

Online Tracking, What Can Be Done About it, and Who's Doing it

Pete Snyder

Senior Privacy Researcher, Director of Privacy

pes@brave.com



Hi, I'm Pete 🙌

- **Grew up in Chicago**
...actual Chicago
- **Law school -> freelance web stuff**
Started: Anchorage, AK
Ended: Judge Judy Show invitation
- University of Illinois at Chicago



Me at Brave

- **Research at Brave**
...privacy, blocking, reliability
- **Co-Chair of PING**
Privacy committee on W3C
- **Research <-> Engineering**
Web compat, filter lists, etc.
- **Academic <-> Industry Collaborations**



Brave in a Slide

- **Privacy focused**
- **Alternative web funding model**
Fix incentive problems
- **Research + Engineering**
- **Not just a browser**
 - search.brave.com
 - talk.brave.com
 - VPN
 - more coming...



Overview

- **Why Privacy Matters**
A sloppy manifesto
- **Defining Tracking**
Abstracting the problem
- **Tracking in Practice**
Methods and defenses
- **Privacy Beyond Tracking**
Other issues and concerns

Overview

- **Why Privacy Matters**
A sloppy manifesto
- **Defining Tracking**
Abstracting the problem
- **Tracking in Practice**
Methods and defenses
- **Privacy Beyond Tracking**
Other issues and concerns



Why Does Tracking Exist?





**MAGAZINE
SUBSCRIPTIONS!**



Welcome to The "First" Banner Ad

Yes, this site is supposed to look this way. After all, this is what most web pages looked like back on October 27, 1994 -- the day that Wired Magazine flipped the switch on its first website, hotwired.com, starting a revolution in web content and advertising that still reverberates today.

This site is dedicated to showing off one of the ads that ran on that site. No, it wasn't the "first" as there were a handful of other ads that ran on various sections of hotwired.com. This site is also here to tell the story of how that ad came to be, how it succeeded beyond anything we had imagined, and how we tried to set an example for how corporations could communicate with their audiences.

This site launched on October 27, 2014. It is being constantly updated, so please check back again soon for more. In the meantime, get started by clicking your mouse in the banner ad above explore these other options:



[Previous](#) | [Next](#) | [Random](#) | [List Sites](#)

Chicago Tribune: Chicago news, X +

← → X chicagotribune.com ☆ UD ⌂ :

SECTIONS SEARCH ENEWSPAPER WEATHER NEWSLETTERS BEST REVIEWS \$2 FOR 20 WEEKS SALE ENDS 11/4 LOG IN

CPS STRIKE IS OFFICIALLY ON AS CHICAGO TEACHERS UNION SAYS THERE IS NO LAST-MINUTE DEAL X

TOP LOCAL NEWS SOURCE OCTOBER 16, 2019

Chicago Tribune \$2 FOR 20 WEEKS ENDS 11/4 51°F

BREAKING NEWS SPORTS BUSINESS POLITICS OPINION ENTERTAINMENT

ADVERTISEMENT

HSBC

CPS STRIKE IS ON

CPS strike is officially on as teachers union, Chicago

SPECIAL SALE ONLY \$2 FOR 20 WEEKS Get stories that impact you

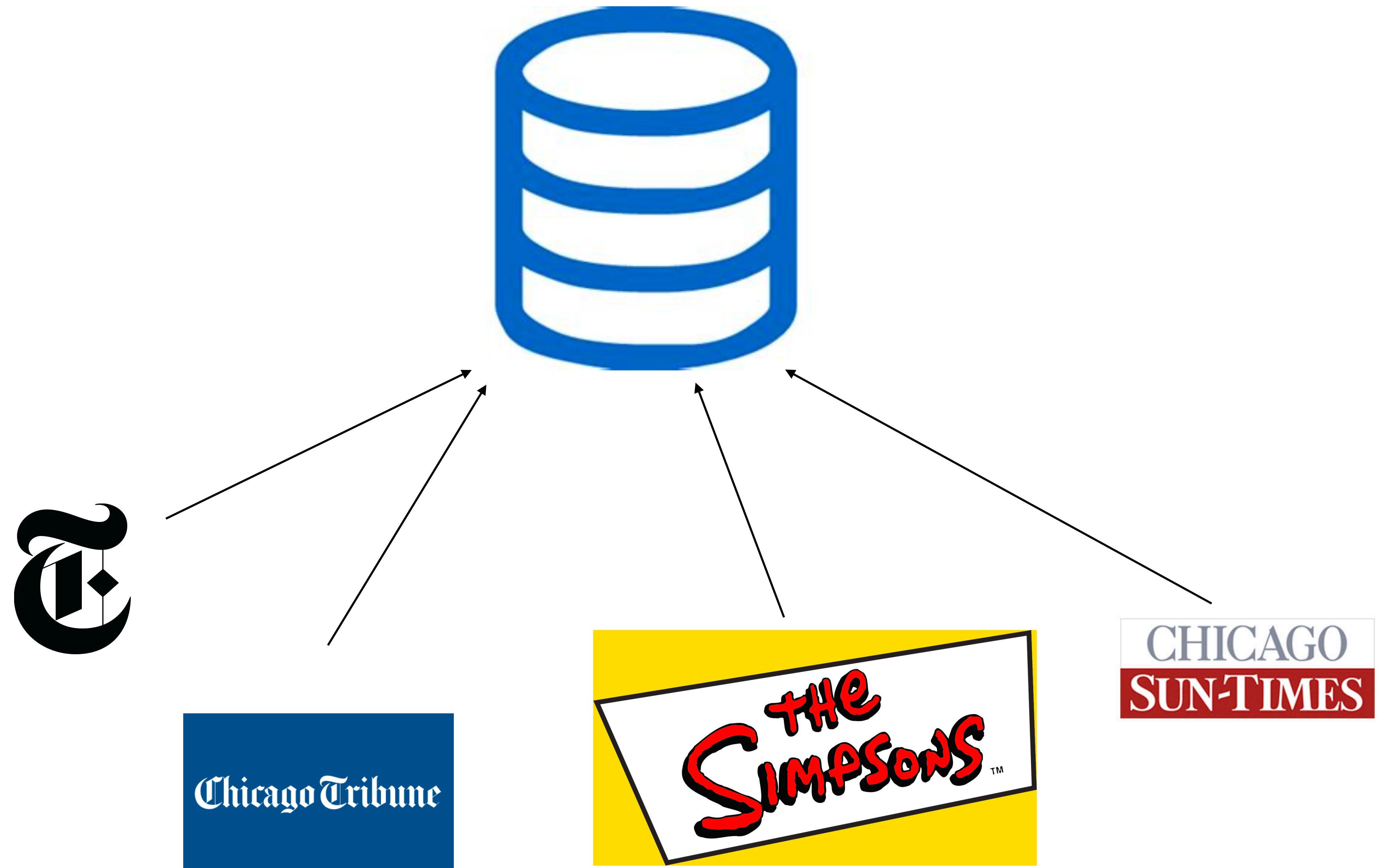
MORE CPS STRIKE COVERAGE >

CPS strike live updates: Chicago teachers reject city offer, will walk off job Thursday

Chicago Park District workers reach contract

SAVE NOW

Waiting for securepubads.g.doubleclick.net...



The World's Worst Website

Gratuitous use of frames is a common mistake of web designers.

Many older browsers do not support frames. They disrupt the flow of the website and can be difficult to anticipate where a page may appear when a link is clicked. [Click here](#) for an example of a frames page which is opening in the wrong window. Use your browser's 'Back' button to escape.

Check out these links to websites whose opinions about frames is self evident:

[The "I Hate Frames" Frames Page](#)

Another [I Hate Frames Page](#)

[The International I Hate Frames Club](#)

[Why Frames Suck \(Most of the Time\)](#)

The screenshot shows the homepage of 'The World's Worst Website'. At the top, there's a blue header bar with the 'Angelfire' logo and a link to 'Build your own FREE website at Angelfire.com'. Below the header, a banner for 'neboweb' features a small image of a vintage airplane and the text 'MARKETING IS MORE FUN WHEN YOU SMASH STUFF' with a 'VIEW THE WORK' button. A red oval highlights this banner area. The main title 'Welcome to the World's Worst Website!' is displayed prominently in yellow text. Below the title, a green callout box contains the text: 'Where am I and where are the links to other pages? An easy to use navigation structure is essential to any well designed website! Important information should never be more than 2 clicks away.' To the left of this box is a small icon of an eye. Further down, another callout box contains: 'Keep your backgrounds simple. White or light colors usually work best. Your background should not compete with the content of the page for the users attention. If you would like to use a background picture, select a picture that uses muted colors or format your picture as a watermark. Select text colors which will contrast well with the background picture.' At the bottom, there's a cartoon illustration of a character holding a bomb, with the text: 'Constantly running animations can be distracting when used excessively.' To the right of the character is a small image of a fire.



CHICAGO
SUN-TIMES

\$ \$ \$

¢ ¢ ¢





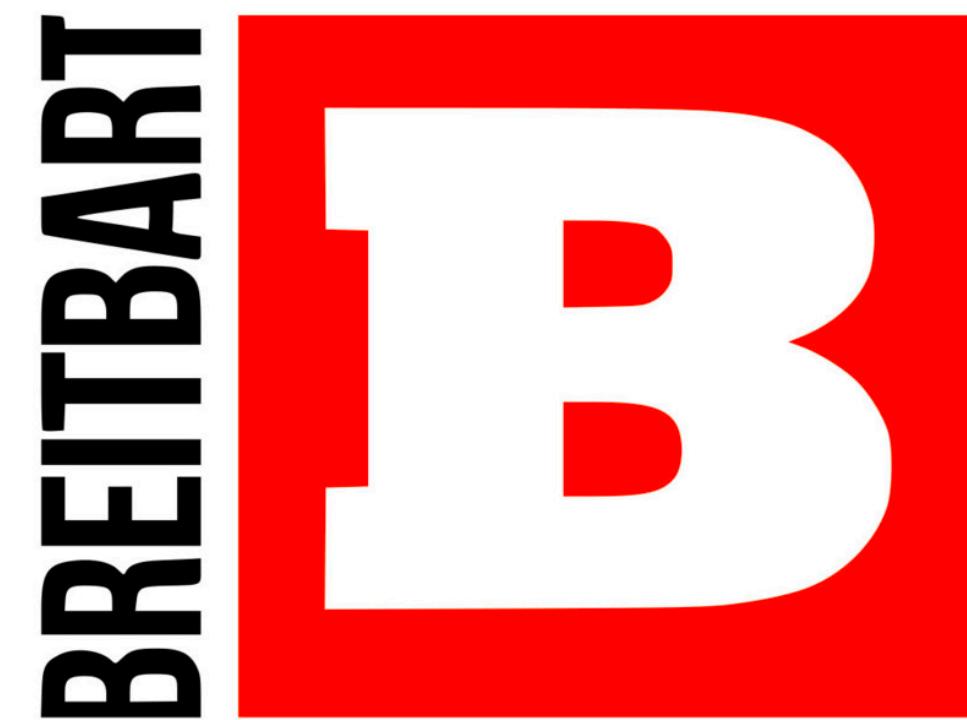
CHICAGO
SUN-TIMES

\$



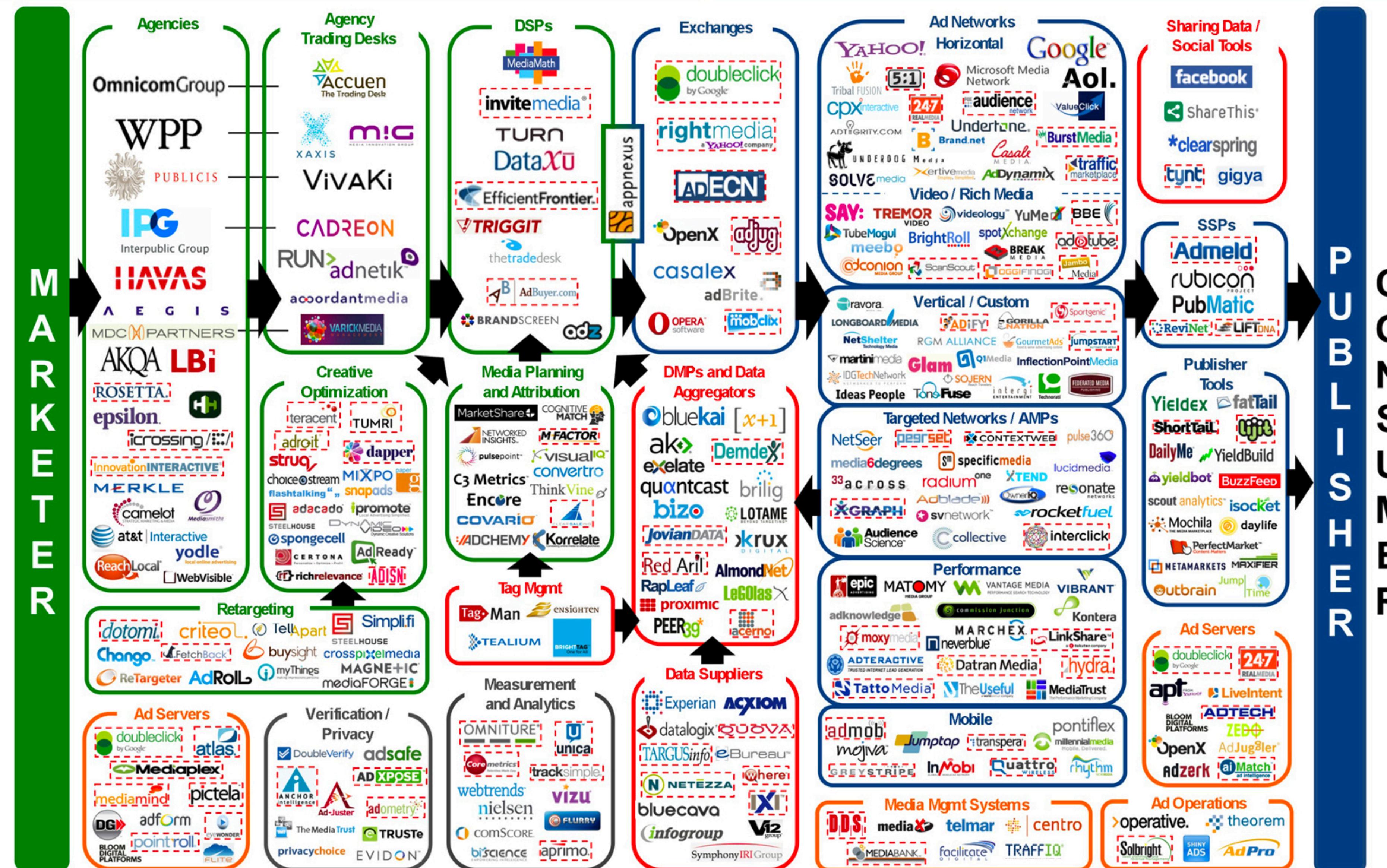


**Identify “expensive”
people here**



**Pay a little to advertise
to them here**

DISPLAY LUMAescape



Marketing Technology Landscape

The Martech 5000

Total Solutions 8,000

Advertising & Promotion

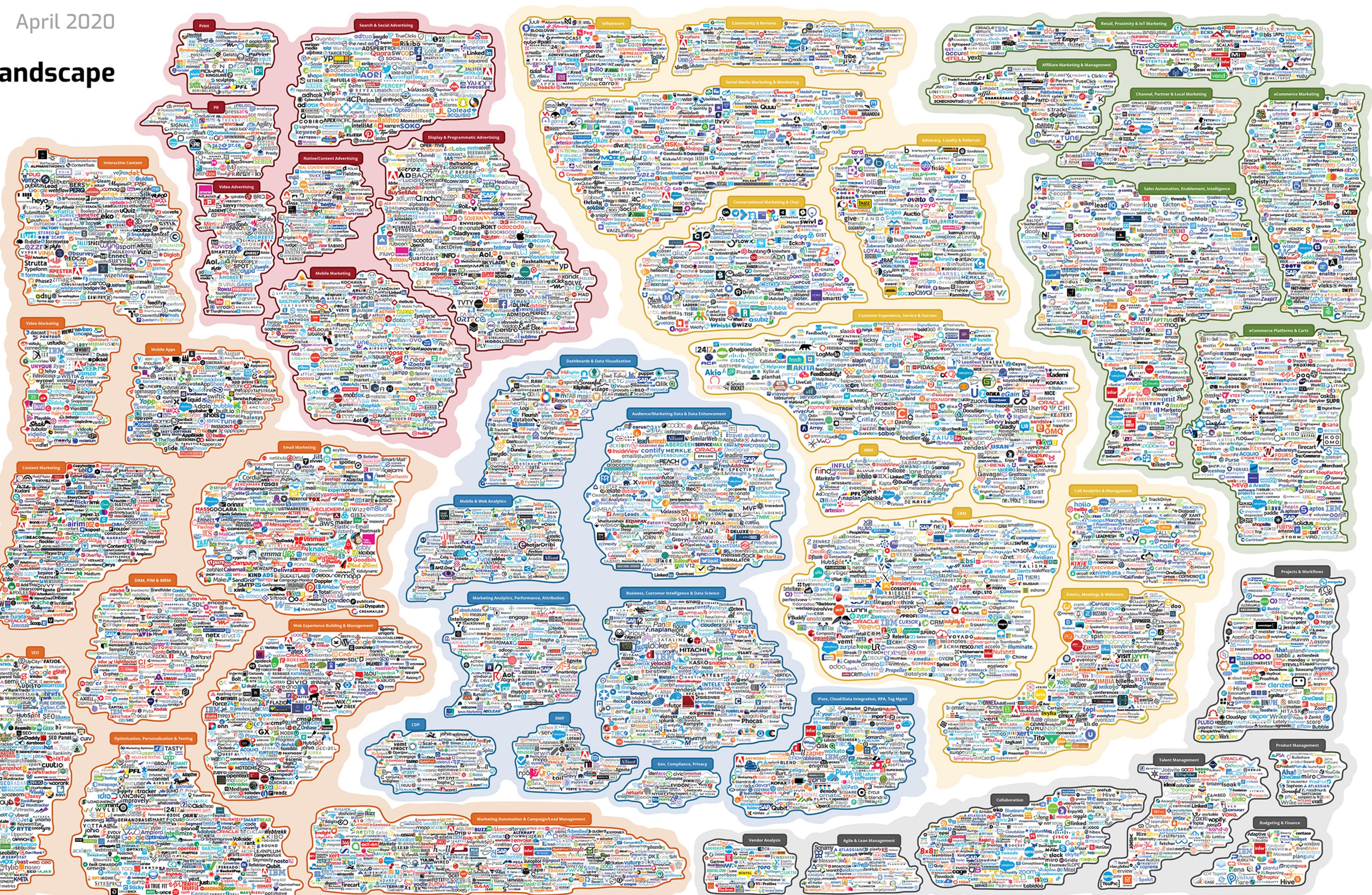
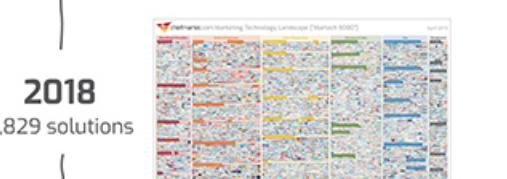
Content & Experience 1.9

Social & Relationships

Commerce & Sales

Data 1,258

Access all the data of this landscape & more at martech5000.com



Summarizing: Why Does Tracking Matter

- Incompatible with dignity
- Power and control
- Transfers wealth from value-creators to attention-attractors

Overview

- **Why Privacy Matters**

A sloppy manifesto

- **Defining Tracking**

Abstracting the problem



- **Tracking in Practice**

Methods and defenses

- **Privacy Beyond Tracking**

Other issues and concerns

Definitions

- **Website:** eTLD+1 (determined by public suffix list)
e.g., brave.com != mozilla.org
e.g., talk.brave.com == search.brave.com
e.g., ted.github.io != betty.github.io
- **Origin:** The full DNS host name serving a site
- **First-party:** Site of the top level document
- **Third-party:** any other site

Definitions (more)

- **DOM Storage:** Explicit storage APIs
e.g., cookies, localStorage, IndexDB
- **Network State:** All other storage
e.g., caches (v8, DNS, HTTP)
e.g., Header instructions (HSTS, ALT-SRV, etc)
- **Online Tracking**
Its trickier...

A Rough Definition of Tracking

- **Linking activities...**
e.g., being “followed”
- **across boundaries...**
e.g., temporal, geographic,
conceptual
- **In a way not expected or desired.**
e.g., ignorance or non-consent

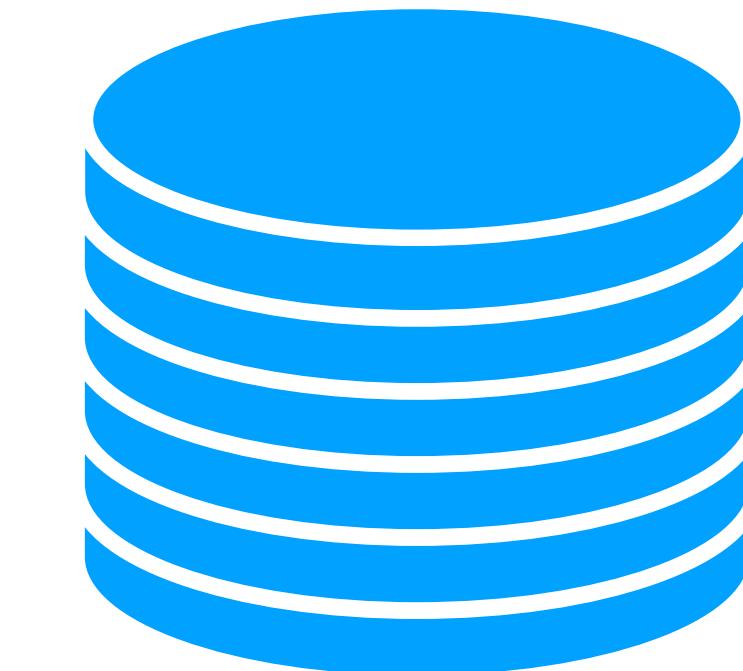
Tracking in Context



A Rough Definition of Tracking

- **Linking activities...**
e.g., being “followed”
- **across boundaries...**
e.g., temporal, geographic,
conceptual
- **In a way not expected or desired.**
e.g., ignorance or non-consent

Question One

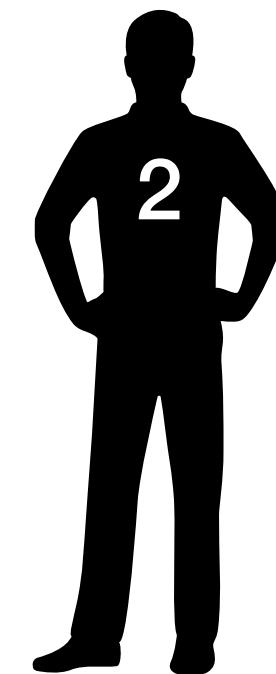
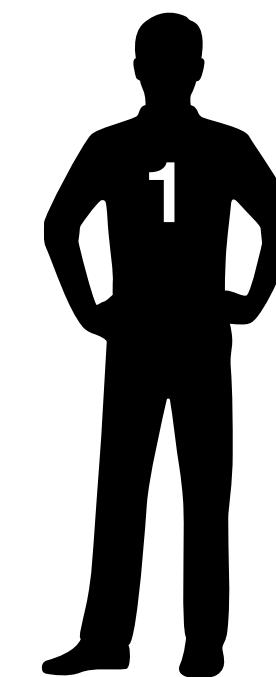


some-site.example

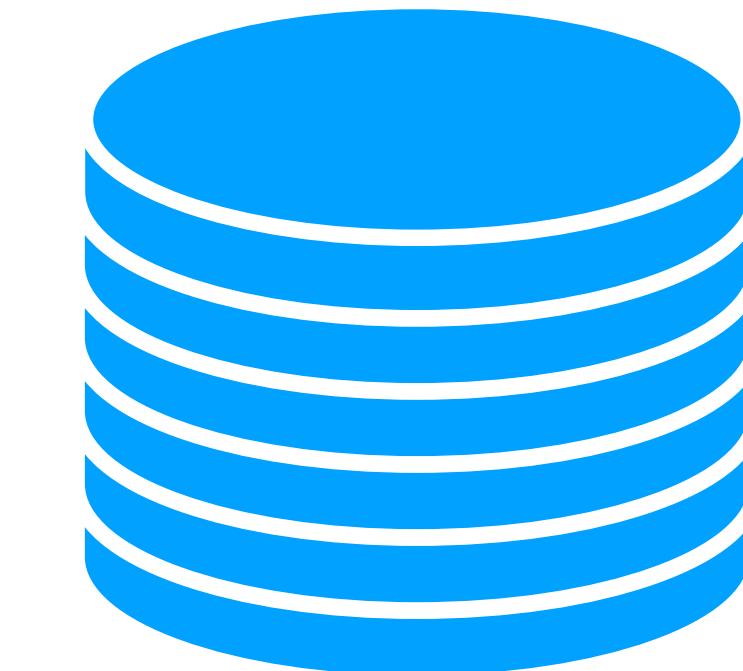
One day...



other-site.example



Question One

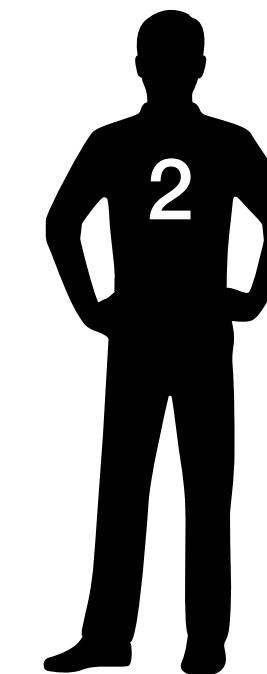
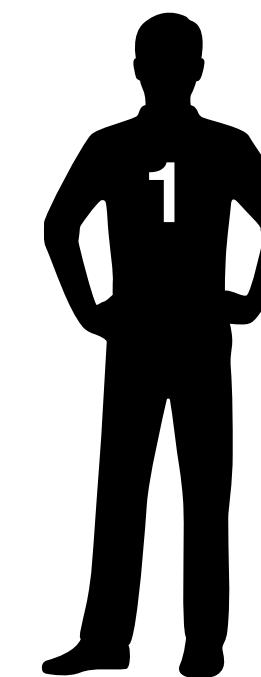


some-site.example

One day...



other-site.example

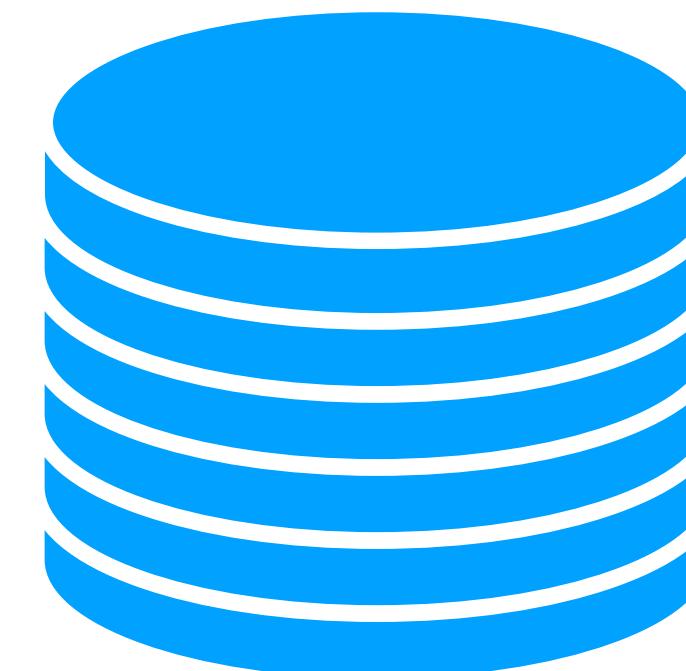


Question Two

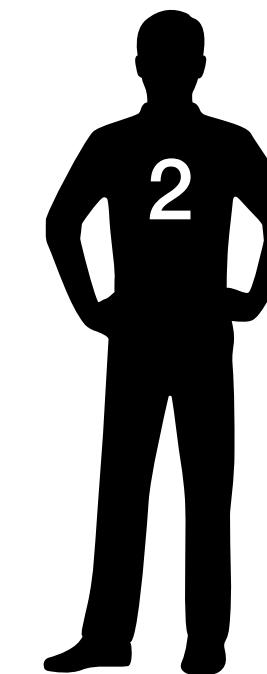
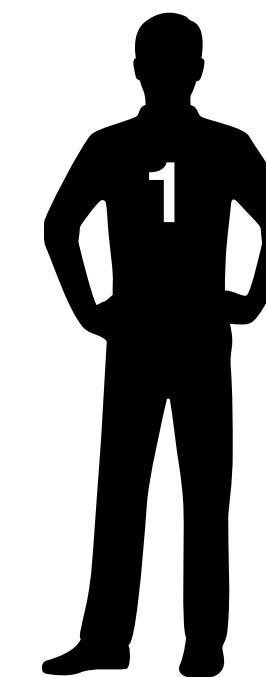


some-site.example

One day...



some-site.example

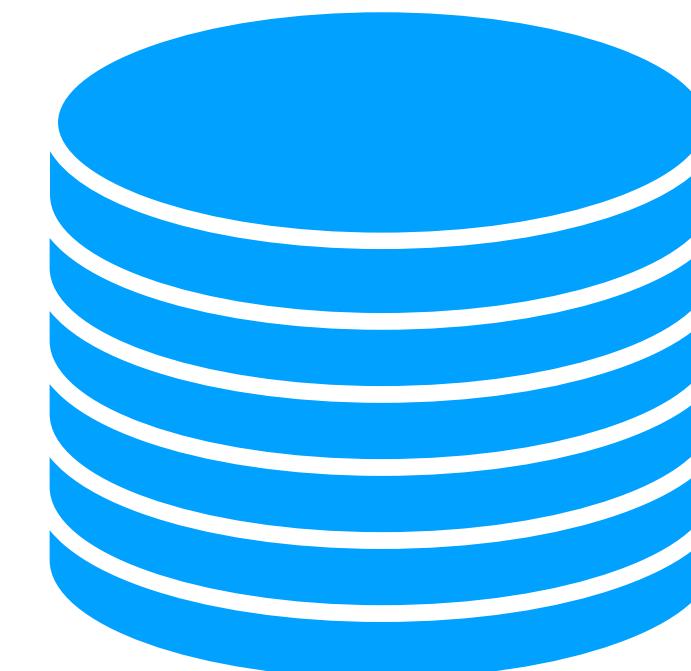


Question Two

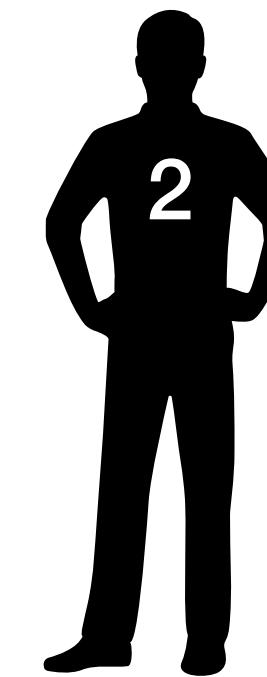
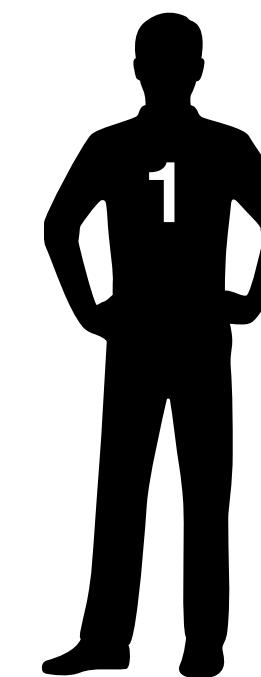


some-site.example

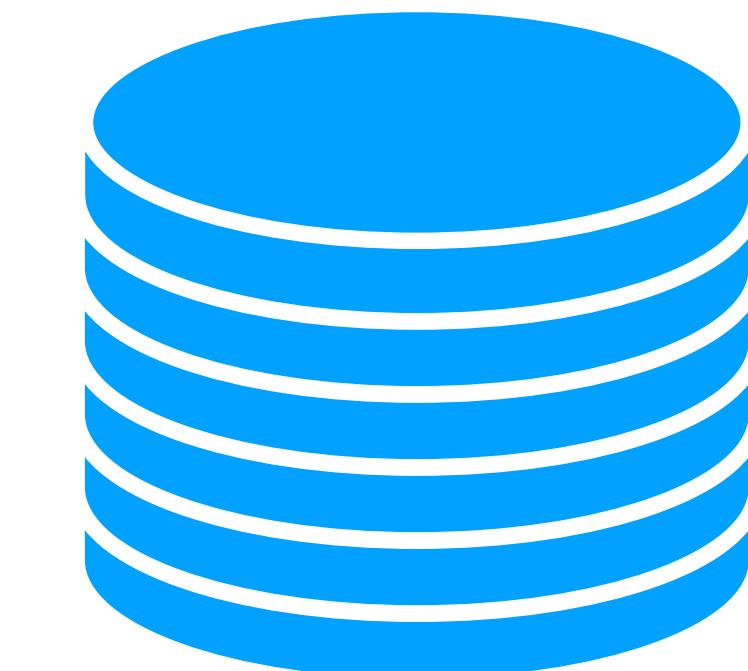
One day...



some-site.example

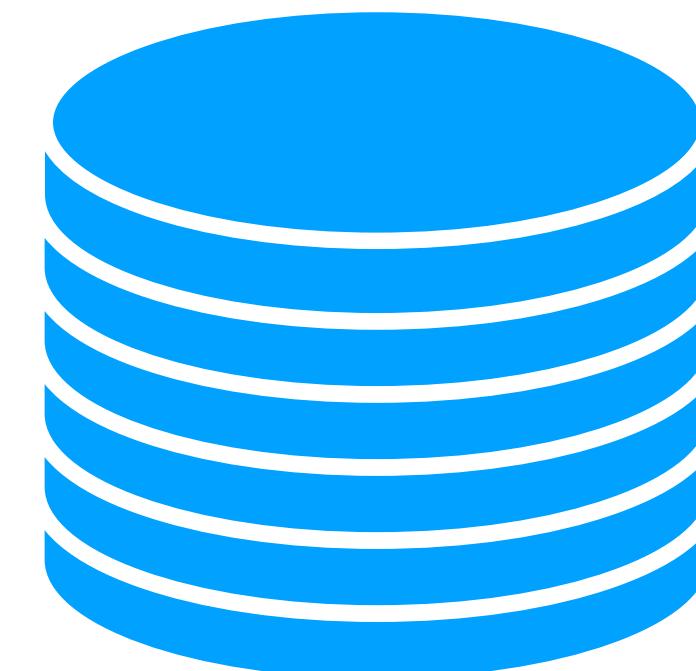


Question Three

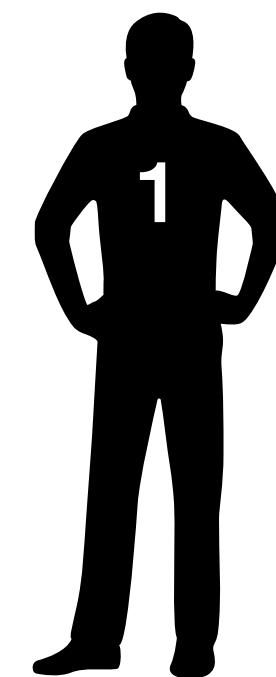
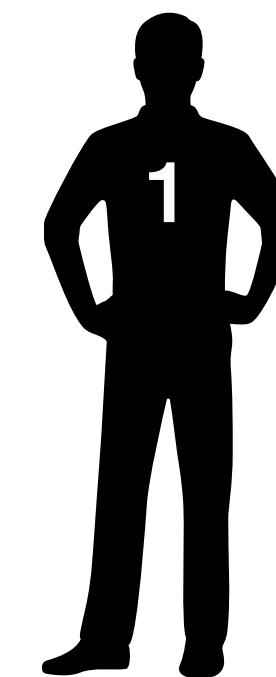


some-site.example

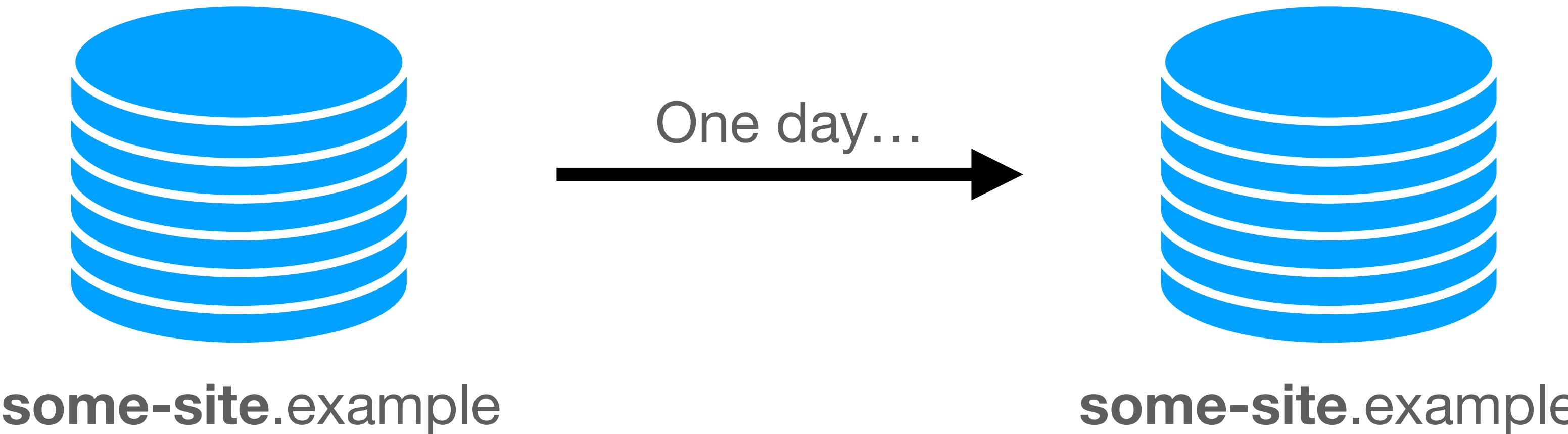
One day...



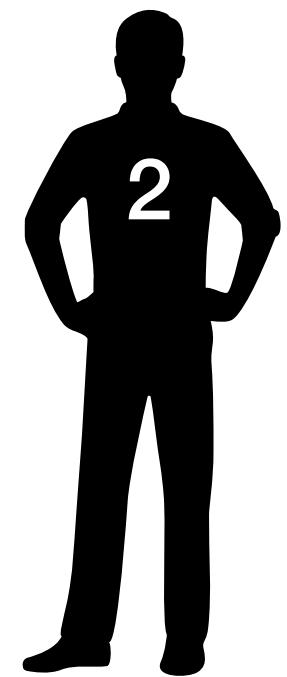
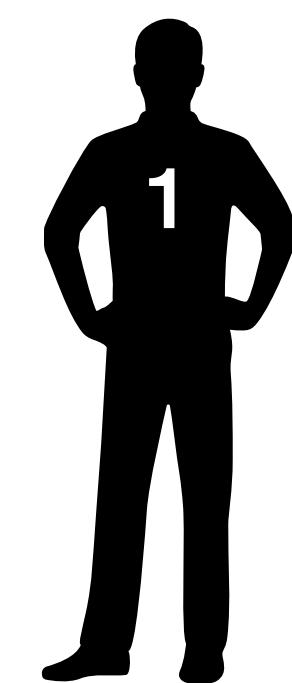
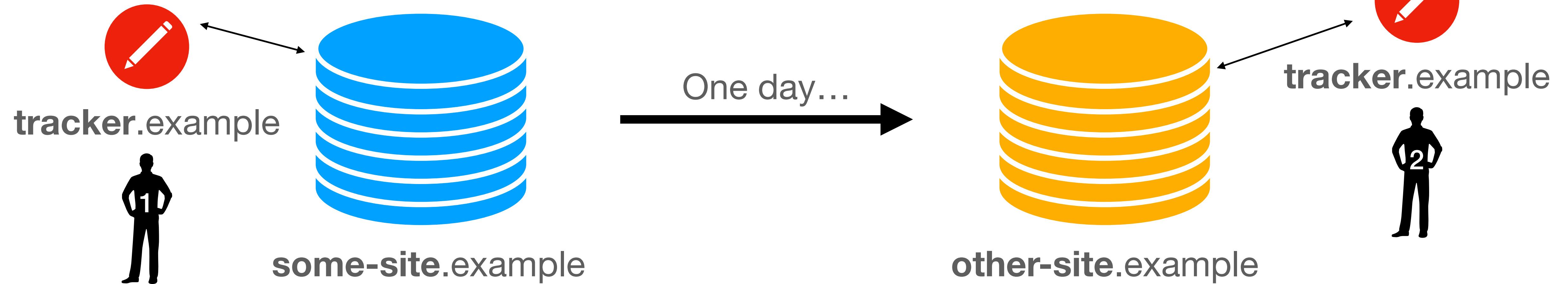
some-site.example



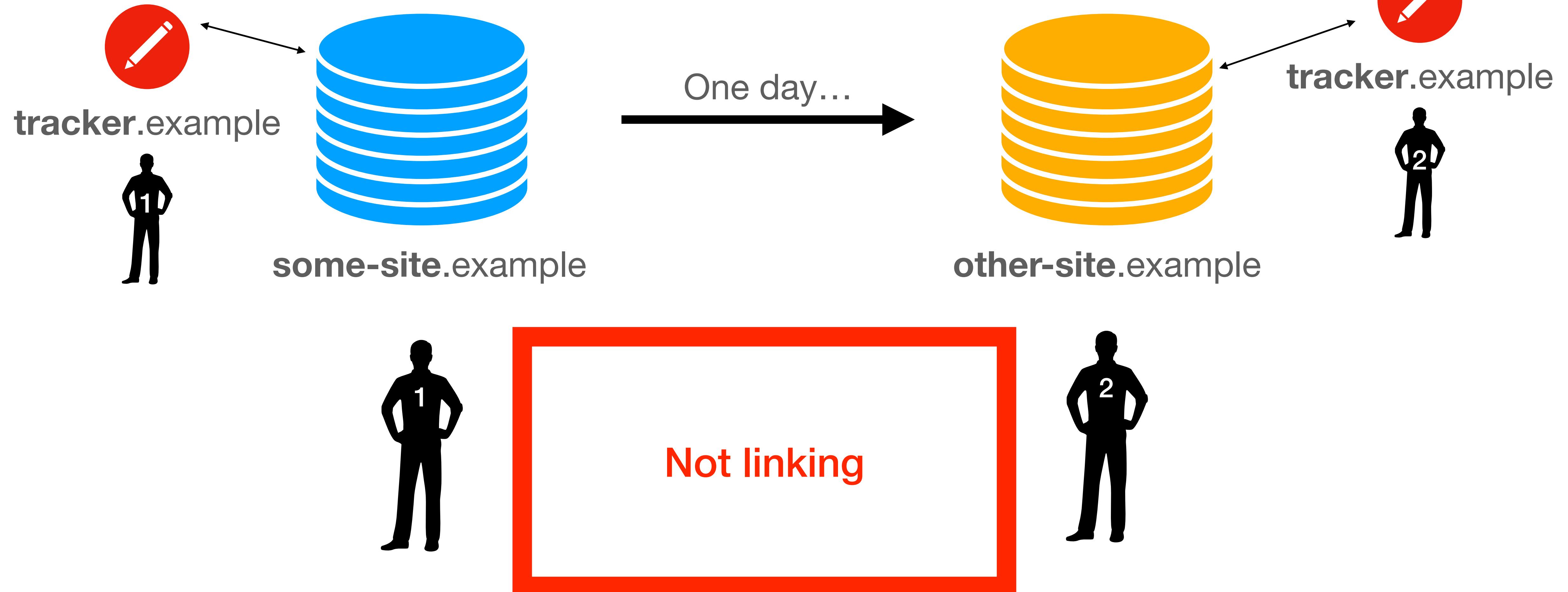
Question Three



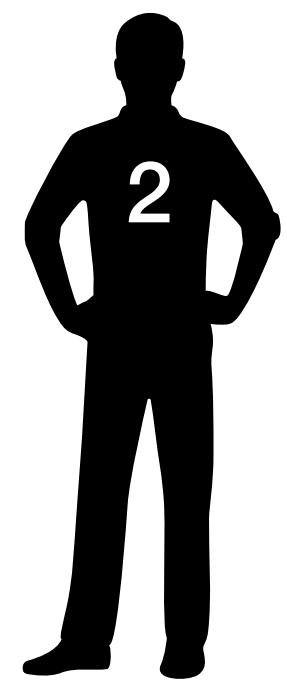
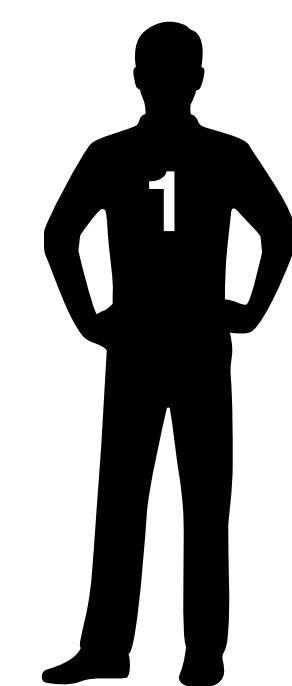
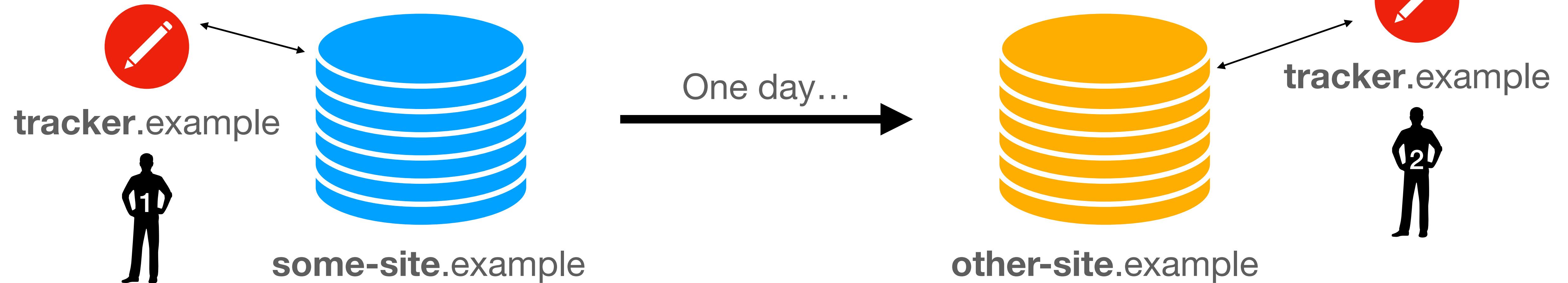
Question Four



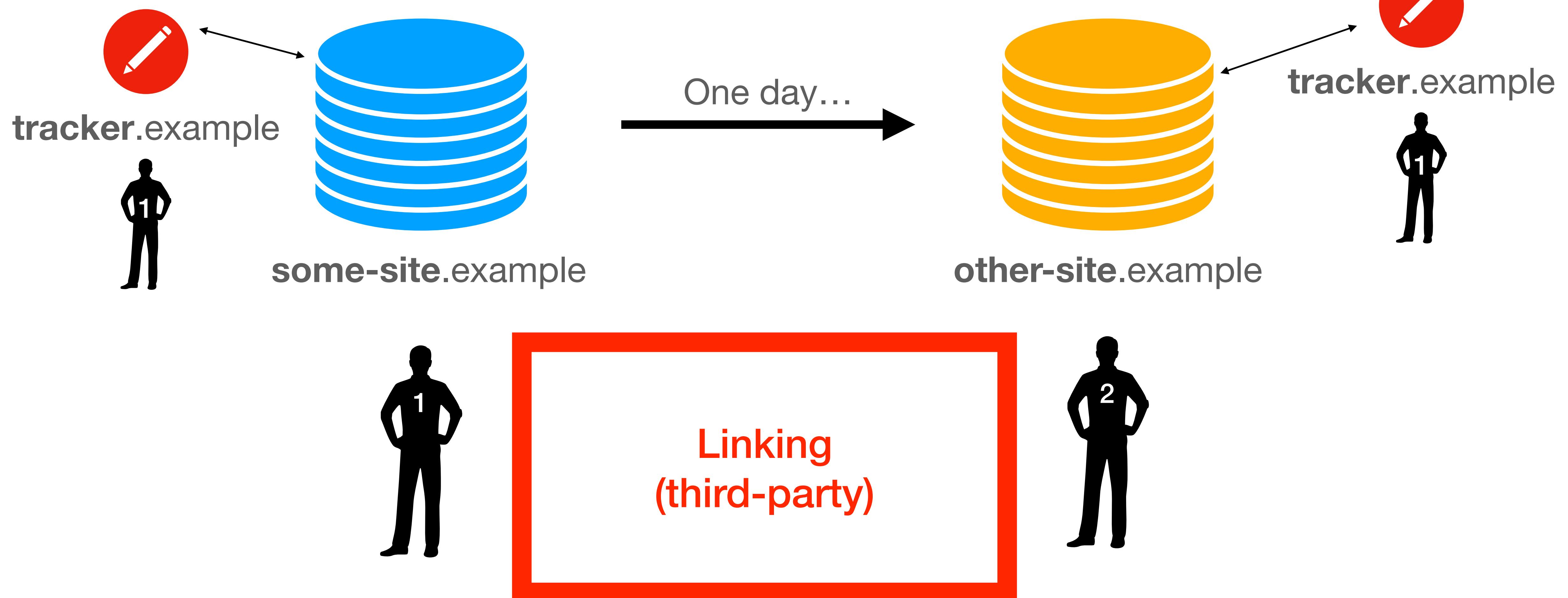
Question Four



Question Five



Question Five



Tracking: Linking...

- **Tying behaviors to same identity**
Could be pseudonymous, or a “real world” identity
- **Probabilistic or deterministic**
For some definition of “probable enough”

A Rough Definition of Tracking

- **Linking activities...**
e.g., being “followed”
- **across boundaries...**
e.g., temporal, geographic,
conceptual
- **In a way not expected or desired.**
e.g., ignorance or non-consent



Tracking: ...across boundaries...

- **Organizational boundaries**
e.g., eTLD+1, origin, “first-party set”
- **Temporal boundaries**
e.g., tying something done last year to something done today
- **Profile boundaries**
e.g., private browsing, different browsers, accounts

A Rough Definition of Tracking

- **Linking activities...**
e.g., being “followed”
- **across boundaries...**
e.g., temporal, geographic,
conceptual
- **In a way not expected or desired.**
e.g., ignorance or non-consent



Tracking: ...expectations

- **Expectations differ across platforms**
Facebook inapp browser vs Tor Browser Bundle
- **Expectations differ across people's expertise**
e.g., my dad vs Dworkin
- **Expectations differ across backgrounds**
e.g., outlook.com vs microsoft.com vs github.com
- **Consent is (sometimes) fuzzy**
Terms of service <-----> Storage Access API

A Rough Definition of Tracking

- **Linking activities...**
e.g., being “followed”
- **across boundaries...**
e.g., temporal, geographic,
conceptual
- **In a way not expected or desired.**
e.g., ignorance or non-consent

Overview

- **Why Privacy Matters**

A sloppy manifesto

- **Defining Tracking**

Abstracting the problem

- **Tracking in Practice**

Methods and defenses



- **Privacy Beyond Tracking**

Other issues and concerns

Tracking Techniques

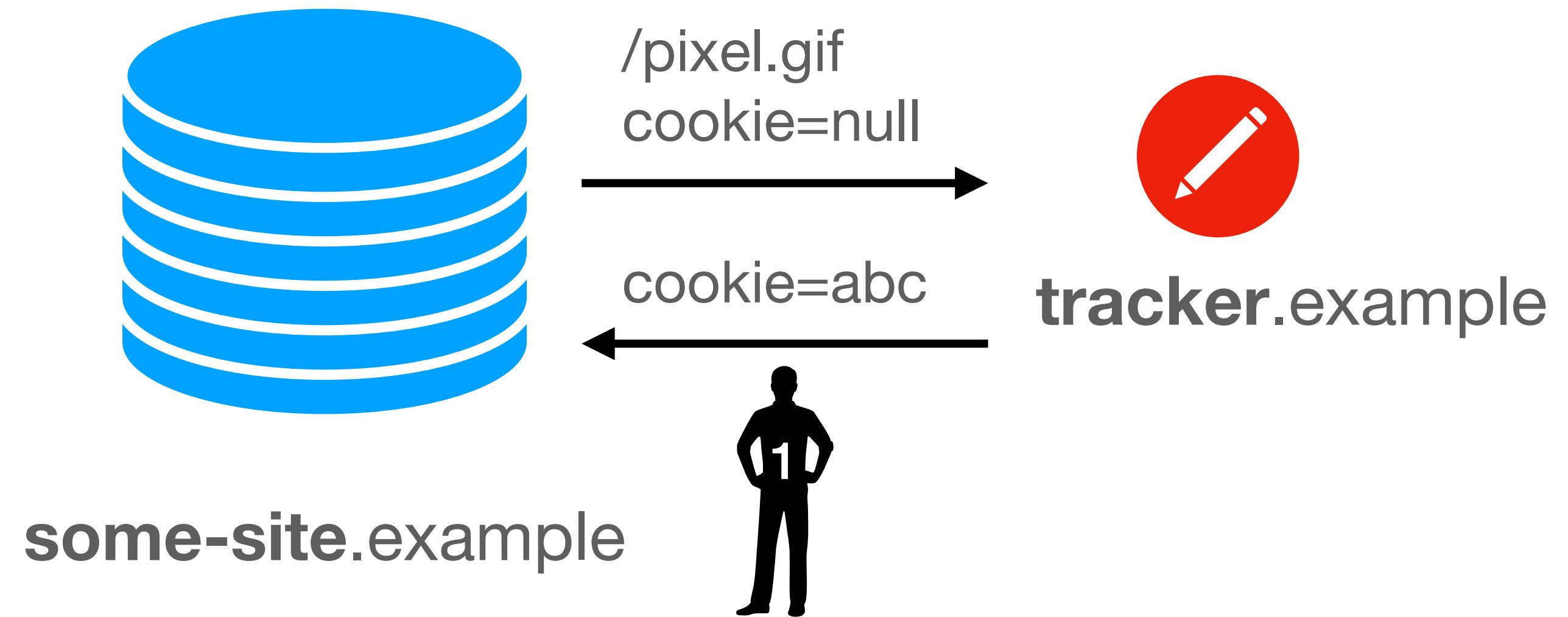
- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers

Tracking Techniques

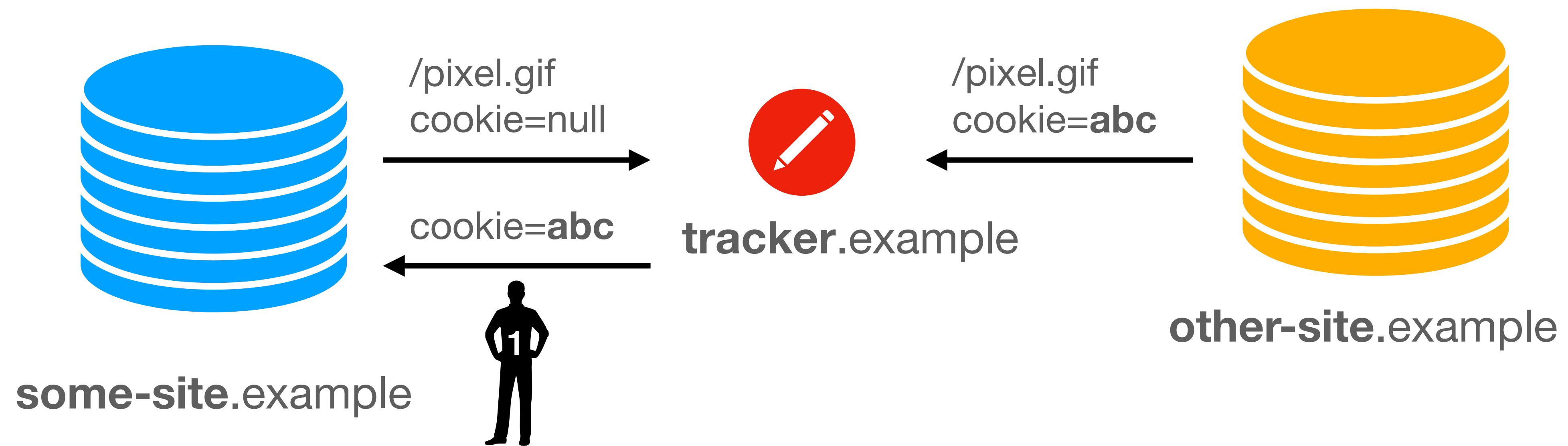
- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



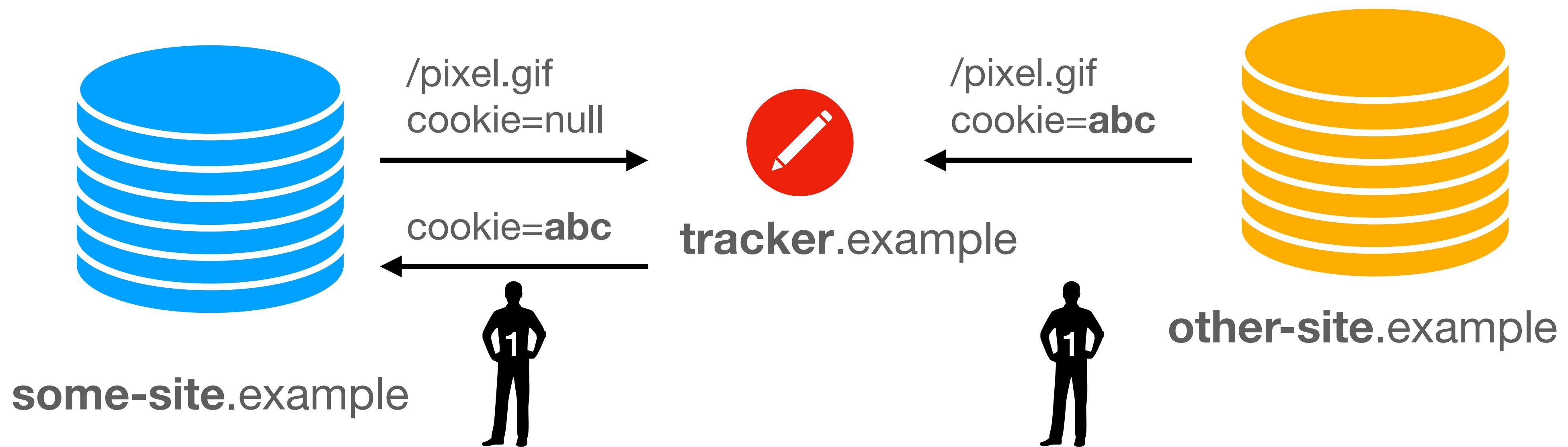
Third-party DOM storage



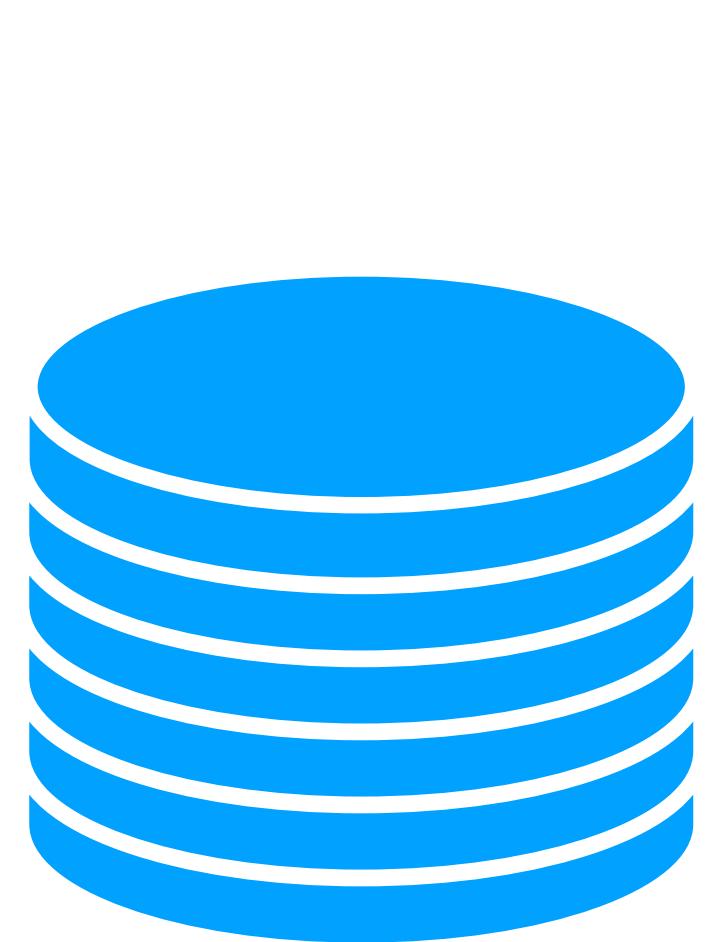
Third-party DOM storage: cookies



Third-party DOM storage: cookies



Third-party DOM storage: iframe



some-site.example

```
<iframe src="//tracker.example">

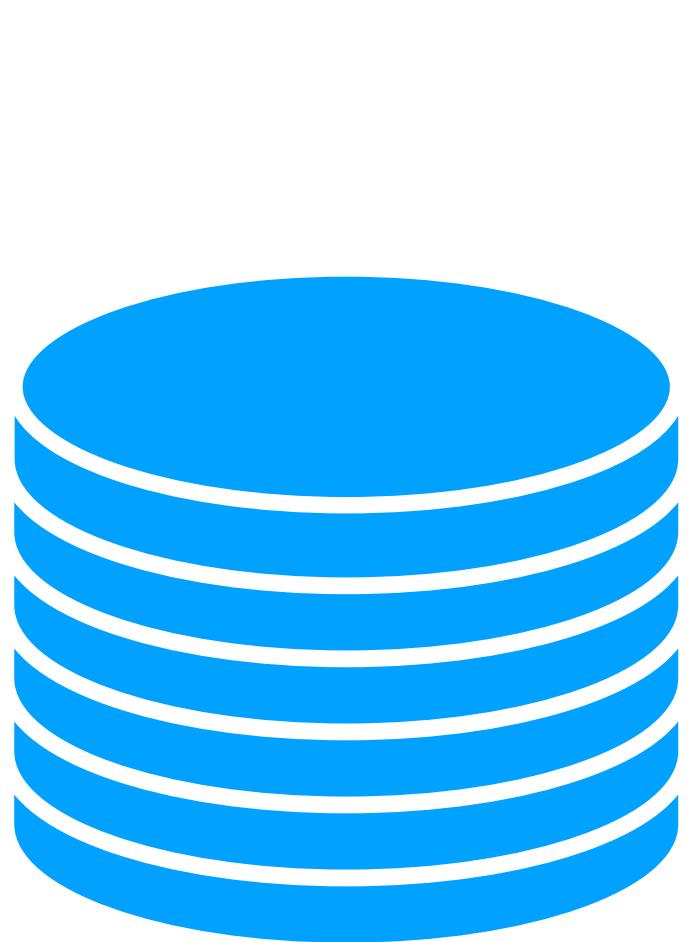
const LS = localStorage

if (LS['id']) {
    // I re-identified a person
} else {
    // new person, assigning ID
    LS['id'] = Math.random()
}

fetch(`/record?id=${LS['id']}`)

</iframe>
```

Third-party DOM storage: iframe



some-site.example

```
<iframe src="//tracker.example">

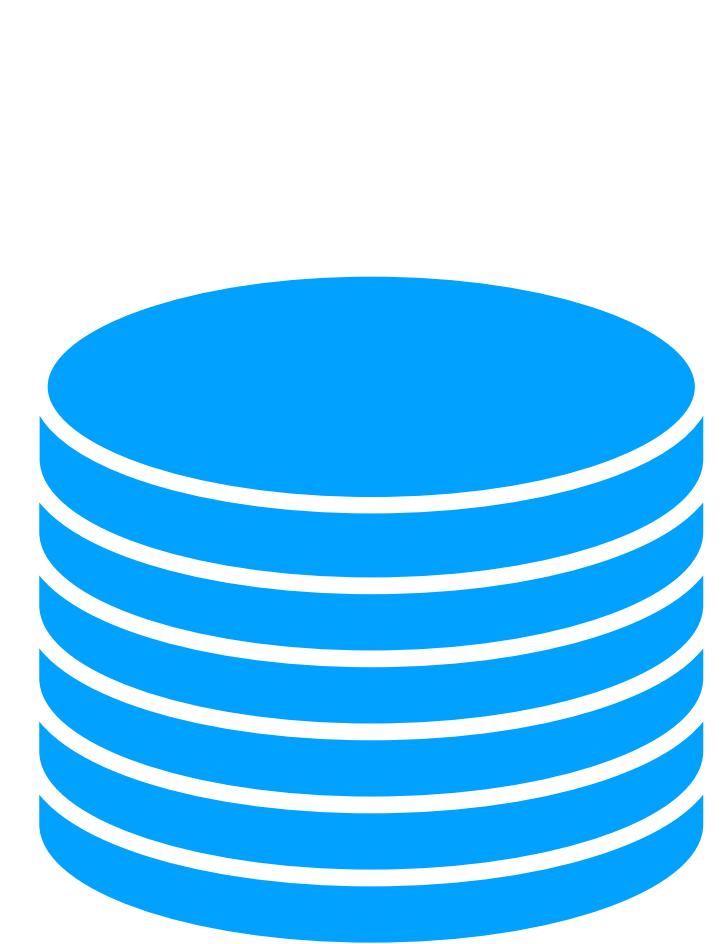
const LS = localStorage

if (LS['id']) {
    // I re-identified a person
} else {
    // new person, assigning ID
    LS['id'] = Math.random()
}

fetch(`/record?id=${LS['id']}`)

</iframe>
```

Third-party DOM storage: iframe



some-site.example

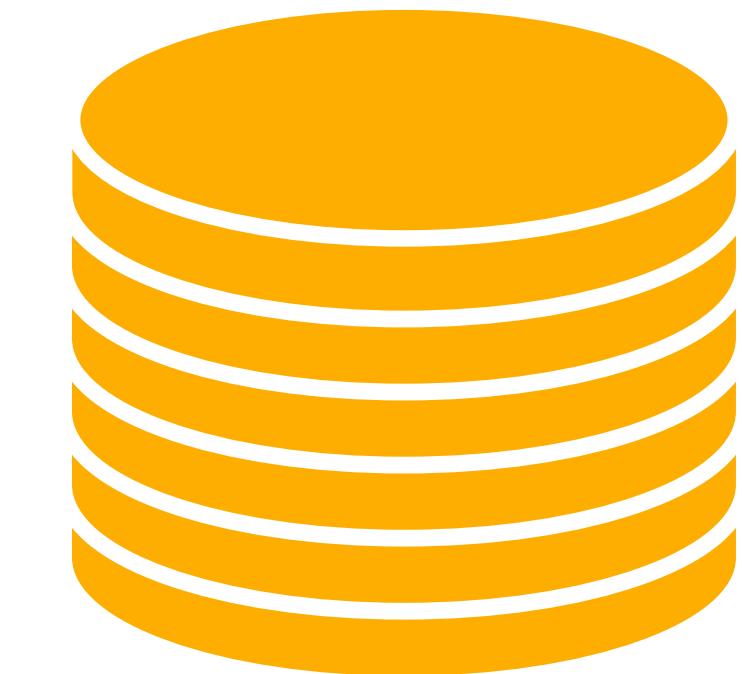
```
<iframe src=/tracker.example>

const LS = localStorage

if (LS['id']) {
    // I re-identified a person
} else {
    // new person, assigning ID
    LS['id'] = Math.random()
}

fetch(`/record?id=${LS['id']}`)

</iframe>
```



other-site.example

Third-party DOM storage: partitioning

- Third-party storage is not shared across sites

- Sometimes called “dual-keying”

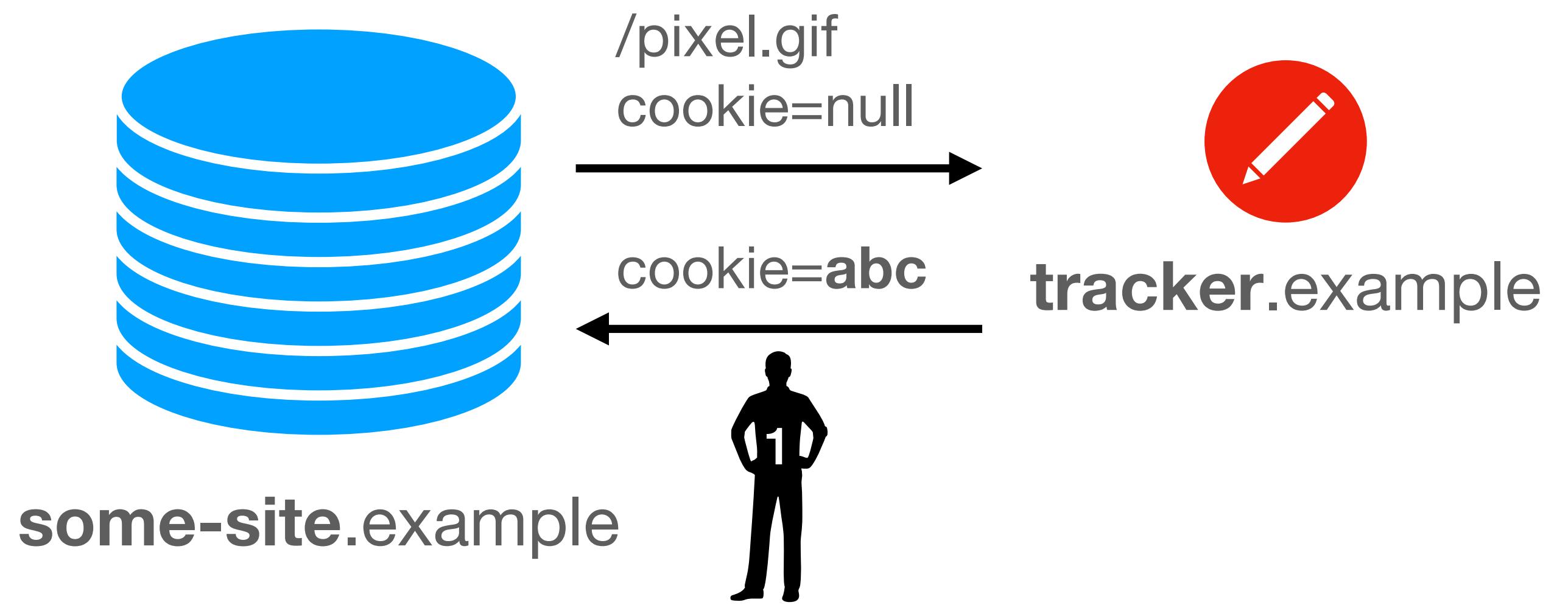
- Previous:

```
storage_data = browser_storage[<requested eTLD+1>]
```

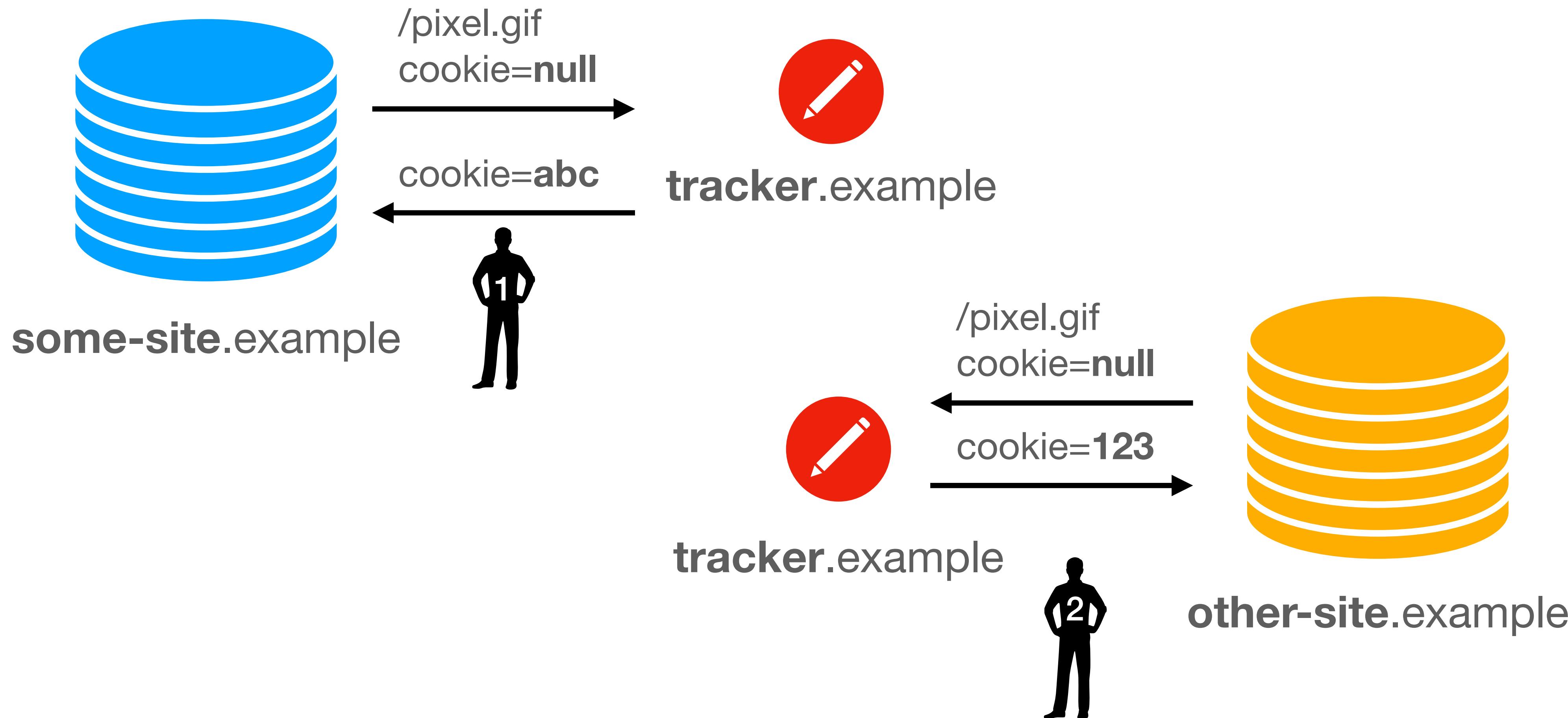
- Partitioning:

```
storage_data = browser_storage[<first-party eTLD+1>][<requested eTLD+1>]
```

Third-party DOM storage: partitioning



Third-party DOM storage: partitioning



Third-party DOM storage: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Block third-party cookies					
Partition storage					
Ephemeral partitions					
List based defenses					

Tracking Techniques

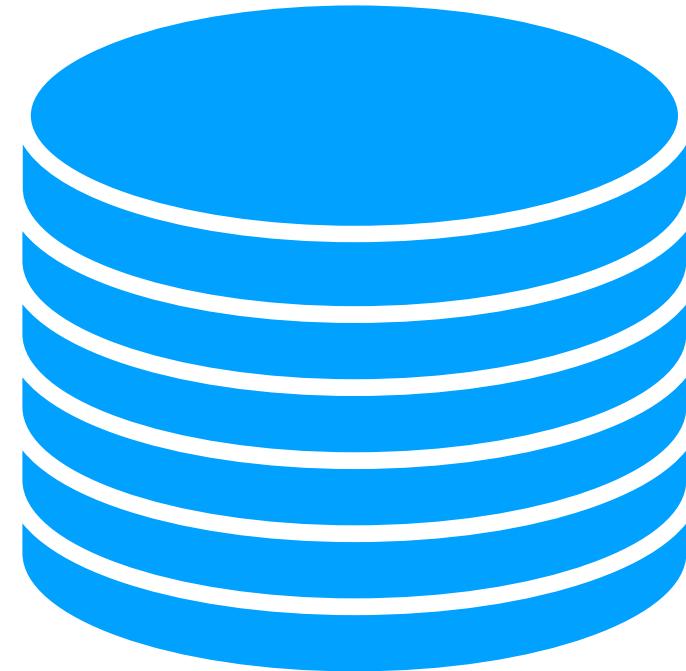
- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



Network State Example: HTTP Cache

- **Browsers cache things for speed**
Images, JavaScript, etc.
- **Caches are generally unpartitioned**
- **Anything unpartitioned can be a linking key**

HTTP Cache Tracking



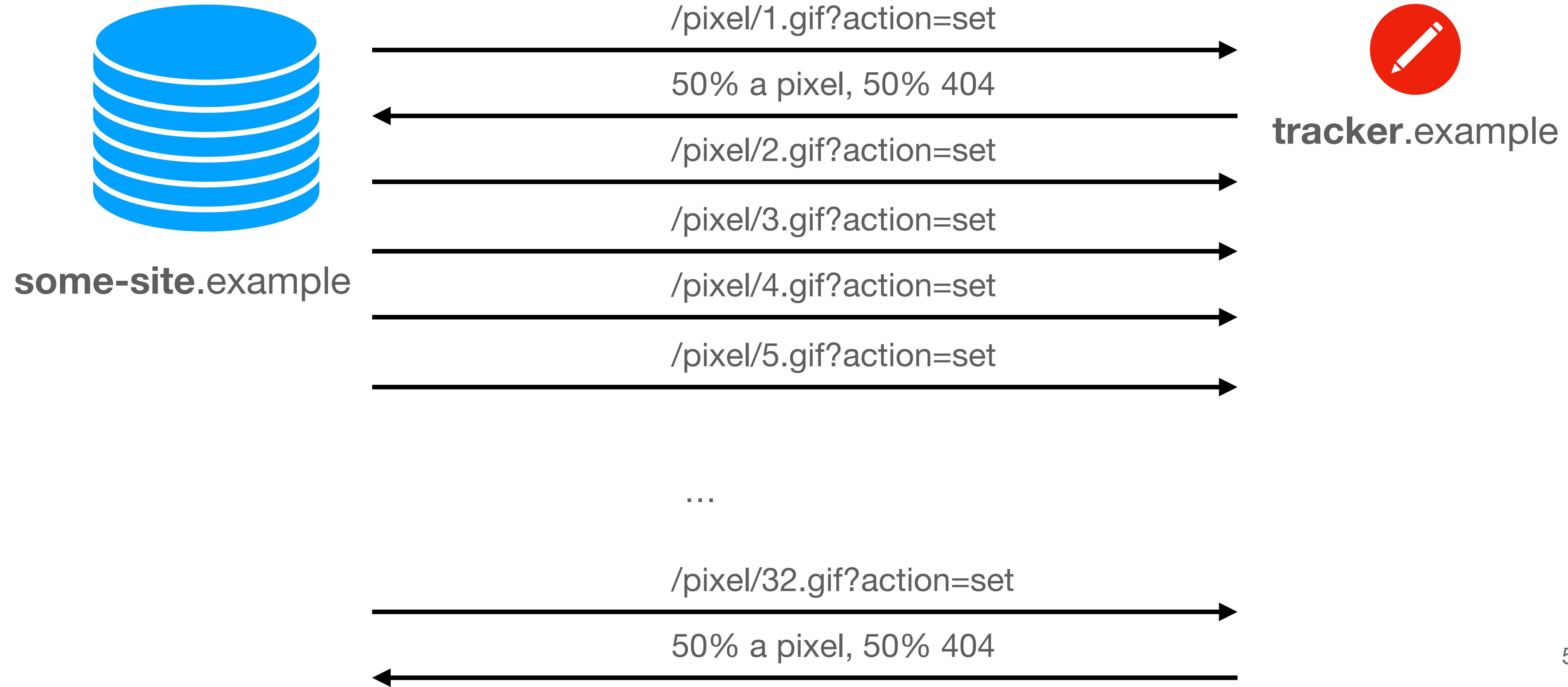
some-site.example



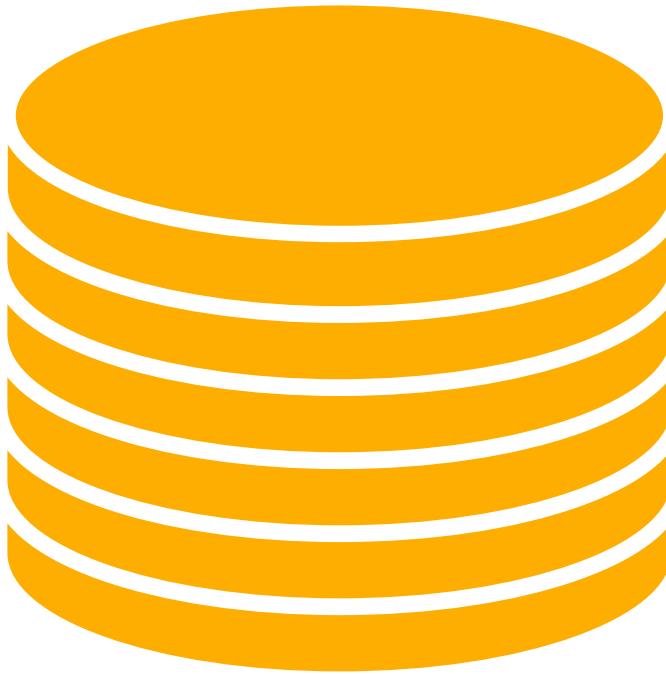
tracker.example

- /pixel/#.gif?action={set, read}
- action=set: 50% return pixel : 50% 404
- action=read: 100% 404

HTTP Cache Tracking



HTTP Cache Tracking



other-site.example

<https://tracker.example/script.js>

```
const identifier = []
for (let i = 0; i < 32; i += 1) {
  try {
    const url = `/pixel/${i}.gif?action=read`
    await fetch(url)
    // We hit the cache
    identifier[i] = 1
  } catch (_) {
    // We missed the cache
    identifier[i] = 0
  }
}
// identifier is now a unique 32 bits
```



tracker.example

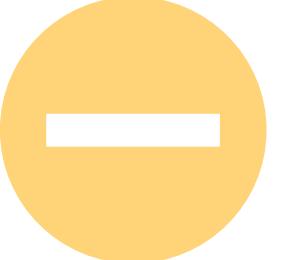
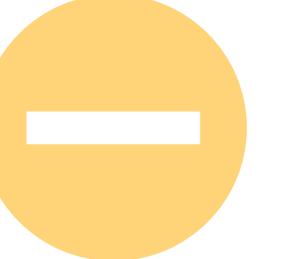
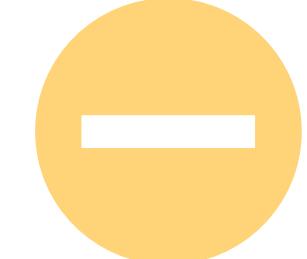
HTTP Strict Transport Security (HSTS)

- **Website Says “only HTTPS, forever”**
e.g. persistent storage
- **Automatic Upgrade**
`http://example.org -> https://example.org`
- **How to leverage?**

HSTS Tracking

- **example.org**
- **a.example.org**
- **b.example.org**
- **a.a.example.org**
- **b.a.example.org**

Network state: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Partition network state					
List based defenses					

Tracking Techniques

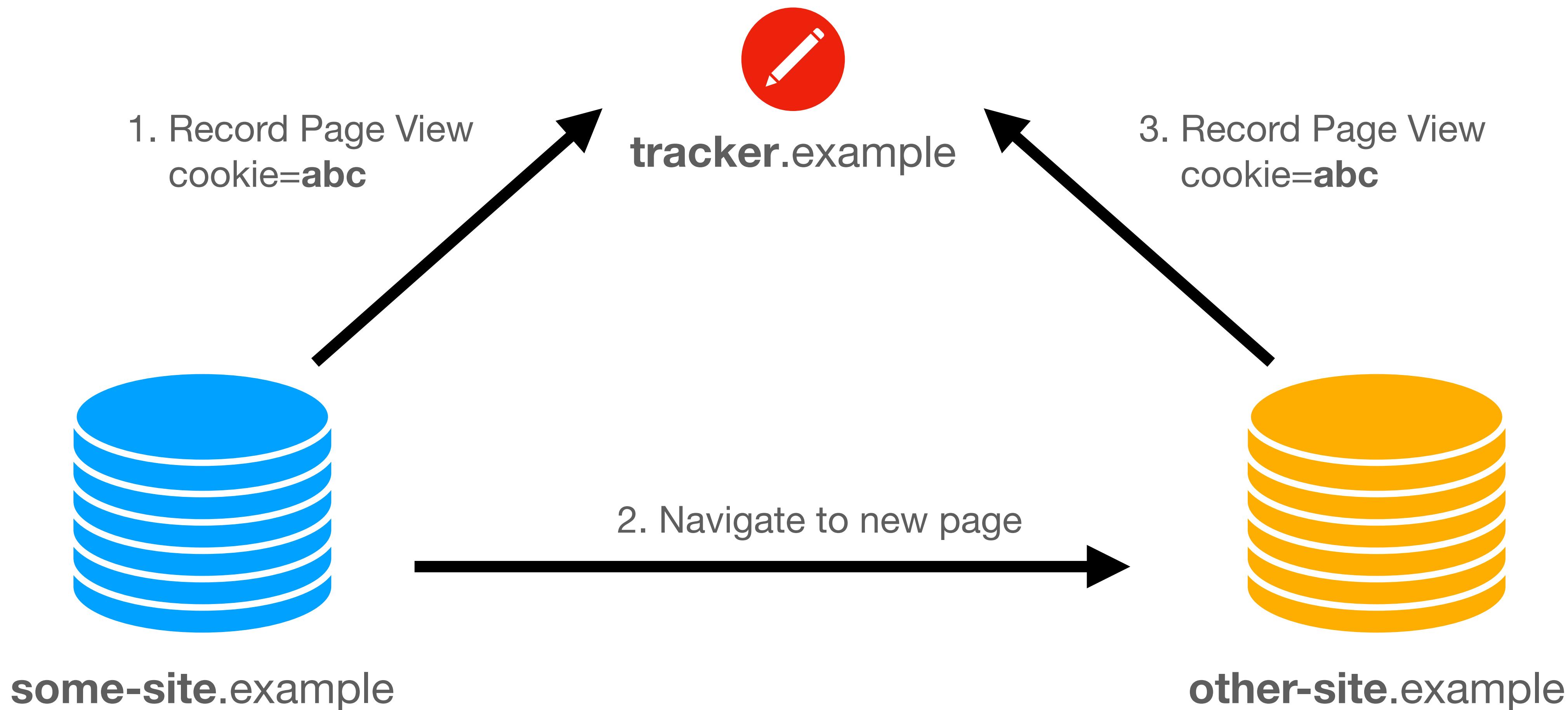
- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



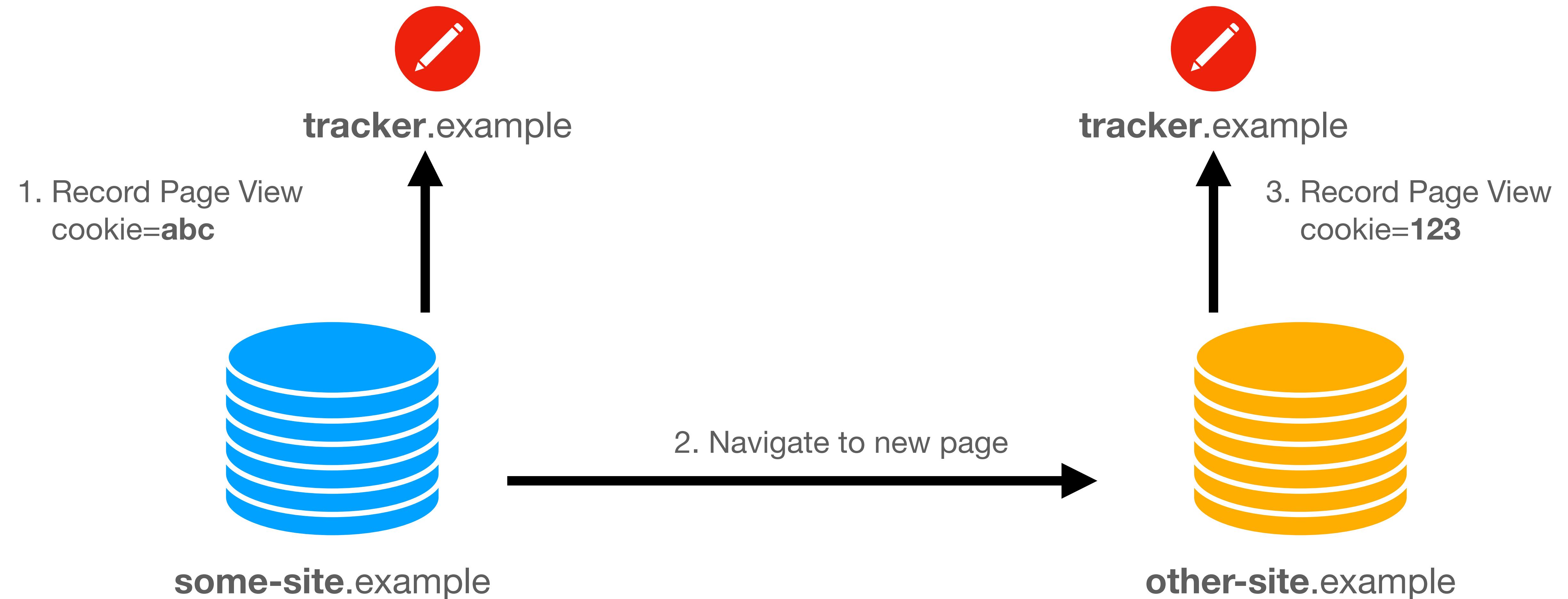
Bounce Tracking

- Response to partitioning
- Third parties use first-parties to track
- Growing in importance as partitioning is more common

Pre-partitioning



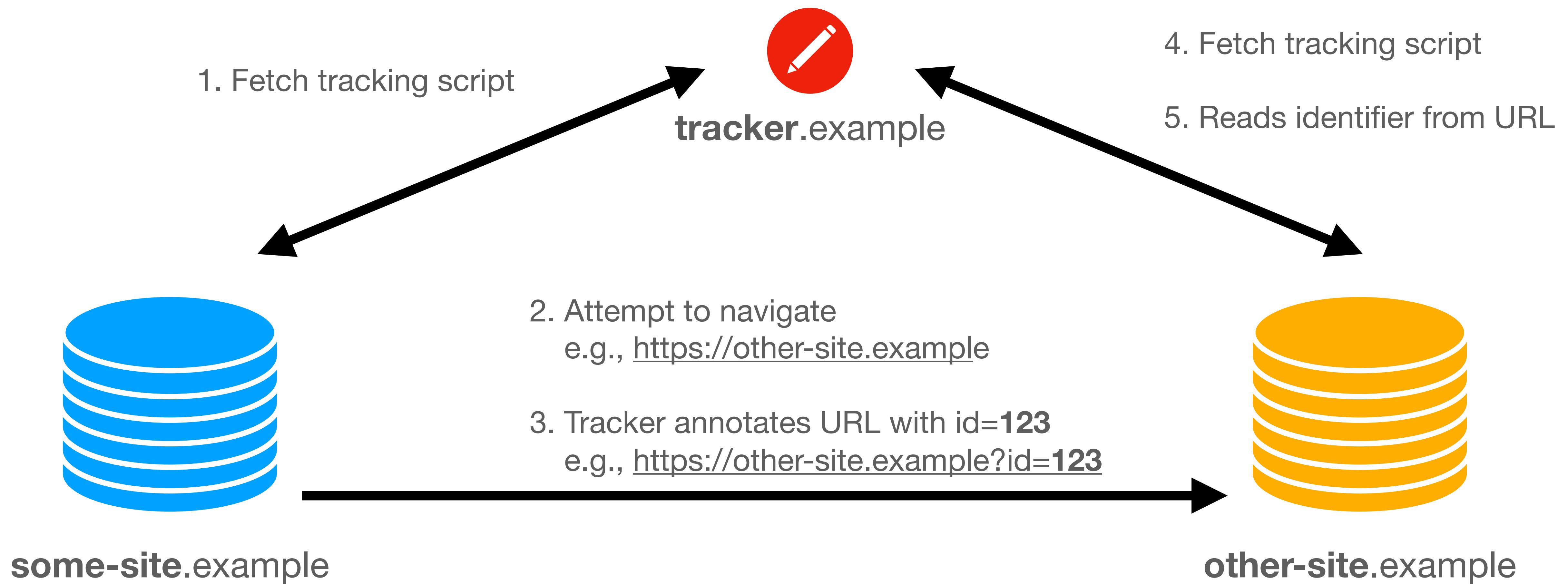
Storage partitioning



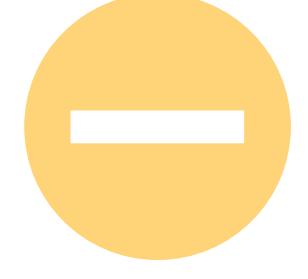
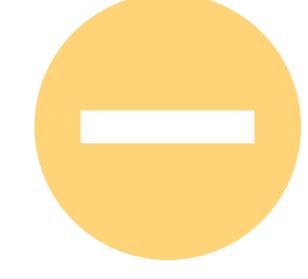
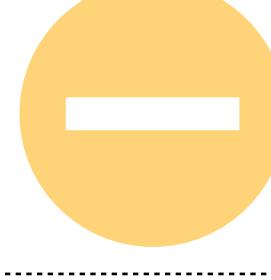
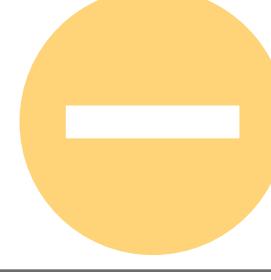
Bounce tracking



Navigation tracking



Bounce and Navigation Tracking: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Limit storage		heuristic 		List 	
“Debounce”					List 
Warn user					List 

Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



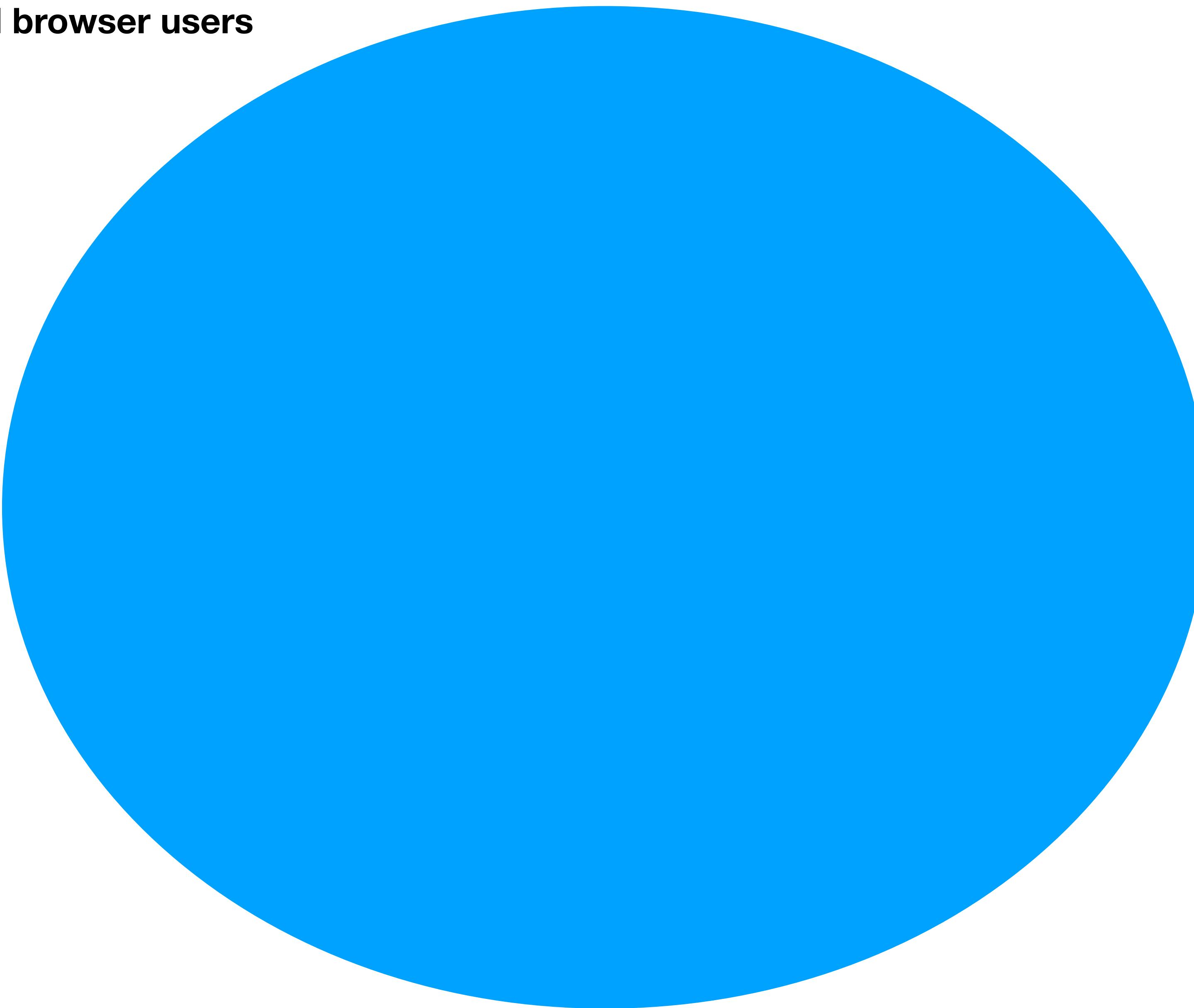
Fingerprinting, contrasted

- **Classic tracking**
 - Website stores an id on the client
 - The client returns the id to the server (cookie or JS)
 - The id is what allows re-identification
 - “Stateful”
- **Fingerprinting / passive tracking**
 - Website finds things different about each visitor
 - Tracker derives the identifier from minor browser differences
 - “Stateless”

Fingerprinting, how?

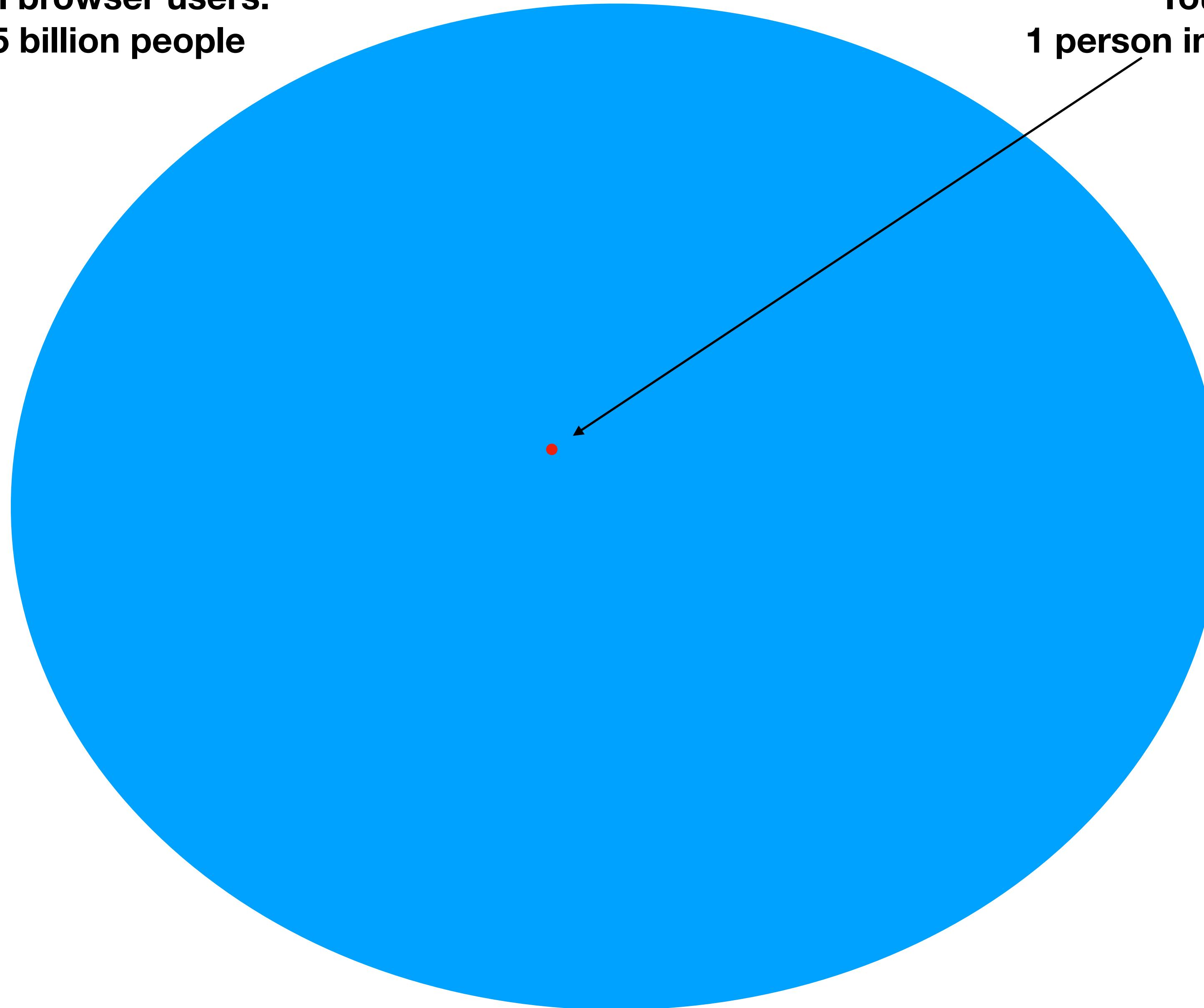
- **Large number of semi-identifiers**
 - Browser size
 - Extra fonts
 - Audio hardware
 - Video hardware
 - Installed plugins
 - Color depth
- **Add the semi identification up...**

All browser users

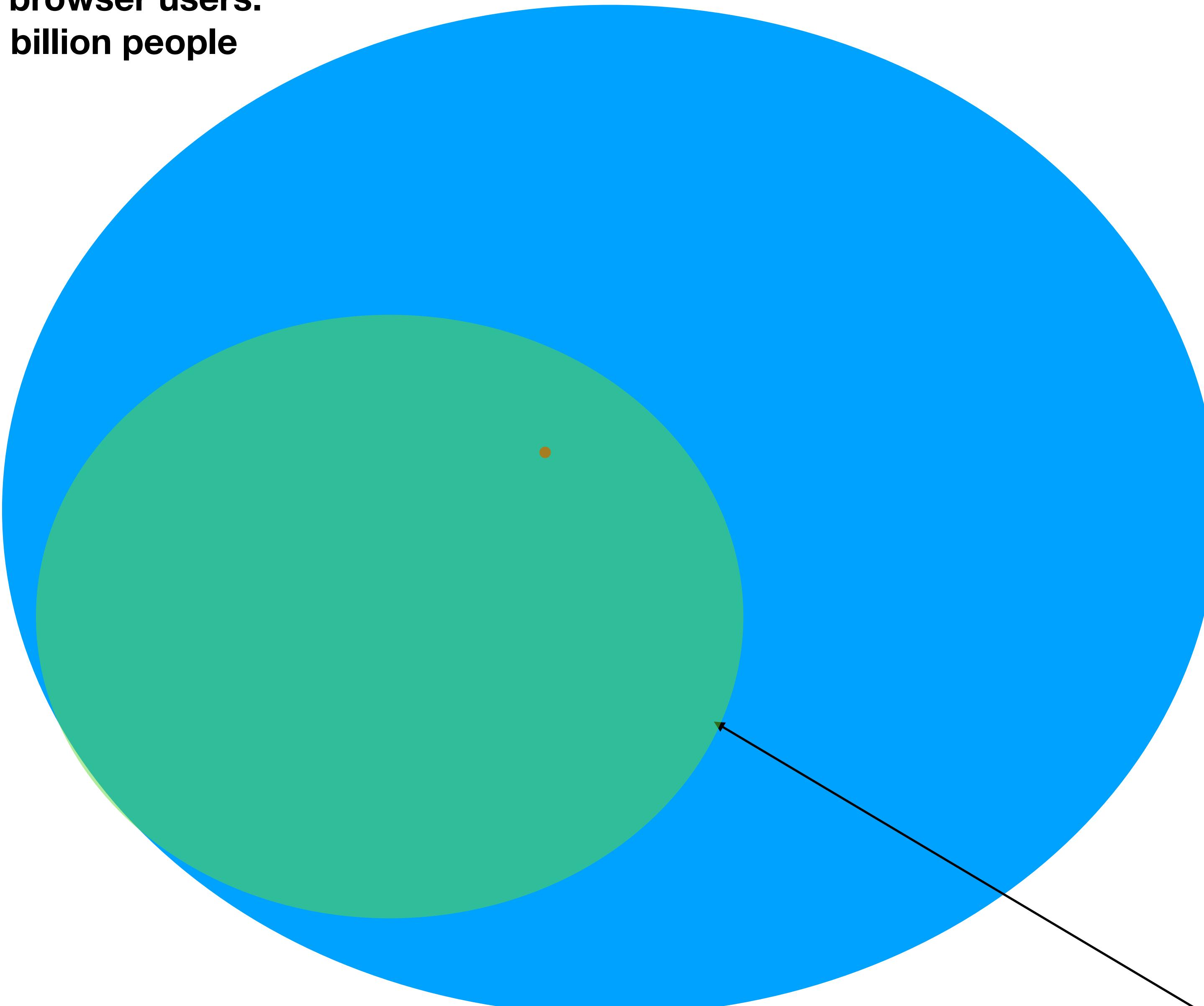


**All browser users:
5 billion people**

**You
1 person in 5 billion**

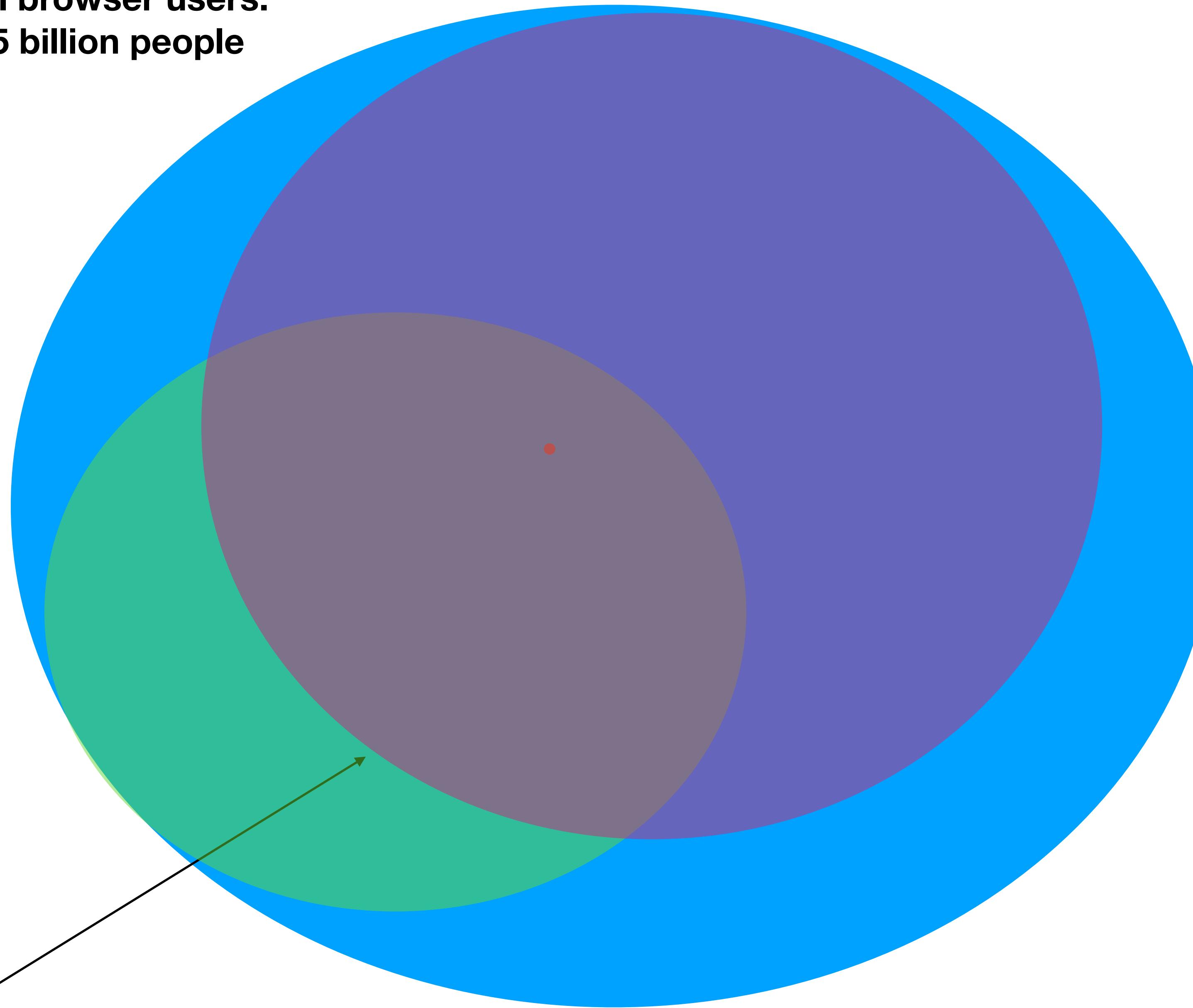


**All browser users:
5 billion people**



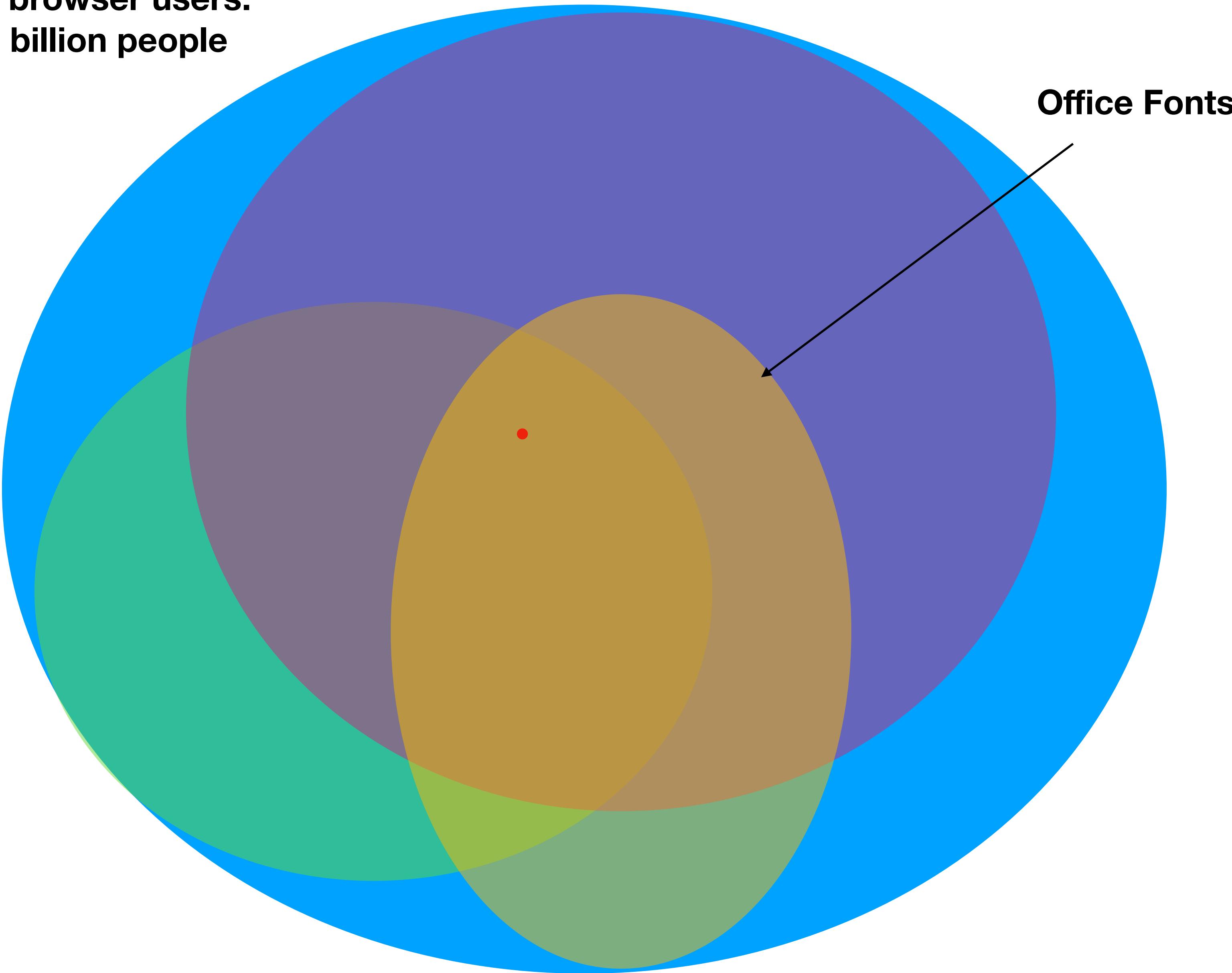
**Firefox
Users**

**All browser users:
5 billion people**



Windows users

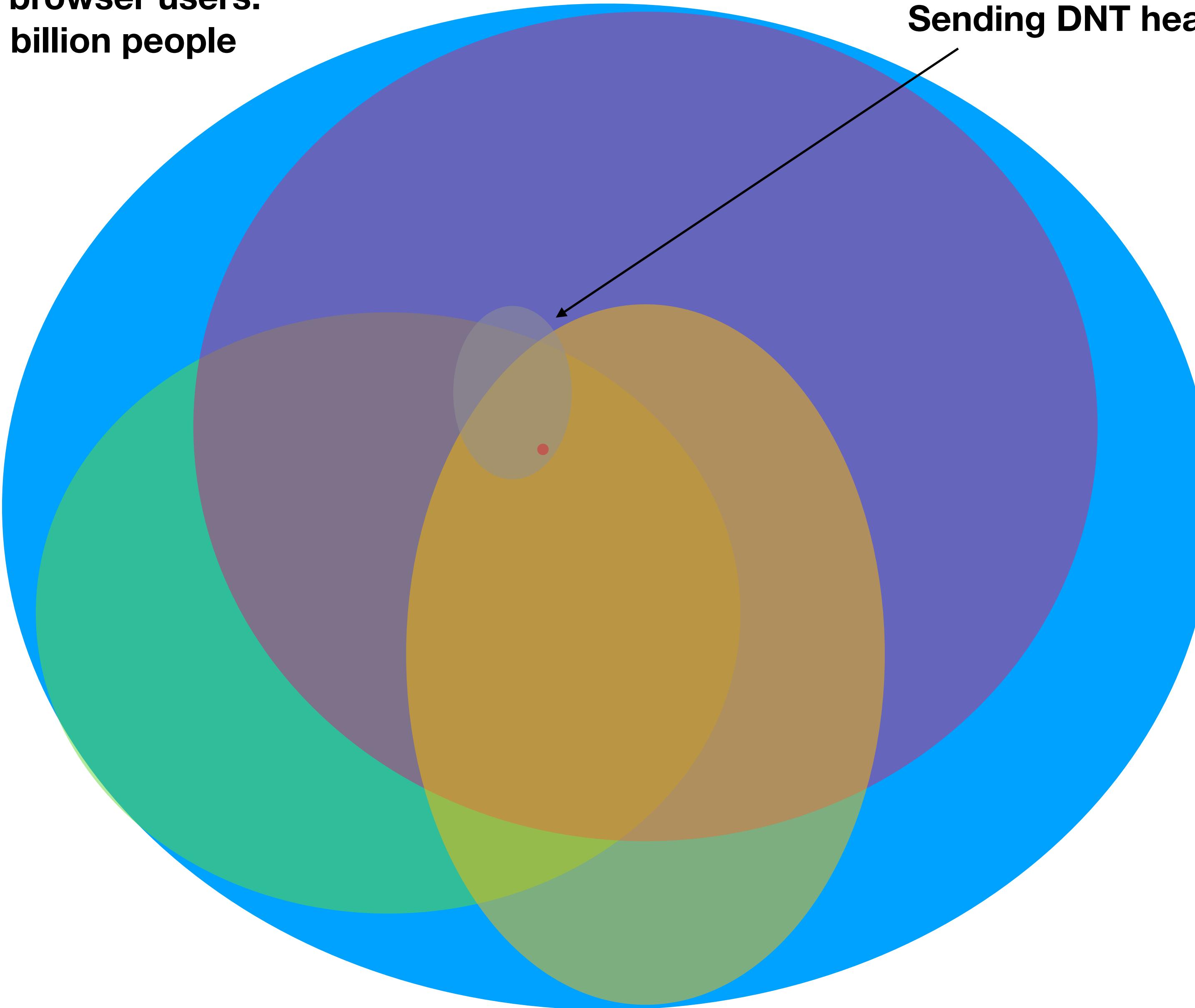
**All browser users:
5 billion people**



Office Fonts

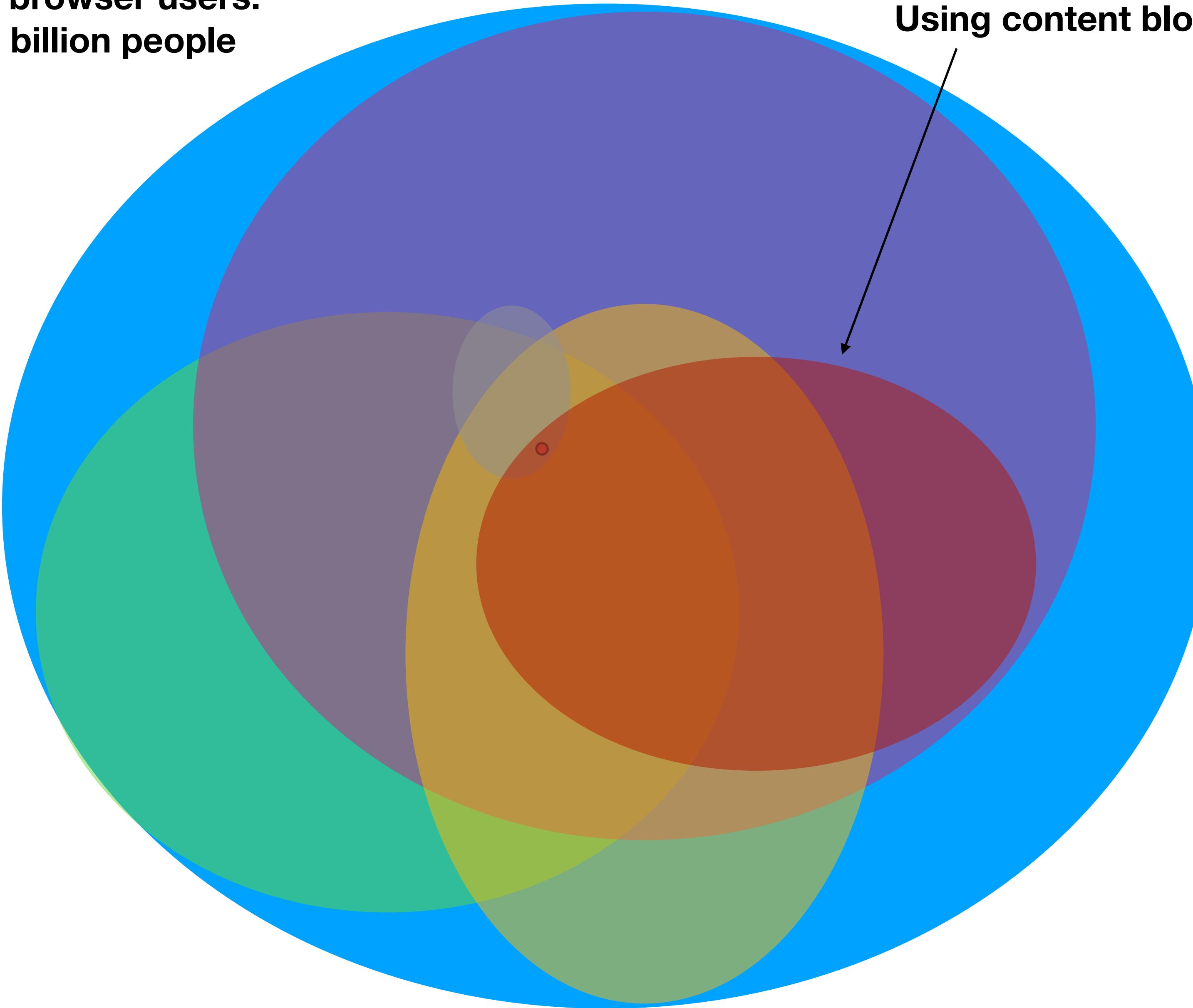
**All browser users:
5 billion people**

Sending DNT header



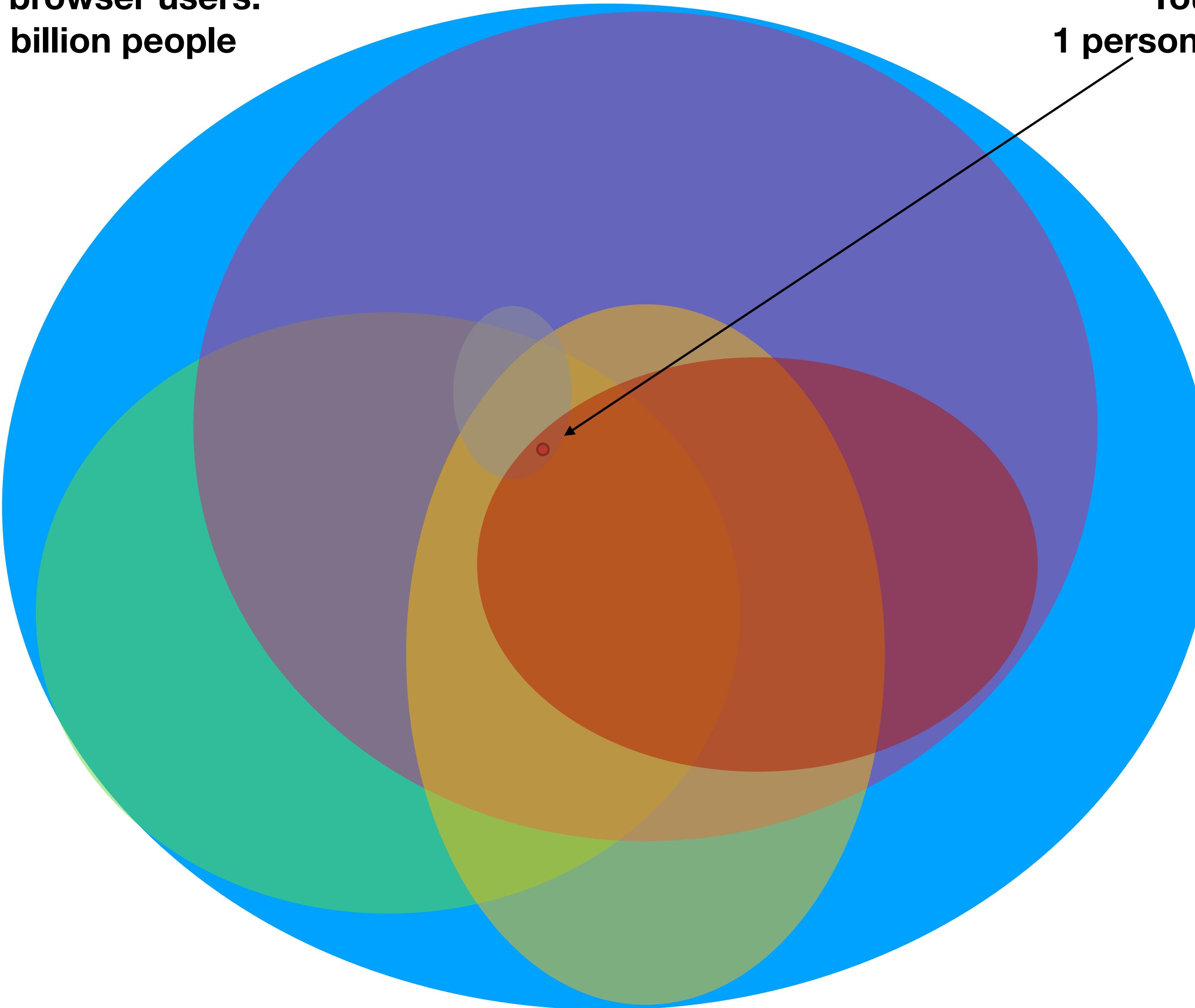
**All browser users:
5 billion people**

Using content blocker



**All browser users:
5 billion people**

**You
1 person in 100**



Fingerprinting, abstracted

- **Still needs a common value across boundaries**
Sites, sessions, time, etc
- **Value needs to be unique**
Otherwise it mixes you up with others
- **Value needs to be consistent**
Otherwise it doesn't (re)identify you

Possible Defenses

- **Try to make browsers look similar**
Reduce the “bits” available to fingerprinters
- **Try to block bad parties**
Keep the “bad folks” out
- **Privacy budgets**
Only allow sites to do so much identifying, e.g., 10 bits but not more
- **Randomization**
Make browser look intentionally different, within each boundary

Fingerprinting: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Restricted hardware					
Feature selection / removal					
Block fingerprinters					
Randomization					

Tracking Techniques

- Third-party DOM storage
- Network state
- Bounce tracking
- Browser fingerprinting
- IP address
- Personal identifiers



IP Addresses are pretty unique!

- ...especially if you look for clusters

The 3 ips you most commonly connect from is very unique

- IPv6 makes it a lot worse

Obviously... :-/

- Four general approaches

- Contracts / promises
- proxies
- mix nets
- block bad parties

IP Addresses Defenses

	Chrome	Safari	Edge	Firefox	Brave
Websites promise					
Proxies		Private relay 		Optional VPN 	Optional VPN
Mix networks					Optional Tor
Block bad parties					

Tracking Techniques

- Third-party DOM storage
 - Network state
 - Bounce tracking
 - Browser fingerprinting
 - IP address
 - Personal identifiers
- 

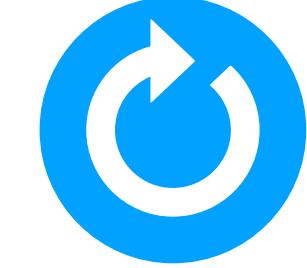
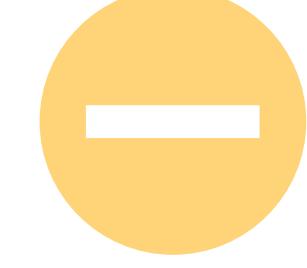
Personal Identifiers

- **Names, email addresses, CCN, etc**
“old school”
- **Can be combined with offline sources**
Credit agencies, public legal records, tax documents, etc
- **Baked into the web**
Authentication, user accounts, etc

Partitioning to the Rescue (?)

- **User holds the “true” value**
e.g., true email address
- **Browser holds a secret**
e.g., secret = rand()
- **Derive per site identities**
e.g. hash(email + secret + eTLD+1) + @private-email.com
- **Applicable to a range of identifiers**
Email, CCN, Crypto addresses

Personal Identifiers: Defenses

	Chrome	Safari	Edge	Firefox	Brave
Partition email					
Partition Web3					
Block scripts					

Overview

- **Why Privacy Matters**

A sloppy manifesto

- **Defining Tracking**

Abstracting the problem

- **Tracking in Practice**

Methods and defenses

- **Privacy Beyond Tracking**

Other issues and concerns



Privacy is more than Absence of Tracking

- **Browsers shouldn't share information unless its helpful to user**
e.g., FLoC
- **Browsers should serve users first and exclusively**
e.g., Reporting API, FLEDGE
- **Browsers shouldn't introduce capabilities that remove user choice**
WebBundles
- **Browsers shouldn't confuse users!**
First-party sets, SXG
- **First-parties are suspect too...**

Other privacy protections

- **Governments increasing provide legal protections**
GDPR, CCPA, etc
- **Browsers can help users assert their privacy rights**
e.g., GlobalPrivacyControl
- **Authored by activists, academics, New York Times, DuckDuckGo, Brave**
Implemented in Brave and DDG
- **Beware of conflating with “consent management” systems**

A final plea...

- You are all plainly, amazingly smart people
- You'll be able to (mostly) choose your job
- Privacy harms are particularly difficult to remediate
- Consider the privacy implications of a job before you take it

A final plea...

- You are all plainly, amazingly smart people
- You'll be able to (mostly) choose your job
- Privacy harms are particularly difficult to remediate
- Consider the privacy implications of a job before you take it

Thanks!