FUTURE OF PRIVACY FORUM
fpf.org | February 2018

Session Replay Scripts: A Guide for Privacy Professionals

Researchers at Princeton University's Center for Information Technology Policy (CITP) recently published a study demonstrating that many website operators are using third-party tools called "session replay scripts" to track visitors' individual browsing sessions, including their keystrokes and mouse movements.[1] These "session replay scripts," typically used as analytics tools for publishers to better understand how visitors are navigating their websites, were found on 482 of the 50,000 most trafficked websites, including government (.gov) and educational (.edu) websites, and websites of major retailers.[2]

Without adequate privacy safeguards, session replay scripts can raise serious privacy concerns, ranging from inadvertent collection of data to security vulnerabilities. Privacy professionals who advise publishers on whether and how to use session replay scripts should carefully evaluate terms and privacy policies, consider whether it is appropriate to implement session replay scripts on particular webpages, and ensure that strong technical safeguards are in place. Privacy and security checks should be ongoing in order to ensure policies, practices, and safeguards continue to be effective.

What are Session Replay Scripts?

A script is a "sequence of instructions carried out (or interpreted) by another program rather than by the computer processor."[3] Many website operators rely on third-party analytics scripts to learn information about how website visitors interact with their websites, including the total time users spend on a site, the sites users previously visited, and even types of browsers they use.

A session replay script is a particular type of analytics script that collects data pertaining to an individual's interactions within a website (e.g., mouse position, data typed into forms). This data can later be overlaid upon an image of the website to produce a video-like reproduction of a user's experience online for operators to view.

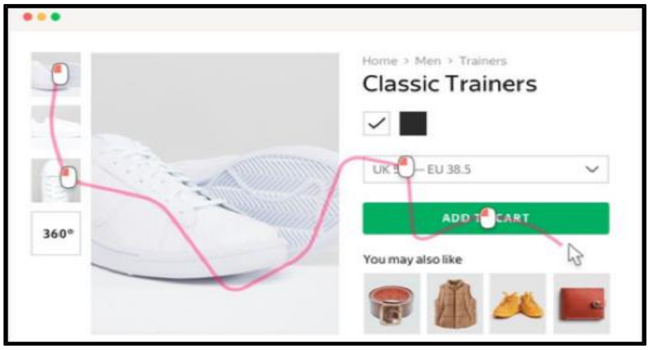


Figure 1. Session replay scripts collect precise mouse movements (source: Yandex Metrica).[4]

Session replay scripts allow website operators to understand how users navigate and interact with their sites, which in turn helps them determine what changes should be made. For example, an operator could detect user experiences that show fruitless interactions ("dead clicks") or understand why users might be giving up ("form abandonment"), and use that information to redesign their websites to increase user engagement.

Privacy and Security Pitfalls

Without adequate privacy safeguards, session replay scripts can inadvertently cause personal information to be transmitted to script providers. While providers generally offer technical measures to protect users' sensitive information, these measures are not always effective. Scripts can raise particular concerns if running on a page that prompts users to input sensitive information, such as credit card numbers, passwords, social security numbers, or health information.

Transmission of personal information to an analytics company, particularly without a user hitting "send" or "submit," can often violate a website's privacy policy. Additionally, session replay scripts can be unexpected or surprising to users because of the perceived similarities to video "recording" (which is not the case, as described above). Security is also a key issue, as data collection can increase exposure to breaches. If analytics software is running on the operator's server, for example, the script provider can still receive and read the data even if it is encrypted in transit.

Checklist for Privacy Professionals

With the right privacy safeguards, limited implementation of session replay scripts can be part of a range of ordinary, useful web analytics. In deciding whether and how to implement this kind of third-party analytics script, privacy professionals should evaluate script providers' terms and privacy policies, carefully select which sites may or may not be appropriate for their use, and continue to assess the strength of technical safeguards over time.

1. Evaluate and inquire into script providers' terms and privacy policies.

Privacy professionals should work with internal product development and marketing teams to evaluate potential vendors before implementing session replay scripts. In evaluating policies, look for provisions that show the script provider's commitment to data privacy, including:

- providing robust **automated redaction** tools (described below) to ensure that the script provider does not collect visitors' personal information;
- an option to respect users' Do Not Track requests;
- clear data retention policies; and
- descriptions of the provider's response plans for data breaches or other security or privacy incidents

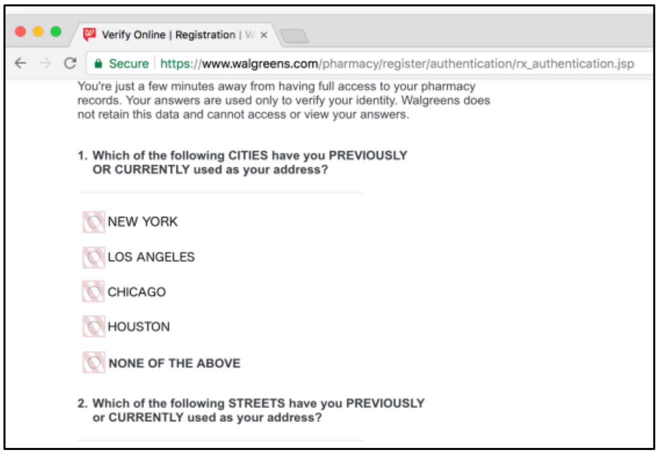
Privacy policies that place the burden solely on the publisher to protect visitors' privacy are not ideal, in part, because publishers often update their websites in ways that render technical safeguards ineffective (described below).

2. Carefully select which pages on a site are appropriate for session replay scripts.

Session replay scripts might be appropriate for some pages within a site and not others. In particular, it is best to avoid these scripts, or ensure the strength of technical safeguards, for any sites that are dynamic or **interactive**, i.e. that either *display* or *collect* individual users' information. In contrast, privacy concerns may not arise from session replay scripts running on static pages displaying non-personalized information.

On sites where users may submit information, technical safeguards are necessary to prevent users' information from being transmitted inadvertently to the script provider. These include, for example: checkout or registration points, identity verification processes, subscription sign-ups, interactive chat features, and surveys.

In addition, privacy risks can occur on sites that *display* user-specific or personal information. For example, it may be best to avoid session replay scripts within a sub-section of a user's personal account, where her name or other personal details may appear on-screen.


Verify Online | Registration | W ×
Secure
https://www.walgreens.com/pharmacy/register/authentication/rx_authentication.jsp

You're just a few minutes away from having full access to your pharmacy records. Your answers are used only to verify your identity. Walgreens does not retain this data and cannot access or view your answers.

1. Which of the following CITIES have you PREVIOUSLY OR CURRENTLY used as your address?

○ NEW YORK

○ LOS ANGELES

○ CHICAGO

○ HOUSTON

○ NONE OF THE ABOVE

2. Which of the following STREETS have you PREVIOUSLY or CURRENTLY used as your address?

Figure 2. In the example above (identified by Englehardt, et al), an identity verification page inadvertently reveals personal information.[5] *Even though the operator has redacted the values of the radio buttons, the names of the cities where the user may have lived appear and mouse movement patterns may reveal where users clicked.*

3. Technical safeguards should be on the side of *collection* rather than *use*.

Technical safeguards that prevent personal information from transmitting to the script provider (client-side) are preferable to use-based privacy protections that occur after the data is sent to the script provider (server-side).

The most common and effective technical safeguards are redaction tools, which can be automated or manual. **Automated redaction** involves the script provider automatically blocking or suppressing data from known input types (such as form fields with attributes set to "cc-number" or "password"). Although useful, automated redaction can still fail, for example if a script provider and a website operator have labeled or classified input elements in ways that do not match.

Manual redaction involves a website operator manually labeling all personally identifiable information on a page. Although manual redaction can supplement automated methods, privacy professionals should be cautious that any script implementation requiring heavy use of manual redaction may be susceptible to vulnerabilities when sites are updated. If sites are updated frequently (as most are), manual redaction can "break" or become ineffective over time.

Key potential vulnerabilities include:

- **Sensitive information** – including passwords, prescriptions, or health information. Input fields that allow users to freely input comments should also be redacted;[6]
- **Individualized content** – such as information displayed within a user's logged in account or built into an automated identity verification process;
- **Mobile displays** – should be reviewed for any key differences with web displays, such as mobile-friendly login boxes that offer to display passwords in clear-text;[7]
- **Metadata** – such as the length of an otherwise redacted password or mouse movement that reveals a user's selection from among choices.

4. Continue evaluating the effectiveness of policies, practices, and technical safeguards over time, by reviewing site implementation and report analytics.

After installing and running session replay scripts, it is important to cross-check against the analytics reports provided by the script provider to ensure that they align with expectations. Privacy professionals should look for outliers in the reporting, and may wish to implement a program of ongoing routine privacy and security checks to account for future changes to their websites.

In evaluating possible uses for session replay scripts, privacy professionals should also keep in mind that they may present a good opportunity to evaluate the effectiveness of privacy notices. Are users engaging with any privacy tools, choices, or notices that the site provides? These may be useful things to understand as part of ongoing website improvement.

[1] Steven Englehardt, *No Boundaries: Exfiltration of Personal Data by Session-replay Scripts*, Freedom to Tinker (Nov. 12, 2017), https://freedom-to-tinker.com/2017/11/15/no-boundaries-exfiltration-of-personal-data-by-session-replay-scripts/.

[2] Steven Englehardt, *No Boundaries: Exfiltration of Personal Data by Session-replay Scripts*, Freedom to Tinker (Nov. 12, 2017), https://freedom-to-tinker.com/2017/11/15/no-boundaries-exfiltration-of-personal-data-by-session-replay-scripts/.

[3] *See generally*, Susan Sharpless Smith, Web-based Instruction: A Guide for Libraries 160 (2006), goo.gl/RBZGo8.

[4] Yandex Metrica, https://metrica.yandex.com/about/info/behavior (Last visited Feb. 26, 2018); *see also* Steven Englehardt et al., *No Boundaries: Exfiltration of Personal Data by Session-replay Scripts*, Presentation to Future of Privacy Forum Location & Ad Practices Working Group (Feb. 2, 2018), https://senglehardt.com/presentations/2018-02_fpf_session_replay.pdf.

[5] Steven Englehardt, *No Boundaries: Exfiltration of Personal Data by Session-replay Scripts*, Freedom to Tinker (Nov. 12, 2017), https://freedom-to-tinker.com/2017/11/15/no-boundaries-exfiltration-of-personal-data-by-session-replay-scripts/.

[6] *See, e.g.*, *In the Matter of Practice Fusion, Inc.* (2016) (involving a company website that solicited users' feedback on their doctors, and did not anticipate users sending personal details regarding their health and prescription information), https://www.ftc.gov/enforcement/cases-proceedings/142-3039/practice-fusion-inc-matter.

[7] Steven Englehardt, *No Boundaries for Credentials: New Password Leaks to Mixpanel and Session Replay Companies*, Freedom to Tinker (Feb. 26, 2018), https://freedom-to-tinker.com/2018/02/26/no-boundaries-for-credentials-password-leaks-to-mixpanel-and-session-replay-companies/.