Resent-From: public-tracking@w3.org
From: David Singer <singer@apple.com>
Subject: 'do not cross-site track' response to Aleecia's outline
Date: April 6, 2012 9:48:03 PM PDT
To: "public-tracking@w3. org Group WG" <public-tracking@w3.org>

Friends

this is my 'homework' response. I am still not sure if I *advocate* this, but I do see the advantages, and think it is worthy of discussion. I am concerned that it represents a major change of a basis, and might cause delay as we review it. Indeed, pressure of time means it hasn't had broad review inside my own company, even.

- - - - - - - -

Contributors to this proposal: Dave Singer, inspired by Roy (who is nonetheless blameless, and was on vacation so wasn't even asked to help edit)

Basic Concept:

Instead of trying to define 1st/3rd parties, we abandon the 1st/3rd distinction, and instead define restrict tracking in such a way that the 1st/3rd party distinction is irrelevant. (We still need a definition of 'party' in general, which this does not address.) Basically, we restrict "cross-site" tracking.

Draft definition of tracking (the "tunnel vision"):
"Tracking is the retention by a party (site),
-- after a user's transaction is complete (served),
-- of data records that can or do associate that user with either
a) any other party (site), or
b) with data not collected from the user's direct transaction with the party (site) performing the transaction."

This says "party (site)" because, as today, we need to permit data to flow as long as obligations and liability flow with it, within an organization. This document does NOT have specific proposals on how to manage that data flow, or how to define "party" or "site". That problem remains.

So this definition allows:
* knowing about another site *during* a transaction (e.g. 'please supply an advert to the BogVille Chronicle')
* retaining records of what happens between your site and someone who requests data from it ('Dave was served an ad for dishwashers') ("tunnel vision")
* retaining the results of user interaction with any site, for those sites to remember that interaction and its results
* using real-time data for targetting (e.g. geo-location from IP address, determining time-of-day at that location, and so on), *during* the transaction
* retaining *separate* records related to the user ('Dave was served an ad') and the site ('an ad was served for the BogVille Chronicle site') as long as these records are unlinked and unlinkable

It does NOT allow:
* exampleAd.com remembering Dave was on the BogVille Chronicle site and that's why Dave was served a dishwasher ad
* retaining the full source URL the BogVille Chronicle used to get the ad, when that URL conveys information from, or identifying, another site, or info passed about the user
* retaining referer information
* combining the data from a user transaction with other data to work out who the user is, or facts about the user, and then retaining that
* social widgets recording your browsing history (without permission)

The big wins:
* no more squirrelly language on trying to guess the site the user thought they were visiting (1st party)
* no more squirrelly language on what constitutes 'interaction' to get promoted from 3rd to 1st party
* the specification is formally 'testable' without the 1st/3rd judgment call (given access to a site's retained information)
* no more worrying about re-directs being (from the browser's point of view) 1st, but 3rd from the user's point of view
* no more worrying about embedded/framed sites, or mash-ups; who is the 1st party and who are 3rd?
* we don't have to worry about raw log files that could easily be converted into user-tracking (as all the data is there); the restriction is a minor one on what gets logged in the first place
* no restricting 'ordinary' logging practices even for sites that offer embeddable content (e.g. a web badge, an embeddable widget) as long as they take care not to record either or both of
a) information that could identify another site
b) information that could identify a user

I think that the last is huge: for 'ordinary sites' the amount they have to do to comply is proportionate to the amount of logging they do that is 'cross-site'. If their logs remember only about their own site, they're statically fine. In general, the amount of work for anyone to comply is proportionate to the amount of logging of user AND other-site info that they do.

For re-direction services, there may be work to do, since almost any logging (e.g. of the URLs) would involve identifying another site (so this means they had better not remember data that can identify users).

Part I: Parties

A. A party is…
whatever we define it as. This document doesn't address that question.

B. A first party is…a party.
C. A third party is…a party.

There is NO first/third distinction.

Part II: Business uses /* or whatever we wind up calling this -- feel free to suggest something different */

Note: unless you specifically document otherwise, this section is understood to ONLY APPLY TO THIRD PARTIES.
This section applies to everyone, as there is no 1st/3rd party distinction.

For each of the seven potential business uses below, please indicate if:
A. this particular use is never allowed under DNT
B. this particular use is allowed with retention limits (describe)
C. this particular use is allowed without retention limits (describe any other limitations)

For any permission in our specification to go beyond the definition, if the data ever gets used for a purpose other than the exception, that's a non-permitted use, and laws (e.g. liability) may apply.

Explicit standard permissions needed:
* outsourcing, as before; if your records involve data that is about another site/party, the data is only available for use by that other site/party (and hence, not by you)
* user-granted exceptions: (e.g. "you may track my visits to other sites while I am logged on to TrackMyReading.com")

1. Frequency Capping - A form of historical tracking to ensure the number of times a user sees the same ad is kept to a minimum.
C. (As long as you only remember data about the ad-serving site, which is all you need.)
(also true for story-boarding)

2. Financial Logging - Ad impressions and clicks (and sometimes conversions) events are tied to financial transactions (this is how online advertising is billed) and therefore must be collected and stored for billing and auditing purposes.
C. As long as you either take care to lose the user-identify information (e.g. IP address, user ID, and so on), or other-site information, or both.

3. 3rd Party Auditing - Online advertising is a billed event and there are concerns with accuracy in impression counting and quality of placement so 3rd party auditors provide an independent reporting service to advertisers and agencies so they can compare reporting for accuracy.
If user-information and other-site information are both recorded, then B, else C. Retention until the audit has been performed (?).

4. Security - From traditional security attacks to more elaborate fraudulent activity, ad networks must have the ability to log data about suspected bad actors to discern and filter their activities from legitimate transactions. This information is sometimes shared across 3rd parties in cooperatives to help reduce the daisy-chain effect of attacks across the ad ecosystem.
B, I suspect, as both user-identify and other-site information will be recorded.

5. Contextual Content or Ad Serving: A third-party may collect and use information contained with the user agent string (including IP address and referrer URL) to deliver content customized to that information.
C. Real-time data in the transaction is all fair game. It's retention that's not.

6. Research / Market Analytics
C. As long as you use data you learned directly from the user, and not about or derived from another site or elsewhere.

7. Product Improvement, or, more narrowly, Debugging
C. As long as you use data you learned directly from the user, and not about or derived from another site or elsewhere.

David Singer
Multimedia and Software Standards, Apple Inc.

Resent-From: public-tracking@w3.org
From: Erica Newland <enewland@cdt.org>
Subject: CDT's proposals re: template for parties and business uses
Date: April 6, 2012 10:08:23 PM PDT
To: public-tracking@w3.org

Contributors to this proposal: **Justin Brookman, Erica Newland**

This proposal seeks to address Issues 10, 17, 19, 22, 24, 25, 31, 49, and 73

Summary of compromise suggestion:

- **Discoverable affiliate definition of party used**
- **Market research removed from permitted uses unless data de-identified within two weeks**
- **Product improvement removed from permitted uses in favor of debugging**
- **All permitted uses require retention only as reasonably required by exempted purpose, and clear statement of data retention periods**
- **Data collected by a company as a first-party may be used by that company to customize content in a third-party settings (logged-in state irrelevant, though could be justified)**
- **"Clear and prominent" notice and consent required for user-granted exceptions (logged-in state irrelevant)**

Part I: Parties

A. A party is . . .

[CDT would prefer a test based on reasonable expectations of a user, but would be willing to compromise on the Amy/Shane definition of discoverable affiliates if concessions are made on other issues (see below)]

Discoverable Affiliates

Normative Discussion
A party is any commercial, nonprofit, or governmental organization, a subsidiary or unit of such an organization, or a person. For unique corporate entities to qualify as a common party with respect to this standard, those entities must be commonly owned and commonly controlled, and must make their parent affiliation (if any) easy discoverable to users.

Non-Normative Discussion
This may be accomplished in many ways, including but not limited to, prominent and common branding on site pages, "one click away" within Privacy Policies, and, if available, a programmatic list of domains that share common ownership (affiliation).

Example 0: If a user visits flickr.com, which is branded "from Yahoo!", are Flickr and Yahoo one party? CDT: YES

Example 1: If a user visits google.com, are other parts of Google, Inc. (adwords, analytics, YouTube, gmail, Google Maps) also the same party as google.com? CDT: YES

Example 2: If a user visits geico.com, is See's Candies also the same party? CDT: YES, although this must be easily discoverable per the specification.

Example 3: If Mozilla and Opera form a jointly-owned and controlled company called Moperilla, and a user visits Moperilla, are Mozilla and Opera part of the same party as Moperilla? CDT: NO, Mozilla, Opera, and Moperilla all have different owners and control structures. Only Moperilla is a first party, and Mozilla and Opera may not use that data as a first party.

B. A first party is . . .

CDT recommends that the definition capture the idea that a first-party site is the site that the user intended to visit. This definition is intended to distinguish between a link that the user intended to visit as opposed to a link shortening service, or the news site that a user typed into a browser as opposed to the operator of any widgets on that site. The "meaningful interaction" test can still turn a third party into a first party.

Meaningful User Interaction: **A "first party" is any party, in a specific network interaction, that can infer with high probability that the user knowingly and intentionally communicated with it. Otherwise, a party is a third party.**

To comply with DNT, a first party MUST...

[CDT: no affirmative obligations]

To comply with DNT, a first party MUST NOT…

CDT: share information about the communication with a third party in a form that can allow the information to be correlated with the same user's activity on other third-party domains, UNLESS the first party ensures that the third party provide technical or legal assurances that it will honor the same obligations to protect the information that the first party is required to honor, OR the sharing is a permitted uses as defined in this standard or pursuant to specific, user-granted exceptions.

CDT: The language above is designed to ensure that first parties cannot help third-party ad networks circumvent a DNT signals through backchannel means.

C. A third party is…

CDT: any entity that is not a first party or the end user.

To comply with DNT, a third party MUST...

CDT: If the operator of a third-party domain receives a communication to which a [DNT-ON] header is attached:

- **that operator MUST NOT collect, share, or use information related to that communication outside of either the permitted uses or any explicitly-granted exceptions, as provided in accordance with the requirements of this standard;**
- **that operator MUST NOT use information about previous communications in which the operator was a third party, outside of the permitted uses as defined within this standard; and**
- **that operator SHOULD NOT retain information about previous communications in which the operator was a third party, outside of the permitted uses as defined within this standard.**
- **that operator, upon receiving data from a distinct, standard-compliant party, must treat that data with at least the level of protection that the other party was required to afford that data under this standard.**

D. A third party acting as a first party (as an agent) is:

[CDT: We suggest this section be moved under the definition of first party:]

Outsourcing Partner of First-Party:

A third-party service may operate as a first-party site if all the following conditions hold:

- **the third party's data collection, retention, and use practices comply with at least the requirements for first-parties;**
- **the data collected by the third party is available only to the first party, and the third party has no independent right to use the data (unless that data is deidentified or aggregated);**
- **the third party makes commitments that are consistent with compliance with this standard and they do so in a form that is legally enforceable (directly or indirectly) by the first party, individual users, and regulators; data retention by the third party must not survive the end of this legal enforceability;**
- **the third party undertakes reasonable technical precautions to prevent the retention of data that could be correlated across first parties.**

Examples and use cases:
ExampleAnalytics collects analytic data for ExampleProducts Inc.. It operates a site under the DNS analytics.exampleproducts.com. It collects and analyzes data on visits to ExampleProducts, and provides that data solely to ExampleProducts, and does not access or use it itself, although it may use and sell aggregate reports about generic user behavior on ExampleProducts.com.

Part II: Permitted Uses

Note: unless you specifically document otherwise, this section is understood to ONLY APPLY TO THIRD PARTIES.

For each of the seven potential business uses below, please indicate if:

A. this particular use is never allowed under DNT
B. this particular use is allowed as long as data is "unlinkable" as described in section 0
C. this particular use is allowed with retention limits (describe)
D. this particular use is allowed with aggregation (describe)
E. this particular use is allowed (describe any other limitations that apply)

As needed, feel free to define and scope the potential business uses.

0. "**Reasonably** Unlinkable" data:

CDT: E (allowed, no limits) if the following definition applies: A party holds information about a communication that is "reasonably unlinkable" to an individual or device provided that this party:

1. **takes reasonable measures to ensure that the data is de-identified — this includes removing IP address or persistent device ID;**
2. **publicly commits not to try to re- identify the data; and**
3. **contractually prohibits downstream recipients from trying to re-identify the data.**

Information that is not "reasonably unlinkable" is referred to as "reasonably linkable" information. A set of precise geo-location data points compiled over time, for example, is likely to be reasonably linkable absent de-identification efforts.

[This definition tracks the specific language of the recent FTC report and also tracks the language in Recital 23 of the European Data Protection Regulation legislation. If data meets this definition, it does not matter to what purpose it is used or how long it is retained.]

1. Frequency Capping - A form of historical tracking to ensure the number of times a user sees the same ad is kept to a minimum.

- **CDT: OK if the data will satisfy the criteria in B/D within two weeks OR**
- **CDT: OK so long as the data is only retained for as long as is reasonably required for this purpose and the third-party's privacy policy (or equivalent, readily-discoverable document) explains with reasonable precision how soon personal information will be deleted or rendered unidentifiable ("C" under the template above)**

2. Financial Logging - Ad impressions and clicks (and sometimes conversions) events are tied to financial transactions (this is how online advertising is billed) and therefore must be collected and stored for billing and auditing purposes.

- **CDT: OK if the data will satisfy the criteria in B/D within two weeks OR**
- **CDT: OK so long as the data is only retained for as long as is reasonably required for this purpose and the third-party's privacy policy (or equivalent, readily-discoverable document) explains with reasonable precision how soon personal information will be deleted or rendered unidentifiable ("C" under the template above)**

3. 3rd Party Auditing - Online advertising is a billed event and there are concerns with accuracy in impression counting and quality of placement so 3rd party auditors provide an independent reporting service to advertisers and agencies so they can compare reporting for accuracy.

- **CDT: OK if the data will satisfy the criteria in B/D within two weeks OR**

- **CDT: OK so long as the data is only retained for as long as is reasonably required for this purpose and the third-party's privacy policy (or equivalent, readily-discoverable document) explains with reasonable precision how soon personal information will be deleted or rendered unidentifiable ("C" under the template above)**

4. Security - From traditional security attacks to more elaborate fraudulent activity, ad networks must have the ability to log data about suspected bad actors to discern and filter their activities from legitimate transactions. This information is sometimes shared across 3rd parties in cooperatives to help reduce the daisy-chain effect of attacks across the ad ecosystem.

- **CDT: OK if the data will satisfy the criteria in B/D within two weeks OR**
- **CDT: OK so long as the data is only retained for as long as is reasonably required for this purpose and the third-party's privacy policy (or equivalent, readily-discoverable document) explains with reasonable precision how soon personal information will be deleted or rendered unidentifiable ("C" under the template above)**

5. Contextual Content or Ad Serving: A third-party may collect and use information contained with the user agent string (including IP address and referrer url) to deliver content customized to that information.

- **CDT: OK if the data will satisfy the criteria in B/D within two weeks OR**
- **CDT: OK so long as the data is only retained for as long as is reasonably required for this purpose and the third-party's privacy policy (or equivalent, readily-discoverable document) explains with reasonable precision how soon personal information will be deleted or rendered unidentifiable ("C" under the template above)**

6. Research / Market Analytics

CDT: OK if the data will satisfy the criteria in B/D within two weeks

7. Debugging

- **CDT: OK if the data will satisfy the criteria in B/D within two weeks OR**
- **CDT: OK so long as the data is only retained for as long as is reasonably required for this purpose and the third-party's privacy policy (or equivalent, readily-discoverable document) explains with reasonable precision how soon personal information will be deleted or rendered unidentifiable ("C" under the template above)**

III. Additional potentially relevant, but likely irrelevant, information

Definition of Tracking (Issue-5):

[CDT: We strongly urge silence. We believe this is defined by the spec; it will be incumbent upon implementers to message to users what the spec accomplishes.]

Logged In (Issue-65):

[CDT: We urge silence. If there is an exception for logged-in state, we recommend that it be as a limitation on when third parties can use information they received previously as a first-party:

Under the current language of the compliance specification, a party is allowed to use data that it received as a first-party in a third-party context to deliver content. For example, I often see LinkedIn ads around the web based on my LinkedIn profile but not based on cross-site behavioral data or even the context of the page that I am on.

If we were to have an exception for logged-in state, CDT would prefer that it be limited to allowing the use of first-party data in the third-party context to only those scenarios where the user is in a logged-in state for the first-party: E.g., Facebook may render social widgets based on logged-in state and information it received as first-party on Washington Post, but Weather.com may not render widgets based on first-party data on New York Times for users who never registered for Weather.com]

Definition of Consent/Affirmative, Informed Consent to be Tracked (Issue 69):

CDT: means consent given by an affirmative action such as clicking a consent box in response to a clear and prominent request to ignore a "Do Not Track" setting that is distinct and separate from any other notifications or requested permissions.

[We believe very strongly that a user-stated rule for "Do Not Track" must be circumvented by third parties ONLY with the user's clear and informed permission. This has nothing to do with any first or third parties' preexisting legal obligations under any regime.]

Erica Newland
Policy Analyst
Center for Democracy & Technology
1634 Eye Street NW, Suite 1100
Washington, DC 20006
202.407.8836
enewland@cdt.org
http://www.cdt.org
Follow us on Twitter at @CenDemTech

Contributors to this proposal: John M. Simpson

Part I: Parties

A. A "party" is any commercial, nonprofit, or governmental organization, a subsidiary or unit of such an organization, or a person, that an ordinary user would perceive to be a discrete entity for purposes of information collection and sharing. A party MAY also include affiliates if the affiliates are commonly owned and controlled, and the relationship is clear to consumers through common branding. A party MUST NOT include more than five affiliates.

Example 0: If a user visits flickr.com, which is branded "from Yahoo!", are Flickr and Yahoo one party? Yes.
Example 1: If a user visits google.com, are other parts of Google, Inc. (adwords, analytics, YouTube, gmail, Google Maps) also the same party as google.com? Yes.
Example 2: If a user visits geico.com, is See's Candies also the same party? No.
Example 3: If Mozilla and Opera form a jointly-owned and controlled company called Moperilla, and a user visits Moperilla, are Mozilla and Opera part of the same party as Moperilla? No.

B. A "first party" is any party, in a specific network interaction, that can infer with high probability that the user knowingly and intentionally communicated with it. Otherwise, a party is a "third party." If a party cannot infer with a high degree of probability that it is a "first party," it MUST behave as a third party.

To comply with DNT, a first party MUST NOT share data with a third party, outside of permitted uses as defined in this standard or specific user-granted exceptions.
To comply with DNT, a first party MAY take additional privacy enhancing steps, such as treating each session with a user as an entirely new session unless it has been given permission to store her information and use it again.

C. A "third party" is any party, in a specific network interaction, that

cannot infer with high probability that the user knowingly and intentionally communicated with it. If a party does not know its status, it MUST behave as a third party.

To comply with DNT, if the operator of a third-party domain receives a communication to which a [DNT:1] header is attached:

1. that operator MUST NOT collect, share, or use information related to that communication outside of the permitted uses as defined within this standard and any explicitly-granted exceptions, provided in accordance with the requirements of this standard;
2. that operator MUST NOT use information about previous communications in which the operator was a third party, outside of the explicitly expressed permitted uses as defined within this standard;
3. that operator MUST NOT retain information about previous communications in which the operator was a third party, outside of the explicitly expressed permitted uses as defined within this standard.
4. that operator MUST NOT use information associated with the user agent that was gathered and stored when the operator was acting as a first party.

D. A third party acting as a first party (as an agent) MUST be under contract to provide a specific service for the first party.

To comply with DNT, a third party acting as a first party MUST NOT combine any data obtained from the first party to perform the contracted service with any other data.

To comply with DNT, a third party acting as a first party MUST retain the data only as long as necessary to perform the contracted service for the first party.

To comply with DNT, a third party acting as a first party MUST NOT collect data that could be combined across first parties.

Part II: Business uses /* or whatever we wind up calling this -- feel free to suggest something different */

I suggest we simple call this "permitted uses."

Note: unless you specifically document otherwise, this section is understood to ONLY APPLY TO THIRD PARTIES.

I agree this section applies to third parties.

For each of the seven potential business uses below, please indicate if:

A. this particular use is never allowed under DNT
B. this particular use is allowed as long as data is "unlinkable" as described in section 0
C. this particular use is allowed with retention limits (describe)
D. this particular use is allowed with aggregation (describe)
E. this particular use is allowed (describe any other limitations that apply)

As needed, feel free to define and scope the potential business uses.

0. Any use is allowed that uses only unlinkable, aggregated data. Unlinkable data is data that has been de-identified by removing the IP address or persistent device ID. There must be a public commitment not to re-identify the data and a contractual prohibition preventing downstream recipients from trying to re-identify the data.

1. Frequency Capping - A form of historical tracking to ensure the number of times a user sees the same ad is kept to a minimum.

A. This use is not allowed when DNT is enabled.

2. Financial Logging - Ad impressions and clicks (and sometimes conversions) events are tied to financial transactions (this is how online advertising is billed) and therefore must be collected and stored for billing and auditing purposes.

C. This use is allowed, but the data MUST be retained only as long

as is reasonably necessary to fulfill billing and auditing purposes. Data gathered under this permitted use MUST NOT be used for any other purpose.

3. 3rd Party Auditing - Online advertising is a billed event and there are concerns with accuracy in impression counting and quality of placement so 3rd party auditors provide an independent reporting service to advertisers and agencies so they can compare reporting for accuracy.

C. This use is allowed, but the data MUST be retained only as long as is reasonably necessary to fulfill auditing purposes. Data gathered under this permitted use MUST NOT be used for any other purpose.

4. Security - From traditional security attacks to more elaborate fraudulent activity, ad networks must have the ability to log data about suspected bad actors to discern and filter their activities from legitimate transactions. This information is sometimes shared across 3rd parties in cooperatives to help reduce the daisy-chain effect of attacks across the ad ecosystem.

C. Data MAY be collected and shared to the extent reasonably necessary to prevent fraud, when there are reasonable grounds to suspect fraudulent activity. Data gathered under this permitted use MUST be retained only as long as necessary for that purpose and MUST NOT be used for any other purpose.

5. Contextual Content or Ad Serving: A third-party may collect and use information contained with the user agent string (including IP address and referrer url) to deliver content customized to that information.

E. This use is allowed when DNT is enabled so long as data is not retained beyond the immediate transaction.

6. Research / Market Analytics

D. This use is allowed when the data is aggregated and not linked to any user.

7. Product Improvement, or, more narrowly, Debugging

D. This is allowed when the data is aggregated and not linked to any user.

First Parties and Third Parties Draft

Unofficial Draft 07 April 2012

Editors:
Peter Eckersley, Electronic Frontier Foundation
Tom Lowenthal, Mozilla
Jonathan Mayer, Stanford University

Abstract

Abstract, version, and status information are not relevant in this partial draft, but are part of the template. Just ignore these for now.

Status of This Document

This document is merely a public working draft of a potential specification. It has no official standing of any kind and does not represent the support or consensus of any standards organisation.

Table of Contents

1. Drafting Notes

- ISSUE-97: A special rule for URL-shortening services remains an open issue and is not addressed in this proposal.
- ISSUE-26: We have not provided a special rule for widgets. The same first party vs. third party test for static content applies.
- This draft does not, in general, address retention limits. We do believe retention limits should be imposed on almost all third-party information practices.
- This draft does not establish special exceptions for frequency capping, financial logging, third-party auditing, market research, product improvement, or personalization for logged-in users. To the extent these practices are not possible within the bounds of an exception we do provide (i.e. unlinkable data, outsourcing, or permission), they are prohibited.

2. Parties, First Parties and Third Parties

2.1 Parties

2.1.1 Definition

A "party" is any commercial, nonprofit, or governmental organization, a subsidiary or unit of such an organization, or a person, that an ordinary user would perceive to be a discrete entity for purposes of information collection and sharing. Domain names, branding, and corporate ownership may contribute to, but are not necessarily determinative of, user perceptions of whether two parties are distinct.

2.1.2 Non-Normative Discussion

Our definition of what constitutes a "party" is guided by ordinary user expectations. We decline to adopt a conflicting approach that would draw a line at domain names, corporate affiliation, or branding, as discussed below.

2.1.2.1 Domain Names

In an uncomplicated world, a party might be delineated by domain boundaries. In practice, however, the domain approach can emphasize differences that would not matter to ordinary users and would be restrictive for many business uses. Suppose Example Company hosts dynamic content on example.com and static images on example-static.com. An average user would understand both domains are operated by Example Company, but a domain name distinction would say the two domains are different parties. Using domain names to differentiate parties would also impose an unnecessary choice on large websites of either hosting all their content on a single domain or having some of their content considered third party. By adopting a user expectations standard, we allow a

single website to span multiple domains.

A domain name approach can also gloss over relevant differences from a user expectations perspective. Suppose Example Company hires an analytics company and aliases the domain analytics.example.com to the analytics company's website. By user expectations, and corporate affiliation and branding, the analytics company would be a separate party. Moreover, circumventing the limits imposed by this standard would require nothing more than switching domain names. The user expectations standard we adopt recognizes that multiple parties may exist at a single domain.

2.1.2.2 Corporate Affiliation

Corporate families can consist of businesses in completely unrelated industries; users may have limited understanding of how businesses are related by corporate ownership or control. Moreover, by creating affiliates for the purposes of data sharing, organizations could circumvent the limits imposed by this standard. Under the user expectations standards we adopt, a corporate affiliate is not, in general, the same party as an organization.

2.1.2.3 Branding

In many cases, branding aligns with ordinary user expectations. Unrelated websites rarely share branding. In company ownership scenarios, prominent language like "Brand A, provided by Company B" may be sufficient for the average user to understand that Brand A is owned by Company B and information shared with Brand A may also be shared with Company B.

But, in some cases, branding does not align with user expectations. Suppose Example Search owns a video sharing website, Example Video. Most users are aware that Example Video is a subsidiary of Example Search, and that the Example Video website differs from the Example Search website for historical reasons. The Example Video home page does not, however, include any branding reference to Example Search. Under a branding test, Example Search and Example Branding would be different parties. The user expectations test allows for factors, other than branding, that influence user understanding.

Branding may also fall short in informing user expectations. If most users have never heard of Company B, language like "Brand A, provided by Company B" may not be adequate for the average user to understand the relationship between Brand A and Company B. A user expectations test recognizes there may be instances where even conspicuous branding is inadequate to inform users.

2.2 Network Interaction

2.2.1 Definition

A "network interaction" is an HTTP request and response, or any other set of logically related network traffic.

2.2.2 Non-Normative Discussion

Determination of a party's status is limited to a single transaction because a party's status may be affected by time, context, or any other factor that influences user expectations.

2.3 First Parties and Third Parties

2.3.1 Definitions

A "first party" is any party, in a specific network interaction, that can infer with high probability that the user knowingly and intentionally communicated with it. Otherwise, a party is a third party.

A "third party" is any party, in a specific network interaction, that cannot infer with high probability that the user knowingly and intentionally communicated with it.

2.3.2 Non-Normative Discussion

2.3.2.1 Overview

We draw a distinction between those parties an ordinary user would or would not expect to share information with, "first parties" and "third parties" respectively. The delineation exists for three reasons.

First, when a user expects to share information with a party, she can often exercise control over the information flow. Take, for example, Example Social, a popular social network. The user may decide she does not like Example Social's privacy or security practices, so she does not visit examplesocial.com. But if Example Social provides a social sharing widget embedded in another website, the user may be unaware she is giving information to Example Social and unable to exercise control over the information flow.

Second, we recognize that market pressures are an important factor in encouraging good privacy and security practices. If users do not expect that they will share information with an organization, it is unlikely to experience market pressure from users to protect the security and privacy of their information. In practice, moreover, third parties may not experience sufficient market pressure from first parties since increasingly third parties do not have a direct business relationship with the first party websites they appear on. We therefore require a greater degree of user control over information sharing with such organizations.

Last, third parties are often in a position to collect a sizeable proportion of a user's browsing history – information that can be uniquely sensitive and easily associated with a

user's identity. We wish to provide user control over such information flows.

We recognize that, unlike with a bright-line rule, there can be close calls in applying our standard for what constitutes a first party or a third party. But we believe that in practice, such close calls will be rare. The overwhelming majority of content on the web can be classified as first party or third party, with few cases of ambiguity in practice.

We require a confidence at a "high probability" before a party can consider itself a first party. Where there is reasonable ambiguity about whether a user has intentionally interacted with a party, it must consider itself a third party. Our rationale is that, in the rare close cases, a website is in the best position to understand its users' expectations. We therefore impose the burden of understanding user expectations on the website. We also wish, in close cases, to err on the side of conforming to user expectations and protecting user privacy. If the standard is insufficiently protective, ordinary users have limited recourse; if the standard imposes excessive limits, websites retain the safety valve of explicitly asking for user permission.

2.3.2.2 Common Examples and Use Cases

1. A user accesses an Example News article. The page includes an advertisement slot, which loads content from many companies other than Example News. Those companies are third parties.
2. A user accesses an Example News article. The page includes an analytics script that is hosted by Example Analytics, an analytics service. Example Analytics is a third party.
3. A user accesses an Example News article. It includes a social sharing widget from Example Social, a popular social network. Example Social is a third party.
4. A user visits Example Diary, which is hosted by the free blogging service Example Blog Hosting but located at examplediary.com. Example Blog Hosting is a third party.
5. A user launches Example Application, an app on a mobile device. The app includes a library from Example Advertising Network that displays ads. Example Advertising Network is a third party.

2.3.2.3 Multiple First Parties

There will almost always be only one party that the average user would expect to communicate with: the provider of the website the user has visited. But, in rare cases, users may expect that a website is provided by more than one party. For example, suppose Example Sports, a well known sports league, collaborates with Example Streaming, a well known streaming video website, to provide content at www.examplesportsonexamplestreaming.com. The website is prominently advertised and branded as being provided by both Example Sports and Example Streaming. An ordinary user who visits the website may recognize that it is operated by both Example Sports and Example Streaming.

2.3.2.4 User Interaction with Third-Party Content

A party may start out as a third party but become a first party later on, after a user interacts with it. If content from a third party is embedded on a first party page, the third party may become an additional first party if it can infer with high probability that the average user knowingly and intentionally communicated with it. If a user merely moused over, closed, or muted third-party content, the party would not be able to draw such an inference.

2.3.2.4.1 Examples and Use Cases

Example: Example Weather offers an unbranded weather widget that is embedded into websites, including Example News. The widget contains small links to Example Weather's website and privacy policy. A user visits Example News and scrolls through the weekly forecast in the Example Weather widget.

Discussion: Example Weather is a third party. The user has interacted with Example Weather's widget, but an ordinary user would not expect that scrolling through the widget involves communicating with Example News.

Example: Example Social, a popular social network, hosts a social sharing button that other websites can embed. The button is colored and styled in the same fashion as Example Social's website, contains descriptive text that is specific to Example Social, includes Example Social's logo, and very frequently appears on Example Social's website. Example News embeds the Example Social button, and a user clicks it.

Discussion: Example Social is a first party once the user clicks its embedded social sharing button. The average user would understand that by clicking the button she is communicating with Example Social.

3. Information Practices

3.1 First Party

A first party MUST NOT share information with a third party that the third party is prohibited from collecting itself.

Best Practice: A first party may voluntarily take steps to protect user privacy when responding to a Do Not Track request.

3.2 Third Party

3.2.1 General Rule

A third party may not collect, retain, use, or share any information related to communication with a user or user agent. There are exceptions to this general rule as defined in the following sections.

3.2.2 Exceptions

3.2.2.1 Protocol Information

A third party may collect, retain, and use protocol information for the purpose of communicating with a user agent.

Drafting note: We look forward to discussing with the group how to effectively impose retention limits on protocol information.

3.2.2.2 Unlinkable Data

3.2.2.2.1 Definition

N-unlinkability is the special case of K-anonymity where all values are considered part of the pseudo-identifier.

A dataset is "unlinkable" when there is a high probability that it contains only information which, for a skilled analyst, is 1024-unlinkable with respect to particular users, user agents, or devices.

3.2.2.2.2 Collection

A third party may collect non-protocol information if it is, independent of protocol information, unlinkable data.

Example: Example Analytics sets a language preference cookie that takes on few values and is shared by many users.

3.2.2.3 Outsourcing

A first party MAY outsource website functionality to a third party, in which case the third party may act as the first party under this standard with the following additional restrictions.

3.2.2.3.1 Technical Precautions

3.2.2.3.1.1 Operative Text

Throughout all data collection, retention, and use, outsourced service providers MUST use all feasible technical precautions to both mitigate the linkability of and prevent the linking of data from different first parties.

Structural separation ("siloing") of data per first party, including both

1. separate data structures and
2. avoidance of shared unique identifiers

are necessary, but not necessarily sufficient, technical precautions.

3.2.2.3.1.2 Non-Normative Discussion

3.2.2.3.1.2.1 Siloing in the Browser

Outsourcing services should use browser access control features so that stored data specific to one first party is never accessed or collected when the user visits another first party.

####### 3.2.2.3.1.2.1.1 Same-Origin Policy

The same-origin policy silos stored data by domain name. An outsourcing service can use a different domain name for each first party.

Example: Example Analytics provides an outsourced analytics service to Example News and Example Sports, two unrelated websites. Example Analytics stores its cookies for Example News at examplenews.exampleanalytics.com, and it stores its cookies for Example Sports at examplesports.exampleanalytics.com.

####### 3.2.2.3.1.2.1.2 Cookie Path Attribute

The HTTP cookie path can be used to silo data to a first party.

Example: Example Analytics stores its cookies for Example News with

"Path=/examplenews", and it stores its cookies for Example Sports with "Path=/examplesports".

3.2.2.3.1.2.1.3 Storage Key

For key/value storage APIs, such as Web Storage and Indexed Database, an outsourcing service can use a different key or key prefix for each first party.

Example: Example Analytics stores data for Example News at window.localStorage["examplenews"] and data for Example Sports at window.localStorage["examplesports"].

3.2.2.3.1.2.2 Siloing in the Backend

3.2.2.3.1.2.2.1 Encryption Keys

An outsourcing service should encrypt each first party's data with a different set of keys.

3.2.2.3.1.2.2.2 Access Controls

An outsourcing service should deploy access controls so that only authorized personnel are able to access siloed data, and only for authorized purposes.

3.2.2.3.1.2.2.3 Access Monitoring

An outsourcing service should deploy access monitoring mechanisms to detect improper use of siloed data.

3.2.2.3.1.2.3 Retention in the Backend

An outsourcing service should retain information only so long as necessary to provide necessary functionality to a first party. If a service creates periodic reports, for example, it should delete the data used for a report once it is generated. An outsourcing service should be particularly sensitive to retaining protocol logs, since they may allow correlating user activity across multiple first parties.

3.2.2.3.2 Internal Practices

3.2.2.3.2.1 Operative Text

Throughout all data collection, retention, and use, outsourced service providers MUST use sufficient internal practices to prevent the linking of data from different first parties.

3.2.2.3.2.2 Non-Normative Discussion

3.2.2.3.2.2.1 Policy

An outsourcing service should establish a clear internal policy that gives guidance on how to collect, retain, and use outsourced data in compliance with this standard.

3.2.2.3.2.2.2 Training

Personnel that interact with outsourced data should be familiarized with internal policy on compliance with this standard.

3.2.2.3.2.2.3 Supervision and Reporting

An outsourcing service should establish a supervision and reporting structure for detecting improper access.

3.2.2.3.2.2.4 Auditing

External auditors should periodically examine an outsourcing service to assess whether it is in compliance with this standard and has adopted best practices. Auditor reports should be made available to the public.

3.2.2.3.3 Use Direction

An outsourced service

1. MUST use data stored on behalf of a first party ONLY on behalf of that first party, and
2. MUST NOT use data stored on behalf of a first party for their own business purposes, or for any other reasons.

3.2.2.3.4 First-Party Requirements

3.2.2.3.4.1 Representation

A first party's representation that it is in compliance with this standard includes a representation that its outsourcing service providers comply with this standard.

3.2.2.3.4.2 Contract

A first party MUST enter into a contract with an outsourcing service provider that requires that outsourcing service provider to comply with these requirements.

3.2.2.4 User Permission

A website my engage in practices otherwise prohibited by this standard if a user grants permission. Permission may be attained through the browser API defined in the companion Tracking Preference Expression document. A website may also rely on "out-of-band" consent attained through a different technology. An "out-of-band" choice mechanism has the same effect under this standard as the browser exception API, provided that it satisfies the following bright-line requirements:

1. **Actual presentation:** The choice mechanism MUST be actually presented to the user. It MUST NOT be on a linked page, such as a terms of service or privacy policy.
2. **Clear terms: The choice mechanism must use clear, non-confusing terminology.**
3. **Independent choice:** The choice mechanism MUST be presented independent of other choices. It MUST NOT be bundled with other user preferences.
4. **No default permission:** The choice mechanism MUST NOT have the user permission preference selected by default.

An "out-of-band" choice mechanism must additionally satisfy the following high-level standard:

An ordinary user would know that the choice overrides his or her privacy protections under this standard.

3.2.2.5 Contextual Personalization

A third party may temporarily use a referrer URL for the purpose of contextually personalizing content.

3.2.2.6 Geolocation

A third party may temporarily use an IP address for the purpose of coarse geolocation.

3.2.2.7 Security

3.2.2.7.1 Operative Text

A third party may collect, retain, and use data about a particular user or user agent for the purpose of ensuring its security, provided that there are reasonable grounds to believe the user or user agent was attempting to breach the party's security at the time the data was collected.

3.2.2.7.2 Non-Normative Discussion

This exception grants third parties (e.g. advertising networks) some latitude to mitigate security risks. Websites that users store sensitive personal information on (e.g. financial services and webmail) are all first-party; they are able to collect, retain, and use information about all users for security purposes.

3.2.2.8 Fraud Prevention

3.2.2.8.1 Operative Text

A third party may collect, retain, and use data about a particular user or user agent for the purpose of preventing fraud, provided that there are reasonable grounds to believe the user or user agent was attempting to commit fraud at the time the data was collected.

3.2.2.8.2 Non-Normative Discussion

When a user meaningfully interacts with third-party content (e.g. clicking an ad), the third party can collect, retain, and use information for fraud prevention. Third parties can also use protocol logs for fraud prevention (subject to appropriate retention limits, which this draft does not address). This exception provides an additional capability to, in certain

circumstances, track impressions for fraud prevention.

A. References

A.1 Normative references

No normative references.

A.2 Informative references

No informative references.

Parties and Necessary Business Uses

We appreciate all the hard work being put in by the W3C, the co-chairs, and all of the stakeholders participating within the Tracking Protection Working Group. The ultimate objective is a standard that will be implemented by a significant portion of the ecosystem. A standard that is not adopted does not benefit consumers and that is everyone's objective – a practical, easy-to-use tool that will enhance consumers' ability to express preferences about certain data collection and use. In order to make that objective possible, the following proposal is put forward regarding exemptions as an attempt to introduce additional important aspects of the DAA Self-Regulatory Principles for Multi-site Data to the existing discussion on permitted data uses for necessary business activities when a user expressly turns on Do-Not-Track (DNT:1).

Part I: Parties

Definitions

A. A party is any commercial, nonprofit, or governmental organization, a subsidiary or unit of such an organization, or a person.
B. For unique corporate entities to qualify as a common party with respect to this standard, those entities MUST be commonly owned and commonly controlled (Affiliates).
C. A First Party is the party that owns the Web site or has control over the Web site the consumer visits. A First Party also includes the owner of a widget, search box or similar service with which a consumer interacts.
    a. NOTE: If a user merely moused over, closed, or muted third-party content, that is not sufficient interaction
D. A Third Party is any party other than a First Party or a user.

Rules

If a user has not granted an exception (via browser agent or out-of-band consent) AND if an activity is not allowed under Permitted Uses, THEN the following general party rules apply when a user expressly sets their tracking preference to DNT:1:

- 1st parties MAY collect and profile in the context of the 1st party experience.

- 3rd parties MUST NOT use data across multiple, non-affiliated websites.

NOTE: Data collected by a 3rd party MUST be segregated according to the 1st party from which it was collected. A 3rd party MUST NOT aggregate, correlate, or use together data that was collected on different 1st party sites.

- 3rd parties MUST NOT add collected data to a "profile" of a user.

- 3rd parties MUST NOT leverage previously collected transactional data to profile a user or to alter a user's experience.

- 3rd parties MUST NOT attempt to personally identify a user.
- A party MUST NOT share (send or receive) collected data or profiles with another party (unless that party is ONLY working on the behalf of that specific party – aka Service Provider relationship).

NOTE: (Outside of DNT Context): Data legitimately collected and received from a party MAY be combined with existing 1st party profile data.

- A party MAY choose to remove any previously profiled data.
- All permitted data uses for necessary business activities apply in all cases.
- User granted exceptions (through DNT standard or out-of-band) supersede these rules.

Part II: Permitted Data Uses for Necessary Business Activities when DNT:1

For all of these permitted uses, the complying entity must make reasonable data minimization efforts to ensure that only the data necessary for the permitted use be retained. This is described under the draft heading "4.4 Usage-based Permitted Uses." The option to designate that restriction was not provided by this template so the restriction on scope is highlighted here and then also applied as an "E" limitation below.

1. **Frequency Capping -** Data MAY be collected and used for the limited purpose of frequency capping. Restricting the number of times a user agent displays ads prevents a user from having to see repetitive ads, prevents publishers from displaying repetitive ads, and prevents advertisers from harming the reputation of their clients.

- Examples of important data uses include, but are not limited to:
  - Reach and frequency metrics
  - Ad performance
  - Logging the number and type of advertisements served on a particular Web site(s).

  (For additional details see DAA Self-Regulatory Principles for Multi-site Data: Reporting)

E. This particular use is allowed with reasonable data minimization efforts

2. **Financial Logging -** Data MAY be collected and used for the limited purpose of billing or product or service fulfillment.

Comment: Ad impressions and clicks (and sometimes conversions) events are tied to financial transactions.

E. This particular use is allowed with reasonable data minimization efforts

3. **3rd Party Auditing -** Data MAY be collected and used for the limited purpose of 3rd Party Auditing. Online advertising is a billed event and there are concerns with accuracy in impression counting and quality of placement so 3rd party auditors provide an independent reporting service to advertisers and agencies so they can compare reporting for accuracy.

> **Comment:** This data use serves an important business purpose in preventing fraud and reasonable data minimization efforts can insure privacy for users
>
> **E.** This particular use is allowed with reasonable data minimization efforts

4. **Security -** Data MAY be collected and used for the limited purpose of security. Security data is any data reasonably necessary for enabling authentication/verification, providing fraud prevention, or bolstering security.

> **Comment**: Restrictions on security efforts would certainly harm users. We do no want to mistakenly turn the DNT:1 signal into a signal for user vulnerability. (For additional details see the DAA Self-Regulatory Principles for Multi-site Data: Authentication, Verification, Fraud Prevention and Security & Compliance, & Public Purpose and Consumer Safety)
>
> **E.** This particular use is allowed with reasonable data minimization efforts

5. Contextual Content or Ad Serving - Data MAY be collected and used for the limited purpose of contextual content or ad serving (examples: serving advertising or content based on the Web page content, search query, time of day or general geographic location detected from current interaction only) as long as the data is used by a party with which the user interacts and is not collected and used for the purpose of advertising on Web sites of non-Affiliate parties.

> **Comment:** Depending on the definition of tracking, defined in Section 3.7, this exemption may not need to be included because the serving of contextual ads will not be within the scope of the definition.
>
> This particular use is allowed without qualification

6. Research / Market Analytics - Data MAY be collected and used for the limited purpose of research & market analytics as long as collection and use are limited in scope to the analysis of**:**

- the characteristics of a market or group of consumers; or
- the performance of a product, service or feature, in order to improve existing products or services or to develop new products or services.

> Data used for this limited purpose is allowed with aggregation. (For additional details see the DAA Self-Regulatory Principles for Multi-site Data: Market Research & Product Development)

D. This particular use is allowed with aggregation where the data may not be re-identified to market directly back to, or otherwise re-contact a specific computer or device.

7. Product Improvement, or, more narrowly, Debugging – Data MAY be collected and used for the limited purpose of product improvement. This includes data used for the express purpose of product improvement related to debugging to specific events, devices, or site locations.

E. This particular use is allowed with reasonable data minimization efforts

8. Legal Compliance & Public Purpose: Data MAY be collected and used for the limited purpose of legal compliance and public purpose. This includes, but is not limited to, intellectual property protection or using location data for emergency services.

E. This particular use is allowed with reasonable data minimization efforts

9. "Unlinkable" data – we believe this is already covered by general anonymization and aggregation approaches that are tied to a specific identifiable individual or device.

Proposed Definition: The FTC defines "linkable" as "consumer data that can be *reasonably linked* to a specific consumer, computer, or other device." [Emphasis added] This reflects a scaled approach rather than a bright line distinction for determining privacy protection. Data is "unlinkable" if it goes through a de-identification process. A de-identification process is sufficient when an entity has taken *reasonable steps* to ensure that the data cannot reasonably be re-associated or connected to an individual or connected to or be associated with a particular computer or device. (For additional details see the DAA Self-Regulatory Principles for Multi-site Data: De-Identification Process Definition)