Secure Software Design and Engineering
(CY-321)

Online Tracking: Who's Doing it?

Dr. Zubair Ahmad

Why Does Tracking Matter

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

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

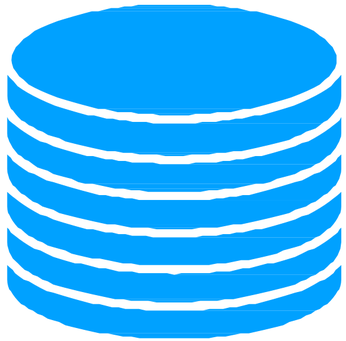
A Rough Definition of Tracking

- **Linking activities…**
  e.g., being “followed”

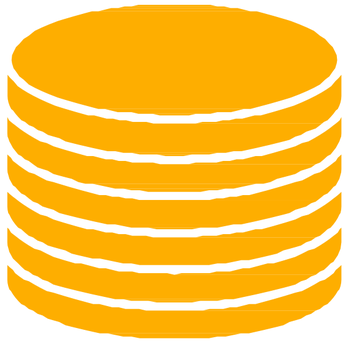
- **across boundaries…**
  e.g., temporal, geographic, conceptual

- **In a way not expected or desired.**
  e.g., ignorance or non-consent

Question One

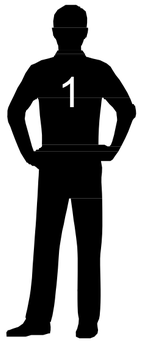


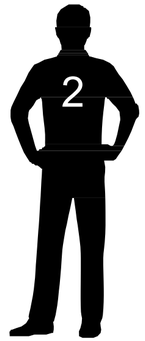
One day…



some-site.example

other-site.example



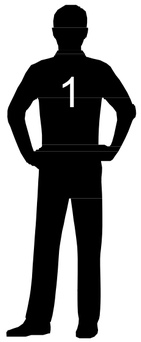


Question One

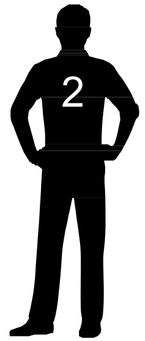
One day…

some-site.example

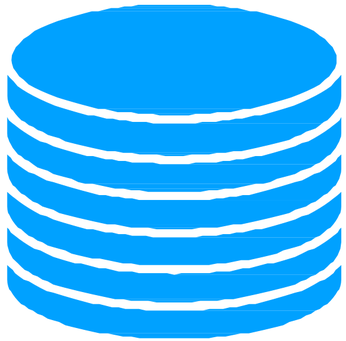
other-site.example



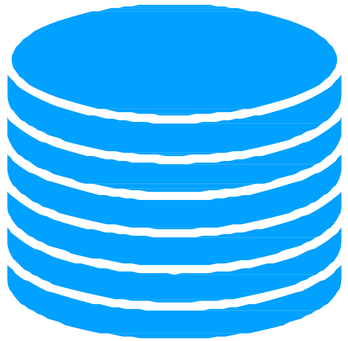
Not linking



Question Two

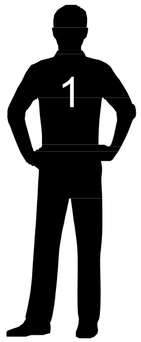


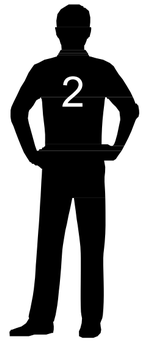
One day…



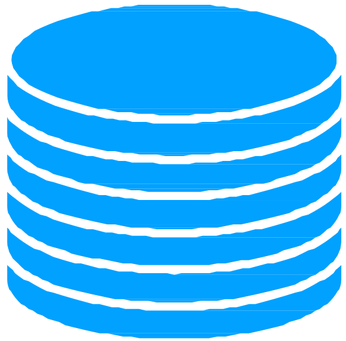
some-site.example

some-site.example

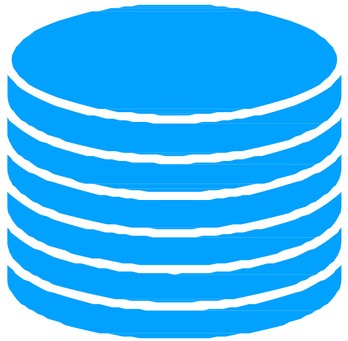




Question Two

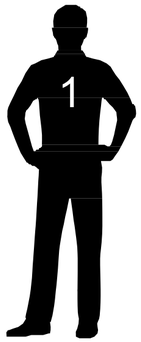


One day…

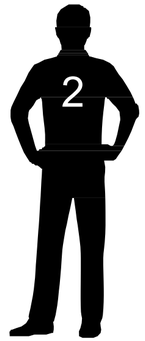


some-site.example

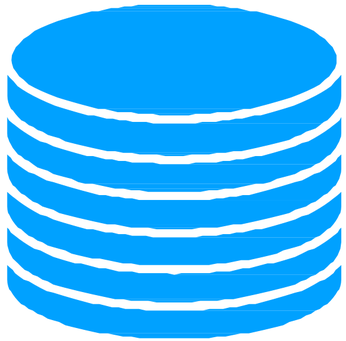
some-site.example



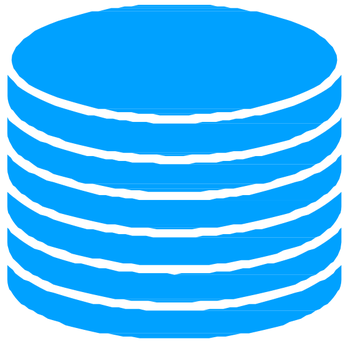
Not linking



Question Three

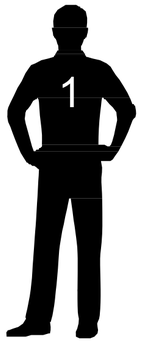


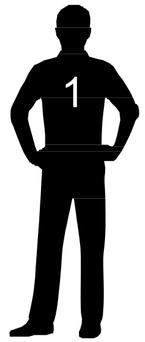
One day…



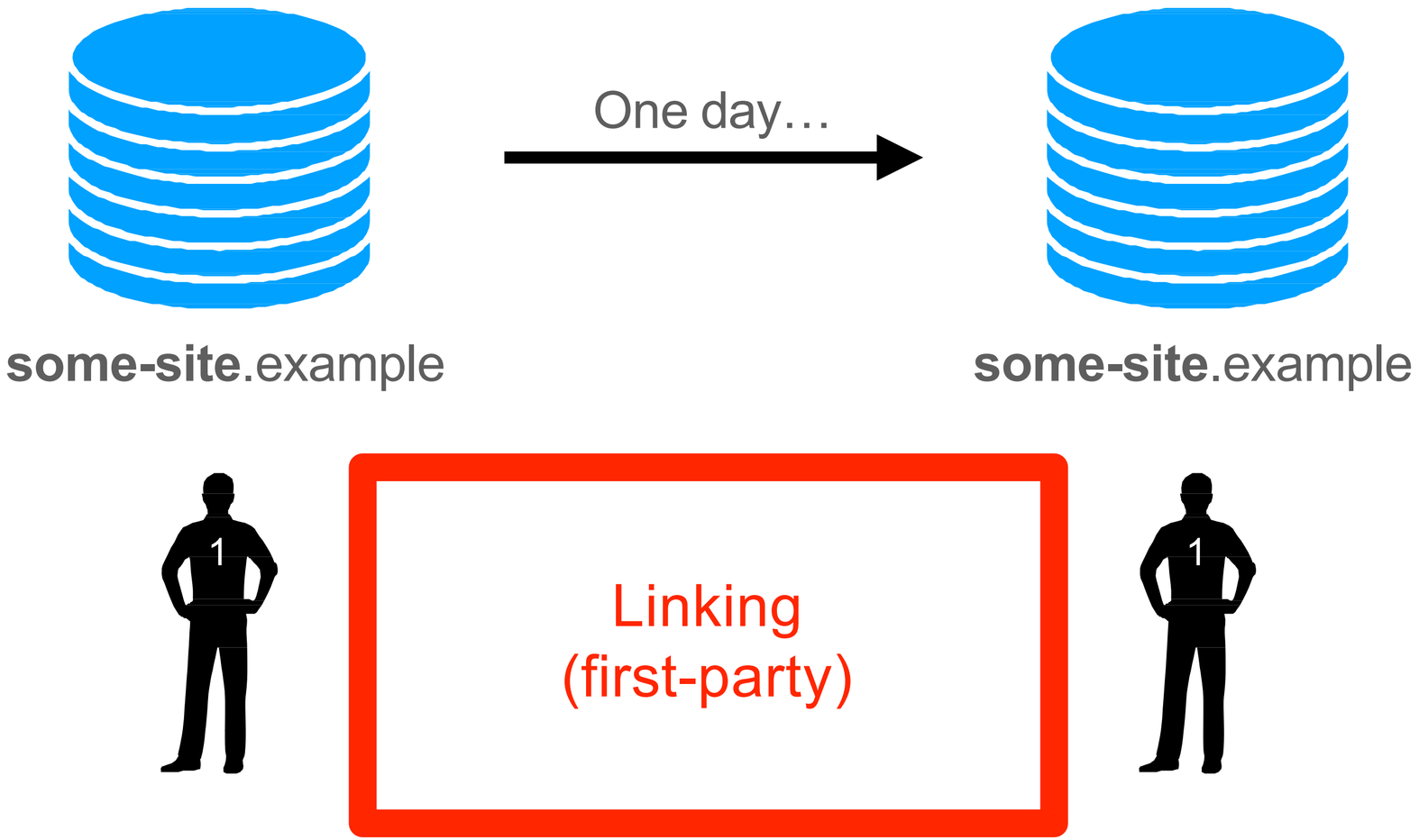
some-site.example

some-site.example

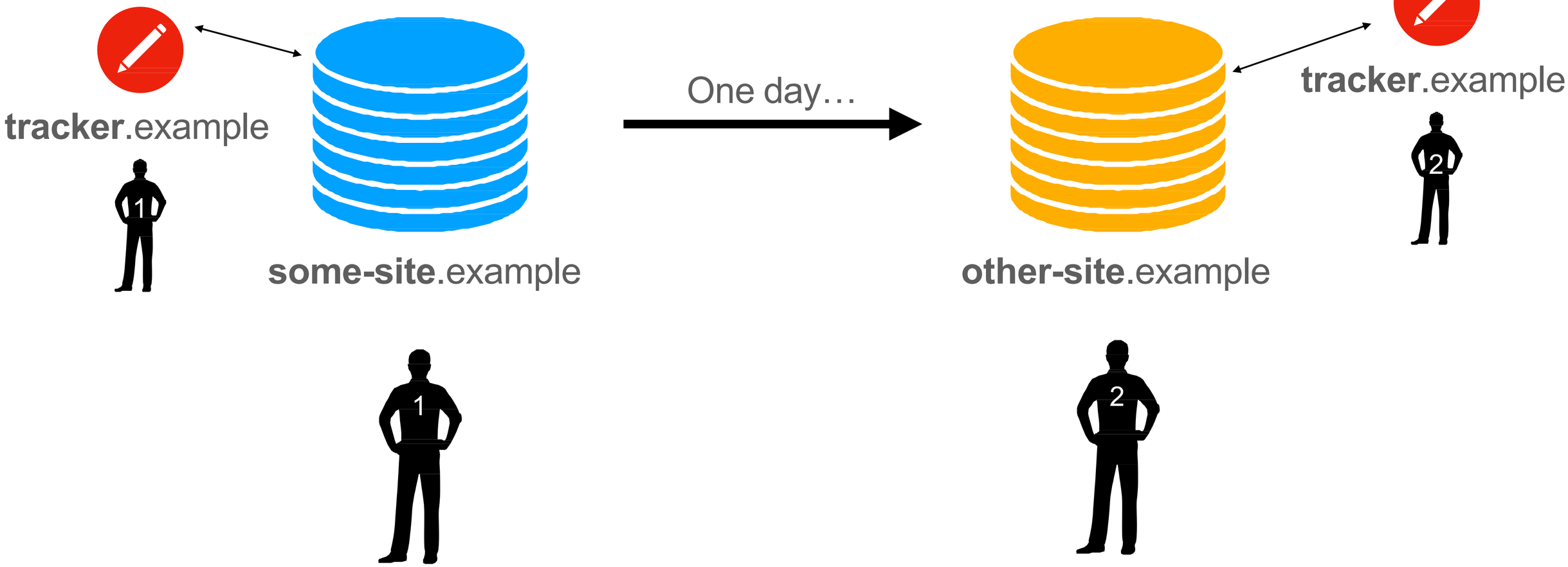




Question Three

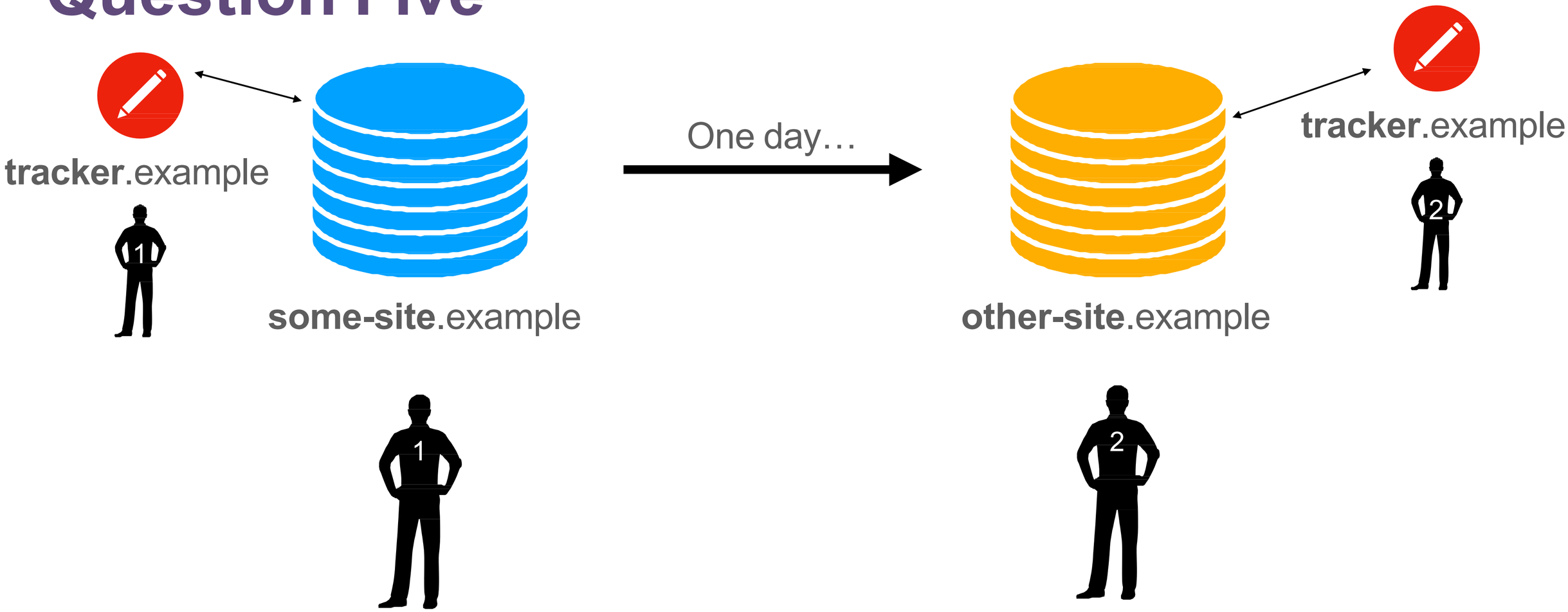


Question Four

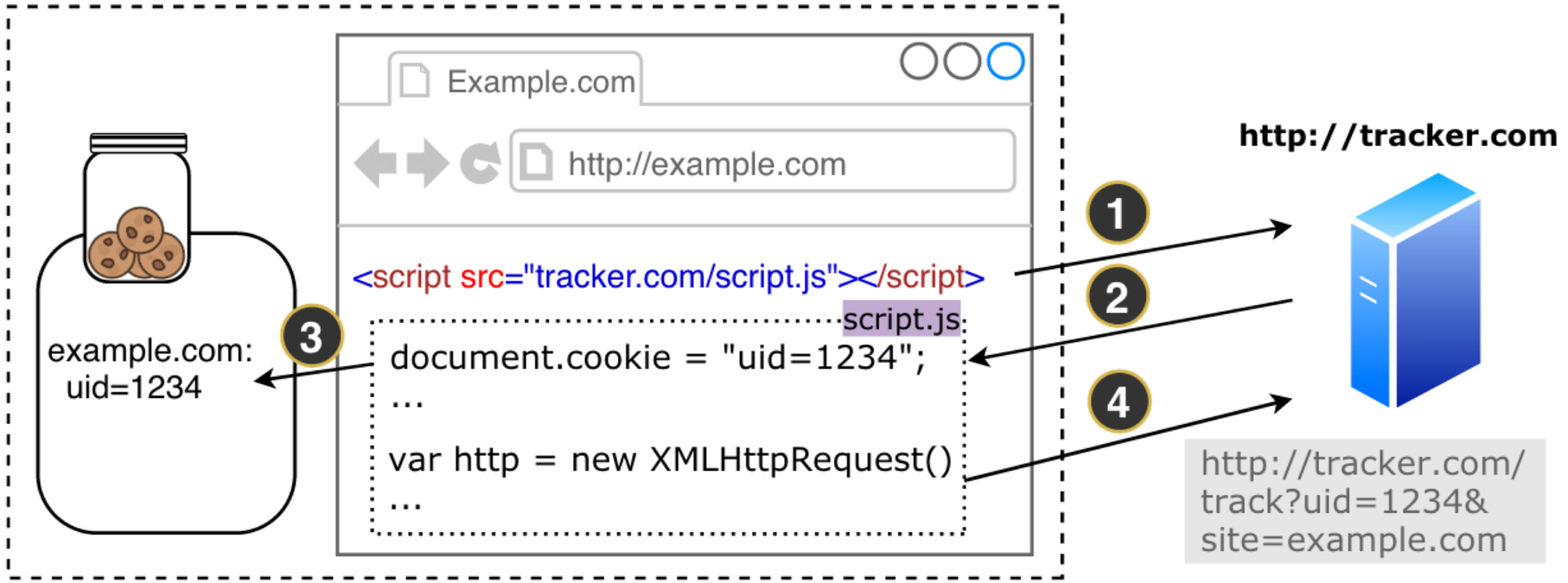


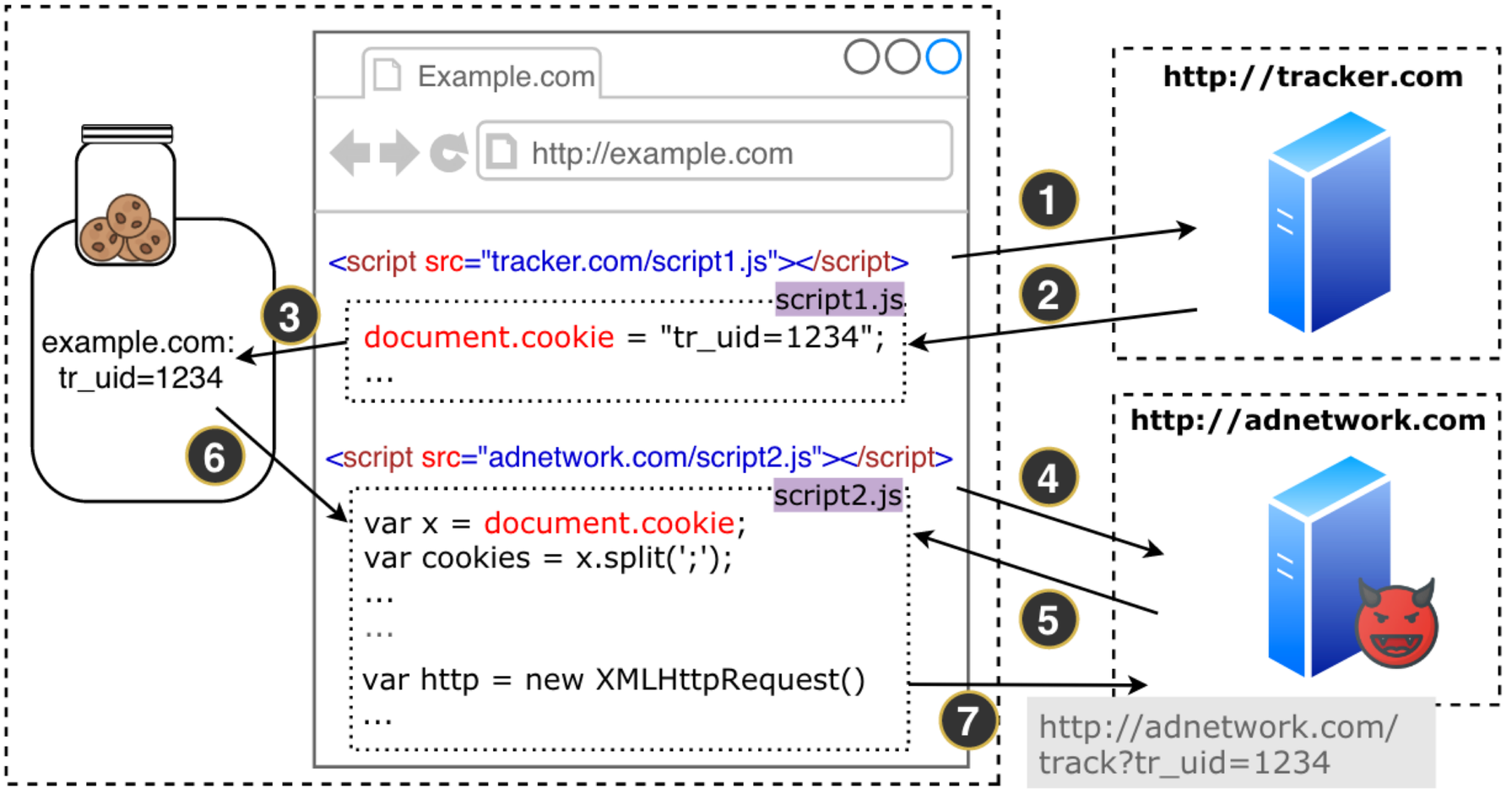
Question Four

Question Five



Question Five


Example.com
http://example.com
<script src="tracker.com/script.js"></script>
script.js
document.cookie = "uid=1234";
...
var http = new XMLHttpRequest()
...
example.com:
uid=1234
3
1
2
4
http://tracker.com
http://tracker.com/
track?uid=1234&
site=example.com


Example.com
http://example.com
<script src="tracker.com/script1.js"></script>
script1.js
document.cookie = "tr_uid=1234";
...
<script src="adnetwork.com/script2.js"></script>
script2.js
var x = document.cookie;
var cookies = x.split(';');
...
...
var http = new XMLHttpRequest()
...
example.com:
tr_uid=1234
http://tracker.com
http://adnetwork.com
http://adnetwork.com/
track?tr_uid=1234
1
2
3
4
5
6
7

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

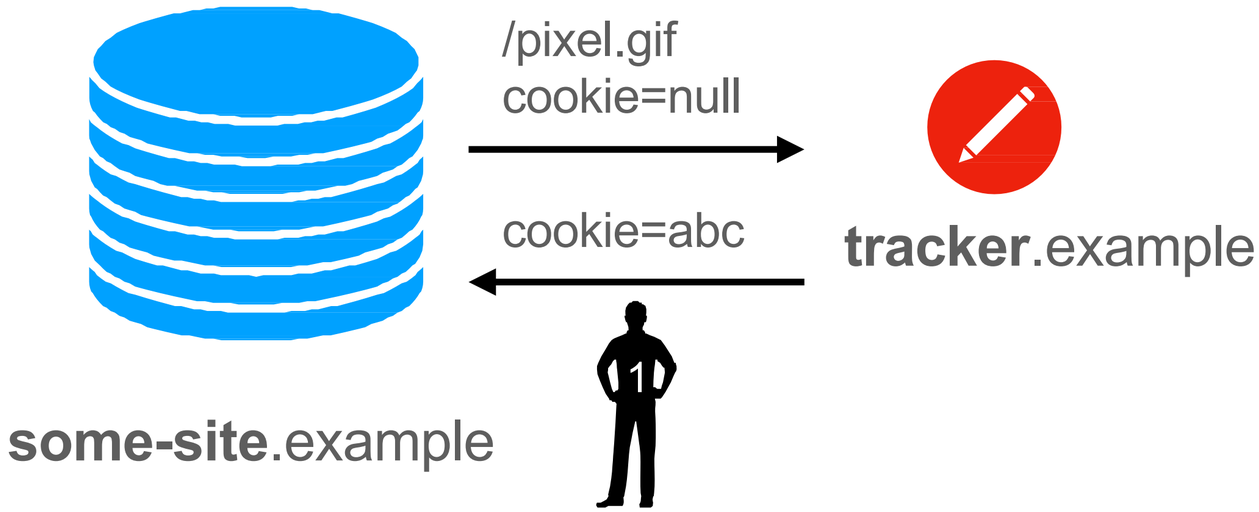
A Rough Definition of Tracking

- **Linking activities…**
  e.g., being “followed”

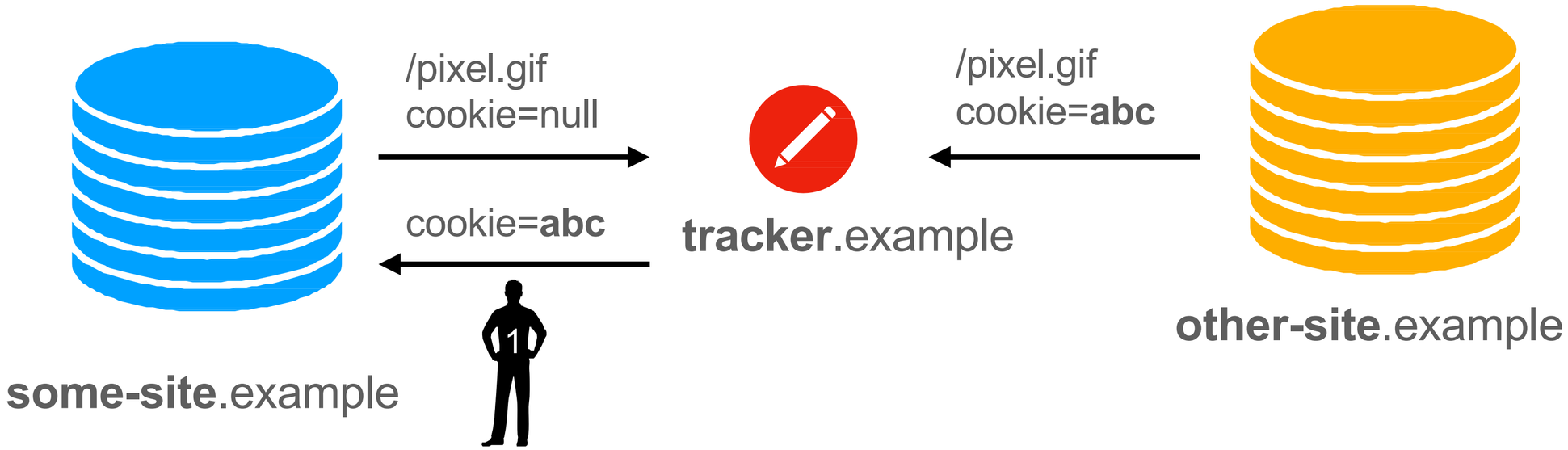
- **across boundaries…**
  e.g., temporal, geographic, conceptual

- **In a way not expected or desired.**
  e.g., ignorance or non-consent

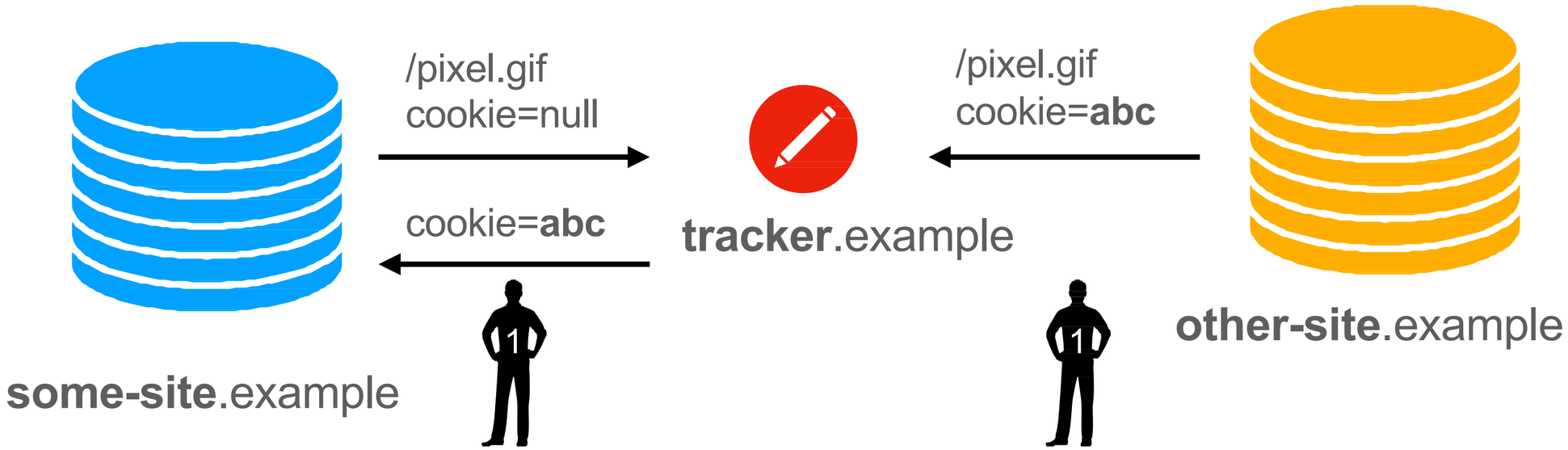
Third-party DOM storage



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

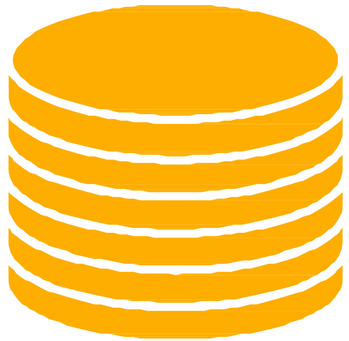
some-site.example

```
<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

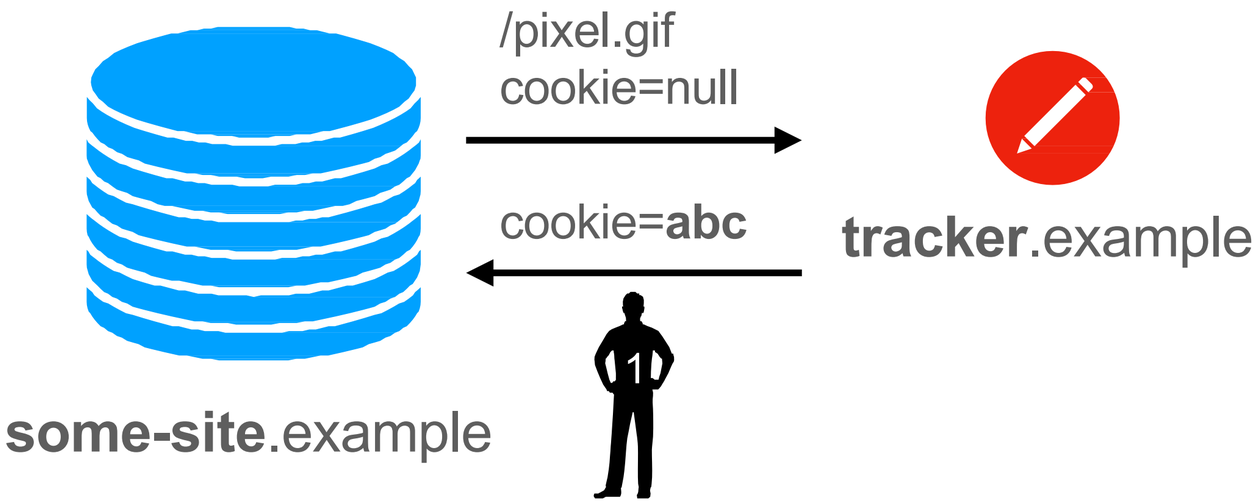
```
<iframe src=//tracker.example>

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>
```

some-site.example

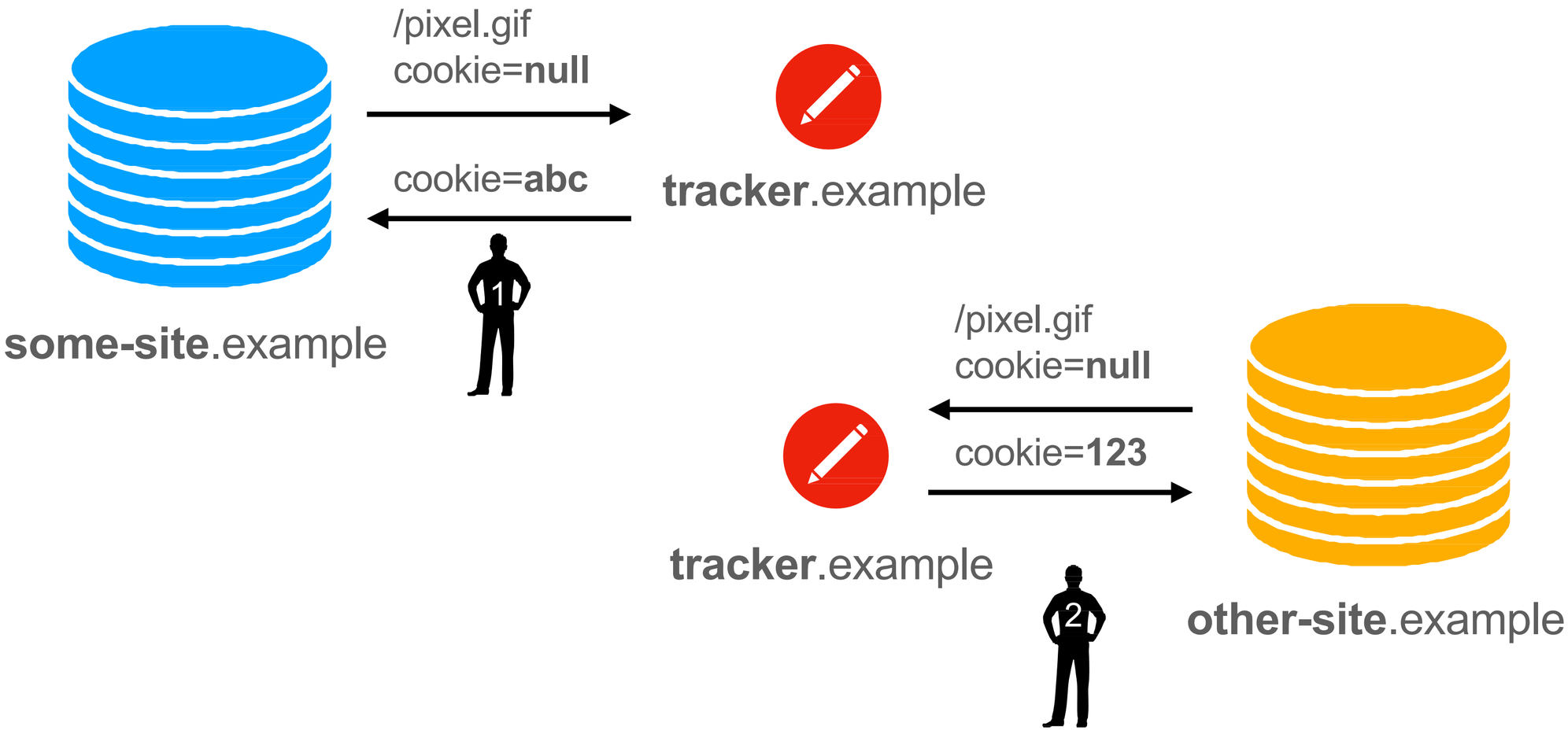


other-site.example

Third-party DOM storage: partitioning



Third-party DOM storage: partitioning



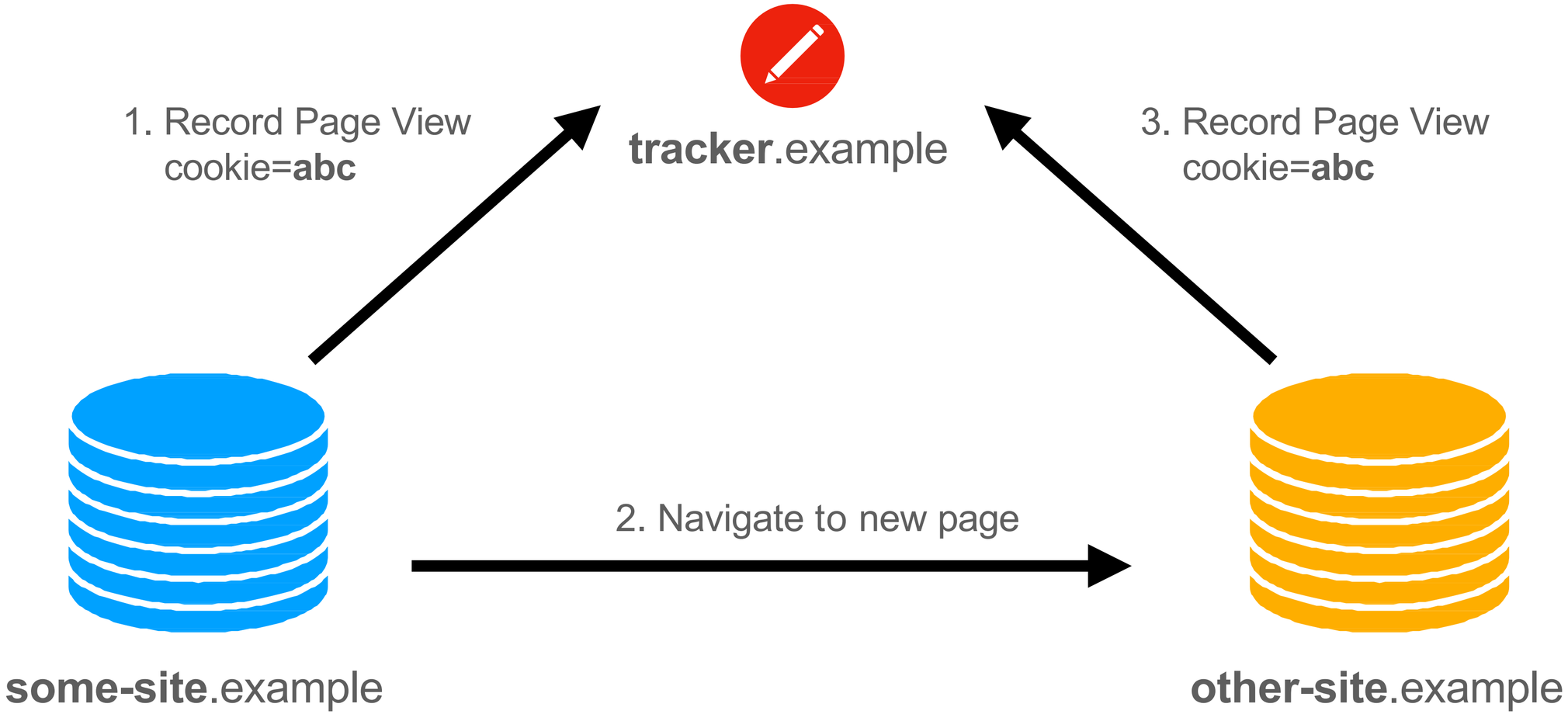
Third-party DOM storage: Defenses

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

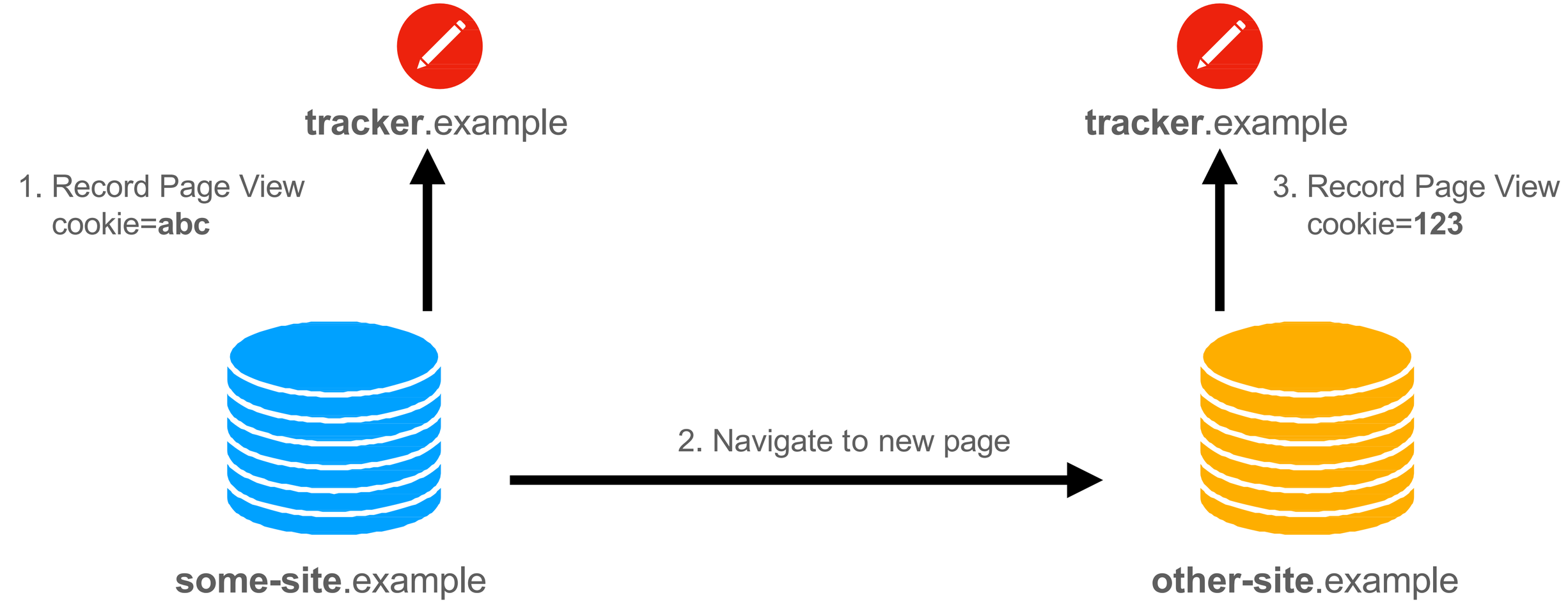
Bounce Tracking

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

Pre-partitioning

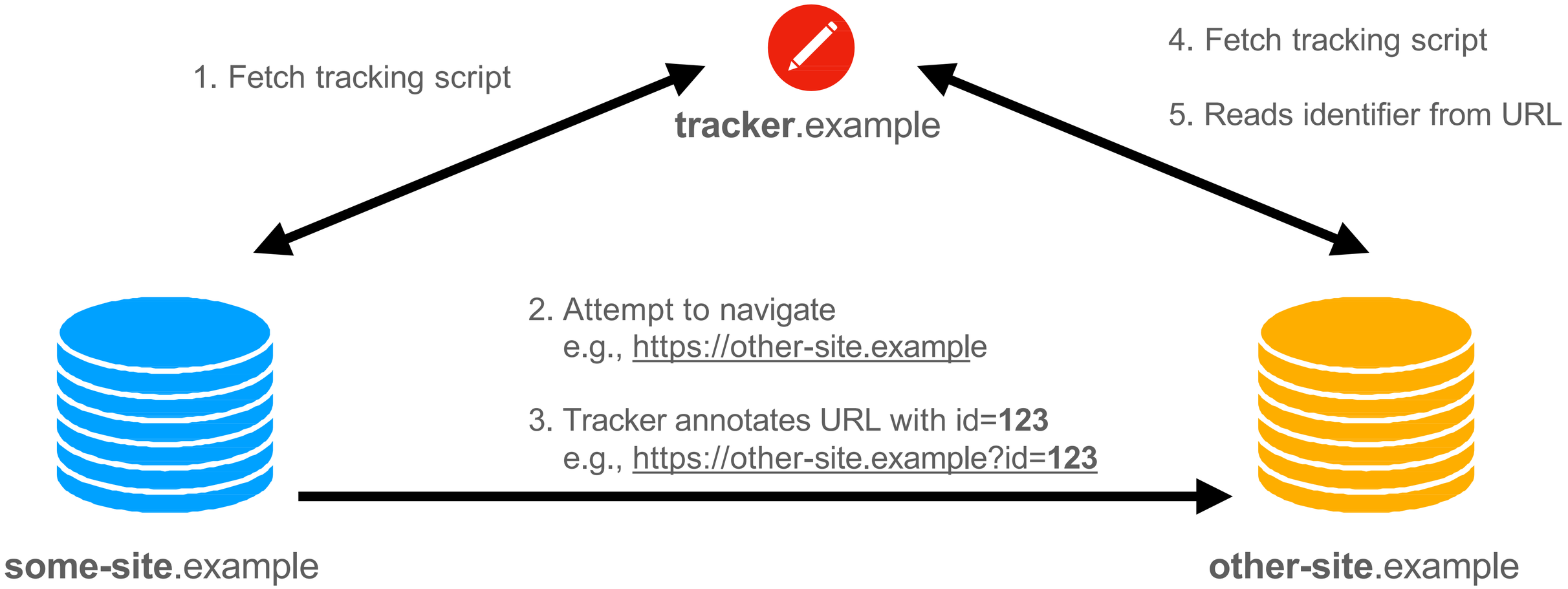


Storage partitioning



Bounce tracking

Navigation tracking



Bounce and Navigation Tracking: Defenses

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

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”

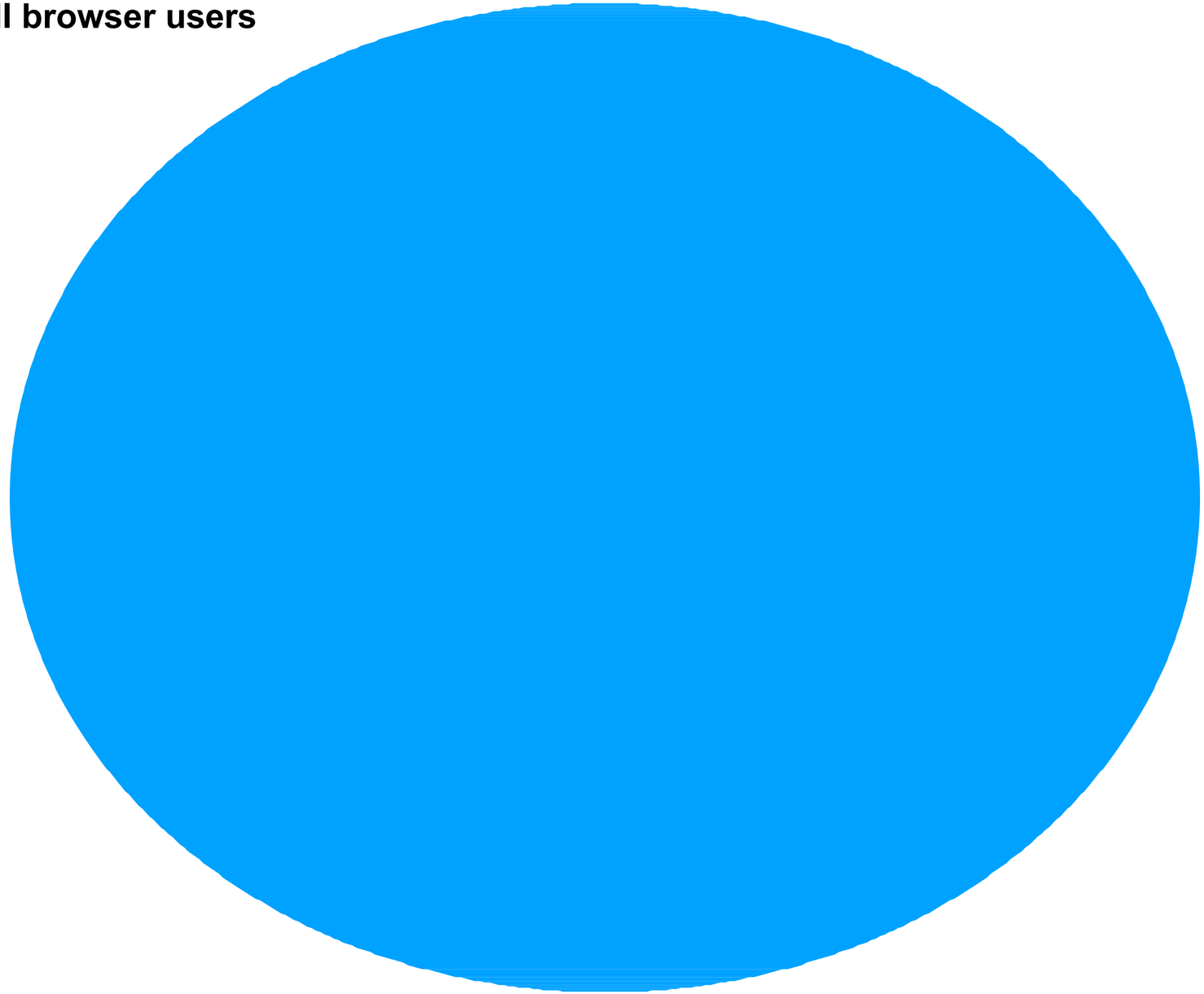
- **Fingerprinting / passive tracking**
  - Website finds things different about each visitor
  - Tracker derives the identifier from minor browser differences
  - “Stateless”

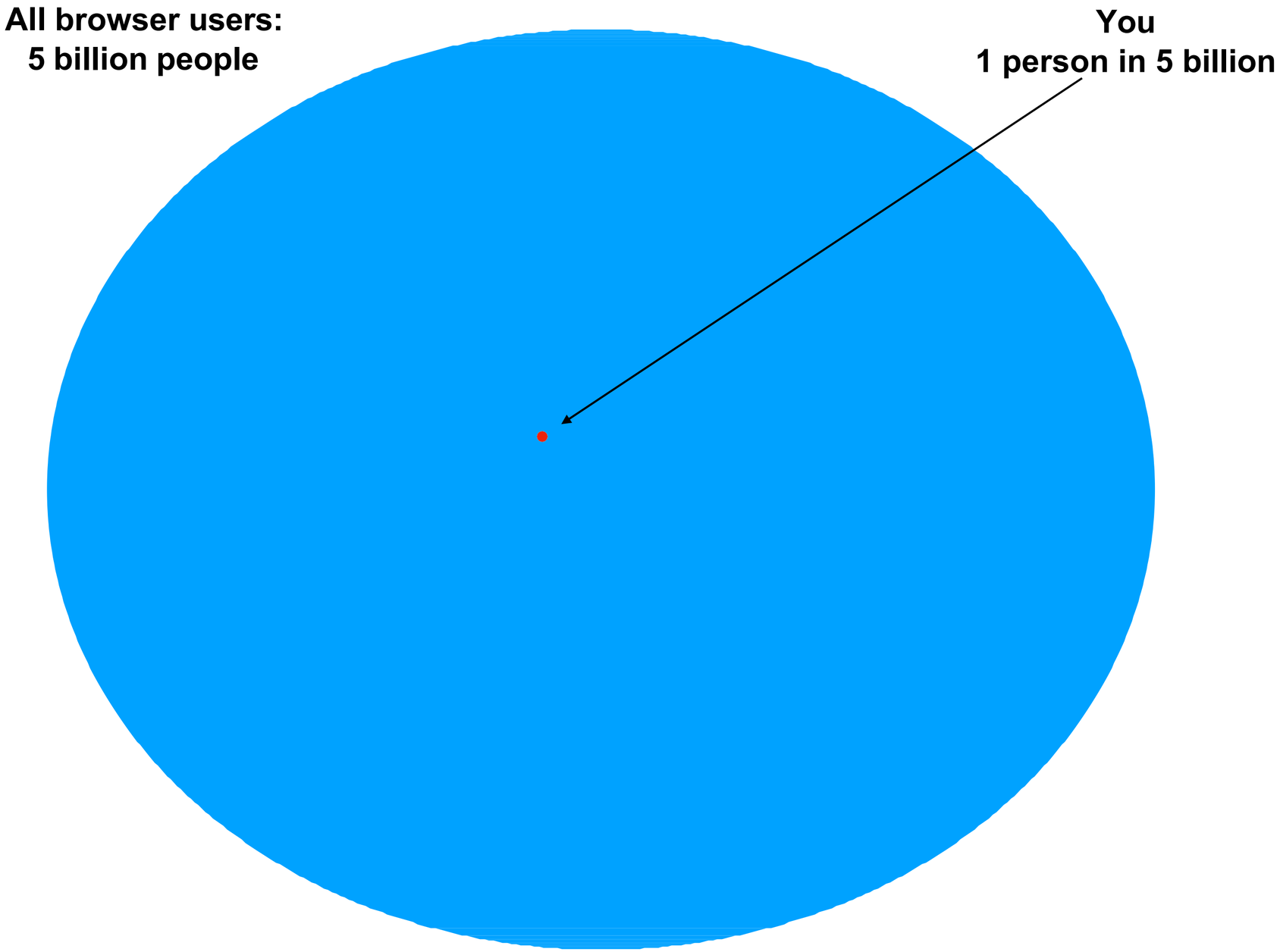
Fingerprinting, how?

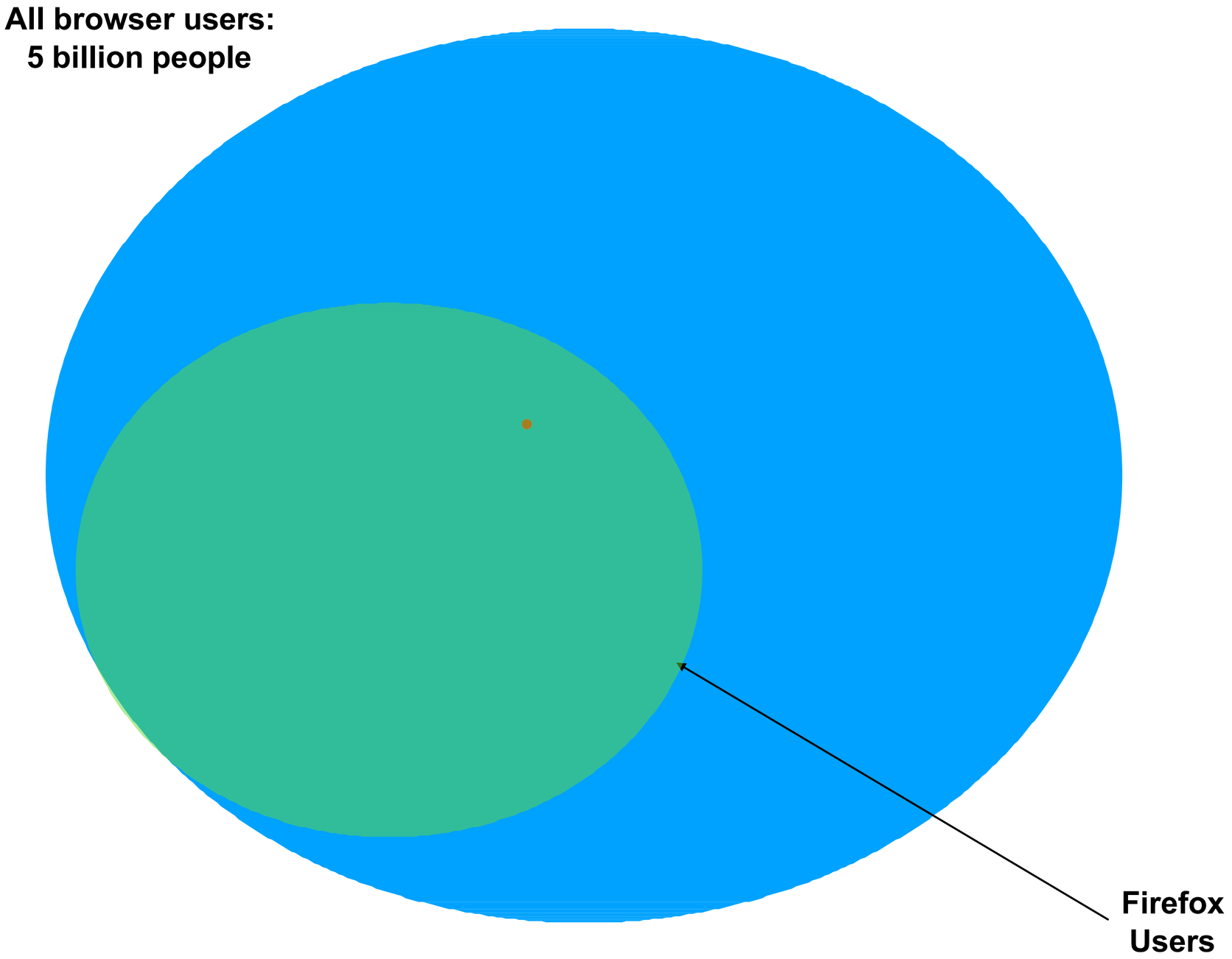
- **Large number of semi-identifiers**
  - Browser size
  - Extra fonts
  - Audio hardware
  - Video hardware
  - Installed plugins
  - Color depth

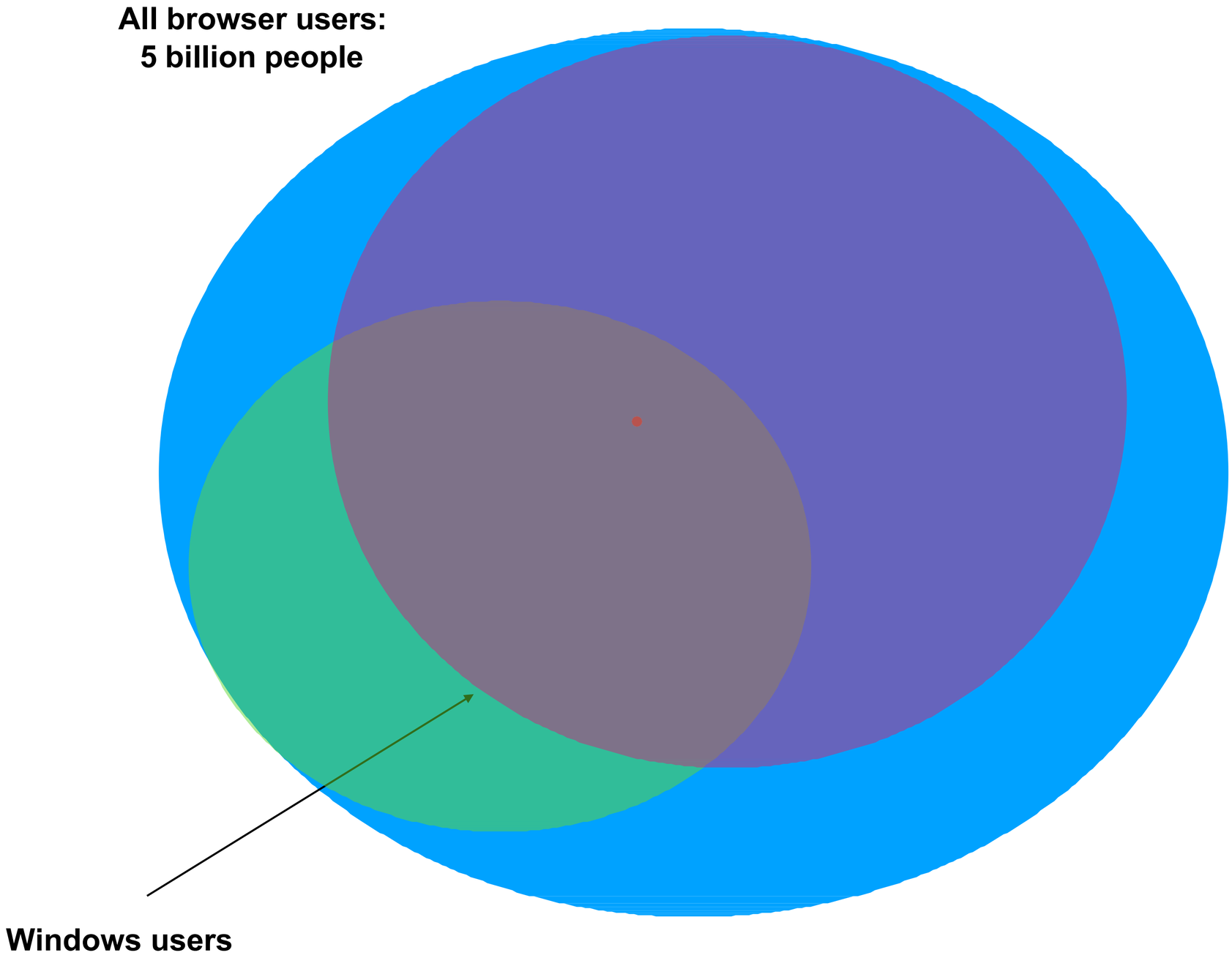
- **Add the semi identification up…**

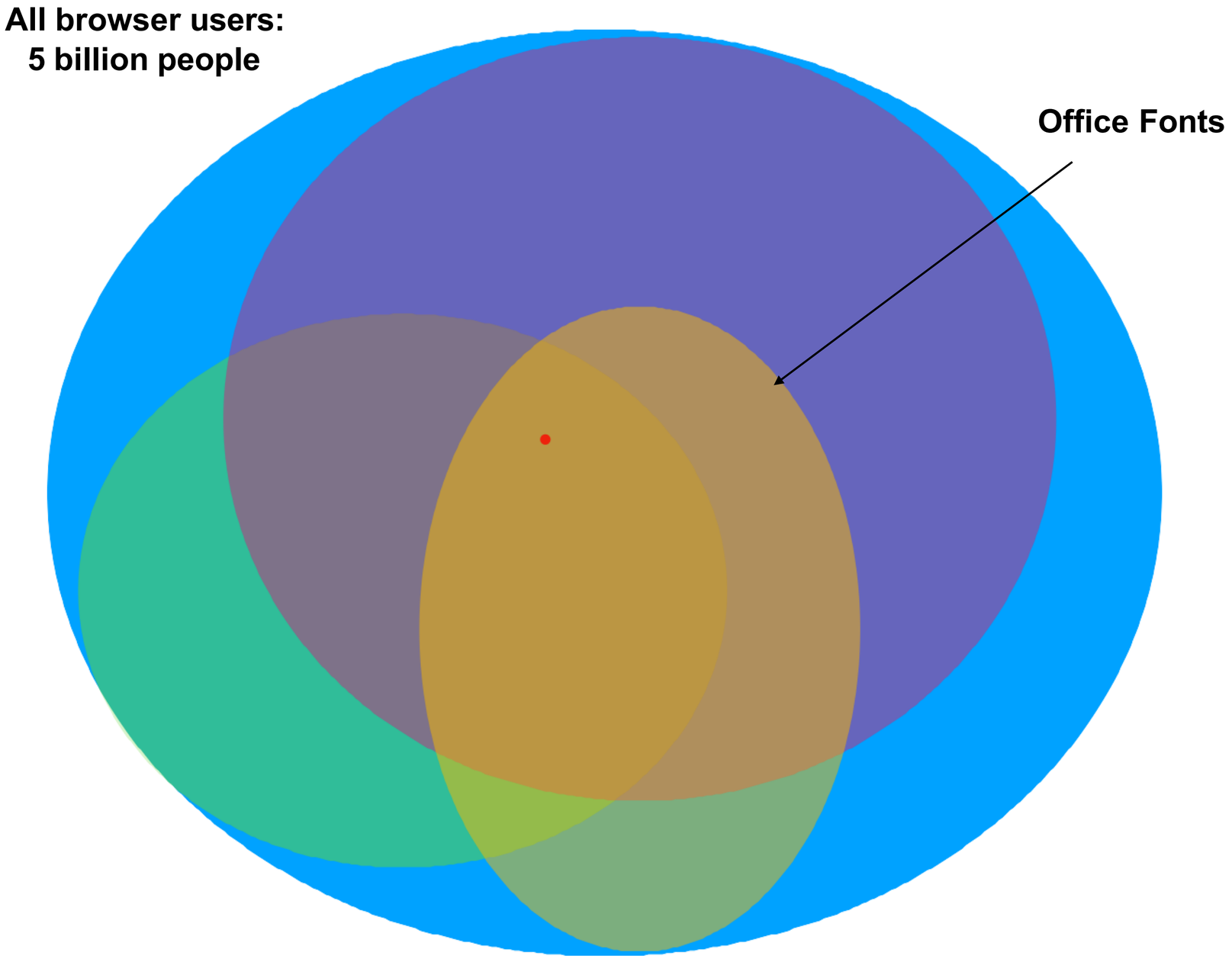
All browser users

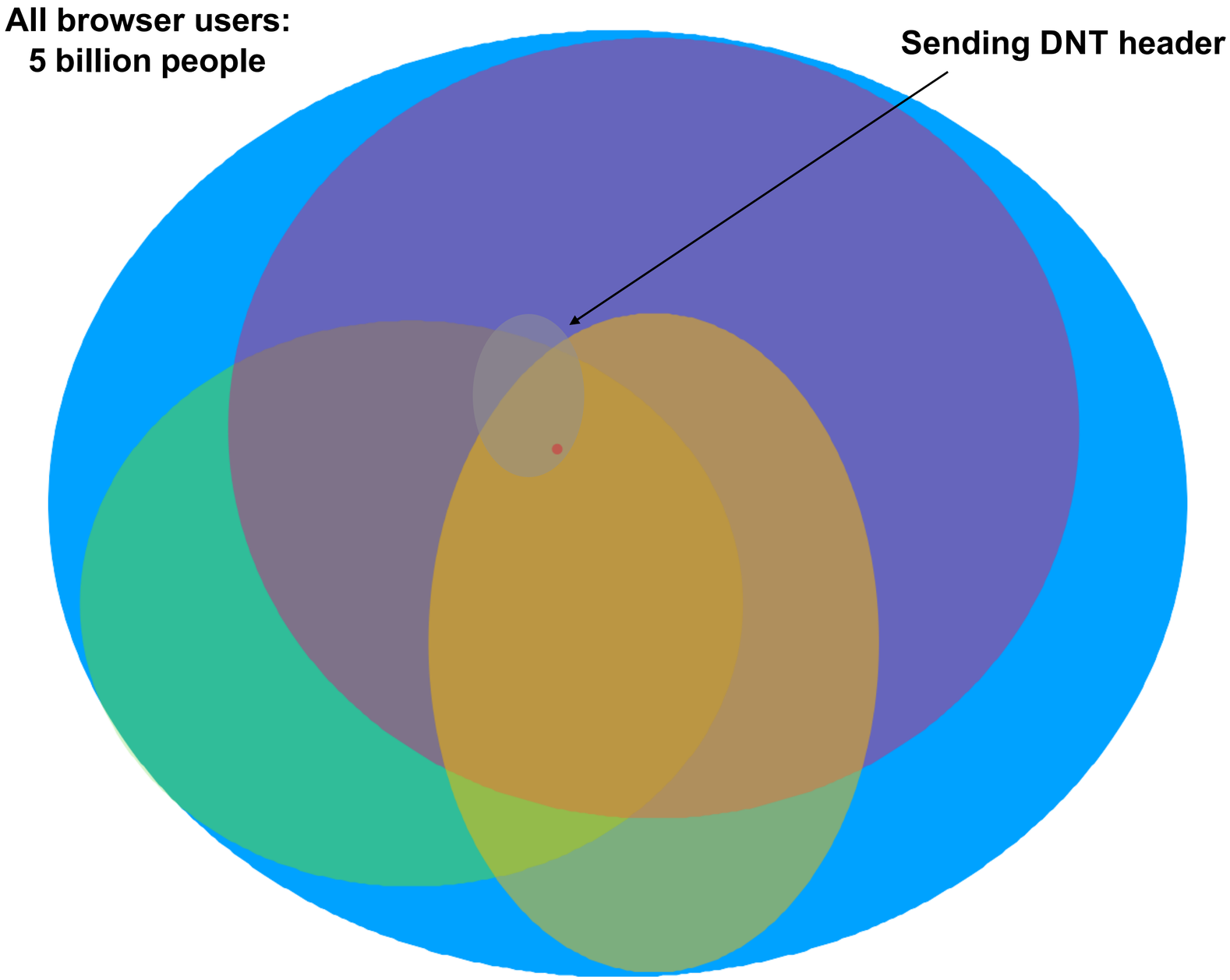


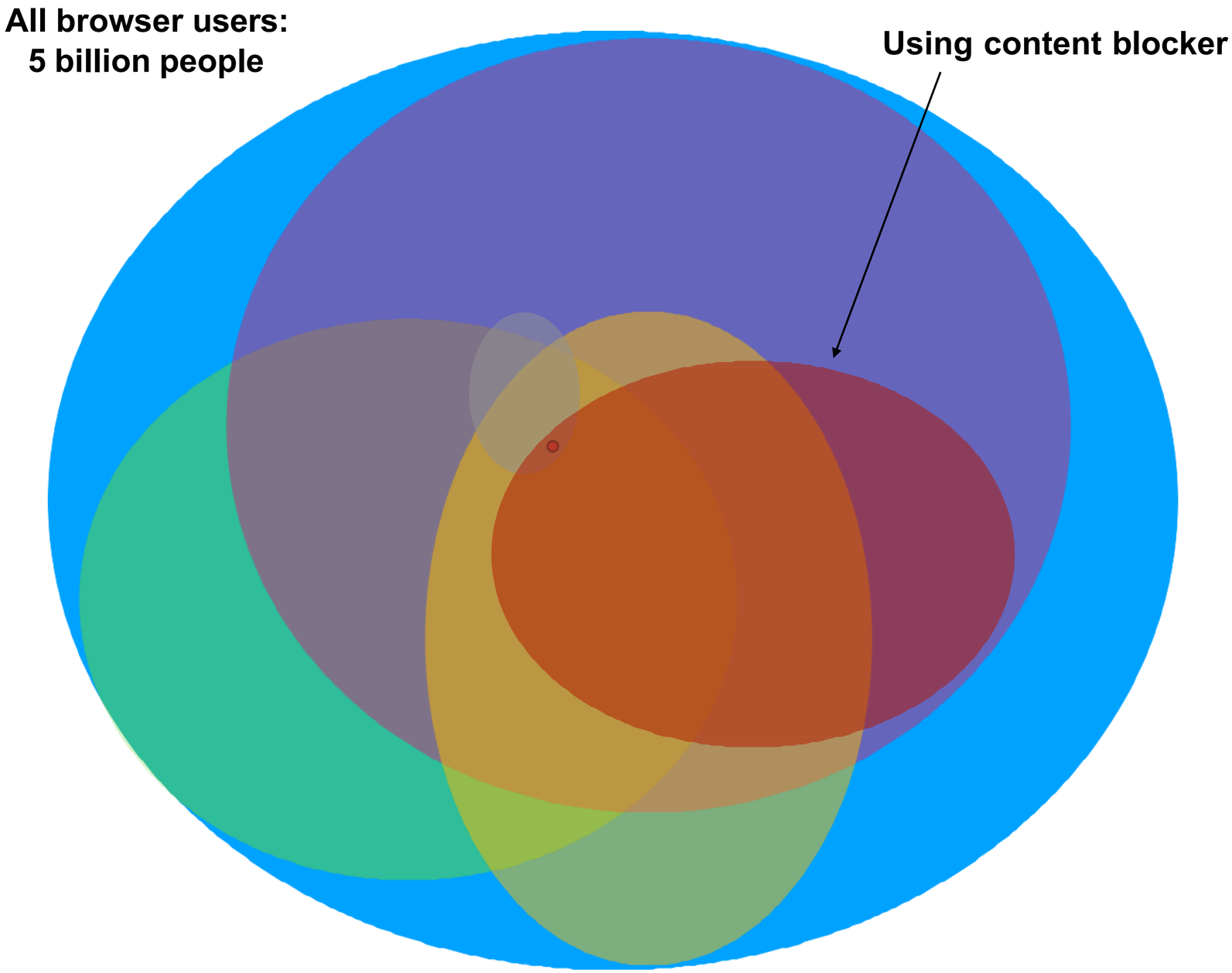

All browser users:
5 billion people
You
1 person in 5 billion

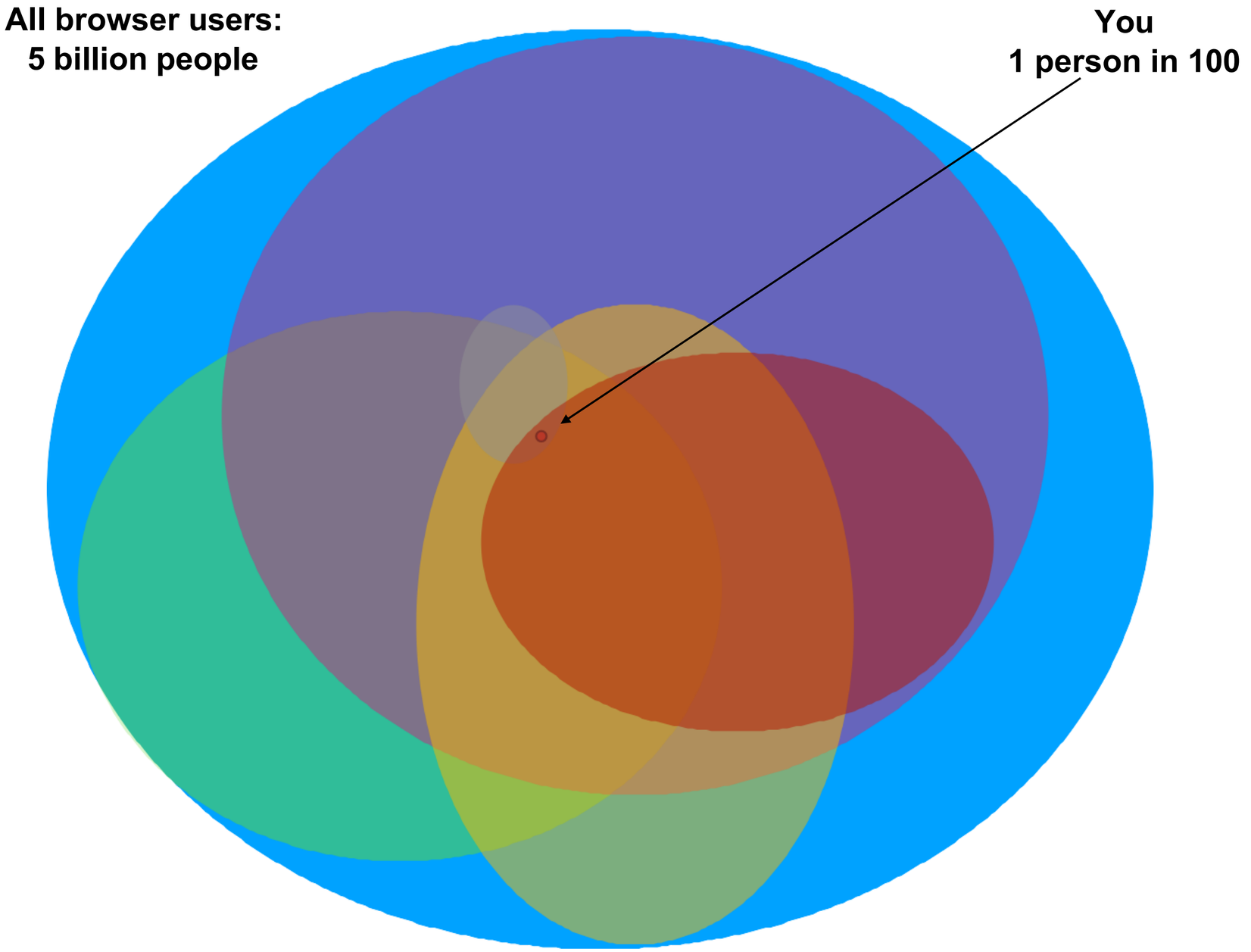

All browser users:
5 billion people
Firefox
Users


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 iIntentionally different, within each boundary

Fingerprinting: Defenses

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