



Point of View : **Browser-aware Data Collection**

Reza Jalili, Head of Architecture Strategy, Adobe Digital Marketing



Instrumentation And data collection

events and capturing of semantics, actions, intentions, and preferences with or without data being collected off a page; in other words, data and signals are generated and sent to a processor somewhere. As well, data from the page can be collected, collated, transformed, and sent to a processor.

Practical Observation

Data collection is done with vendor/
collector-specific javascript.

Zen Problem

Browsers do not know the reasons for javascript communication with remote servers.

Technical Problem

Technical barriers to open combination of data collection technologies

- each library uses its own means of collection and transmission
- the names, labels, and semantics of what is collected vary
- vendor data collection endpoints vary and they cannot communicate without one another's formats and URL(s)

Customer Problem

Limited by the current technologies for data collection, site owners are missing out on potential gains of an open data collection interface.

- a. combining data collection services could add value to a site owner's property
- b. flexibility of changing vendors is limited

Problem: Objectives

For example, a site might have data collection libraries running on a client's browser for the following purposes from different non-interoperable vendors:

1. to gather basic site navigation information for usability purposes
2. to gather content attention information for monetization of content and advertising
3. to personalize the user's experience

Proposal

Objectives:

- Know what is being collected
- Understand the entities legally involved
- Guide user with respect to privacy and data sharing options.
- Give control over **which** data processors are used.
- Allow users to modify the data being collected, if they so choose

More?

We envision, for example, either a to be created standard browser mechanism or Browser-enabled standard JS library

- that can be configured by a site owner to enable data collection
- that can be used by any javascript library to deliver collected data
- that can be data format agnostic or allow for a standard data object (e.g. see the work
- of the w3c customer experience data collection working group)
- and which might further enable economic and social models to emerge

What if

Sites declared:

- I am _____
- _____ collects data for me.
- My privacy statement is here
- Set your preferences here

"manifest"

(A software term)

And

relation to privacy and what we are after:

*readily perceived by the senses and
especially by the sense of sight*

Might look like ...

| attribute / parameter / member | value | comments / notes |
|--------------------------------|--|--|
| dataCollector | {label, URL to about, URL to policy} | |
| dataProcessor | {label, URL to about, URL to policy} | |
| options | URL to human-readable UI to select opt-in options and view current selected options if/when possible | Opt-in choices: URL to page to allow those settings and view current choices |
| scripts | [{URL1, [DOM scopes]}, {URL2, [DOM scopes2]}, ...] | Scripts: URL of data collection script (make this array of URLs?) and DOM Scopes: identifiers of classes or use jquery selectors to indicate which dom elements are being instrumented |
| deviceScopes | [labels] | Device Scopes: What is being tracked about the device |
| userScopes | [labels] | User Scopes: What is being tracked about the user |
| appScopes | [labels] | Application Scopes: What is being tracked about the application usage |
| extensions | [{URL1, [DOM scopes]}, {URL2, [DOM scopes2]}, ...] | Extensions: URLs of where extension definitions are loaded from |
| collectedData | {} | events and other collected data objects |

And ...

For communication of the data collection manifest to the browser / device

use HTTP headers for data collection manifest

transmit the json manifest in an httpd- text header or one header per field of the manifest of same name

add a new <instrumentation> tag with attributes and properties

embed the instrumentation manifest and specific instrumentation declarations inside a new tag to be defined

use the <head> section

use a new attribute to the <Script> tag to indicate that a data collection manifest is created by the script

data collection and instrumentation will be used interchangeably; preference is to move to "instrumentation"

For actual instrumentation of UI elements and actions with semantics

use extensions to CSS to define instrumentation, and source and sink of events

.cis for instrumentation file that similar to CSS is a Cascading Instrumentation Sheet

<head> <link rel="stylesheet" type="text/cis" href="theme.ins"> </head>

.cis for instrumentation file that similar to CSS is a Cascading Instrumentation Sheet



Adobe