Variations in Tracking In Relation To Geographic Location

Nathaniel Fruchter
Hsin Miao
Scott Stevenson
Rebecca Balebako

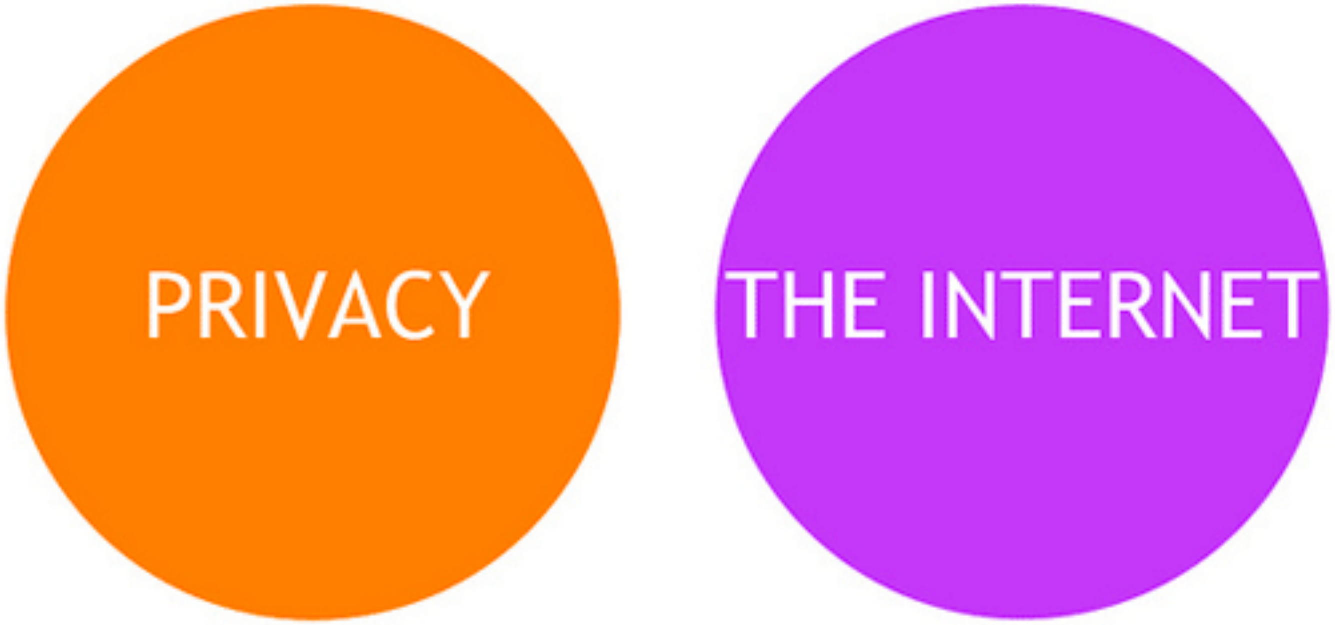
W2SP 2015

Carnegie Mellon University

The short version

- An empirical, automated method of measuring web tracking across countries
- Deployed in four countries representing three regulatory styles
- Significant differences found in amount of tracking
    - Where do these come from? Site > user.

Privacy and regulation



PRIVACY

THE INTERNET

A HELPFUL VENN DIAGRAM

Privacy

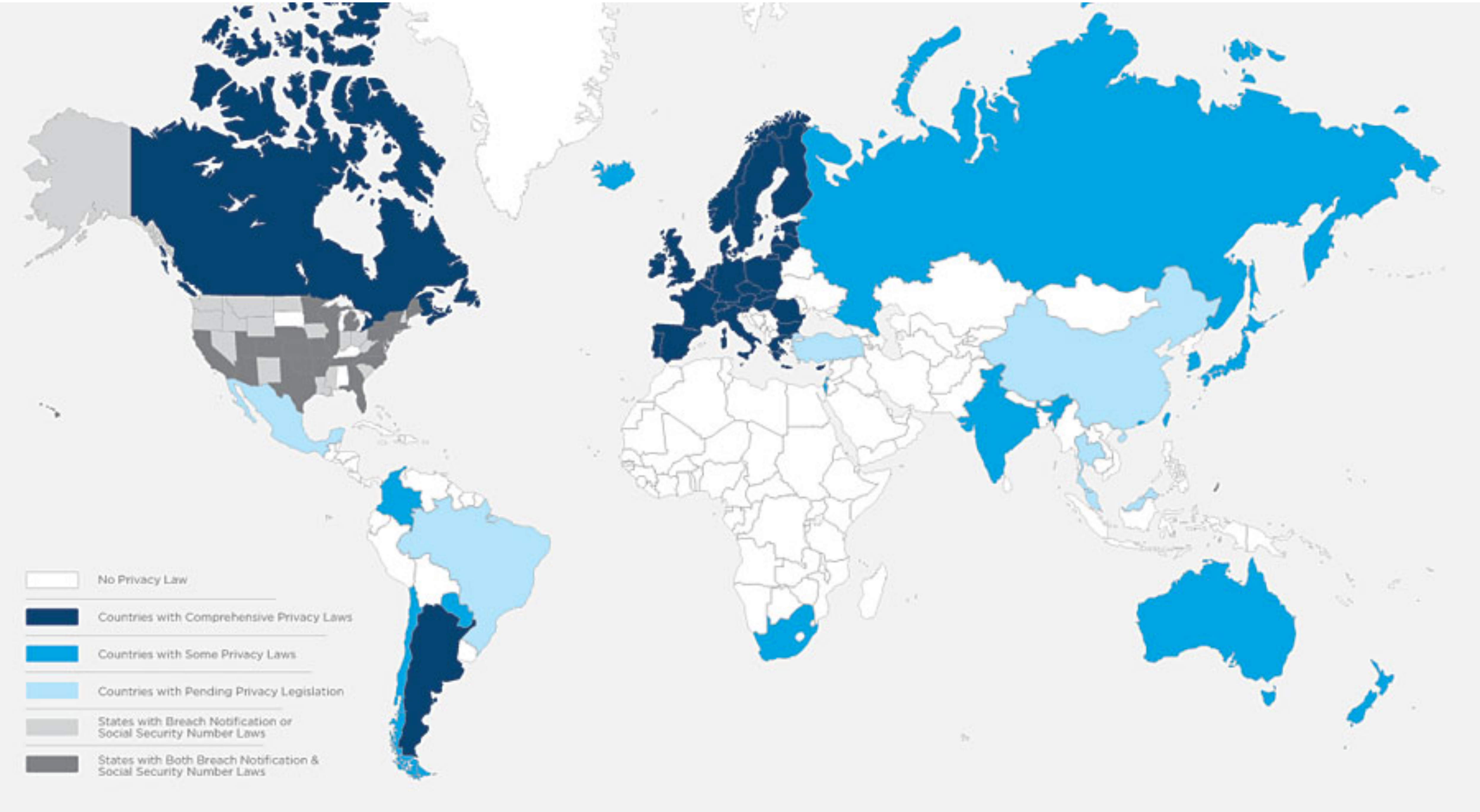
- It's **hard to define**.
- It's an **incredibly relative concept**: culturally, personally, technologically…
- It's an **incredibly dynamic concept** that changes along with many social and technological factors.

“Privacy is a value so complex, entangled in competing and contradictory dimensions, so engorged with various and distinct meanings… that I sometimes despair whether it can be usefully addressed at all.”

—Robert C. Post

Three Concepts of Privacy, 89 GEO. L.J. 2087, 2087 (2001).

This doesn’t really make for the easiest landscape when it comes to regulatory action…


No Privacy Law
Countries with Comprehensive Privacy Laws
Countries with Some Privacy Laws
Countries with Pending Privacy Legislation
States with Breach Notification or Social Security Number Laws
States with Both Breach Notification & Social Security Number Laws

Behunin & Associates, P.C.
http://sunsigndesigns.com/prod/behuninassociates/privacy.html

Regulatory Regimes

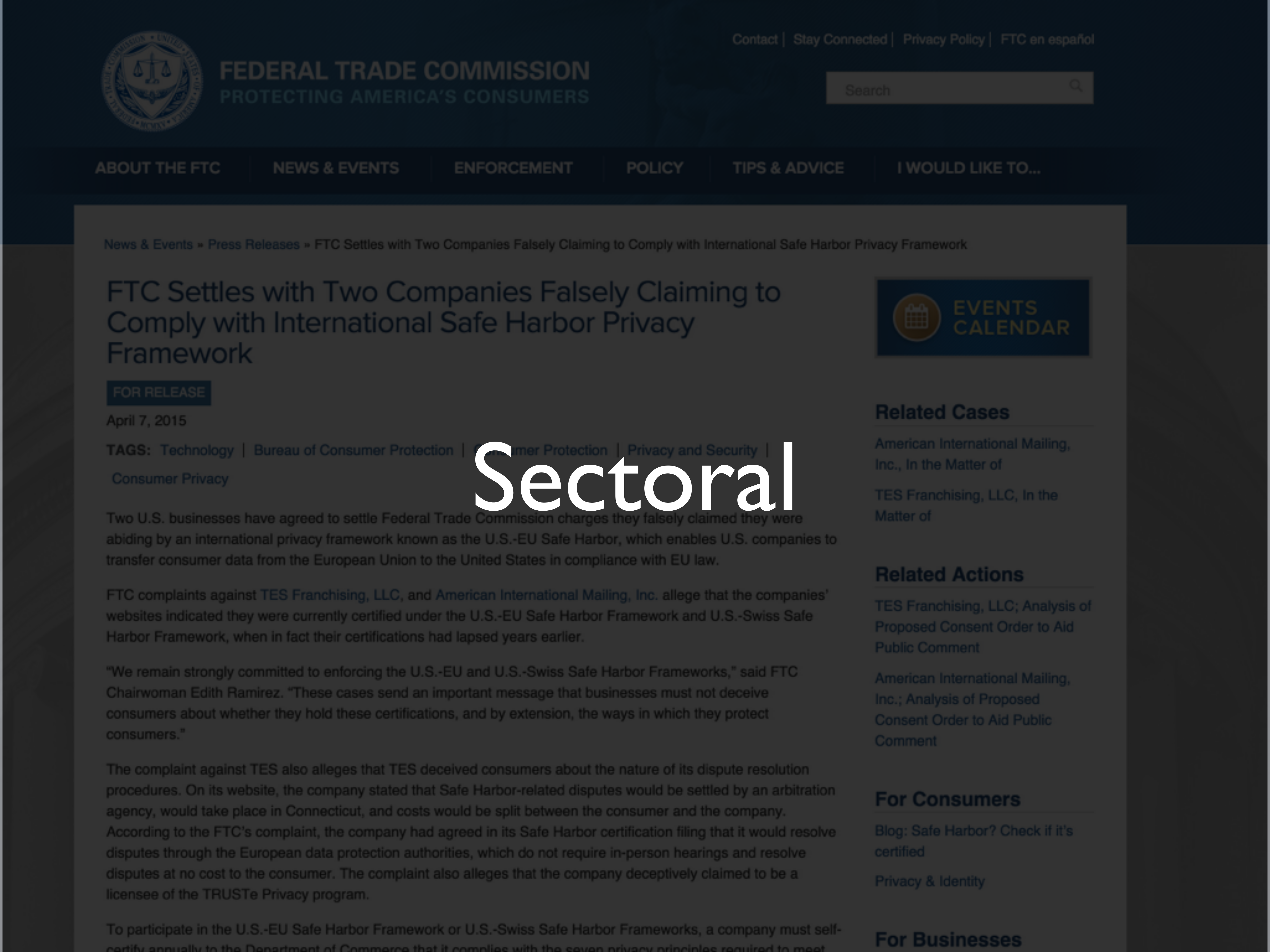
- Contrasting models of digital privacy regulation
  - Comprehensive (“European”)
  - Sectoral (“American”)
  - Co-regulatory
  - None/other
- Different philosophies and methods!

Comprehensive

Regulatory Regimes

- Comprehensive
  - Privacy is a fundamental right.
  - Legislated, top-down restrictions on collection, use, and disclosure.
  - Enforced by dedicated regulatory bodies.

Sectoral



News & Events » Press Releases » FTC Settles with Two Companies Falsely Claiming to Comply with International Safe Harbor Privacy Framework

FTC Settles with Two Companies Falsely Claiming to Comply with International Safe Harbor Privacy Framework

FOR RELEASE

April 7, 2015

TAGS: Technology | Bureau of Consumer Protection | Consumer Protection | Privacy and Security | Consumer Privacy

Two U.S. businesses have agreed to settle Federal Trade Commission charges they falsely claimed they were abiding by an international privacy framework known as the U.S.-EU Safe Harbor, which enables U.S. companies to transfer consumer data from the European Union to the United States in compliance with EU law.

FTC complaints against TES Franchising, LLC, and American International Mailing, Inc. allege that the companies' websites indicated they were currently certified under the U.S.-EU Safe Harbor Framework and U.S.-Swiss Safe Harbor Framework, when in fact their certifications had lapsed years earlier.

"We remain strongly committed to enforcing the U.S.-EU and U.S.-Swiss Safe Harbor Frameworks," said FTC Chairwoman Edith Ramirez. "These cases send an important message that businesses must not deceive consumers about whether they hold these certifications, and by extension, the ways in which they protect consumers."

The complaint against TES also alleges that TES deceived consumers about the nature of its dispute resolution procedures. On its website, the company stated that Safe Harbor-related disputes would be settled by an arbitration agency, would take place in Connecticut, and costs would be split between the consumer and the company. According to the FTC's complaint, the company had agreed in its Safe Harbor certification filing that it would resolve disputes through the European data protection authorities, which do not require in-person hearings and resolve disputes at no cost to the consumer. The complaint also alleges that the company deceptively claimed to be a licensee of the TRUSTe Privacy program.

To participate in the U.S.-EU Safe Harbor Framework or U.S.-Swiss Safe Harbor Frameworks, a company must self-certify annually to the Department of Commerce that it complies with the seven privacy principles required to meet

Related Cases

American International Mailing, Inc., In the Matter of

TES Franchising, LLC, In the Matter of

Related Actions

TES Franchising, LLC; Analysis of Proposed Consent Order to Aid Public Comment

American International Mailing, Inc.; Analysis of Proposed Consent Order to Aid Public Comment

For Consumers

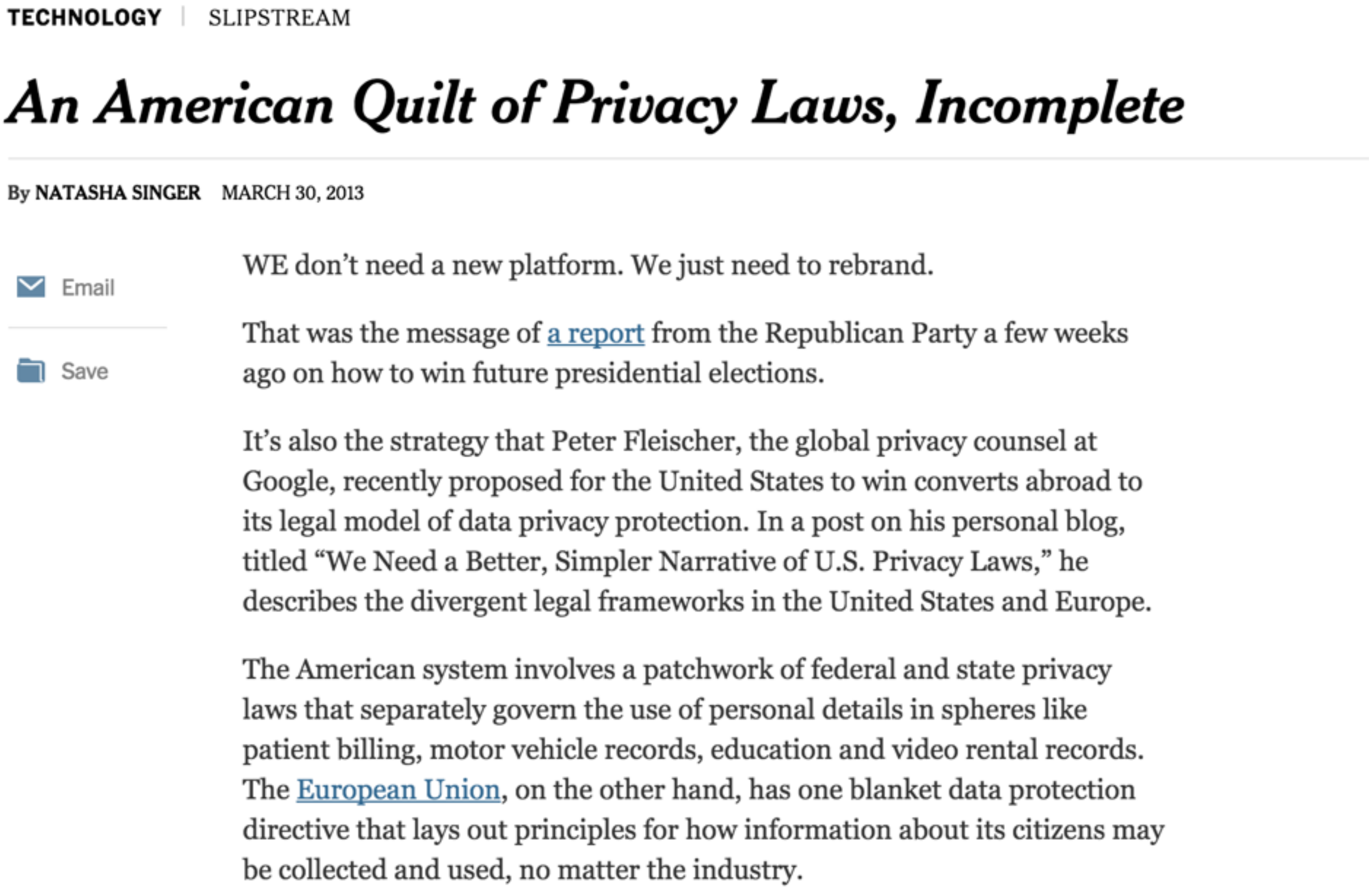
Blog: Safe Harbor? Check if it's certified

Privacy & Identity

For Businesses

Regulatory Regimes

- **Sectoral**
  - Fewer fundamental protections.
  - Privacy where it's deemed to be needed: more of a patchwork.
    - Health (HIPAA), children (COPPA)—differences between US states.
  - Emphasis on industry self-regulation and cooperation: "notice and choice"



TECHNOLOGY | SLIPSTREAM

An American Quilt of Privacy Laws, Incomplete

By NATASHA SINGER MARCH 30, 2013

WE don’t need a new platform. We just need to rebrand.

That was the message of a report from the Republican Party a few weeks ago on how to win future presidential elections.

It’s also the strategy that Peter Fleischer, the global privacy counsel at Google, recently proposed for the United States to win converts abroad to its legal model of data privacy protection. In a post on his personal blog, titled “We Need a Better, Simpler Narrative of U.S. Privacy Laws,” he describes the divergent legal frameworks in the United States and Europe.

The American system involves a patchwork of federal and state privacy laws that separately govern the use of personal details in spheres like patient billing, motor vehicle records, education and video rental records. The European Union, on the other hand, has one blanket data protection directive that lays out principles for how information about its citizens may be collected and used, no matter the industry.

Co-regulatory

Australian Government

Office of the Australian Information Commissioner

Privacy fact sheet 17

Australian Privacy Principles

January 2014

From 12 March 2014, the Australian Privacy Principles (APPs) will replace the National Privacy Principles and Information Privacy Principles and will apply to organisations, and Australian Government (and Norfolk Island Government) agencies.

This privacy fact sheet provides the text of the 13 APPs from Schedule 1 of the *Privacy Amendment (Enhancing Privacy Protection) Act 2012,* which amends the *Privacy Act 1988*. For the latest versions of these Acts visit the ComLaw website: www.comlaw.gov.au.

Part 1—Consideration of personal information privacy

Australian Privacy Principle 1—open and transparent management of personal information

1.1 The object of this principle is to ensure that APP entities manage personal information in an open and transparent way.

(b) how the entity collects and holds personal information;

(c) the purposes for which the entity collects, holds, uses and discloses personal information;

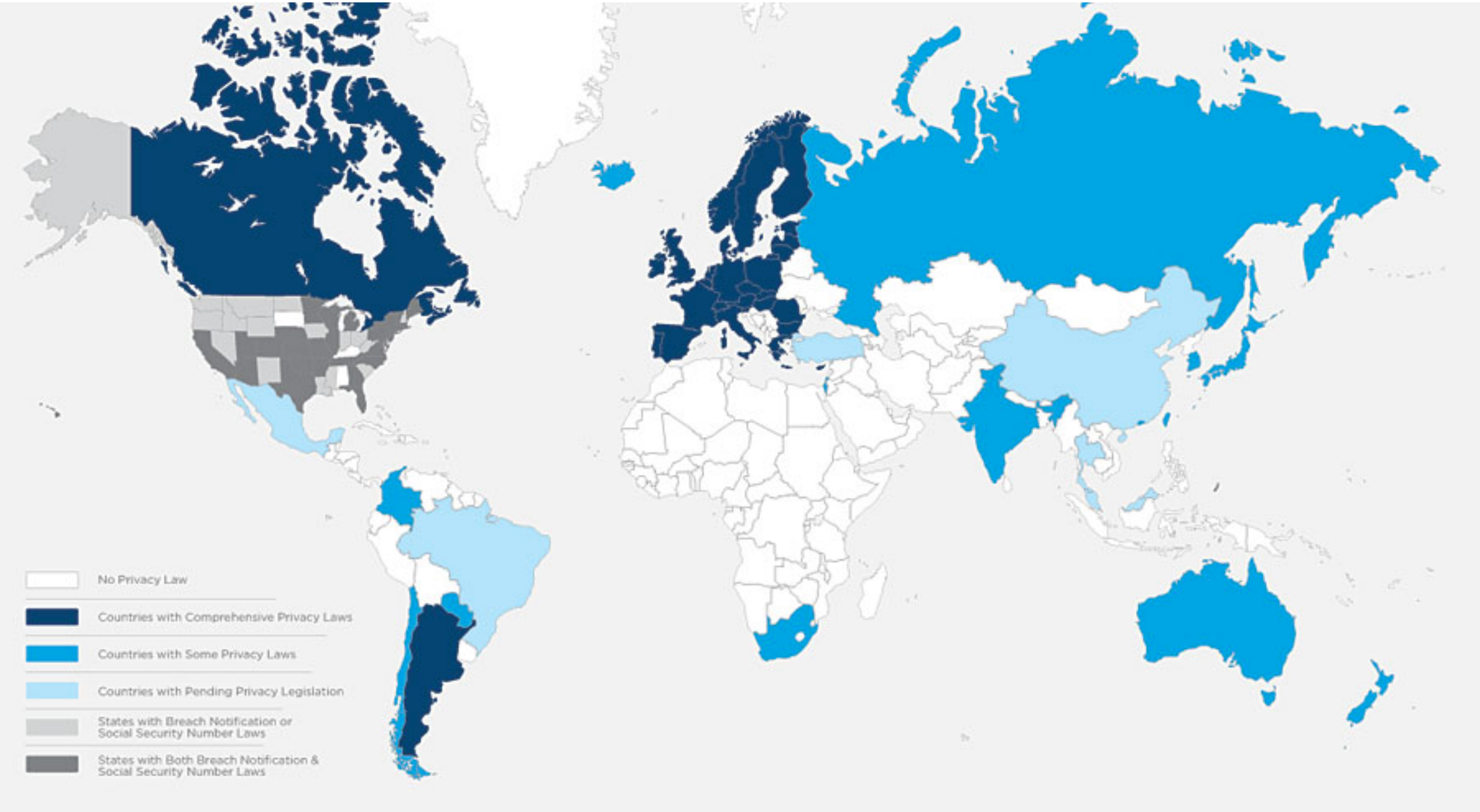
(d) how an individual may access personal information about the individual that is held by the entity and seek the correction of such information;

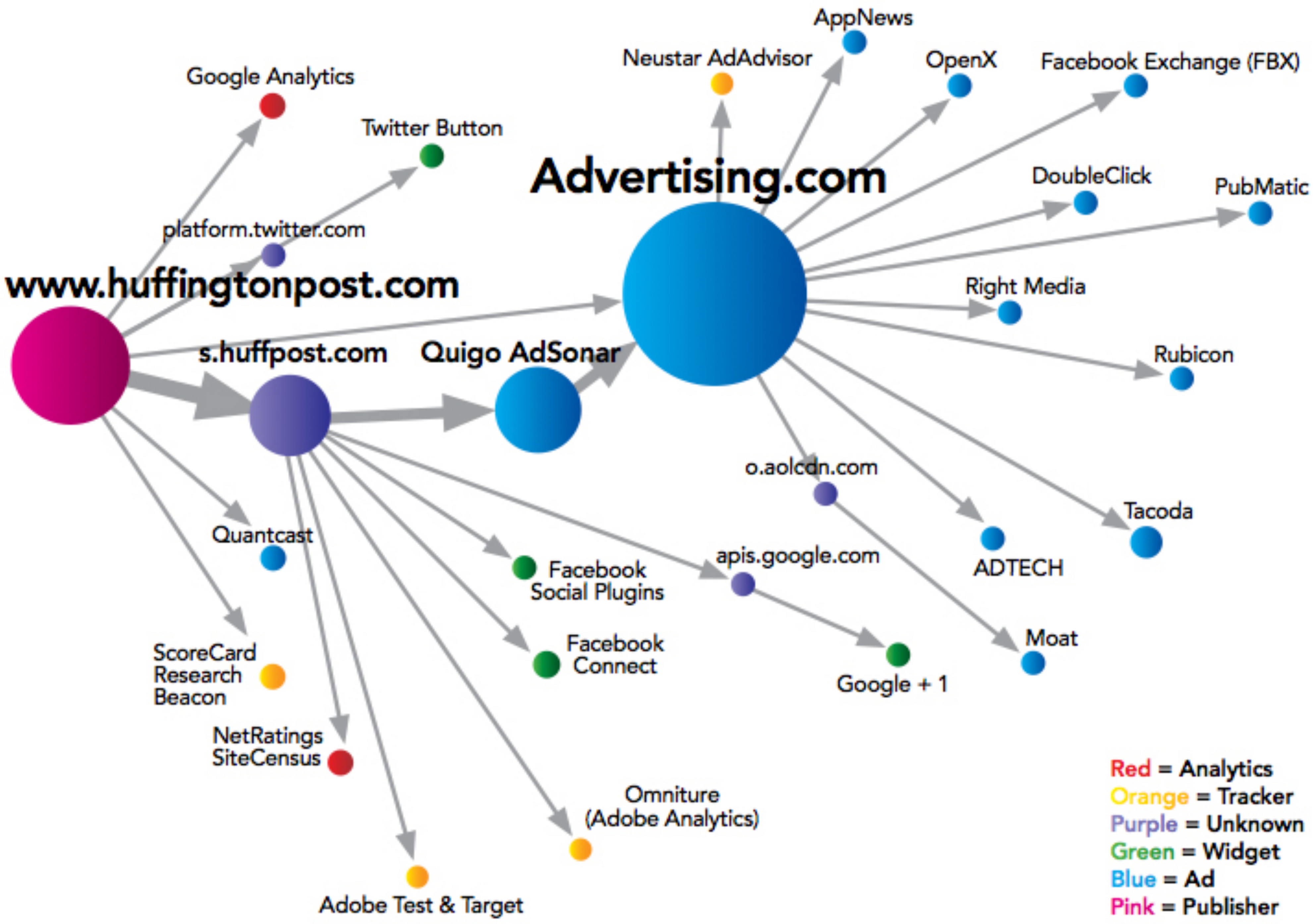
Regulatory Regimes

- **Co-regulatory**
  - Reliance on industry self-regulation with a government "backstop"
  - Industry bound to create enforceable codes
  - Most notably in Australia.

Regulatory Regimes

- **No regulation**
  - Lack of effective legislated privacy law


No Privacy Law
Countries with Comprehensive Privacy Laws
Countries with Some Privacy Laws
Countries with Pending Privacy Legislation
States with Breach Notification or Social Security Number Laws
States with Both Breach Notification & Social Security Number Laws



Evidon / Ghostery Enterprise, 2014

Do these regulatory (and geographic) differences lead to any quantifiable impact?

Do these regulatory (and geographic) differences lead to any quantifiable impact?

What is driving these differences?

Web measurement methods

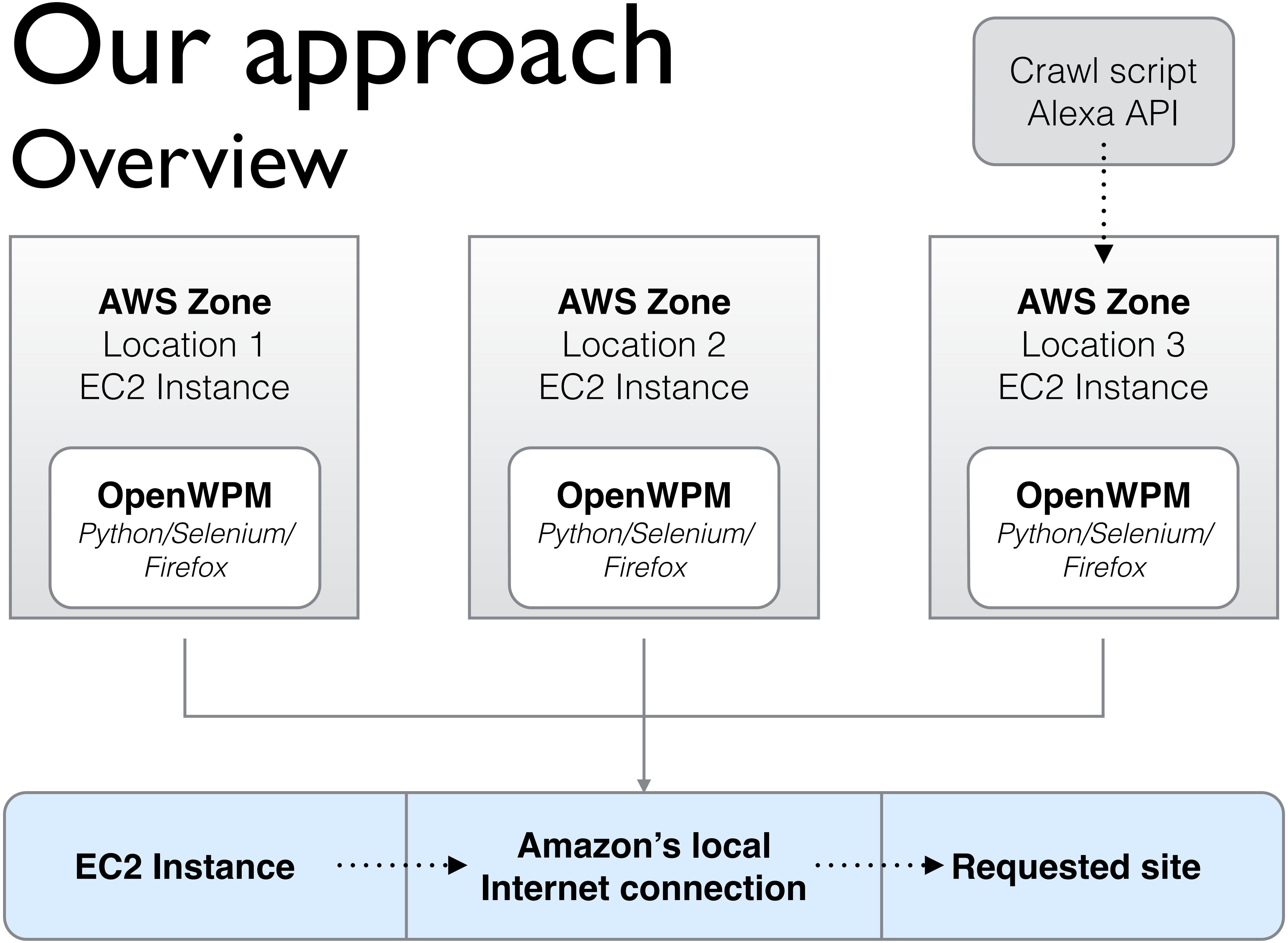
Web measurement

- Measuring what the user (and their browser) actually sees and receives
- Assessing and quantifying what happens "in the wild" in a variety of situations
- Challenges: automation, control, randomization, consistency

Our approach
Overview

- Standardized
  - Python + OpenWPM library
- Reproducible
  - Open source, scripted
- Empirical
  - Controlled, automated, no humans
- Realistic*
  - Flash, JavaScript, Firefox engine

Our approach
Overview

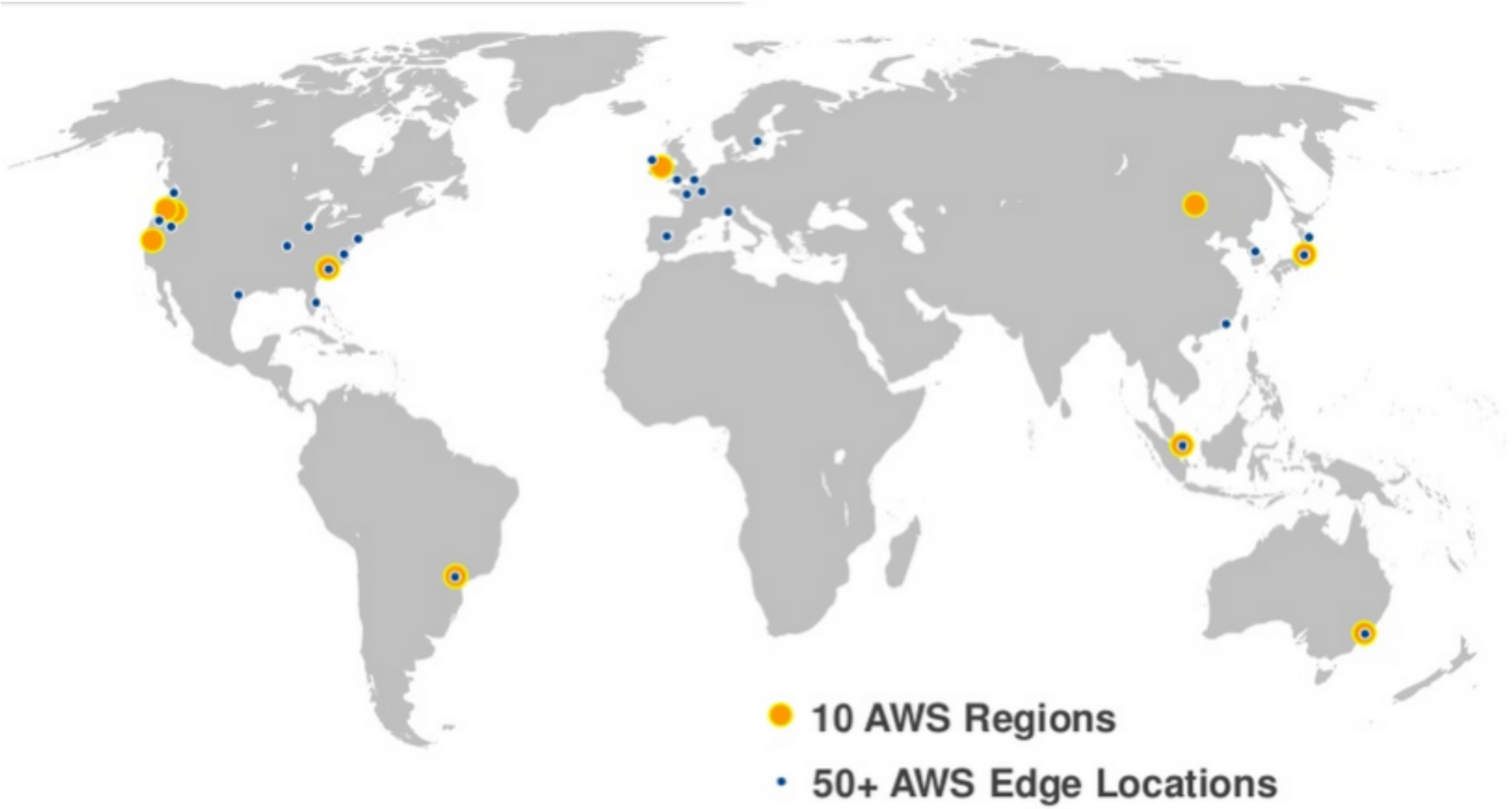


Our approach
Network infrastructure

- How do you source a network endpoint in different countries?
- Tor is a possibility, but messy to work with
- Sourcing VPNs is an unreliable process
- Both introduce extra confounds into the measurement process

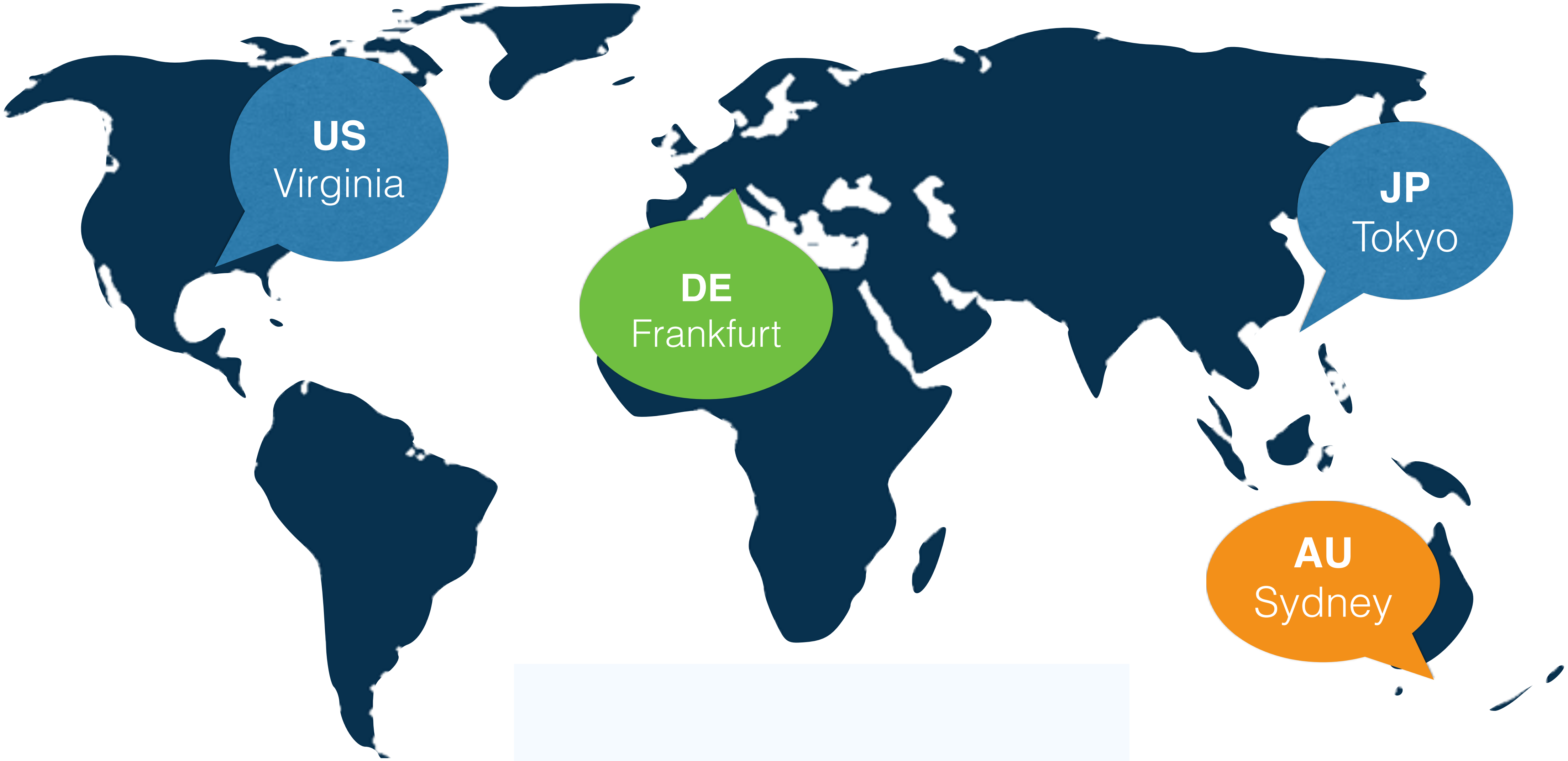
Our approach

Network infrastructure



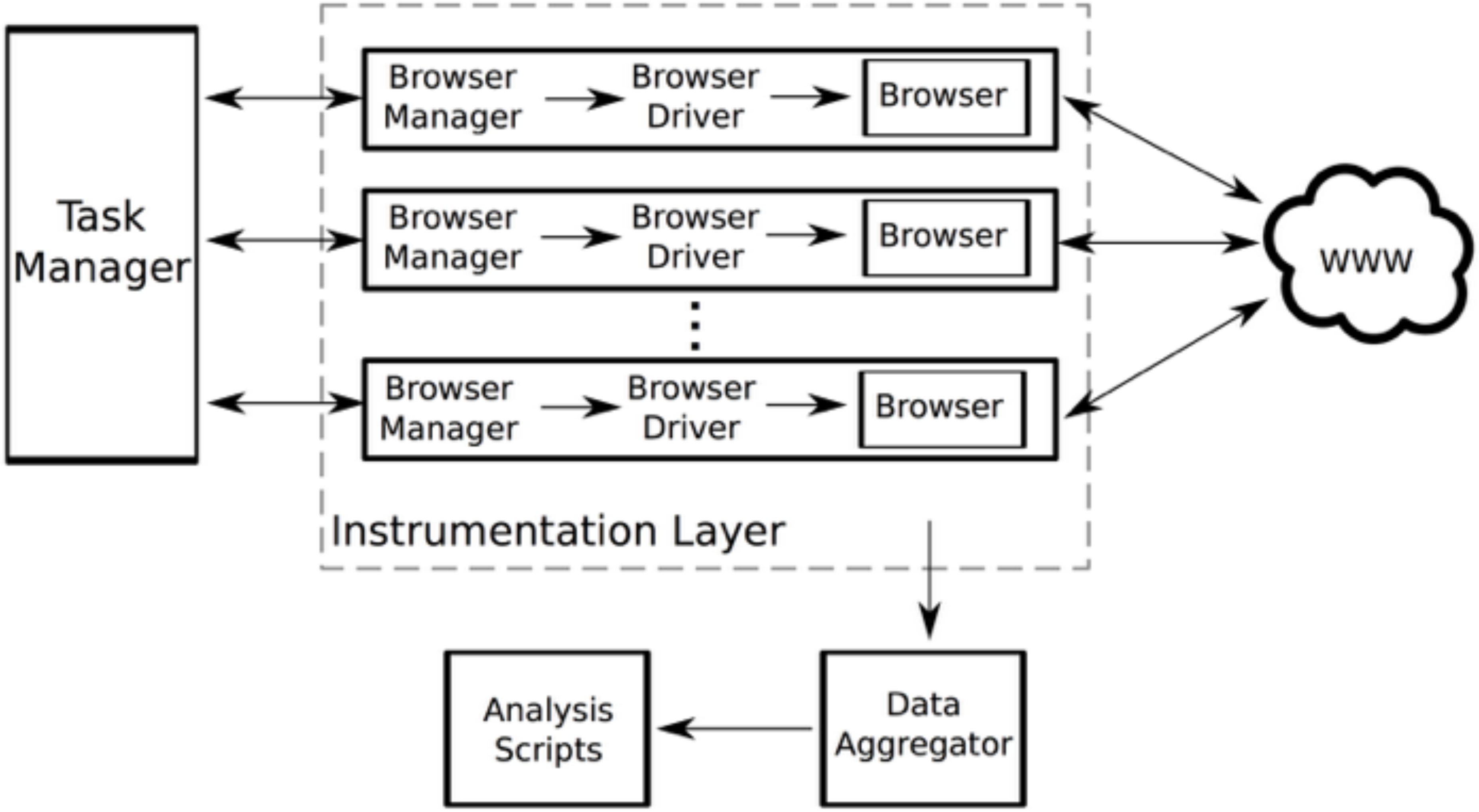
Our approach

Network infrastructure



Sectoral | Comprehensive | Co-regulatory

OpenWPM 0.2.1
(Engelhardt et al, 2014)



http://randomwalker.info/publications/WebPrivacyMeasurement.pdf

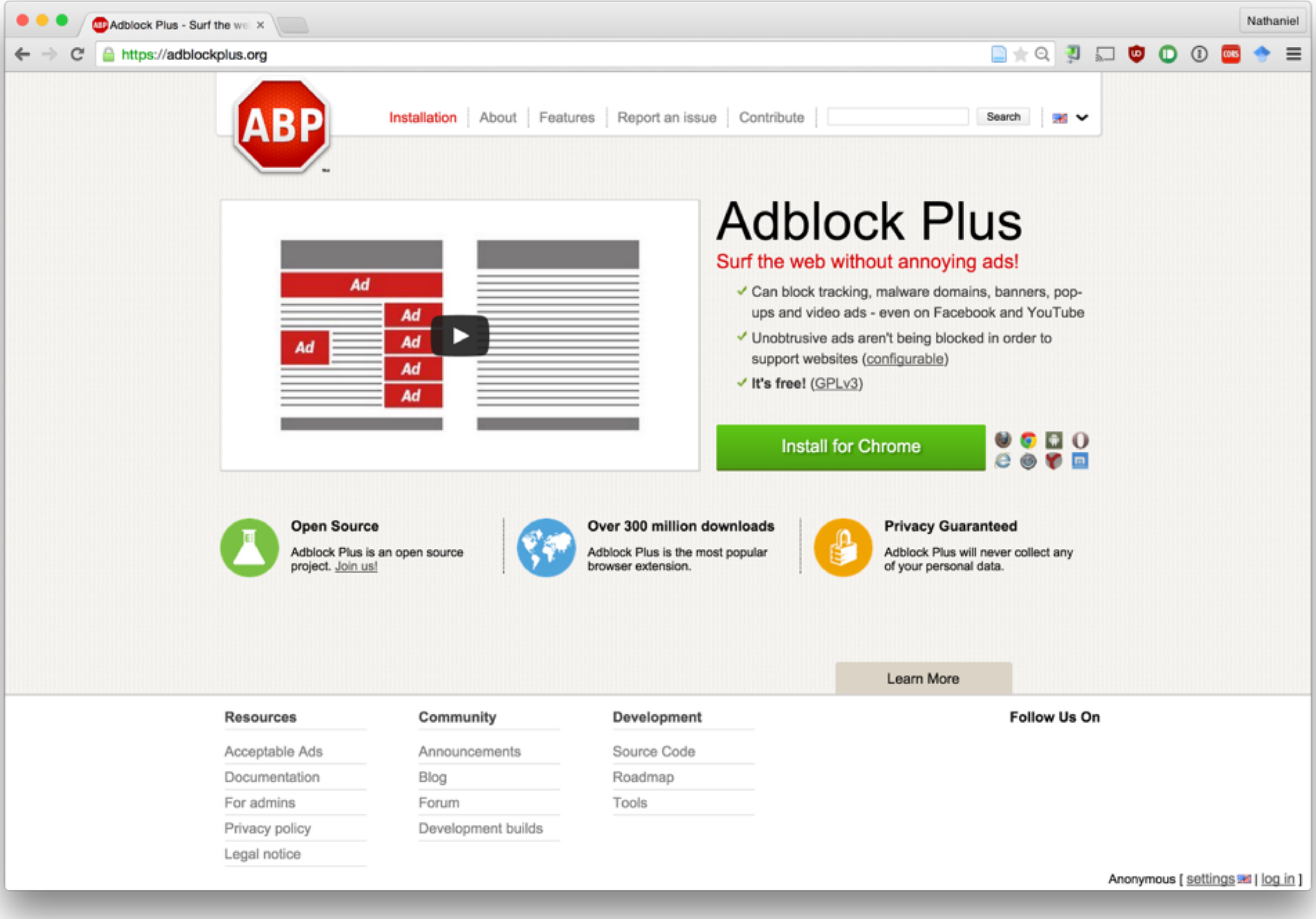
Our approach
Web crawling

- What do you crawl?
  - Alexa "Top Sites" API - Globally and by country
  - Some overlap (google.com), some localized (google.de), some local (spiegel.de)
- What do you record?
  - OpenWPM lets you do everything!

Our approach
Heuristics

- Approach A: **third-party HTTP requests and cookies.**
  - Rough metric, but can be representative
  - First-party requests have been exempted from definition of tracking/advertising (Do Not Track specification*)
- Approach B: match against a large **database of web assets** generally agreed upon as tracking

**McDonald and Peha (2011), "Track Gap: Policy Implications of User Expectations for the `Do Not Track' Internet Privacy Feature"*



Adblock Plus

Surf the web without annoying ads!

- ✓ Can block tracking, malware domains, banners, pop-ups and video ads - even on Facebook and YouTube
- ✓ Unobtrusive ads aren't being blocked in order to support websites (configurable)
- ✓ **It's free!** (GPLv3)

Open Source

Adblock Plus is an open source project. Join us!

Over 300 million downloads

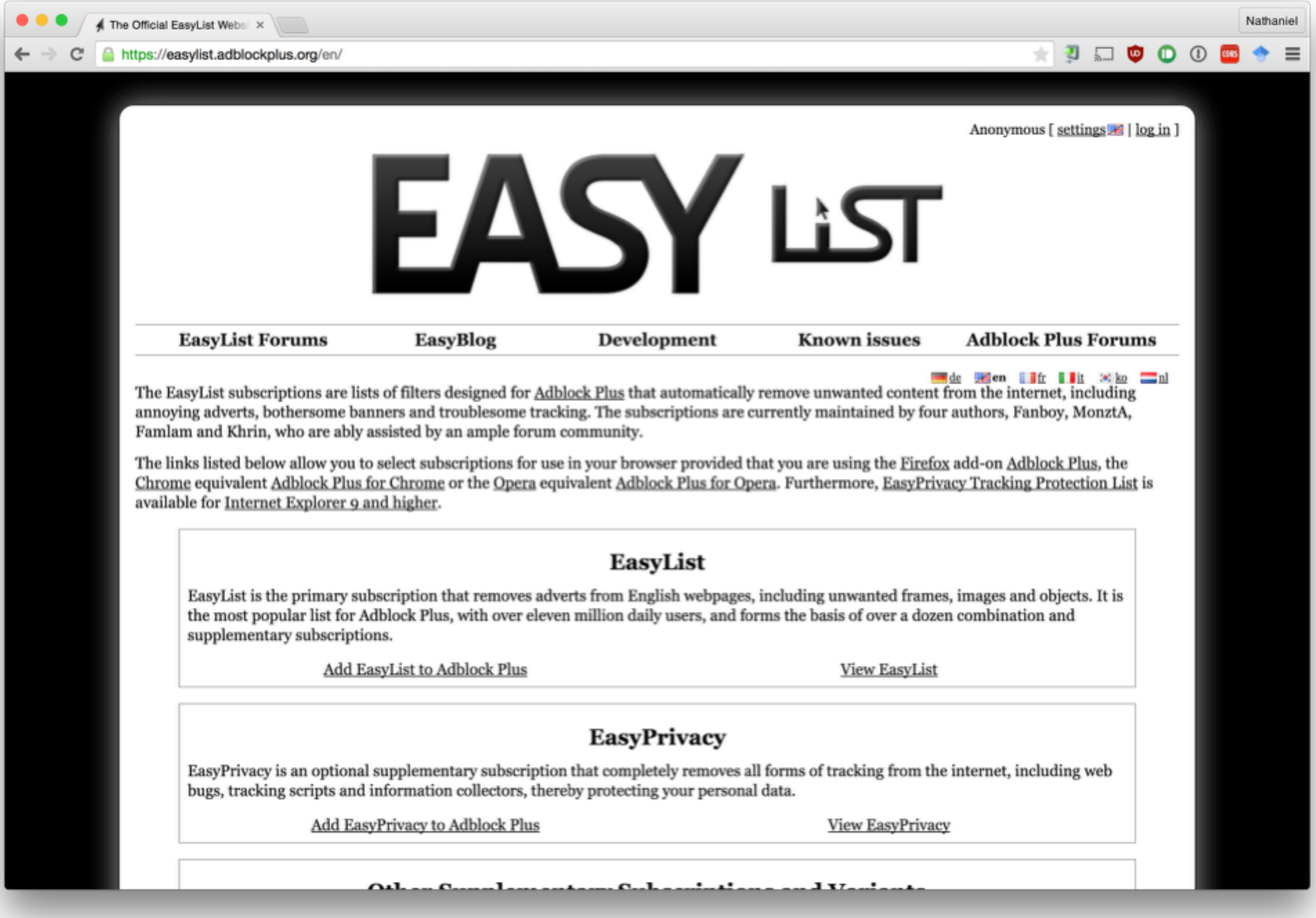
Adblock Plus is the most popular browser extension.

Privacy Guaranteed

Adblock Plus will never collect any of your personal data.

Resources	Community	Development	Follow Us On
Acceptable Ads	Announcements	Source Code	
Documentation	Blog	Roadmap	
For admins	Forum	Tools	
Privacy policy	Development builds		
Legal notice			

Anonymous [settings | log in]



The EasyList subscriptions are lists of filters designed for Adblock Plus that automatically remove unwanted content from the internet, including annoying adverts, bothersome banners and troublesome tracking. The subscriptions are currently maintained by four authors, Fanboy, MonztA, Famlam and Khrin, who are ably assisted by an ample forum community.

The links listed below allow you to select subscriptions for use in your browser provided that you are using the Firefox add-on Adblock Plus, the Chrome equivalent Adblock Plus for Chrome or the Opera equivalent Adblock Plus for Opera. Furthermore, EasyPrivacy Tracking Protection List is available for Internet Explorer 9 and higher.

EasyList

EasyList is the primary subscription that removes adverts from English webpages, including unwanted frames, images and objects. It is the most popular list for Adblock Plus, with over eleven million daily users, and forms the basis of over a dozen combination and supplementary subscriptions.

EasyPrivacy

EasyPrivacy is an optional supplementary subscription that completely removes all forms of tracking from the internet, including web bugs, tracking scripts and information collectors, thereby protecting your personal data.

Other Supplementary Subscriptions and Variants

Our approach
Heuristics

- Approach B: parse and match against **open-source ad blocking rulesets**
  - We chose EasyList, the most commonly used and distributed AdBlock list
    - EasyList Ads and EasyPrivacy list
    - Over 50,000 regex-based rules
  - *adblockparser* Python module*

* *https://github.com/scrapinghub/adblockparser*

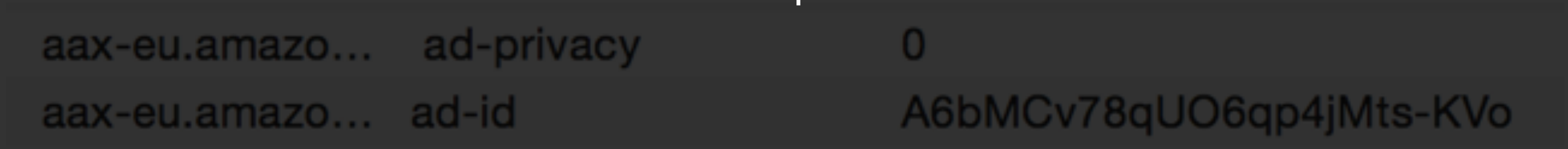
Our approach

Analysis

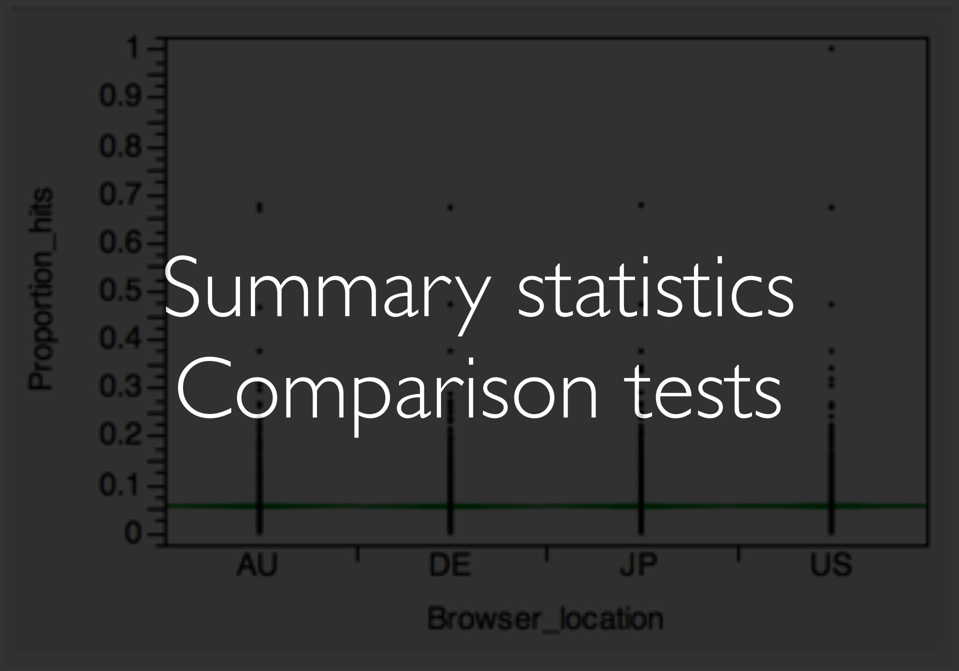
```
ssl-images-amazon.com/images/js/live/adSnippet._V142890782_.js
```

+

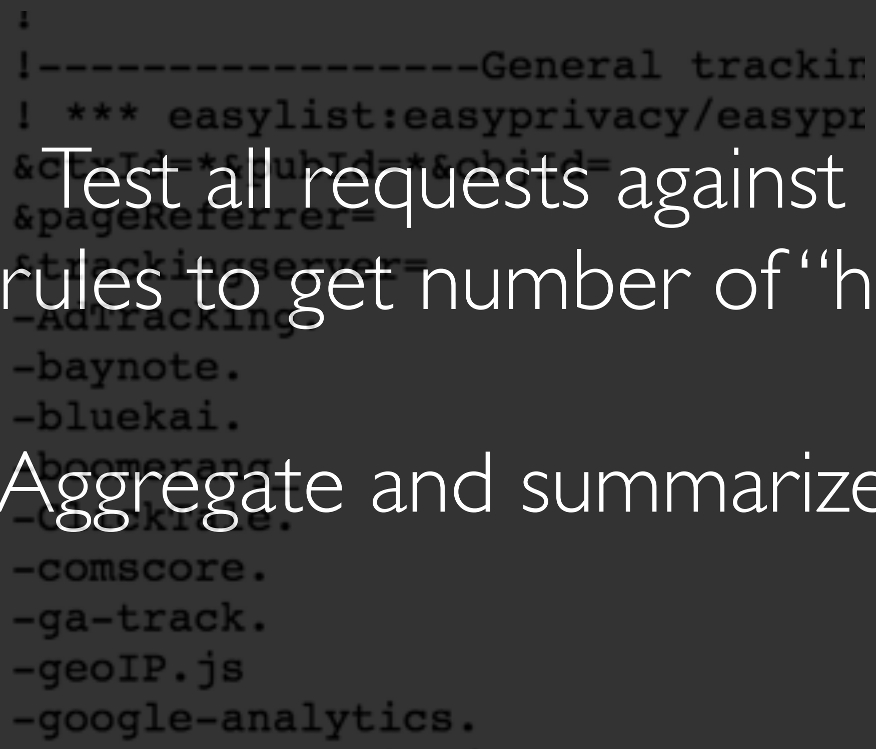
Extract full URLs from HTTP requests, domains from set cookies



aax-eu.amazo...	ad-privacy	0
aax-eu.amazo...	ad-id	A6bMCv78qUO6qp4jMts-KVo



Summary statistics
Comparison tests



```
!
!-----------------General trackin
! *** easylist:easyprivacy/easypr
&ctxId=*&pubId=*&cbId=
&pageReferrer=
&trackingserver=
-AdTracking.
-baynote.
-bluekai.
-boomerang.
-clicktale.
-comscore.
-ga-track.
-geoIP.js
-google-analytics.
```

Test all requests against
all rules to get number of “hits”

Aggregate and summarize

Key observations

Third-party requests/cookies

- Rank test against totals and normalized ratios

Requests	
US	1
AU	2
DE	3
JP	4

p < 0.0005

n.s.

p < 0.0005

Cookies	
US	1
DE	2
AU	3
JP	4

p < 0.05

all n.s.

Third-party requests/cookies

- The United States has significantly more activity across both metrics
- Interesting differences across countries and models
  - Caveat: sample representativeness

Ad blocking rules

Origin-dependent activity

- Does tracking activity change depending on the origin of the user *or* the origin of the website?
- How much do we need to control for geographic factors?
- Synchronized crawl of top 500 global websites (same sites from different locations)
- No significant differences!

Ad blocking rules

Country-level results

Country	*Average* **requests/page**	*Average* **hits/page**	*Average* **% hits**
AU	99.2	6.8	**6%**
DE	121.0	5.7	**5%**
JP	103.2	4.1	**5%**
US	120.6	9.3	**8%**

Ad blocking rules

Country-level results

Country A	Country B	Z	p	95% CI For Change
US	JP	10.42	<.0001	**[0.028, 0.040]**
US	DE	7.77	<.0001	**[0.018, 0.031]**
US	AU	2.57	<.02	**[0.001, 0.014]**
JP	DE	-3.64	<.0005	**[-0.013, -0.002]**
DE	AU	-5.29	<.0001	**[-0.021, -0.009]**
AU	AU	-8.33	<.0001	**[-0.031, -0.019]**

Ad blocking rules

Results

- Trackers accounted for **1.5 - 2.1% more** requests compared to advertisements
  - Considering that both make up less than 6% of total page assets…
  - User awareness

Ad blocking rules

Results

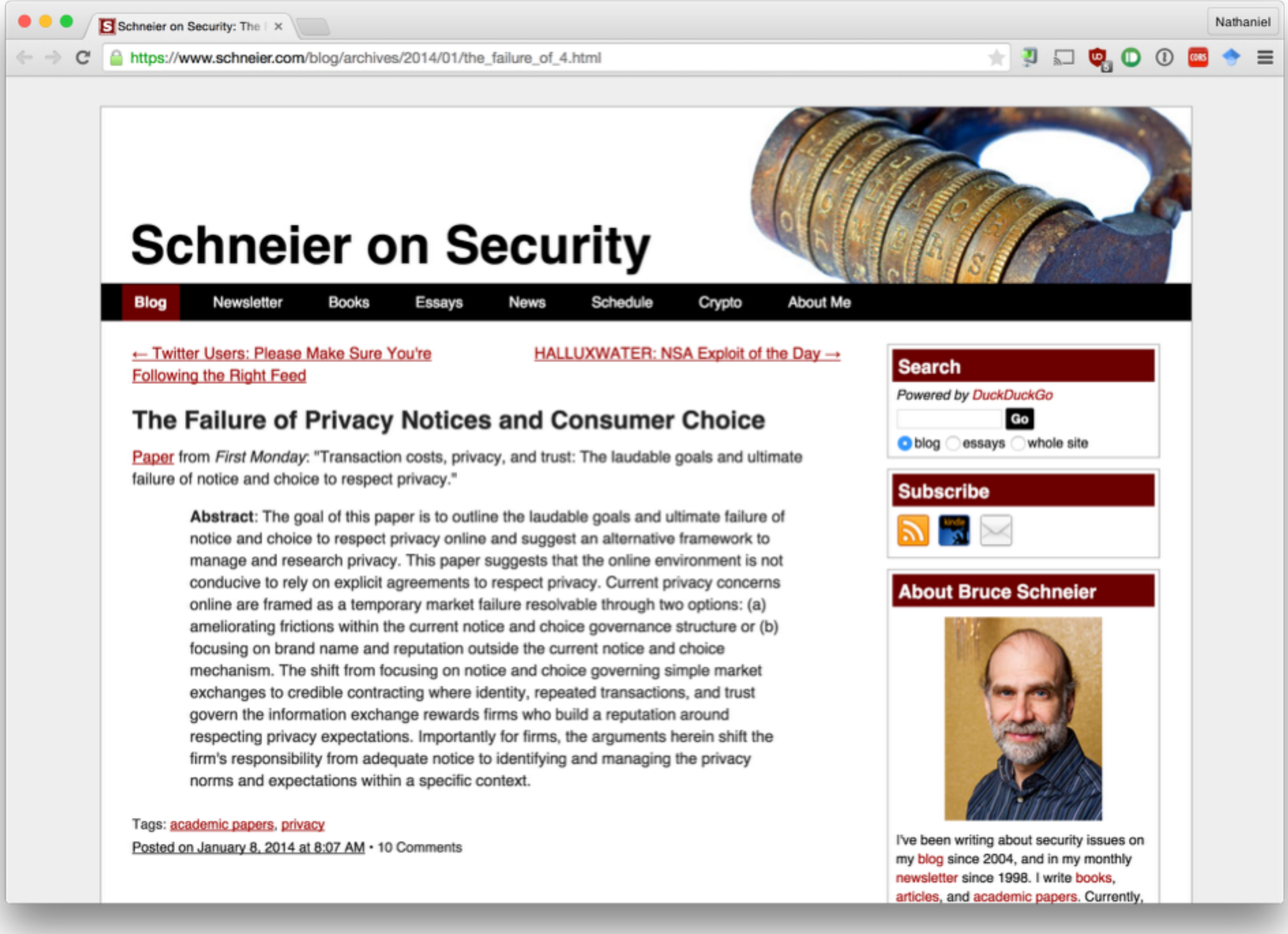
- Significant differences between all pairs of countries
  - United States: more activity in all cases
    - 0.1% compared to Australia
    - 4% compared to Japan
- 4% x ~100 average requests = 4+ tracking elements

Challenges

The policy lifecycle

- **Development**: Recognize, diagnose, identify institutions, evaluate options
- **“In the wild”**: Implement, enforce, monitor (the hard part)

Wheelan (2010)


Schneier on Security: The F
https://www.schneier.com/blog/archives/2014/01/the_failure_of_4.html
Nathaniel

Schneier on Security

Blog
Newsletter
Books
Essays
News
Schedule
Crypto
About Me

← Twitter Users: Please Make Sure You're Following the Right Feed

HALLUXWATER: NSA Exploit of the Day →

The Failure of Privacy Notices and Consumer Choice

Paper from *First Monday*: "Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy."

> **Abstract**: The goal of this paper is to outline the laudable goals and ultimate failure of notice and choice to respect privacy online and suggest an alternative framework to manage and research privacy. This paper suggests that the online environment is not conducive to rely on explicit agreements to respect privacy. Current privacy concerns online are framed as a temporary market failure resolvable through two options: (a) ameliorating frictions within the current notice and choice governance structure or (b) focusing on brand name and reputation outside the current notice and choice mechanism. The shift from focusing on notice and choice governing simple market exchanges to credible contracting where identity, repeated transactions, and trust govern the information exchange rewards firms who build a reputation around respecting privacy expectations. Importantly for firms, the arguments herein shift the firm's responsibility from adequate notice to identifying and managing the privacy norms and expectations within a specific context.

Tags: academic papers, privacy

Posted on January 8, 2014 at 8:07 AM • 10 Comments

Search

Powered by DuckDuckGo

Go

● blog ○ essays ○ whole site

Subscribe

About Bruce Schneier

I've been writing about security issues on my blog since 2004, and in my monthly newsletter since 1998. I write books, articles, and academic papers. Currently,

https://www.schneier.com/blog/archives/2014/01/the_failure_of_4.html

Policy challenges

- Are these regulatory models doing what they're supposed to?
- Is this (admittedly narrow) viewpoint where we would see the effect? If not, where else?
- How do you define a privacy standard? How do you translate it?

Cultural challenges

- US vs. Japan: sectoral vs. sectoral
  - Why does the US have more tracking?
  - Cultural practices, business norms, "Internet ecosystem", what's popular
- Website business models
  - Outliers: news websites? (6000+ cookies!)

Cultural challenges

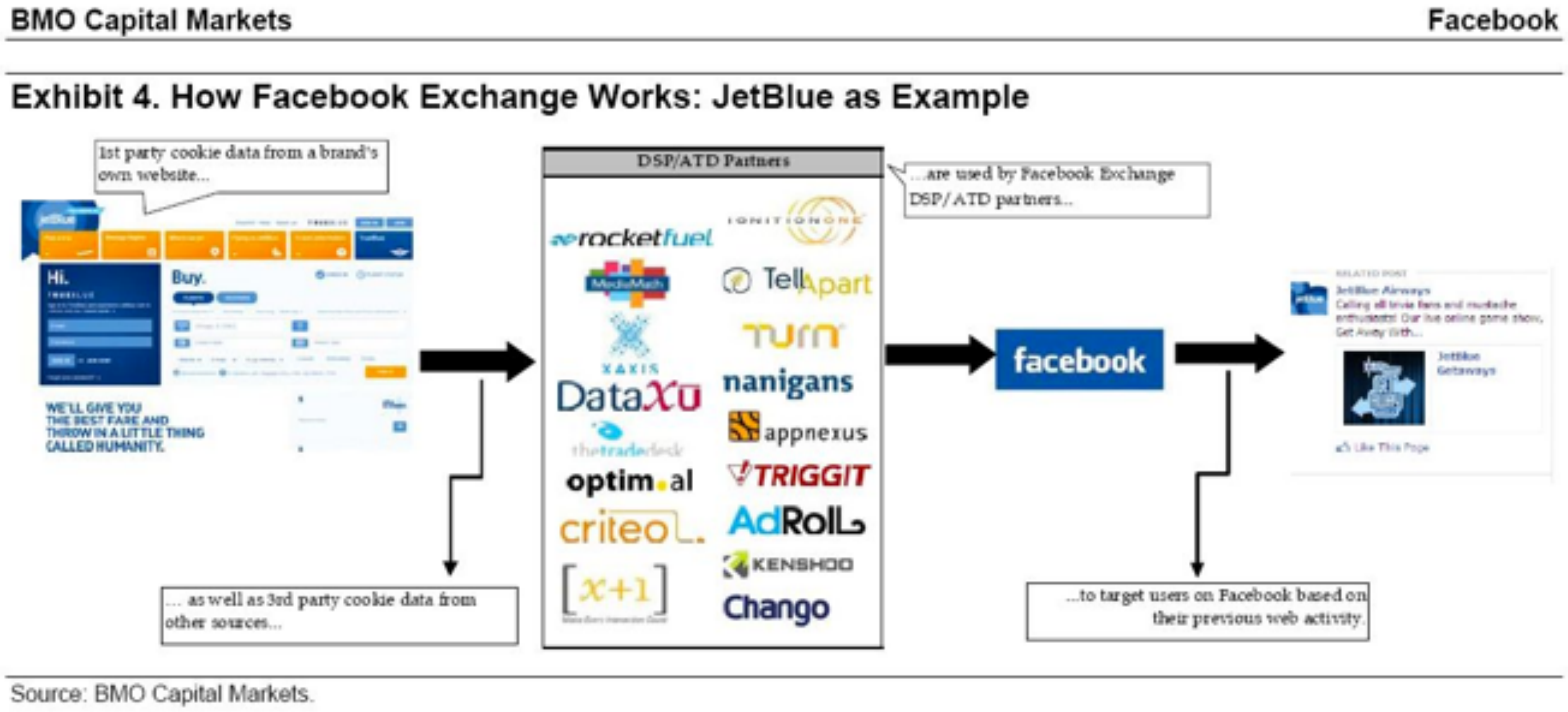
- How does culture affect Internet use?
- How do we intersect this with businesses' data collection habits?

Technical challenges

- What if the Internet looked a bit different?
  - China, other "interesting places"

Technical challenges

- Is first-party still a relevant distinction?
- Inter-session, inter-device, and more pervasive forms of tracking



BMO Capital Markets Facebook

Exhibit 4. How Facebook Exchange Works: JetBlue as Example

1st party cookie data from a brand's own website...

DSP/ATD Partners

...are used by Facebook Exchange DSP/ATD partners...

... as well as 3rd party cookie data from other sources...

...to target users on Facebook based on their previous web activity.

Source: BMO Capital Markets.

http://www.businessinsider.com.au/how-facebooks-fbx-ad-exchange-works-2013-1

Technical challenges

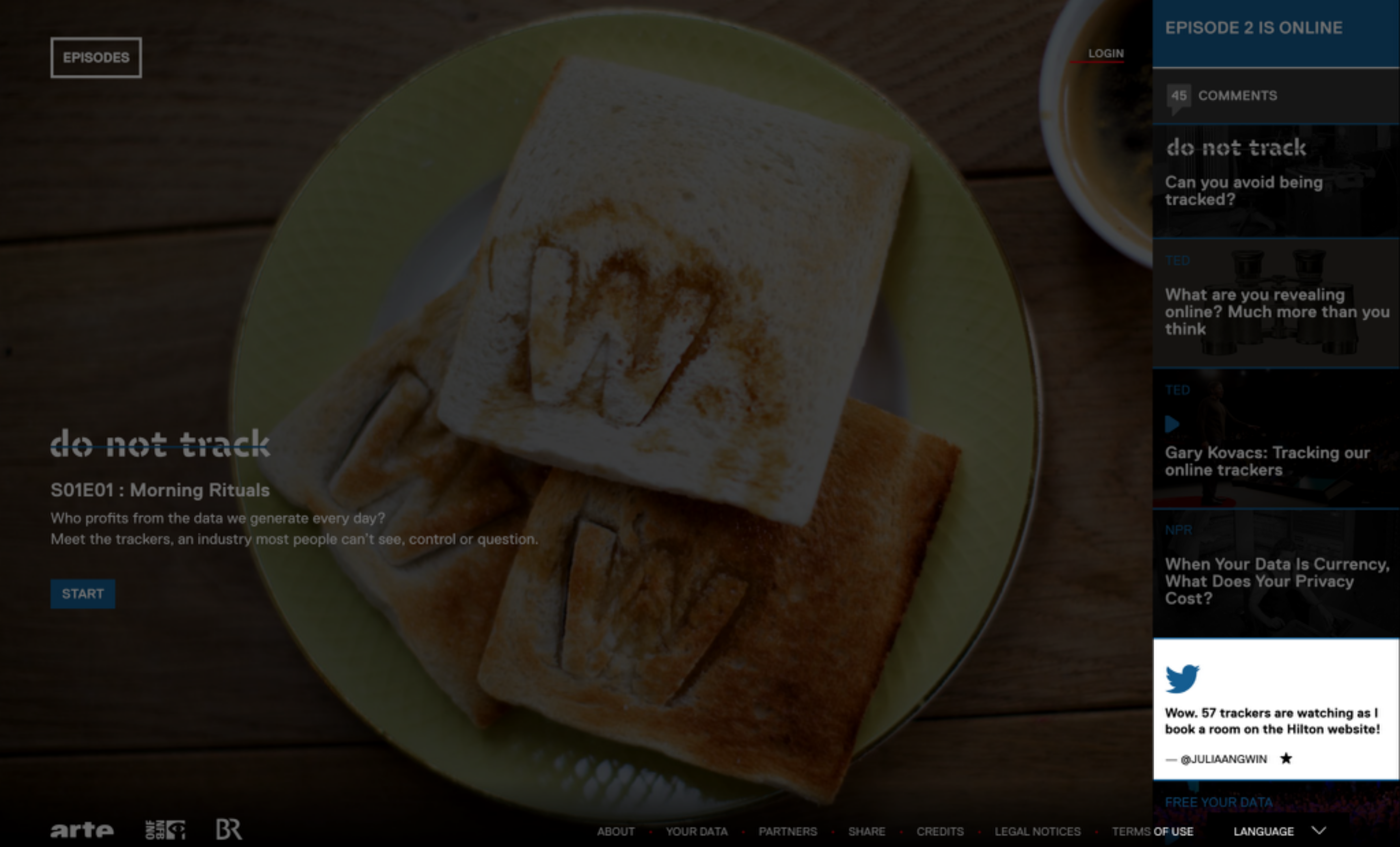
- **Is online / web activity deterministic?**
  - Page loads
  - People
  - Devices
  - Locations
  - Internet connections
  - The list goes on…

Keep in mind…

- Limited sampling base (more internet connections needed!)
- Differences within regulatory models
- You can always use more controls
    - Time of day, changes in sites, ISP policy, browser type, numerous other variables
- Replication!

At the end of the day

- How effective are regulatory models for protecting end users?



do not track

S01E01 : Morning Rituals

Who profits from the data we generate every day?
Meet the trackers, an industry most people can't see, control or question.

do not track

Can you avoid being tracked?

What are you revealing online? Much more than you think

Gary Kovacs: Tracking our online trackers

When Your Data Is Currency, What Does Your Privacy Cost?

Wow. 57 trackers are watching as I book a room on the Hilton website!

— @JULIAANGWIN ★

https://donottrack-doc.com *(April 2015)*

Thank you!

Questions?

Nathaniel Fruchter <fruchter@cmu.edu>
Hsin Miao <hsinm@andrew.cmu.edu>
Scott Stevenson <sbsteven@andrew.cmu.edu>
Rebecca Balebako <balebako@rand.org>

Carnegie Mellon University