

Do Not Track

The Future of Web Privacy

Nick Doty

UC Berkeley, School of Information

World Wide Web Consortium

<http://npdoty.name>

who I am

"future of" a clarification

not that Do Not Track is a solution to all Web privacy problems
or that derivations of this work are going to be the pattern for all future privacy issues
but the technical architecture provides hints at potential directions for Web privacy
and that the process we're going through (and its success/failure) will spell

these comments are my own, certainly not an official position of W3C or its members

therefore you can attribute all scatterbrained ideas to me and all the coherent brilliance to the WG and industry members

Agenda

- How we got here
- The current state of Do Not Track
- Trends for Web privacy
- Call for participation

to see how we got here, let's appropriately start with a few maps

Display Advertising Technology Landscape



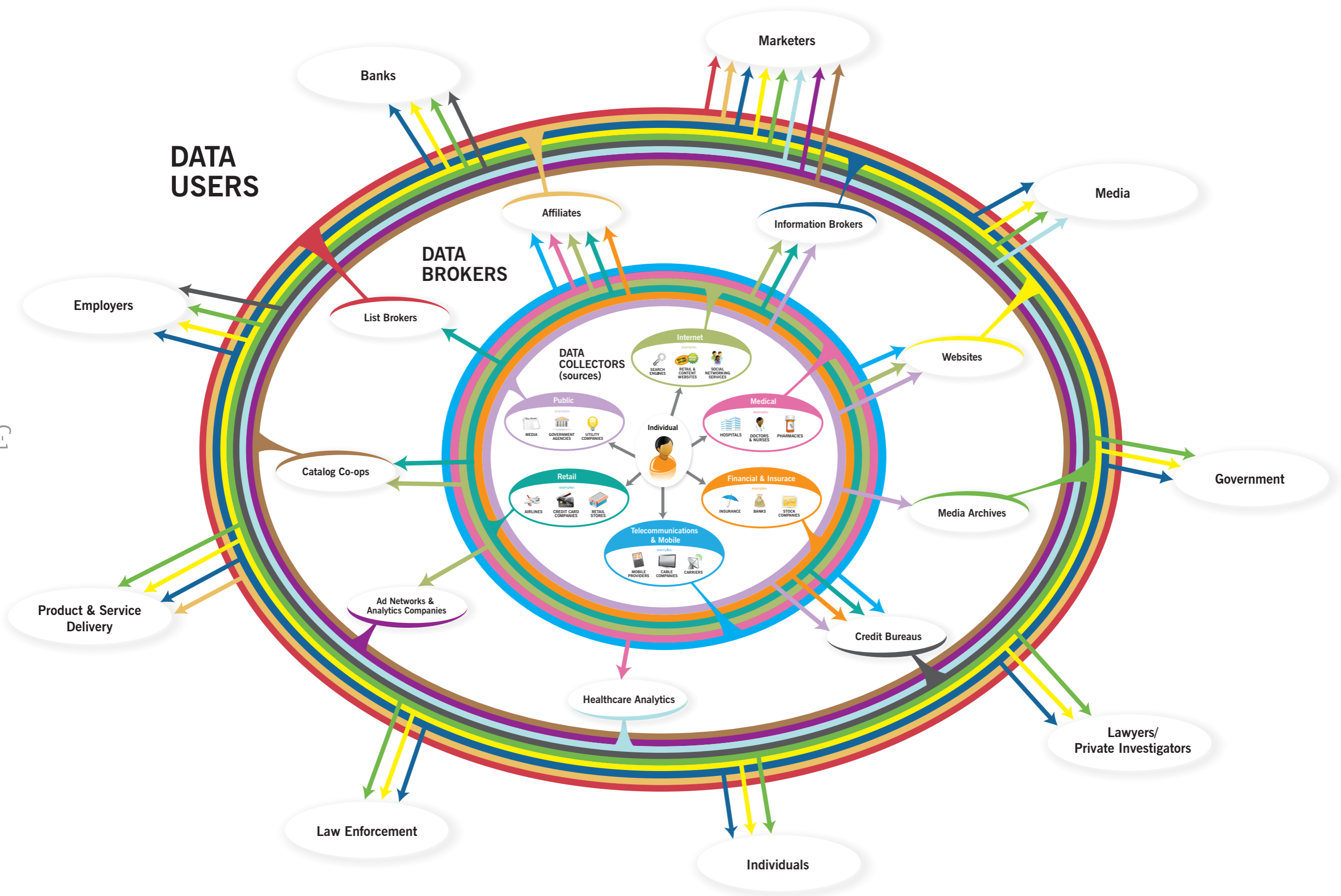
From LUMA Partners, and slightly out of date, this is the 2010 version

the multi-faceted chains of online advertising provide a shocking list of companies involved



In a way this diagram, from the Future of Privacy Forum, gets at the key idea even more clearly, that the user is at the center and while server-to-server communications happen too, the user and their browser is unknowingly in communication with many of these players directly.

Personal Data Ecosystem

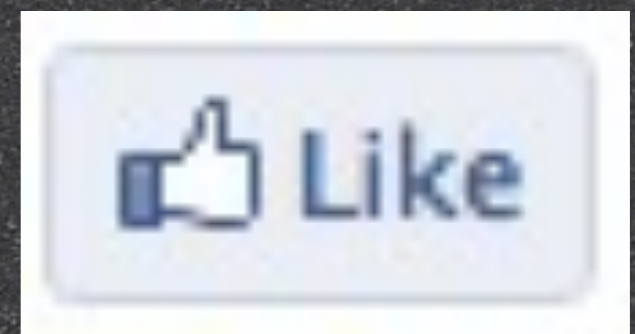
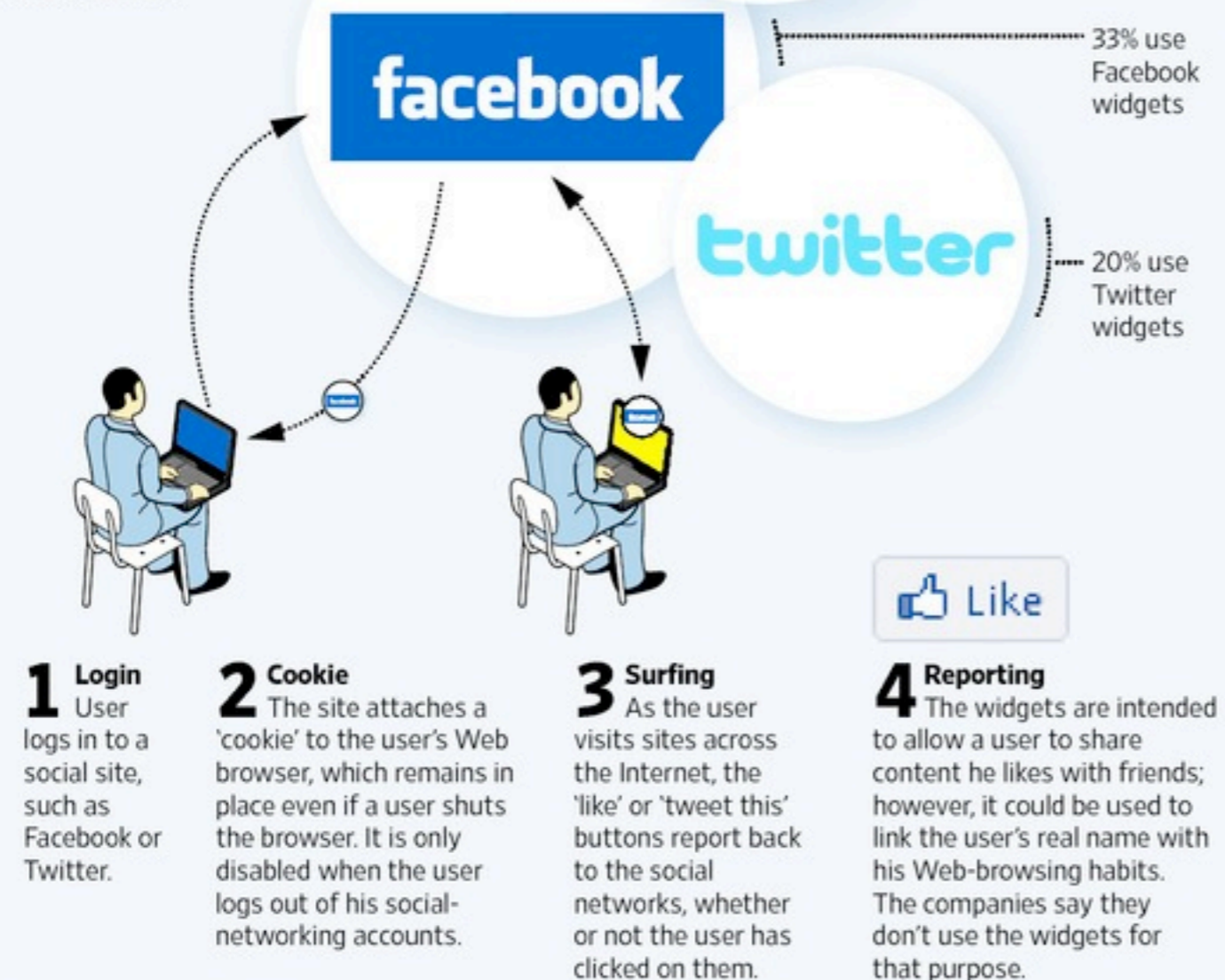


And this proliferation of data and its unclear transmission is of concern to policymakers, including the FTC who presented this diagram in their 2010 report in which they endorsed the creation of a Do Not Track mechanism.

Social Awareness

Social websites know where you've been on the Internet. Behind the scenes, they collect data on users' Web surfing, using the Facebook 'Like' buttons and other widgets embedded in websites.

How it works:



Not just advertising, social networking widgets are another key example (in that case often connected via log-in cookies to your real name).

Diagram from WSJ article one year ago.

Might seem obvious to you all (loading of external resources, authentication cookies, potential logging, etc.) but when I talked about this to a group of lawyers earlier this week at Stanford...



A research project of the **Electronic Frontier Foundation**

Panopticlick

How Unique – and Trackable – Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopticlick tests your browser to see how unique it is based on the **information** it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only **anonymous data** will be collected by this site.



A paper reporting the statistical results of this experiment is now available: **How Unique Is Your Browser?**, Proceedings of the Privacy Enhancing Technologies Symposium (PETS 2010), Springer Lecture Notes in Computer Science.

```
2 /**
3  * @depends jquery
4  * @depends swfobject
5  *
6  * evercookie (
7  *
8  * by samy kamkar
9  *
10 * this api attempts
11 * to essentially
12 *
13 * specifically
14 * - standard
15 * - flash cookies
16 * - silverlight
17 * - png generation
18 * - http etags
19 * - http cache
20 * - window.name
21 * - IE userData
22 * - html5 sessionStorage
23 * - html5 local storage
24 * - html5 geolocation
25 * - html5 database
26 * - css history
27 *
28 * if any cookies are
29 * for example
30 * that cookies are
31 *
32 * !!! SOME OTHER SITES
33 * OTHER SITES
34 *
35 * USAGE:
36
37 var ec = new Evercookie();
38
39 // set a cookie
40 // usage: ec.cookie(
41 ec.set("id", "1234567890",
42
43 // retrieve a cookie
44 ec.get("id",
45
```

Ashkan Soltani

Summary

Abstract—This is a pilot study of popular websites. We find that our sample are using Flash cookies. Some are using HTTP cookies deleted by the user. Some are using the same values as HTTP cookies deleted by the user. Some are using government websites to test policies rarely disclose their controls for effectuating privacy.

Privacy, tracking, usability, online advertising

I.

Advertisers are increasing tracking of users online over 30% of users deleted last month, thus leading to unique visitors to web advertising impressions.

Mindful of this problem, we have attempted to improve methods. In 2005, United Virtualities, an advertising company, and networks use [HTML] but cookies are under a that it had, “developed a web sites, ad network deleted by users. UV’s (PIE) is tagged to the unique ID just like traditional cannot be deleted by spyware, mal-ware, or even function at the Explorer.”[5] (Since “BetterPrivacy”, and a called “Glary Utilities P cookies.)

United Virtualities’ Flash MX: the “local flash cookie.” Flash cookies lead to more persistence cookies can contain up (HTTP cookies only still have expiration dates expire at the end of a longer by the domain stored in a different location

is this just a question of cookie management?

flash cookies

every other local storage technique

browser fingerprinting

an escalating list of management techniques and tracking techniques -- do we expect users to keep up with these?

and in a way, this is worse for all parties -- companies doing legitimate tracking may lose out on data while users never have the comfort of knowing that they won't be tracked (chilling)

in fact, this has been characterized as an “arms race”

GREETINGS PROFESSOR FALKEN

HELLO

A STRANGE GAME.

THE ONLY WINNING MOVE IS
NOT TO PLAY.

HOW ABOUT A NICE GAME OF CHESS?

mutually assured destruction

A brief history

headers proposed in
browser extensions
(2009)

W3C Working
Group formed
(August 2011)

“Do Not Track” registry
(2007)

FTC report
(2010)

Neelie Kroes’
challenge (June)

IE & Firefox
implementations
(2010-11)

Starting with the popular name/idea from advocacy groups in 2007. (Not to scale, but you get the picture.)
Note that this is starting more with “running code” and then getting to “rough consensus”.

Agenda

- How we got here
- The current state of Do Not Track
- Trends for Web privacy
- Call for participation

DNT : 1

How does Do Not Track work? Well, most of it comes down to this.

Divided into technical mechanism and compliance policy documents, but let's start with the technical side, which may be more accessible to this audience.

In some ways this is a pretty straightforward bits on the wire...

Request and response

- DNT: 1
- Tk: {0,1,3,u}
- `/.well-known/dnt/`
- DNT: 0
- `navigator.doNotTrack`

A little more complicated, we're looking at a request and response model. The value of that response is transparency for the user (as the CMU study pointed out, the biggest usability issue may be the doubt that this is being respected) and a "regulatory hook".

Exceptions

- `navigator.doNotTrack.`
`requestSiteSpecificTrackingException()`
- `requestWebWideTrackingException()`
- `removeSiteSpecificTrackingException()`
- `removeWebWideTrackingException()`

user-agent-managed exceptions via JavaScript API

let sites have an explicit negotiation over whether they wish to allow tracking in exchange for a service
... and then manage those exceptions in a single place where they can be monitored and changed

Compliance

- What does it mean to comply with a user's expressed tracking preference?
- What does "tracking" mean?

separation of mechanism and policy... separate documents, but otherwise Do Not Track is confronting this rather directly

Compliance

- Few limitations for first-party interactions
- Restrictions on both collection and use
- Permitted uses under heated debate
- Service providers (collector vs. processor)
- “Unlinkable” data

Process

- Tracking Protection Working Group
- Art of Consensus
- Multistakeholderism



"rough consensus and running code"

Tracking Protection Working Group charter, what the W3C is and a Working Group is

political context (Berlaymont, but also US gov, industry trade associations)

Process

- *“freedom is an endless meeting”*
- 3,122 emails
- 75 participants from 41 organizations
- Four face-to-face meetings



public list, and pretty substantial emails at that not without its frustrations

10 full days of meeting time so far, next meeting scheduled for next month in Seattle
fast, aggressive timeline to attempt this in under a year
graduating maturity of drafts (not yet at Last Call)

Skepticism

Example #1:

P5P: NO-TRACK, PINKY-SWEAR

...specifies that the server should not track the user. The PINKY-SWEAR token is described in the Policy Tokens section below.

...

NO-ADS-IM-SURE-YOU-WILL-FIGURE-OUT-ANOTHER-BUSINESS-MODEL

Indicates that the user does not wish to be shown any form of advertising content, and expresses their earnest belief that the web publisher will find some way to remain in business without an income stream.

some objections to the system that we've heard

<http://pastebin.com/ijjRKvUB>

Skepticism

“The “Do Not Track” HTTP header is useless, equivalent to a “Do not Steal from Me” T-shirt.”

— some commenter on Hacker News

Skepticism

3. Setting the Evil Bit

*There are a number of ways in which the evil bit may be set. Attack applications may use a suitable API to request that it be set. Systems that do not have other mechanisms **MUST** provide such an API; attack programs **MUST** use it.*

— RFC 3514

Skepticism

Privacy in an open society also requires cryptography.

[...]

We cannot expect governments, corporations, or other large, faceless organizations to grant us privacy out of their beneficence.

— Cypherpunk Manifesto

Engineers like solutions that are self-enforcing and Do Not Track is affirmatively not.

To answer some of the common questions, enforcement is done through legal means, or through market means, or even through social norms and ethics. (Regulatory hook, economics of large trackers, etc.)

Agenda

- How we got here
- The current state of Do Not Track
- Trends for Web privacy
- Call for participation

Capabilities, not resources

ee8f6e1260fd5a80cf5f5fb5546beff6c2a01cab

given that users struggle to understand the mechanisms and privacy implications, we should be managing privacy concerns based on the capability rather than the particular tool

"don't track me" not "don't set a cookie for this domain pair"

the Apple UDID controversy

potentially the Android manifest categorization, or research work in that area

Machine-readable policy



DNT is in essence the simplest form of machine-readable policy, a single bit. Hints at the possibility of other machine-readable policy systems.

Anecdote about keeping count of mentions of “creative commons for privacy” at privacy events.



Your data is never bartered or sold.

The site that is collecting data about you is not trading or selling it. It will only share your data with other organizations in order to carry out the intended transaction.



Your data may be bartered or sold.

This means that a website is collecting or trading it with another organization. An example of this is where a shop shares your shopping preferences, frugality info to data aggregators or to other



Your data is never given to advertisers.

Besides the information exposed via on-page advertisement, the site does not share the data it collects about you with advertisers.



Site gives your data to advertisers.

This means that a site either shares marketing or advertising companies collect info about you while on its site



Your data is kept for less than 1 month.



Your data may be kept indefinitely.

Your data is deleted before 1, 3, 6, or 18 months from the date of transmission have elapsed, respectively. Alternatively















Data is given to law enforcement only when legal process is followed.



Data may be given to law enforcement even when legal process is not followed.



Privacy Icons,
Aza Raskin,
Mozilla
2011

| TYPE OF DATA COLLECTED | GENERAL DATA PRACTICES | DATA SHARING |
|--|--|---|
|  <p>contact: name, mailing address, email, or phone number</p> |  <p>ad customization: user data may be used for the purpose of customizing advertising</p> |  <p>affiliates: affiliates and subsidiaries bound by the same privacy practices</p> |
|  <p>computer: IP address, browser type, or operating system</p> |  <p>third party tracking: site allows third parties to place advertisements that may track user behavior</p> |  <p>contractors: third party contractors bound by the same privacy practices</p> |
|  <p>interactive: browsing behavior or search history</p> |  <p>public display: service allows users to contribute information which may be displayed publicly</p> |  <p>third parties: third parties not subject to same data practices</p> |
|  <p>financial: account status or activity, credit information, or purchase history</p> |  <p>user control: users allowed to access and correct personal data collected</p> | |
|  <p>content: contents of personal communications, stored documents or media</p> |  <p>data retention: explicitly stated duration of retention for personal data collected</p> | |

KnowPrivacy
UC Berkeley
2009

Bell Group

information we collect

ways we use your information

information sharing

| | ways we use your information | | | | information sharing | |
|-----------------------------------|--------------------------------------|-----------|---------------|-----------|---------------------|---------------|
| | to provide service and maintain site | marketing | telemarketing | profiling | other companies | public forums |
| contact information | | opt in | | | opt out | |
| cookies | | | | | | |
| demographic information | | opt in | | | opt out | |
| financial information | | | | | | |
| health information | | | | | | |
| preferences | | | | | | |
| purchasing information | | opt in | | | opt out | |
| social security number & gov't ID | | | | | | |
| your activity on this site | | opt in | | | opt out | |
| your location | | | | | | |

Access to your information

This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site

Please email our customer service department

bell.com

5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@bell.com

Privacy Label
CMU
2009-2010

[close]

Site Details

Site Name: www.site.com

Site Owner: Example Site, Inc.

Certified by: TRUSTe



Privacy Policy: <http://www.site.com/privacy>

Privacy Summary [learn more](#)



Secondary Use: **Customization**

User data is used for the purpose of completing the current interaction, or to customize, personalize, or tailor the current user interaction.



Data Sharing: **None**

User data only shared internally within the data collector's organization or with organizations that help the data collector provide the current interaction



Data Retention: **Indefinite**

User data may be retained by the collecting party for an unspecified or indefinite amount of time

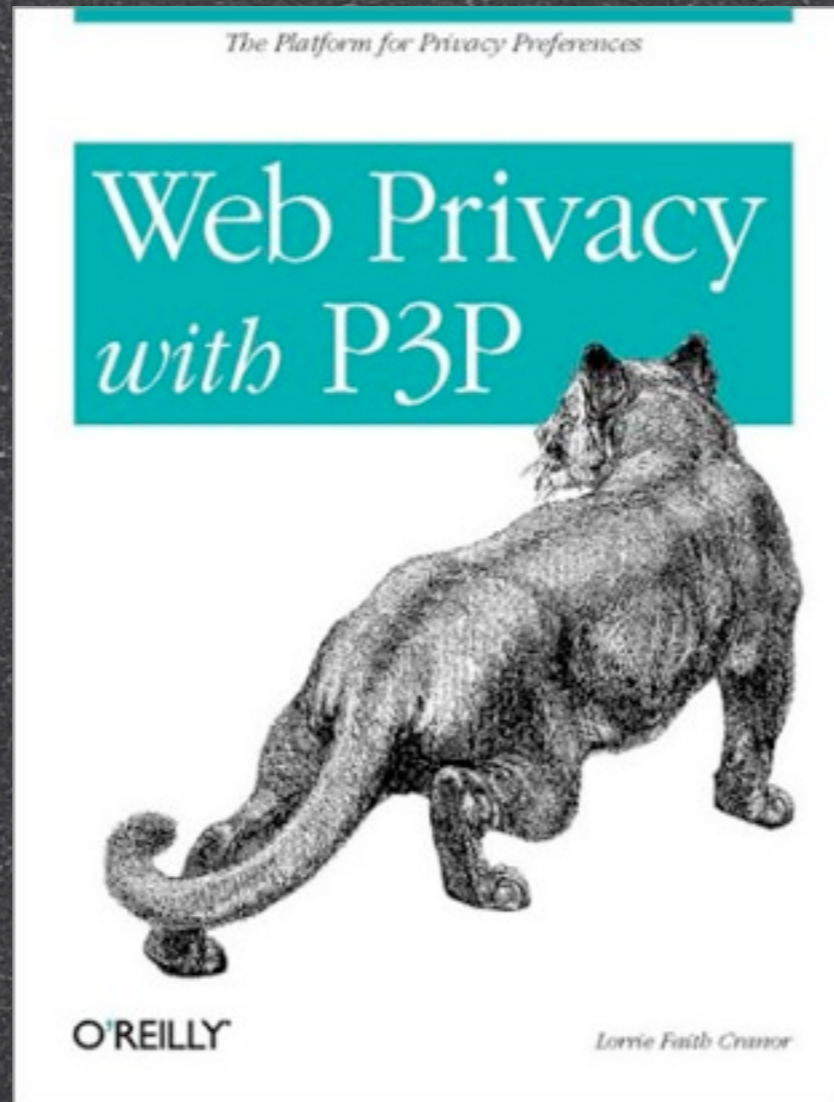
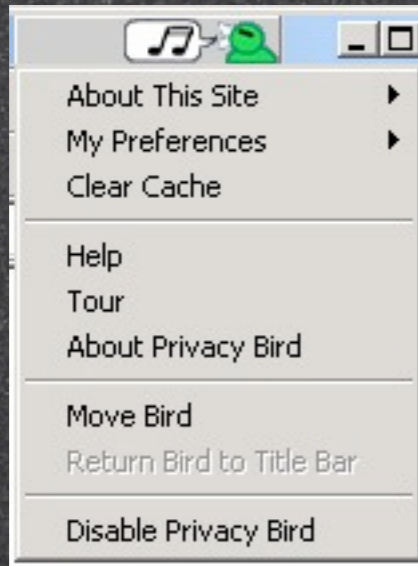


Third Party Tracking: **Trackers Detected**

Third party trackers are present on this site. To learn more or to set tracking preferences, visit the [TRUSTe Tracker Manager](#)

TRUSTe Privacy Short Notice 2011

built on top of XML policy database
Travis worked on the KnowPrivacy example as well



Privacy Bird
AT&T
2002

At least 2002, maybe earlier.
Based on the site's P3P policy, P3P standardized between 1996 and 2002

Machine-readable policy

Rehashing P3P?

An idea whose time has come?

Technology facilitating policy?

Creative Commons

more generically, the Policy Aware Web idea, a dream of the Semantic Web

"policy description with late binding of rules for accountability"

"avoid legal system the way we do in the rest of life"

Multistakeholderism

"Internet policy like the internet itself is best built through collaboration."

both W3C and I personally would like to make the case that the Tracking Protection Working Group is a promising attempt for multistakeholderism in addressing Internet privacy

but you'll hear this term used often enough (if you haven't already) that we may need to be skeptical of it

- like "democracy" something that you can't be against?

- debate over a potential ITU role in Internet governance

- the conditions of multistakeholderism

 - really what we mean is procedural and substantive legitimacy, some normative democratic weight behind decisions that are made

 - in our case consensus and multistakeholderism has the pragmatic aim of needing everyone to agree to find adoption

 - we've tried to make the process as open and involved multiple viewpoints BOTH to get a technically better result and to get a result that will fairly satisfy the community goal

- like democracy, the worst form except for all the alternatives

- government regulation, industry-only self-regulation, standards that aren't implemented

this is a lot of theory, but concretely: MSH is something you'll hear about directly from USG

- NTIA wants to host MSH processes to develop privacy codes of conduct

Agenda

- How we got here
- The current state of Do Not Track
- Trends for Web privacy
- Call for participation



CfP



optimism
we can build technologies that translate privacy implications into human terms and communicate human privacy preferences
building these tools correctly requires understanding both the technology and the human privacy concern
get involved!
NTIA, W3C, IETF, ITU, etc.
and if the available specific work items aren't of interest, we also have the question of considering privacy while building other Web standards...
W3C Privacy Interest Group and IAB privacy programs

Nick Doty

<http://npdoty.name>

npdoty@ischool.berkeley.edu

npdoty@w3.org