

Open Recommendations for the Use of Web Measurement Tools on Federal Government Web Sites

*Center for Democracy & Technology
Electronic Frontier Foundation*

May 2009

▣ Executive Summary

The Internet is an ideal medium for citizen engagement in government, and Federal agency Web sites will have a key role to play in making government more transparent, accountable and participatory. To succeed, the operation and improvement of Federal agency Web sites must be done through the lens of protecting citizen privacy.

As the government pursues new open government strategies, the benefits and risks of each strategy will need to be examined. This paper focuses 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.

The Center for Democracy & Technology (CDT) and the Electronic Frontier Foundation (EFF) have analyzed the existing policy and technology frameworks around Web measurement and produced a set of recommendations for its use on federal agency Web sites.¹ We suggest that agencies should only be permitted to use Web measurement on their sites if they:

¹ The authors also sought an opinion on technical issues from The Sunlight Foundation.

Use data only for measurement

Data collected for Web measurement should only be used for that purpose. Agencies should avoid outsourcing data collection to commercial partners.

Prominently disclose

Federal agencies using Web measurement tools on their sites should provide disclosures in their privacy policies about the tools.

Offer choice

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

Limit data retention

The individual-level data collected for measurement purposes should be retained for no more than 90 days. The retention time frames should be disclosed, correlated to the purpose for which the data was collected, enforced through technology, and explicitly stated in commercial partner contracts.

Limit cross-session measurement

Federal agencies should only use cross-session measurement when single-session measurement cannot be used to obtain the same metric.

Obtain third-party verification

Agencies engaged in Web measurement and their partners should have their privacy compliance procedures regularly verified by their Inspectors General or a designated independent third party.

We also suggest that the current federal policy on the use of persistent tracking technologies be updated to allow Web managers to use persistent tracking technologies for Web measurement purposes if and only if the above six conditions plus several of the existing federal policy conditions are met.

▣ Introduction

The Internet is an ideal medium for citizen engagement in government. Digital technologies have already transformed the dissemination of government information and citizen participation in government activities, and yet the Internet's full potential as an open government tool is still far from being realized. President Obama's pledge to make the government more transparent, accountable and participatory is already spurring 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.

THE ROLE OF WEB MEASUREMENT

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.² As the government pursues new open government strategies, the benefits and risks of each of these techniques will need to be examined. The set of recommendations offered here focuses 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.

² See Godwin, Bev, et al, *Social Media and the Federal Government: Perceived and Real Barriers and Potential Solutions* (Dec. 23, 2008), http://www.usa.gov/webcontent/documents/SocialMediaFed%20Govt_BarriersPotentialSolutions.pdf.

Our recommendations are generally limited to federal agencies' use of Web measurement. We hope to use this process to determine the potential for recommendations that seek to advance both open government and privacy at the same time. Should this model prove successful, we hope to expand our scope to make broader recommendations about techniques beyond Web measurement that could benefit federal agency Web sites but may also raise privacy risks.

IN THIS PAPER

Section II describes what "Web measurement" means and how it works, distinguishing "single-session" measurement, which does not require persistent user tracking, from "cross-session" measurement, which does. Section III provides background about the current federal policy with regard to the use of persistent tracking technologies on government Web sites. Section IV explains some of the motivations that federal Web managers have for using Web measurement. Section V provides our recommendations for the use of Web measurement on federal Web sites.

▣ What is Web Measurement?

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.³

This paper focuses 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

³ Web Analytics Association, *About Web Analytics Association*, <http://www.webanalyticsassociation.org/aboutus/> (last visited Apr. 15, 2009).

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.

In the abstract, Web measurement usually involves the following steps:

1. A Web site owner decides to start using a Web measurement tool. The site owner might run the tool from the site's Web domain, or the tool might be deployed from the third-party domain of the tool provider.
2. When a user visits the site, the measurement tool places a unique identifier on the user's computer. Currently this is usually accomplished using a cookie, although other mechanisms are possible.
3. As the user visits different pages of the site, the measurement tool records standard Web log information about the visits (the user's IP address, a time stamp, the URL of the page, the user's browser and operating system information, the user's cookie(s), and information about the referring page). Each of these logs would be considered individual-level data.
4. At an interval appropriate to the metrics desired by the site owner, the measurement tool aggregates and analyzes this individual-level data into reports that are delivered to the site owner. For example, if the site owner wants monthly reports about site usage, the measurement tool may aggregate and analyze data at the end of each month.

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. As discussed in Section III, federal guidance to agency Web site operators already permits single-session measurement. Thus, the majority of this paper focuses on cross-session

measurement, although we offer a number of recommendations in Section V that apply to both types of Web 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.

Regardless of the type of measurement performed, federal agencies have several choices when it comes to configuring measurement tools on their sites. An agency could purchase and deploy measurement tools itself, collecting individual-level data and performing its own data aggregation and analysis. Or, an agency can contract with a commercial third-party provider of measurement tools. In the latter case, the third party may do all of the individual-level data collection, analysis, and aggregation, providing only aggregate reports to the agency, or the agency may aid the third party with some aspects of the data collection. The recommendations made in this paper aim to address all of these different configurations.

▣ Current Federal Policy

Federal agencies' use of Web measurement tools is governed by federal guidance about persistent tracking technologies.⁴ 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.⁵ 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

⁴ Lew, Jacob J., *Memorandum for the Heads of Executive Departments and Agencies: Privacy Policies and Data Collection on Federal Web Sites* (June 22, 2000), http://www.whitehouse.gov/omb/memoranda_m00-13/.

⁵ See Lindsey, Daryl, *Drug Cookies: Why was the White House drug office monitoring your computer behavior?* (June 23, 2000), <http://archive.salon.com/news/feature/2000/06/23/ondcp/index.html>.

include any technology that can track site visitors beyond a single session.⁶ 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.

Furthermore, the time and resources required to explain persistent tracking to those who do not deal with Web technologies on a frequent basis may be more than agency officials at all levels have at their disposal. Thus, only a small number of agencies have received approval for persistent tracking of any kind, and only a handful of those approvals have been for Web measurement purposes. In Section V we offer recommendations about how the current policy could be updated to accommodate Web measurement while continuing to provide the same high levels of privacy protection as the current policy.

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.

⁶ Bolten, Joshua B., Memorandum for the Heads of Executive Departments and Agencies: OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002 (Sept. 26, 2003), <http://www.whitehouse.gov/omb/memoranda/m03-22.html>.

▣ Motivations for Change

Web measurement provides federal Web managers with extremely valuable data about the usage and success of their Web sites. Whereas commercial Web sites can use commercial measures to determine the success of their Web sites, government Web sites do not generally aim to make a profit. Thus, federal Web managers need other means to justify investment in their sites. Web measurement tools provide Web managers with the capacity to prove that their sites are achieving a certain level of viewership and participation, which in turn is vital to securing additional funds to support increased transparency and more new services on agency sites. In essence, Web managers need to be able to measure the success of their sites in order to justify additional spending on continual improvements.

Although many metrics can be determined from single-session measurement, there are a number of important metric types that can only be derived from cross-session measurement. The most basic of these is the total number of unique visitors to a site. Perhaps more than any other metric, total unique visitors allows Web managers to determine whether their audiences are growing. Cross-session tools can also be used to build granular detail into this metric, revealing how many unique users have a certain browser configuration, for example.

Cross-session measurement also allows for some qualitative analysis about Web site visitors' experiences, which helps Web managers determine which site content is effective and how users flow from page to page within a site. The average number of individual pages viewed per visitor and the average visit length per visitor are examples of metrics that provide insight into how different users experience a site, while accounting for natural variations in visitor behavior. Cross-session tools can also measure site loyalty (how often visitors come back) and recency (the average length of time between visits). All of these metrics provide detail about how users experience the site, which can be a helpful complement to simply knowing the volume of site traffic.

Cross-session tools can also help Web managers determine how site usage changes over time. For example, Web managers may want to be able to determine whether upgraded search functionality on a site helps visitors who previously spent a lot of time searching to find what they are looking for more quickly. Being able to understand, in the aggregate, how user behavior changes over time can help Web managers determine whether changes to Web site content are actually benefiting site visitors.

Thus, cross-session measurement can provide invaluable insights into key aspects of the user experience. Harnessing this data can help federal Web managers justify investment in their sites, which in turn helps to make government sites more open and participatory.

▣ Web Measurement 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.⁷

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, as noted in Section III). 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. As explained in Section II, 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. These configurations fall into three categories, each of which imparts distinct usage limitations on the parties involved. The categories are as follows:

- **Agency-only** – The agency runs the measurement tools itself, collecting individual-level data and using the tools to perform the necessary analysis and reporting. The agency serves persistent user identifiers (through cookies or other means) from its own first-party Web domain.

⁷ Although some server logs will be generated specifically for measurement purposes, others are likely generated and stored for other purposes (site improvement or anti-fraud, for example). Developing a privacy protection plan for logs that serve multiple purposes will likely require a privacy analysis similar to what is provided here in the limited context of Web measurement.

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.

Recommendation: This configuration is highly recommended. It 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.

- **Agency/partner** – The agency serves persistent user identifiers (through cookies or other means) from its own first-party Web domain, but contracts with a commercial partner to collect individual-level data and perform analysis and reporting. The agency receives only aggregate reports. In this configuration, the partner should be limited by contract to only use the collected measurement data for the purpose of delivering reports to the agency. The text of the contract should be publicly available.

Recommendation: This configuration should be used only when an agency-only configuration is infeasible. Although it involves more data collection by a commercial third party than the agency-only model, the partner's ability to use the data it collects for other purposes is limited both contractually and technologically (by browser mechanisms that will only allow persistent identifiers served by the agency to be used on the agency's site).

- **Partner-only** – The agency contracts with a commercial partner to run the measurement tools, collect individual-level data, and use the tools to perform the necessary analysis and reporting. The partner serves persistent user identifiers from a third-party Web domain. The partner thus collects individual-level data and delivers reports in the aggregate to the agency. In this configuration, the partner should be limited by contract to only use the collected measurement data for the purpose of delivering reports to the agency. The text of the contract should be publicly available.

Recommendation: This configuration should only be used when the agency-only and agency/partner models are both infeasible. This configuration provides no technical barrier against the commercial partner's use of the same persistent user identifier across multiple sites (whether they be government agencies or other sites where the third

party's measurement tools are also used).

Other cross-session measurement configurations that do not include use limitations either on the agency or its commercial partner should not be considered.

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 (i) as part of an agency Privacy Impact Assessment (PIA)⁸ published at least 30 days before the agency begins use of the Web measurement tools, and (ii) as part of the agency Web site privacy policy. These disclosures must be made whether agencies engage in single-session measurement, cross-session measurement, or both.

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.⁹

⁸ See *supra* note 5 for OMB guidance to agencies about conducting Privacy Impact Assessments.

⁹ See, e.g., the user interface of Ask.com's AskEraser, <http://about.ask.com/en/docs/about/askeraser.shtml> (last visited Apr. 15, 2009). We believe that users could be given an option as to whether they would prefer their choice status to be stored in a cookie or a small browser plug-in.

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. Agencies and their commercial partners should take the following steps in connection with limiting data retention:

- **Purpose correlation** – 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.
- **Immediate deletion** – 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.
- **Disclosure** – Data retention time frames should be published in agency privacy policies.
- **Technical enforcement** – Expiration time frames for cookies and other technologies that store data on users' computers should be set to match, not exceed, the data retention time frames adopted by agencies and their partners.
- **Partner contracts** – 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 the following conditions are met:

- i. The agency and any commercial Web measurement partners limit their use of measurement data in accordance with Implementation Recommendation 1, and operate under either the agency-only or agency/partner configuration;
- ii. The agency discloses its use of Web measurement in accordance with Implementation Recommendation 2;
- iii. The agency offers site visitors a choice about the use of their data for Web measurement in accordance with Implementation Recommendation 3;
- iv. The agency and any commercial Web measurement partners limit their data retention for Web measurement in accordance with Implementation Recommendation 4;
- v. The agency limits its use of cross-session measurement to cases where single-session measurement is not applicable, in accordance with Implementation Recommendation 5;
- vi. The agency and any commercial Web measurement partners undergo verification in accordance with Implementation Recommendation 6;

- vii. There is a compelling need to gather Web measurement data on the site; and
- viii. Appropriate and publicly disclosed privacy safeguards exist for handling Web measurement data.

This set of conditions is equivalent to the six Implementation Recommendations described above together with the first three conditions in the current federal policy on persistent tracking technologies (only agency head approval is missing). The effect is that agencies that can meet these eight conditions¹⁰ – which we believe provide a comprehensive slate of privacy protections – are not required to obtain agency head approval before deploying cross-session measurement tools. 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.

▣ Conclusion

Web measurement holds much promise for federal Web managers seeking to optimize user experiences on their Web sites. The insight that Web measurement provides could be a crucial tool for federal agencies as they seek to justify increased investments in their Web sites, which in turn could lead to increased government transparency and services on the Web.

But much has to change before federal government Web sites can take full advantage of Web measurement without harming individual user privacy. First and foremost, the providers of measurement tools must build their products to higher privacy standards than what currently exists in the commercial sector. Agencies must craft robust policies to ensure that data collected for measurement purposes is adequately safeguarded. And the federal policy on persistent tracking technologies must be adapted to continue to establish the highest levels of privacy protection while accounting for recent technological advances.

We have provided a set of open recommendations to stimulate public comment and debate about how to facilitate President Obama’s vision of a transparent, accountable, and participatory government that at the same time is protective of individual privacy. We look forward to receiving public feedback, and we are hopeful that this process may serve as a model for developing recommendations beyond the realm of Web measurement.

¹⁰ Recommendation 2 and the condition in the federal policy about disclosure overlap, yielding a total of eight conditions.

FOR MORE INFORMATION

Please contact: Brock N. Meeks, (202) 637-9800, brock@cdt.org