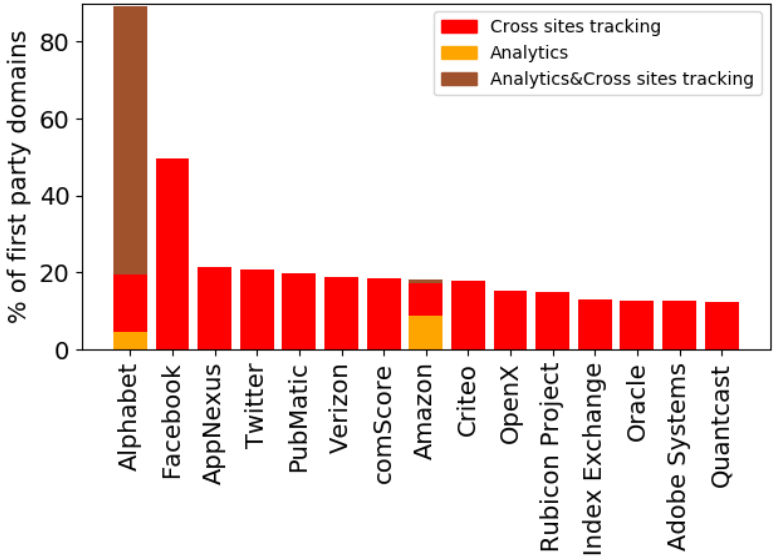


At least one type of tracking found on 92% of domains!

Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels. Imane Fouad, Nataliia Bielova, Arnaud Legout, Natasa Sarafijanovic-Djukic.To appear in Privacy Enhancing Technologies Symposium (PETS 2020).

Third-party tracking domains and companies





Some third parties combine privacy-invasive tracking and analytics behaviors on the same website!

Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels. Imane Fouad, Nataliia Bielova, Arnaud Legout, Natasa Sarafijanovic-Djukic.To appear in Privacy Enhancing Technologies Symposium (PETS 2020).

Cookie respawning

AKA ZOMBIECOOKIES

Cookie respawning

- Cookies **can respawn** even if the user has deleted them
  - HTML5 localStorage (across sessions only)
  - Flash LSOs (across sessions and web browsers)
  - HTTP headers: Etag, LastModified

HTML5 localStorage

- HTML5 localStorage allows to store pairs of strings key + value
- localStorage has no expiration date

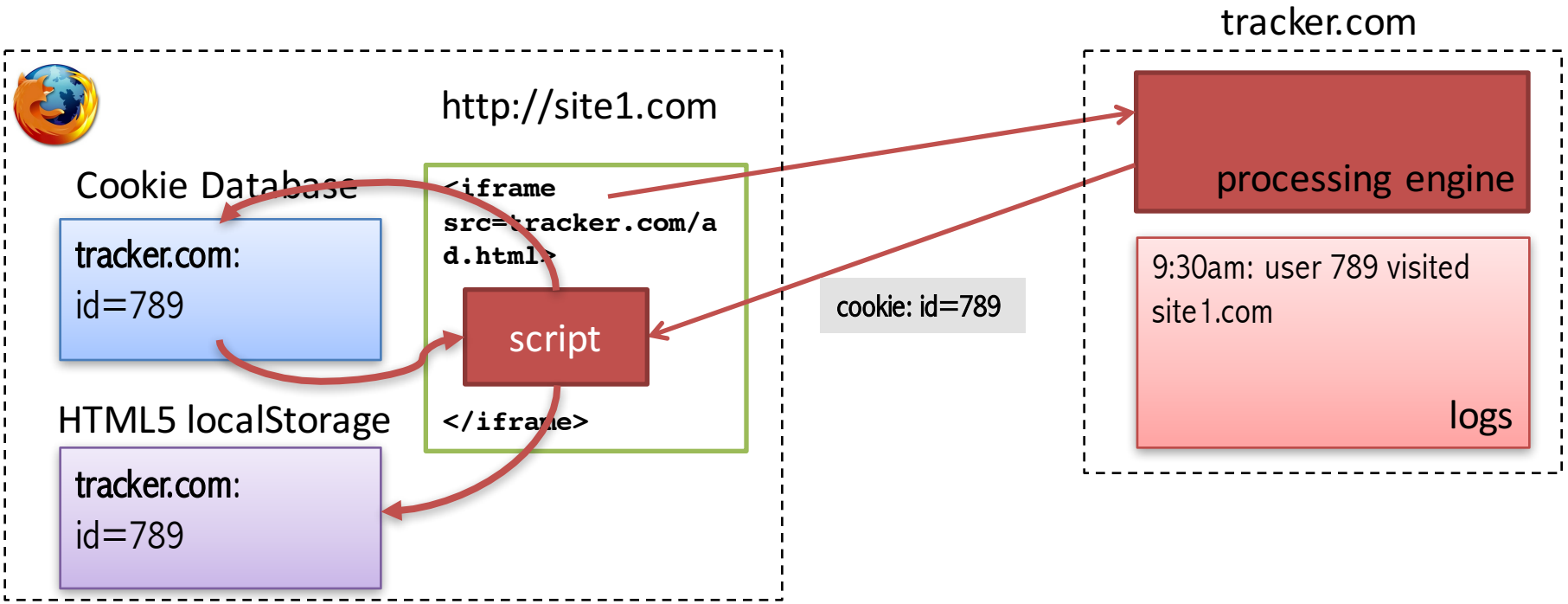
```
localStorage.setItem('key', 'value');

var x = localStorage.getItem('key');

localStorage.removeItem('key');
```

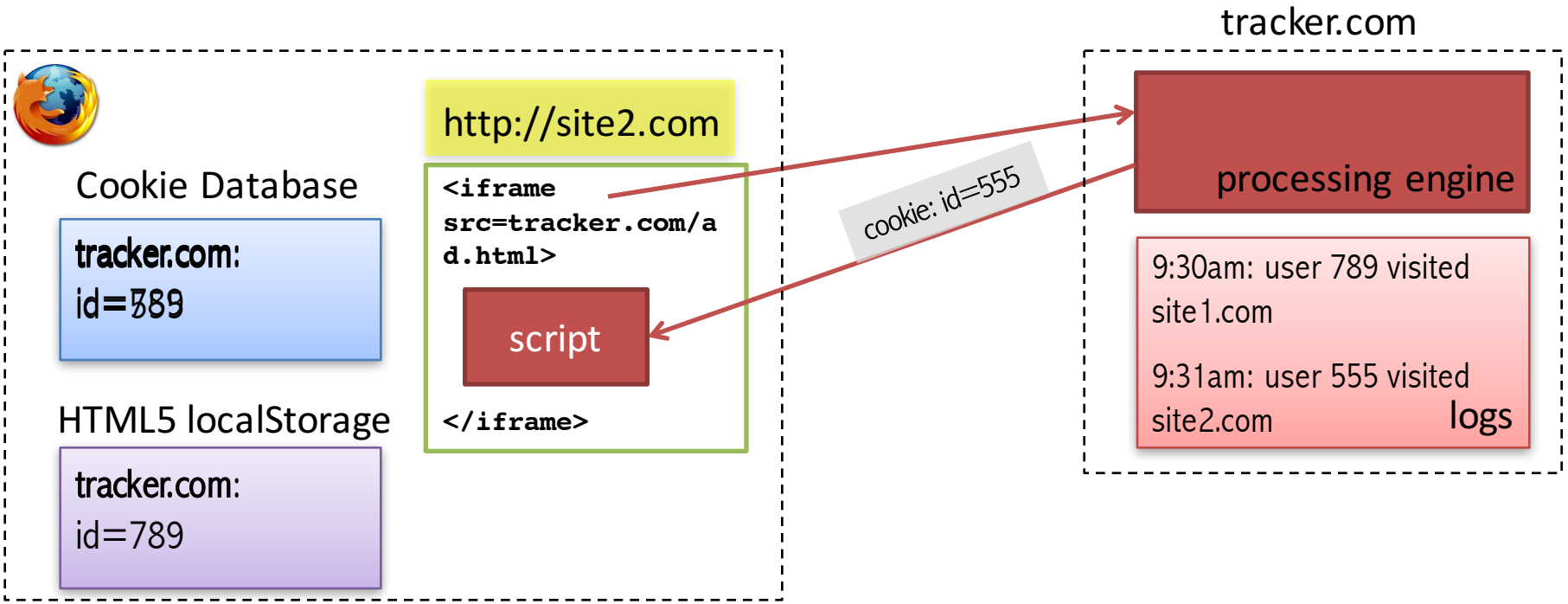
Respawning via HTML5 localStorage

User leaves the page



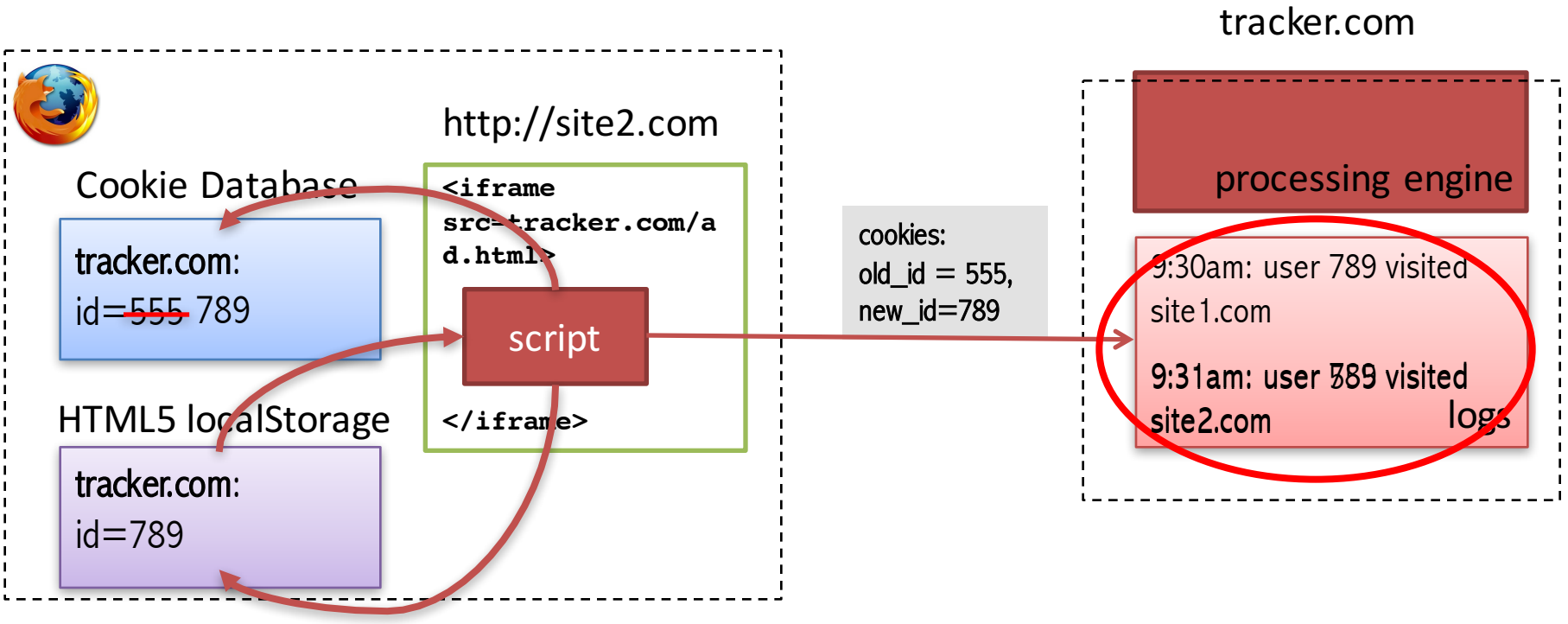
Respawning via HTML5 localStorage

User deletes all the cookies!



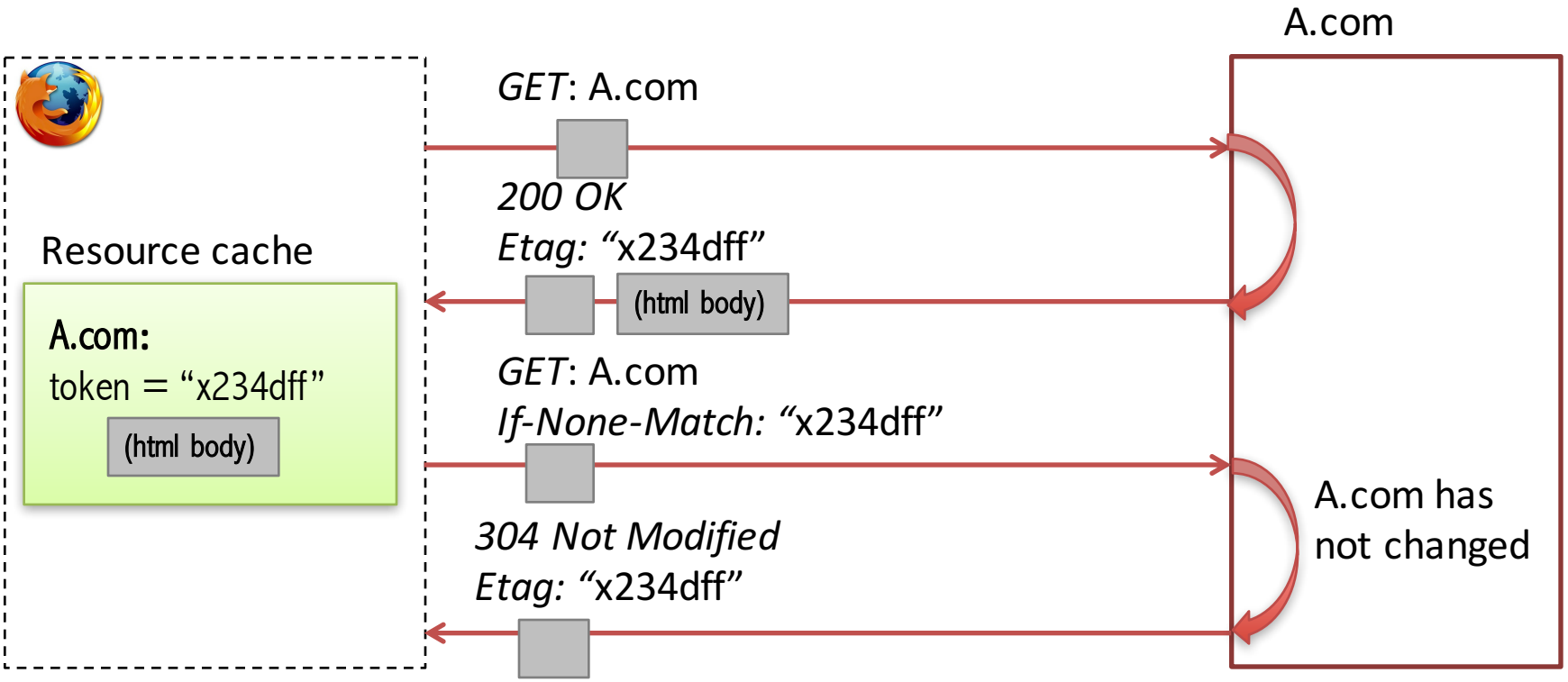
Respawning via HTML5 localStorage

User deletes all the cookies!



Respawning via Etag header

- Etag HTTP header is a caching mechanism



Respawning via Etag header

```
INITIAL REQUEST HEADER:
  GET /i.js HTTP/1.1
  Host: i.kissmetrics.com

INITIAL RESPONSE HEADER:
  Etag: "Z9iGGN1n1-zeVqbgzrlKkl39hiY"
  Expires: Sun, 12 Dec 2038 01:19:31 GMT
  Last-Modified: Wed, 27 Jul 2011 00:19:31 GMT
  Set-Cookie: _km_cid=Z9iGGN1n1-zeVqbgzrlKkl39hiY;
              expires=Sun, 12 Dec 2038 01:19:31 GMT;path=/;

SUBSEQUENT REQUEST HEADER (PRIVATE BROWSING MODE WITH ALL COOKIES BLOCKED):
  GET /i.js HTTP/1.1
  Host: i.kissmetrics.com
  If-None-Match: "Z9iGGN1n1-zeVqbgzrlKkl39hiY"
```

- KissMetrics lawsuit, August 2011

Not Respawning, but Tracking

- **Important detail**:
  - If Etag header, HTML5 localStorage, or Flash LSO didn't store a copy of cookies

  => **tracking would not be detected**!

- **Privacy problem:**
  - All of these storages can be used for tracking without cookies

What I Learned from Fighting a 12-Month-Long Lawsuit

by NEIL PATEL *on* DECEMBER 1, 2014

Entrepreneurs tend to talk about the glory moments. *You know... about raising millions of dollars from investors or selling their companies.*

Sadly, there isn't much you can learn from those glory stories, which is why I rarely discuss them. Instead, I focus on sharing my mistakes because if you can avoid making the ones I made, *you'll increase your chances of succeeding.*

One of the toughest parts about my entrepreneurial journey very few people know about was spending a year fighting a class action lawsuit (it's just a fancy word for multiple lawsuits combined into one) and the Federal Trade Commission.

Download this printable cheat sheet of 8 lessons learned from fighting a 12-month long lawsuit.

Before I get into what I learned, let me give you the back story...

My startup

Over five years ago, my co-founder and I started an analytics company, KISSmetrics. Our goal was to help companies increase the lifetime value of their customers.

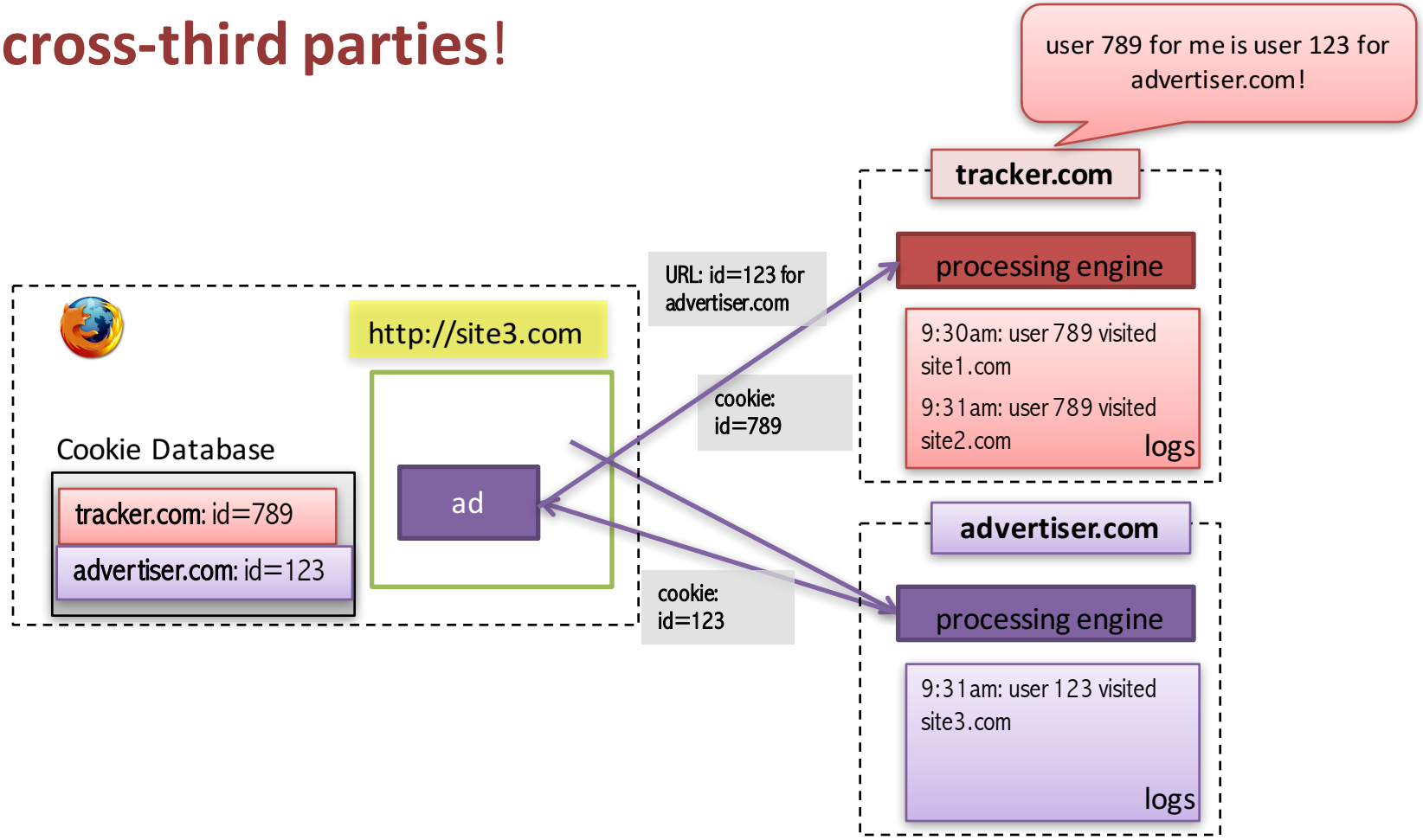
When we launched, we had no competitors. Through our network, we were able to land a few big accounts like Amazon and Microsoft as well as large startups like Hulu and Spotify.

KissMetrics story

Cookie Synchronisation

Cookie synchronization

- **Cookie syncing** is used to merge user's identifier **across-third parties**!

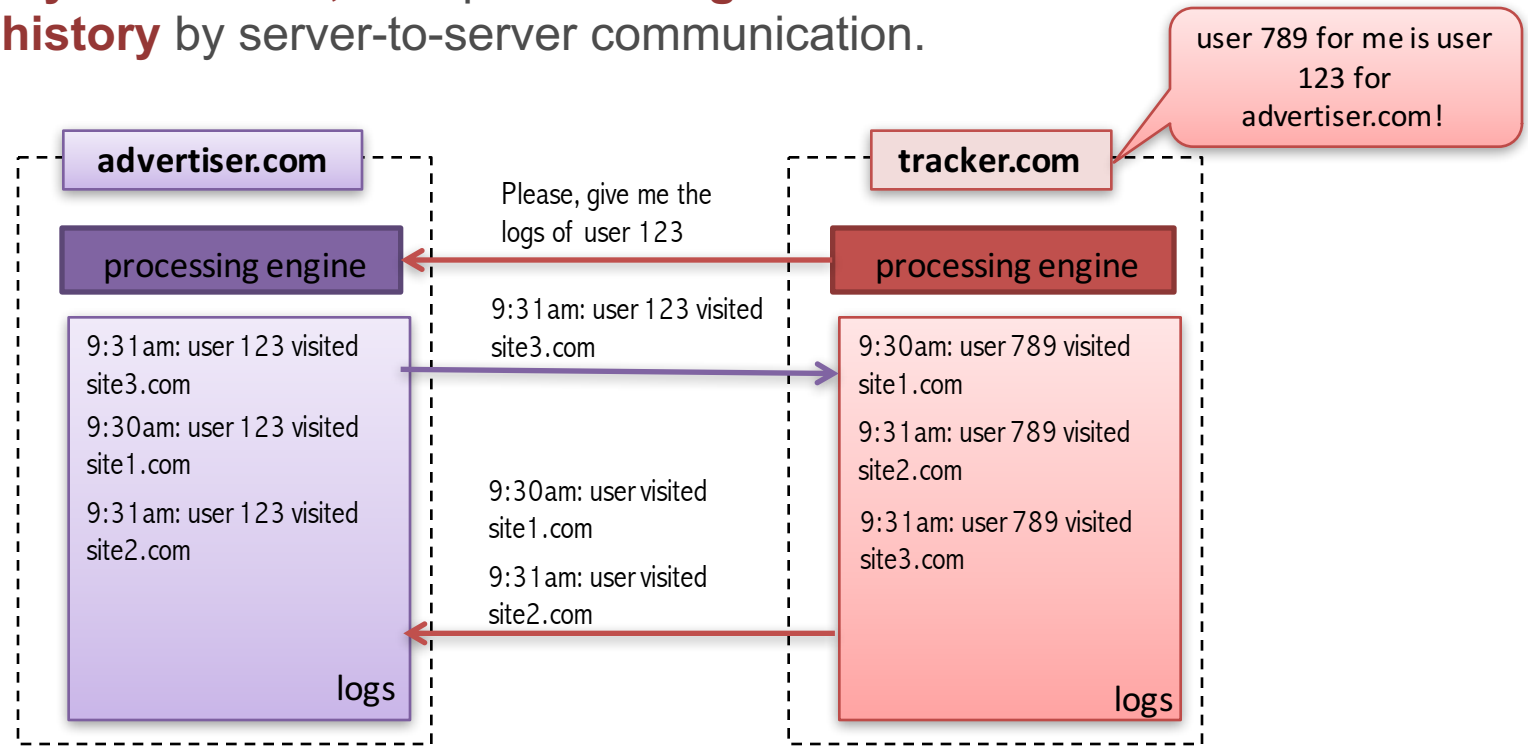


Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels. Imane Fouad, Nataliia Bielova, Arnaud Legout, Natasa Sarafijanovic-Djukic.To appear in Privacy Enhancing Technologies Symposium (PETS 2020).

Based on the slide of Franziska Roesner

Data merging on server-side thanks to cookie syncing

- Once **ID is synchronised,** third parties **merge user's browsing history** by server-to-server communication.

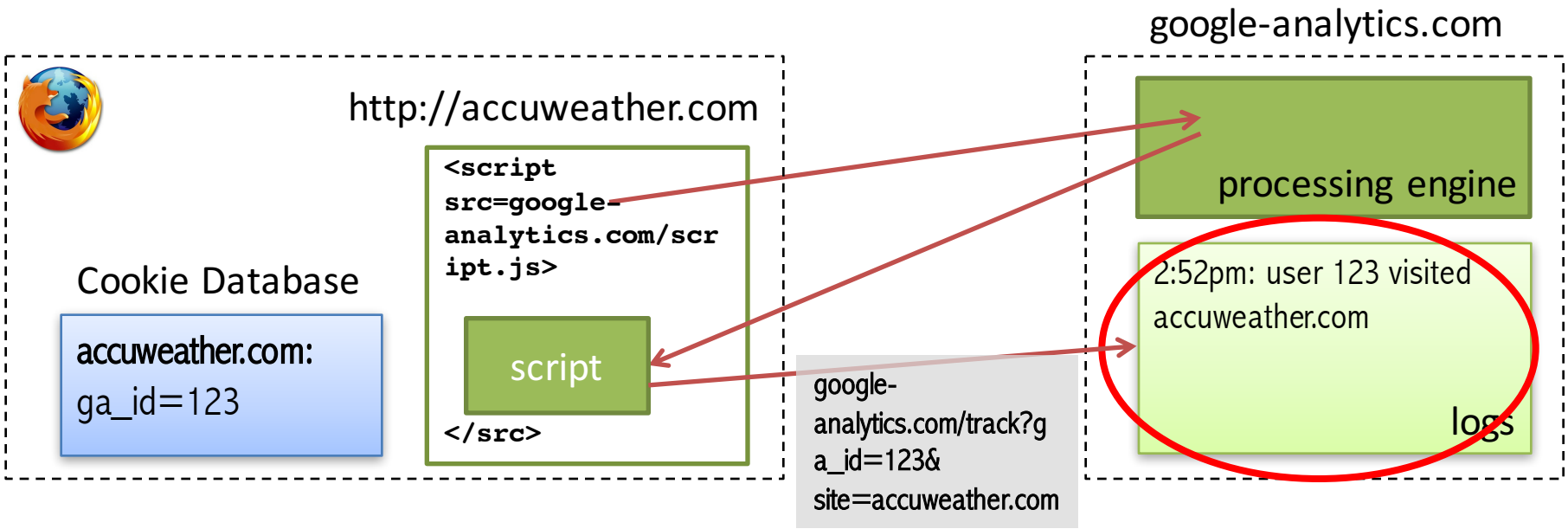


A closer look at Google Analytics

Google Analytics

Is it a within-site or cross-site tracking?

What could go wrong?



Based on the slide of Franziska Roesner

https://support.google.com/analytics/answer/2799357?hl=en

Analytics Help | Describe your issue | Sign in

Configure Analytics to display Demographics and Interests data

Before you can see or work with Demographics and Interests data in Analytics, you need to:

1. Enable Advertising Reporting Features for your property
2. Enable the Demographics and Interests reports for the view

Where Analytics gets the data

Once you update Analytics to support Advertising Reporting Features, Analytics collects Demographics and Interests data from the following sources:

Source	Applies to	Condition	Result
Third-party DoubleClick cookie	Web-browser activity only	Cookie is present	Analytics collects any demographic and interests information available in the cookie
Android Advertising ID	App activity only	You update the Analytics tracking code in an Android app to collect the Advertising ID	Analytics generates an identifier based on the ID that includes demographic and interests information associated with users' app activity
iOS Identifier for Advertisers (IDFA)	App activity only	You update the Analytics tracking code in an iOS app to collect the IDFA	Analytics generates an identifier based on the IDFA that includes demographic and interests information associated with users' app activity

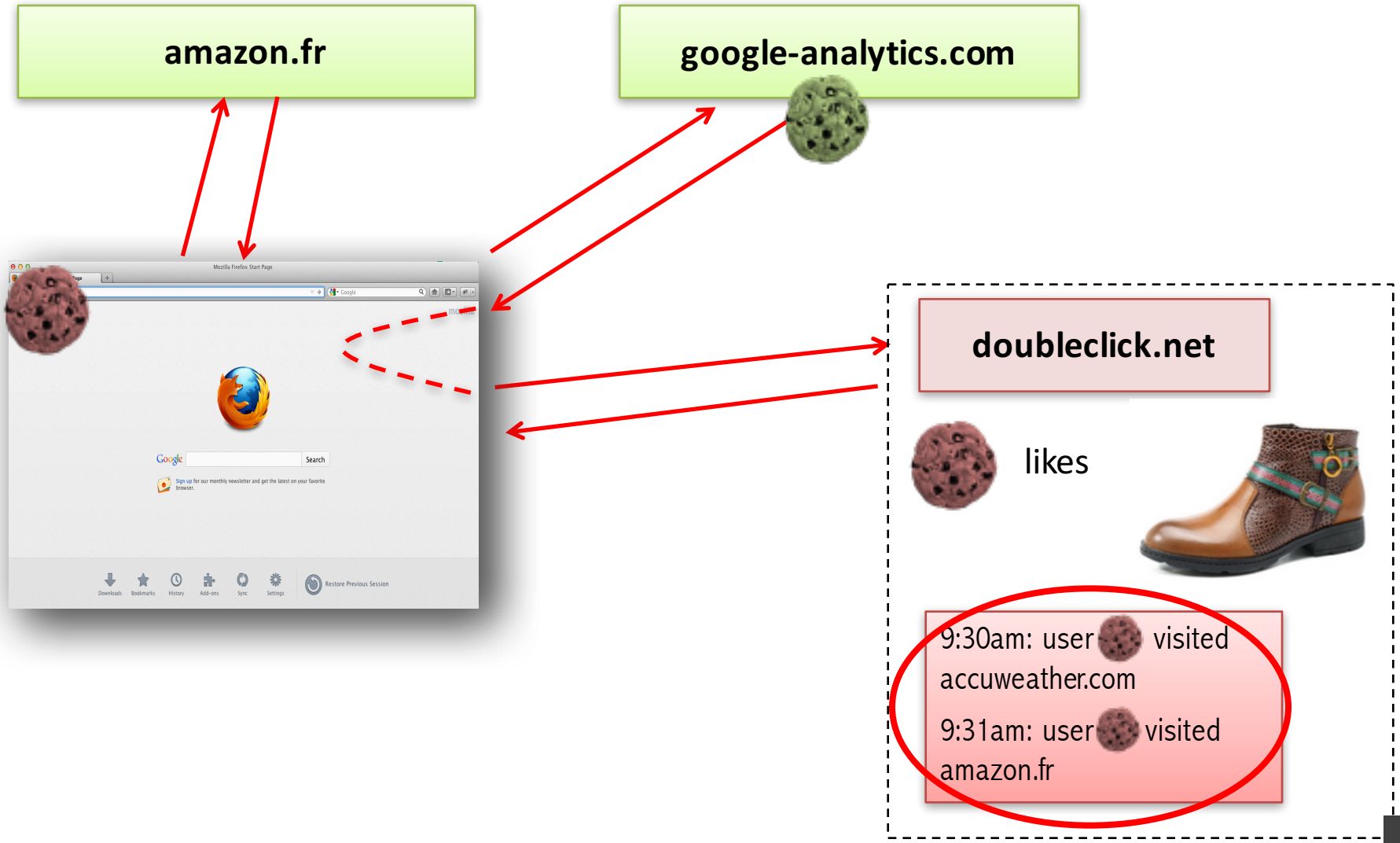
Demographics and interests data may only be available for a subset of your users, and may not represent the overall composition of your traffic: Analytics cannot collect the demographics and interests information if the DoubleClick cookie or the Device Advertising ID is not present, or if no activity profile is included.

Demographics and Interests

- About Demographics and Interests
- Enable Demographics and Interests reports
- Analyze Demographics and Interests data

See https://support.google.com/analytics/answer/2799357?hl=en

Google Analytics extended service



One change in Google Analytics setting…

- **Important detail**:
  - Silent redirection requests make it impossible to notice
  - Analytics company could offer to change settings any time without explaining implications to users' privacy!

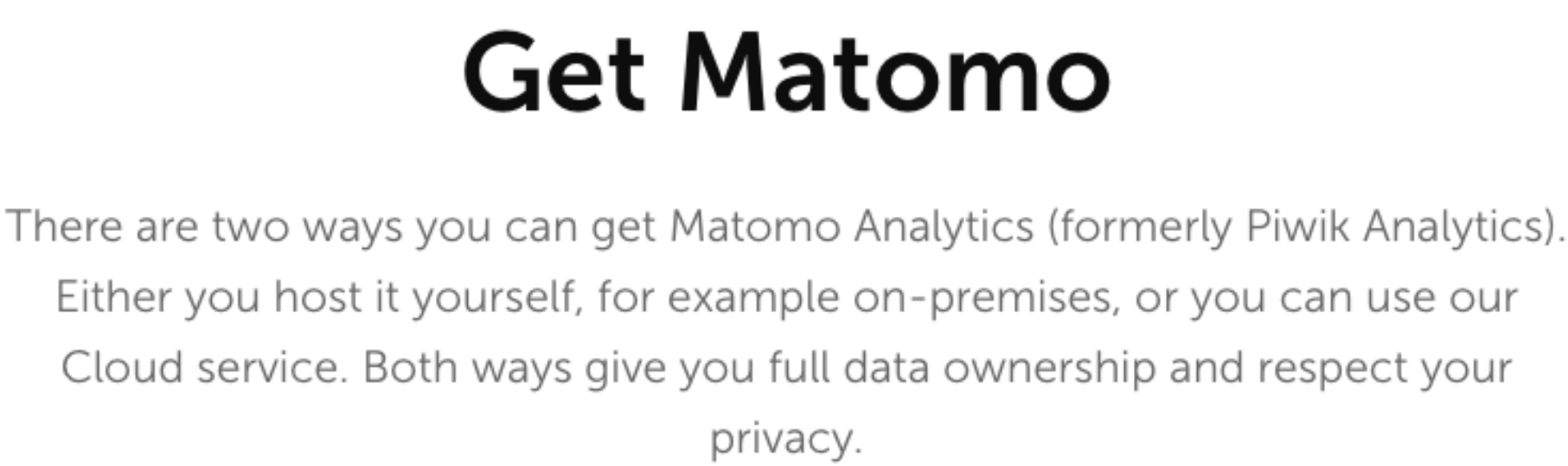
- **Privacy problem:**
  - A within-site analytics script enable cross-site tracking with one redirection request!
  - Works in Internet Explorer/Edge even if third-party cookies are disabled!

Should I notify my users?

- If I change the Google Analytics setting and redirect to doublecliek, should I notify the user?

Matomo (ex. Piwik) analytics

- Can be hosted on your own domain ☺
- Requires installation on your domain ☹



Source: https://matomo.org/download/