

Response to the Call for Objections on ISSUE-5 What definition of “Tracking” to use and where to include it

The Tracking Protection Working Group is chartered “to improve user privacy and user control by defining mechanisms for expressing user preferences around Web tracking.”

In this Call for Objections, the questions before the group were what definition of “tracking“ to choose as the basis for further work and in which specification document to include this definition. Thus far, neither the Charter nor the two specifications Tracking Preference Expression (DNT) and Tracking Compliance and Scope included a consensus definition of the term “Tracking.”

Based on the comments submitted the co-chairs conclude that the Working Group has made the decision to chose Option A as the basis for its future work. The definition *“Tracking is the collection of data regarding a particular user’s activity across multiple distinct contexts and the retention, use, or sharing of data derived from that activity outside the context in which it occurred”* provides the most appropriate outline of what the Working Group has been been trying to address so far and will continue to work on in the future. We also conclude, based on the received comments, that in order to get a comprehensive, adoptable and self-contained Tracking Preference Expression (DNT) specification out to last call for implementation and testing, it is necessary to include the above definition in the technical specification.

This definition of tracking is also most consistent with the history of Do Not Track and this W3C Working Group. “Do Not Track” was first proposed by U.S. consumer groups in 2007 as a means to express a persistent opt-out to cross-site behavioral tracking and advertising. This proposal was endorsed by the Federal Trade Commission in December 2010, when they called on companies to implement a Do Not Track mechanism to allow users to opt out of behavioral advertising and data collection. The Web Tracking Protection submission that led to the creating of this Working Group was also focused entirely on cross-domain web tracking. The working assumption of the Working Group for most of the past two-and-a-half years has been that the Do Not Track signal would apply primarily to companies who track users across different first-party sites; for example, all five proposals presented at the Washington face-to-face meeting focused primarily on limiting cross-context tracking.

For some time, several Working Group members have complained that the Working Group has failed to precisely define the term “tracking.” That complaint has now been resolved by this decision of the Working Group.

History

The Working Group was formed in September, 2011. Currently, the Working Group is chaired by three co-chairs: Justin Brookman (CDT), Carl Cargill (Adobe), and Matthias Schunter (Intel). The Working Group currently has 106 participants, representing consumer groups, browsers, first parties, third parties, regulators, subject matter experts, and others. Since its formation, the group has held weekly teleconferences, met face-to-face regularly, and exchanged thousands of emails on its mailing list, with all proceedings public at <http://www.w3.org/2011/tracking-protection/>.

Since the beginning the Working Group has made several efforts to come up with a consensus definition for the term “tracking.” The discussion at the third Face-to-Face meeting in Brussels could be given as example, where the proposed interpretations of “Do Not Track” ranged from “Do Not Collect” to “Do Not Profile” to “Do Not Personalize.”

The definition of tracking has long been an issue against the Compliance specification. As the co-chairs decided in October 2013 based on the feedback of the Working Group to finalize the TPE specification first and bring it to Last Call, this has raised the question of whether the TPE specification does need some core definitions to be self-contained for Last Call implementation.

In the following discussion, several text proposals were made by members of the Working Group. In a joint effort these have been worked on, merged, and finally limited to two different definitions in normative text that were put before the Working Group to make a decision.

The options

1. “Option A: Tracking across multiple distinct contexts

Tracking is the collection of data regarding a particular user's activity across multiple distinct contexts and the retention, use, or sharing of data derived from that activity outside the context in which it occurred.

2. Option B: Retention/use associated with user, user agent, device

In general terms, tracking is the retention or use after a network transaction is complete, or sharing, of data that is, or can be, associated with a specific user, user agent, or device.

Non-normative text

Tracking may result in the compilation of a database about a person and their online activity, perhaps without their knowledge. Harms from this might include direct ones, such as differential pricing or service provision, through to major ones, including the consequences of public revelation of the database, access to it by persons with criminal intent, or its use by government or other bodies.

Note that the extent to which tracking data may nonetheless be retained in the presence of this signal under some circumstances is defined in the companion specification.

3. Option C: No Definition/Definition Location

No definition; remove from Definitions section, rest of document unchanged.

Additionally, by Option C the participants were asked for objections to including the definition of tracking in either the TPE document or the Compliance document and to describe their objection.

The Call for Objections was open from November 8, 2013 to November 20, 2013. In total 14 Members of the Working Group participated and presented arguments against or in favor of the options. The full results of the questionnaire are public at <https://www.w3.org/2002/09/wbs/49311/twpg-tracking-5/results>.

After consideration, the co-chairs have determined that the group has rejected Option B as too broad and unsuited for the scope of the charter.

Explanatory considerations on the choice of definition

The decision was made by exclusion. Based on assessing the substance of the objections against each option.

Objections against Option B:

Overall the chairs conclude that the Working Group raised the more substantial arguments against Option B.

Shane Wiley voiced the concern that this definition “provides for far too broad of a view of “tracking” including any and all events online”. The breadth of this definition was similarly criticized by Roy Fielding, stating that Option B would “include any retention of personal data, including that by first party sites”. Vinay Goel pointed out was that this definition “completely ignores the intent of the working group to place limits on data collection and use by a company across unaffiliated websites“. The co-chairs agree that the main focus of the Working Group so far has been to put restrictions on data collection and use by third parties (or parties acting in a third-party context). The received comments reflect that Option B did not appropriately portray this understanding within the group.

Several participants, Brooks Dobbs, Brad Kulick, Vinay Goel, and Shane Wiley, raised the objection that the non-normative text accompanying Option B has the potential to lead to misunderstandings of users, as their expression of choice will not affect all data collection online, but only within the limitations of a companion compliance document that will not reach Last Call status simultaneously. Jack Hobaugh criticized the non-normative text as premature and prejudicing the discussion process for the Compliance specification.

Roy Fielding stated that most hardware routers would be tracking users according to this definition because they retain status monitoring and caching data after a network transaction is complete. They “retain IP addresses and/or session pointers in their routing tables until they are replaced by later connection.”

The chairs find these objections to be stronger than the objections posed to Option A. For the last two years it has been the working assumption of the group that first party data collection and use would have much fewer restrictions than third parties. If the TPE specification would define each and every data collection on the web as tracking, the standard might be expected to block data collection and use that a user might contextually understand and want, leading to confusion — or to users turning the setting off. For these reasons, Option B is inconsistent with the Charter and the longstanding working assumption within the group over to whom the Do Not Track signal should apply, and it should not be the definition of tracking that the standard uses.

Objections against Option A:

There were also several objections raised against Option A. Nonetheless, the co-chairs conclude that these objections are not as strong, or they can be resolved within the Working Group.

The concern most often voiced was that Option A relies on the concept of “context.” Brooks Dobbs, Mike O’Neill, David Wainberg, David Singer, John Simpson, Brad Kulick, Jack Hobaugh, and Shane Wiley similarly stated that “context” is ambiguous and equally un-agreed upon as the term tracking. The co-chairs think that a definition of tracking is a significant step towards getting the TPE specification to Last Call. The co-chairs agree, however, that this is a legitimate concern about the Option A definition, and the group may decide that a definition of context is necessary because the meaning is not globally self-evident to individual end users. However, this definition is still more closely scoped to the work and understanding of the Working Group than Option B.

Several participants objected to this definition because it was more narrow than Option B and thus would have a more limited effect on protecting user privacy. Rob van Eijk wrote that “defining the boundaries of context by anything else but the user would seriously undermine the user centric nature of Do Not Track. Context is highly subjective from a user perspective and should therefore not be pinned down by boundaries like common branding or other assumptions.” Additionally, Mike O’Neill and David Singer raised their concern that Option A could allow any collection and use of data within a single context, even by third parties and would therefore be at odds with the current Compliance specification. The co-chairs consider these objections run counter to the Working Group’s understanding that Do Not Track is not fundamentally intended to limit data collection and use by first parties with which the user has a direct relationship. Moreover, we are not convinced

that Option A's definition of tracking runs counter to the current TCS (which is not anyways at consensus), as siloed service providers may collect and use data on behalf of first-party publishers. In any event, if the group decides there is ambiguity about what context means, we can add further language to the document.

David Singer, and similarly John Simpson, argues that this definition may be "not comprehensible to users, whereas the purpose of the definition is to reflect the users' general desires. Users' general objection is to 'keeping records about them' and this is more complex and rather different from that." The chairs disagree that a prohibition on companies "keeping records about them" would necessarily be users' intent in turning on Do Not Track. That would heavily depend on the user agent's interface for the Do Not Track setting, and is considerably broader than anything the group has considered in negotiating the TCS.

Mike O'Neill brought up the issue of compatibility with existing legal regimes. "Restricting the definition of tracking in this way may also reduce the likelihood that this specification could fulfill the requirements for "browser settings" in Recital 66 of the European e-privacy directive or as an automated mechanism to impart a consent or "right to object" signal referred to by the forthcoming general data protection regulation. This could lessen the chance of getting legal backing for the standard in Europe." The Working Group's charter was not to create a legal compliance regime for one particular jurisdiction; it was merely to allow users a mechanism to express a preference to prevent or allow certain tracking behavior. For a long time the group's working assumption for the TCS was to specifically address cross-site tracking, not first-party data collection and use. Whether the expression of a particular preference has legal ramifications in any particular regime is beyond the control of this Working Group, though the issue has been discussed in the Global Considerations sub-group.

Another objection was raised by David Wainberg, who is concerned that "'Data derived' from users' activity is over-broad in that it could include data and data uses that should not reasonably fall within the definition of tracking, including, for example, aggregate data, or other data that is not tracking data". The co-chairs believe that non-normative text could be developed to portray the current consensus of the Working Group regarding the exclusion from scope of de-identified data (though there is disagreement within the group over how strictly data must be de-identified). This objection, however, is not as fundamental as the objections to Option B.

Based on these comments received in the Call for Objections, the co-chairs conclude that *Option A: Tracking across multiple distinct contexts* portrays more adequately the intent and previous work of the Working Group.

Objections against Option C:

There emerged three clusters of answers from Option C. First, to have no definition of tracking, second, to include it in the TPE, third, to not include it in the TPE.

Although several participants objected the inclusion of this (or any) definition in the TPE, the co-chairs believe that the stronger arguments were in favor of defining tracking within the TPE.

Mike O'Neill's statement in favor of the TPE emphasizes that "this is an essential definition because it is fundamental to the whole standard. There needs to be a definition in general terms of tracking, though we can leave it to the TCS to detail how servers should respond to a DNT signal."

David Singer states leaving out the definition of tracking of the TPE would fail "to set the stage, and describe even in rough terms what is being expressed by the 'T' in DNT".

Roy Fielding describes the importance of this definition in the TPE to allow all stakeholders to have a common understanding of what a user's preference expression does mean. "The essence of

standards is to ensure that all parties share the same vocabulary when communicating. Consistency in user expression depends on a common definition and consistency in its implementation by browsers. The way to get that consistency is to include a definition in the protocol document that all user agent implementers need to read to properly implement the protocol, so that user agents will (in turn) describe it consistently to users when offering the configuration option.” “We need a definition of tracking because the protocol is supposed to be expressing a user preference. Hence, the user needs a definition of tracking (or "to track") in order to have an informed preference, and sites need to know what that definition is in order to understand the meaning being expressed by the user and to ensure that their own behavior is consistent with how they have informed users in the exception dialogs and privacy policies”

The co-chairs agree with the substantive technical arguments raised by the TPE editors, David Singer and Roy Fielding. The definition of Tracking within the technical TPE specification is necessary to define the scope of the specification, to define the meaning of the "DNT;1" signal by the user, and to ensure that users and recipients have a common understanding of the DNT signal (independent of how a server later addresses this user preference).

Several members of the Working Group objected the inclusion of any definition in the technical specification TPE. Shane Wiley, and similarly Chris Mejia and Brad Kulick, argue that “TPE fully functions in an accurate and expected manner by allowing Servers to offer up their compliance program (and the definition of tracking within them) in a well-known URI and/or in pointers provided in Server responses. UAs can provide these to users in real-time.” Chris Mejia adds that “no further context [is] required in a pure tech protocol on how a HTTP header is sent by browsers and received by servers. [...] how the receiving server handles a preference statement from the user should be dependent on the compliance regime honored by the called server, and this should be messaged accordingly to the user, so the user understands the choice she is making, or has made.” Based on the aforementioned arguments in favor of inclusion to the TPE the co-chairs respectfully disagree. Without a basic outline or context, servers will have no consistent expectation about what a user is intended to convey when it receives a Do Not Track signal.

Chris Mejia, and also Brad Kulick, brought up the possibility of multiple compliance regimes. “TPE will not stand alone in the marketplace. It will be accompanied by a compliance regime. Definitions in the tech spec creates risks - If these ported definitions were too broad, valuable context could be lost. If these ported definitions were too narrow, technical and policy compliance may be jeopardized.” Technically, the TPE is designed to standardize the expression of a particular preference. It is necessary then to include within this standard what exactly it is that the user is expressing, and to whom. The chairs will keep in mind, however, the fundamental concern about transferring from the Compliance document elements of how third-party servers should *comply* with that standardized request — that is not the intent of the chairs or the members of the Working Group.

Alan Chapell criticized the discussion as not honoring the previous decision on prioritizing TPE. “This working group voted, and the preference of the working group was to work on the TPE document. W3C staff and chairs are attempting to create a de facto Compliance document by importing key compliance-specific terms into the TPE.” The co-chairs do not agree that by adding very few core definitions to the TPE specification to outline the scope and make it self-contained the Working Group will produce a “Compliance light” document. The TPE does not dictate what a server must do in response to a Do Not Track preference; it only standardizes the expression and offers a mechanism by which the server can communicate back to the user. What the server must do in order to comply with the preference expression will subsequently be addressed in the TCS document.

Concerns about the dependency on the compliance specification were raised by Brooks Dobbs and David Wainberg: “Consistency between user expectation and compliance requirements: Any definition of tracking in TPE will necessarily be incomplete and therefore incorrect. If ‘tracking’ is

the activity that users will be able to effect a choice with respect to, but compliance with that choice is to be determined in another document (documents) there is a logical breakdown.” The definition of tracking is not meant to anticipate the complete Compliance specification but to outline a common scope of what a user preference does mean. Leaving the further guidelines on how to comply with this signal to a compliance regime does not render the definition incomplete or useless. It allows the user to make a well-defined choice about her own preferences and to allow servers to respond back about what they are doing in response.

Alan Chapell further criticizes that “the definition for tracking is dependent on other definitions (e.g., collection, retention, use, sharing, and context). As a result, its difficult to assess their impact on the DNT standard in isolation. There is a significant possibility that multiple working group members will look to reopen this issue depending upon how the myriad of dependencies are defined.” The co-chairs agree that no matter where we include the tracking definition, it will lead to further efforts to define the underlying concepts. To make the TPE self-contained but not to overload it with definitions, the Working Group has to carefully consider which definitions are truly necessary. The co-chairs plan to review the TPE document with the Working Group before Last Call and remove all definitions that are not necessary to implement and understand the specification.

Result

In conclusion, the *ISSUE-5, What is the definition of tracking?*, is hereby closed, and the following definition will set the outline for the group’s continued work on TPE and TCS: “Tracking is the collection of data regarding a particular user’s activity across multiple distinct contexts and the retention, use, or sharing of data derived from that activity outside the context in which it occurred.”