Using Analytics in a Privacy-Protective Way on Government Web Sites: CDT Recommendations

May 21, 2009

Policy Posts are in-depth analyses on current tech policy issues from CDT experts. Sign up to receive the latest Policy Posts here:

President Obama's pledge to make the government more transparent, accountable and participatory is already generating increased interest from both the commercial and government sectors in capitalizing on all the Internet has to offer. Federal agency Web sites will have a key role to play in this endeavor. Since the E-Government Act established a mandate to move government services online in 2002, many agencies have accelerated their Web development and some have created interactive tools for their constituents. But the Web has evolved dramatically in recent years. To truly harness the power of today's Web, agency Web sites will require continual improvements.

---

1) The Role of Analytics on Government Web Sites

President Obama's pledge to make the government more transparent, accountable and participatory is already generating increased interest from both the commercial and government sectors in capitalizing on all the Internet has to offer. Federal agency Web sites will have a key role to play in this endeavor. Since the E-Government Act established a mandate to move government services online in 2002, many agencies have accelerated their Web development and some have created interactive tools for their constituents. But the Web has evolved dramatically in recent years. To truly harness the power of today's Web, agency Web sites will require continual improvements.

These improvements will not succeed, however, if they are not viewed through the lens of protecting citizen privacy. Given the government's increasing appetite for citizen data in recent years, the public is rightly skeptical about data collection on government Web sites. Strong federal guidance already exists about how federal agencies may collect data on their Web sites, and a continued focus on privacy is essential as agencies begin to bring their Web sites into the Web 2.0 era.

Federal Web managers are justifiably enticed by the wealth of recent developments on the commercial Web, from personalization and social networking techniques to novel content delivery and Web analytics approaches. As the government pursues new open government strategies, the benefits and risks of each of these techniques will need to be examined. CDT and the Electronic Frontier Foundation (EFF) have crafted a set of recommendations that focus on the practice of "Web measurement:" the collection and analysis of Internet data that is reported in the aggregate and used for the purposes of understanding and optimizing Web usage.

CDT/EFF Recommendations (May 2009) [1]

Social Media and the Federal Government [2]: Perceived and Real Barriers and Potential Solutions (Dec. 23, 2008)

---

2) Web Measurement and Current Federal Privacy Guidance

Since the Web's earliest days, Web site owners have had a strong interest in measuring the performance of their sites. What began as simple counters displaying the number of visits to a

particular page has evolved into a thriving and diverse commercial industry in Web measurement tools and services. The industry that is popularly known as "Web analytics" has become a vital sector of the online economy. The Web Analytics Association defines Web analytics as follows:

- Web analytics – The measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage.

These recommendations focus on "Web measurement," which we define as a subset of Web analytics:

- Web measurement – The collection and analysis of Internet data that is reported in the aggregate and used for the purposes of understanding and optimizing Web usage. Because results of the analysis are reported in the aggregate, the risk of re-identifying an individual using only the reported data is negligible. Any individual-level data collected for the purpose of Web measurement is retained only for a limited time period.

Web measurement is confined to reporting results in the aggregate, whereas Web analytics covers a broader space of practices that may involve reporting individual-level data. We define individual-level data as follows:

- Individual-level data – Data about an individual Web site visit.

For the purpose of discussing Web measurement in the government context, it is useful to distinguish two different types of Web measurement:

- Single-session measurement – Measures a single user's back-to-back interactions with a site within a limited time period (a "session"). Any identifier correlated to a particular user is only used within that session, is not later reused, and, ideally, is deleted after the measurement is performed.
- Cross-session measurement – Measures a single user's site usage over time. Requires the use of a persistent identifier per user, which lasts across sessions.

Because of its use of persistent identification, cross-session measurement raises more privacy risks than single-session measurement. By necessity, Web measurement tools make use of technologies that track individual user behavior on Web sites. For cross-session measurement, persistent cookies have been and continue to be the most common mechanism used to identify and track users, although other mechanisms exist and are likely to see increased uptake in the future.

Federal agencies' use of Web measurement tools is governed by federal guidance about persistent tracking technologies. This policy was originally issued in June 2000 after it was revealed that the Office of National Drug Control Policy had contracted with a commercial ad network (DoubleClick) to use persistent cookies to track users as part of an advertising campaign. In response to criticism of this tracking, the Office of Management and Budget (OMB) released a policy about the use of cookies by federal agencies, explicitly stating a presumption that federal Web sites would not use persistent cookies and explaining what was required of federal agencies when cookie use was deemed necessary. This policy made it difficult, but not impossible, to use persistent cookies on federal Web sites.

In 2003, OMB issued guidance on the E-Government Act privacy implementation, expanding the scope of this existing policy beyond cookies to include any technology that can track site visitors beyond a single session. Based on this update, federal Web sites are currently prohibited from using persistent tracking technologies unless four conditions are met:

i. The site gives clear and conspicuous notice of the use of the technology;

ii. There is a compelling need to gather the data on the site;
iii. Appropriate and publicly disclosed privacy safeguards exist for handling any information derived from the technology; and
iv. The agency head gives personal approval for use of the technology.

This is a highly privacy-protective policy that provides useful safeguards and appropriate guidance with regards to agency disclosure. In practice, however, it has resulted in a near prohibition of persistent tracking technologies, largely because obtaining agency head approval can be an extremely difficult task. In many cases, obtaining this approval requires federal Web managers to work their requests up through the entire agency hierarchy, a process whose success is often based more on the strength of personal connections than the soundness of a request.

The current federal policy also fails to allow for user control and choice, treating all persistent tracking equally regardless of whether it is user-activated or not. Thus, federal agencies have been restricted in their ability to offer users the option of advanced features that are powered by persistent cookies or other tracking technologies, because agency head approval is required whether user controls are offered or not.

CDT Blog: A New Cookie Policy for eGov 2.0 – Part 1 [3]

CDT Blog: A New Cookie Policy for eGov 2.0 – Part 2 [4]

OMB Guidance [5] for Implementing the Privacy Provisions of the E-Government Act of 2002 (Sept. 26, 2003)

---

3) CDT Recommendations

Deploying Web measurement in a privacy-preserving manner on agency Web sites will require both technical and policy-based protections in agencies' measurement tools. The recommendations in the first section below suggest how to go about implementing these protections. These same recommendations could be applied to all server logs generated and stored by agency Web sites.

Even if agencies follow all of these implementation recommendations, the federal policy on persistent tracking technologies may still prevent some agencies from using Web measurement tools (primarily because obtaining approval from the head of the agency can be extremely difficult). The second part of this section addresses how to update current federal policy to allow for privacy-preserving uses of Web measurement.

Implementation Recommendations

1. Use data only for measurement purposes

Data collected for Web measurement should only be used for measurement and protecting against measurement fraud. There are a number of different deployment configurations for cross-session measurement on agency sites where responsibility for collecting and safeguarding measurement data may or may not be shared between an agency and a commercial third party. We recommend that the agency run the measurement tools itself, collecting individual-level data and using the tools to perform the necessary analysis and reporting. The agency should serve persistent user identifiers (through cookies or other means) from its own first-party Web domain. In this configuration, the agency should use measurement data only for measurement purposes. Measurement data should not be disclosed to any other government or commercial entity.

This gives the agency responsibility for safeguarding user data, removing the possibility for secondary use of the data by a commercial third party. It also avoids the possibility that the data collected by a commercial third party be subject to lawful requests or legal processes, because no third party collects data.

2. Prominently disclose

Federal agencies using Web measurement tools on their sites should disclose, at a minimum, the following items:

- The fact that measurement is happening,
- The reasons for conducting the measurement,
- The type of measurement used (single-session or cross-session),
- The technologies used to measure,
- The identities of all third-party vendors involved in the measurement process,
- How site visitors can exercise choice about having their behavior measured,
- The data retention policies of the agency and all third-party vendors involved, and
- How measurement data is safeguarded.

These disclosures should be made both (a) as part of an agency Privacy Impact Assessment (PIA) published at least 30 days before the agency begins use of the Web measurement tools, and (b) as part of the agency Web site privacy policy.

3. Offer choice

Site visitors should be offered choices about having their data collected for cross-session measurement. The choice mechanism(s) and the visitor's choice status should be clearly visible on every page of the agency site.

For example, an agency could provide a simple on/off switch on each page of its site, with one option highlighted to indicate the user's current status and the other option provided as a link to allow the user to switch his or her status at any time. This kind of persistent choice indicator is already in use on commercial sites. Site visitors should be given detailed information about how the choice mechanisms work and other means to stop persistent tracking, such as links to descriptions about how to use cookie blocking and deletion tools.

4. Limit data retention

The individual-level data collected for measurement purposes should only be retained for a limited period of time. The time frame for which individual-level data is retained should be no longer than 90 days and should be correlated to specific measurement goals. For example, if an agency is looking to measure site usage on a monthly basis, individual-level data should be deleted each month. Individual-level data associated with single-session measurement should be deleted soon after session completion.

Elements of individual-level data logs that are not relevant to measurement analysis and reporting should be deleted as soon as possible after the data is collected. IP addresses, for example, should be deleted (and possibly replaced with their corresponding geographic or ISP information) soon after collection, if not immediately.

If an agency contracts with a commercial partner for measurement tools, the data retention time frames that apply to individual-level data collected by the partner should be explicitly stated in the contract.

5. Limit cross-session measurement

Federal agencies should only use cross-session measurement when single-session measurement cannot be used to obtain the same metric. Single-session measurement has many uses and may suffice for many agencies. Examples of metrics that can be calculated using only single-session measurement include: measuring how often users take a particular navigation path through a site; measuring the "bounce rate" (how often users hit an agency page and immediately navigate away); and taking aggregate measures of which other sites serve as sources of traffic to an agency site (by analyzing referring pages).

6. Obtain third-party verification

Prior to beginning use of Web measurement tools, the agency office responsible for privacy should review the agency's published Privacy Impact Assessment to ensure compliance with the above five recommendations.

Agencies engaged in Web measurement and their partners should also regularly review their systems and procedures to determine if they are in compliance with the above five recommendations. Agencies should be required to report the results of these reviews to, or have these reviews undertaken by, OMB, their Inspectors General and/or a designated independent third party.

Federal Policy Updates

The current federal policy on persistent tracking technologies should be updated to deal specifically with Web measurement. This update should prohibit agencies from using persistent tracking technologies for cross-session measurement purposes unless they can meet all of the implementation recommendations. If they can meet all of these recommendations, then they might forego agency head approval. Any agency that cannot meet all eight conditions would still be required to obtain agency head approval, and in so doing the agency's Web manager would need to explain at each level of the agency hierarchy why it cannot meet all of the conditions.

Source URL:
https://www.cdt.org/policy/using-analytics-privacy-protective-way-government-web-sites-cdt-recommendations

Links:
[1] http://cdt.org/privacy/20090512_analytics.pdf
[2] http://www.usa.gov/webcontent/documents/SocialMediaFed%20Govt_BarriersPotentialSolutions.pdf
[3] http://blog.cdt.org/2009/01/08/a-new-cookie-policy-for-egov-20-%E2%80%93-part-i/
[4] http://blog.cdt.org/2009/01/09/a-new-cookie-policy-for-e-gov-20-part-2/
[5] http://www.whitehouse.gov/omb/memoranda/m03-22.html