NATIONAL WEB PRIVACY FORUM
2018 – 2019

A NATIONAL FORUM ON WEB PRIVACY AND WEB ANALYTICS:

ACTION HANDBOOK

MONTANA STATE UNIVERSITY
LIBRARY

CONTENTS

DOCUMENT DOI

https://doi.org/10.15788/20190416.15446

PUBLICATION DATE

May 1, 2019

OVERVIEW

This is a practice-oriented action handbook that provides background, resources, and good practices to guide libraries in ethically implementing web analytics with a view towards privacy.

This guide contains two main parts, followed by a references section.

- **Part 1—Technical:** focuses on implementation strategies for privacy-aware web analytics
- **Part 2—Social**: focuses on communication strategies for building support for privacy-aware analytics practices.

INSTITUTE of Museum and Library SERVICES

MONTANA STATE UNIVERSITY LIBRARY

NATIONAL WEB PRIVACY FORUM 2018 – 2019

PART 1: TECHNICAL — IMPLEMENTATION STRATEGIES FOR PRIVACY & ANALYTICS

GOOGLE ANALYTICS IMPLEMENTATION

Many libraries have installed Google Analytics with the basic configuration. A few easy-to-implement changes can add benefits to the performance and the privacy of your website:

1. Implement forceSSL
2. Implement anonymizeIP[1]
3. Turn off Data Sharing with other Google properties[2]
4. Turn off Display Features, Remarketing, and Advertising Reporting[3]
5. Set an expiration date for Data Retention[4]
6. Ensure no Personally Identifiable Information is being sent to Google Analytics[5]
7. Enabling Opt-Out[6]
8. Consider accessing data via the Google Analytics API (https://minimalanalytics.com/)

ALTERNATIVE TOOLS

Given the discussion about responsible implementations of Google Analytics, finding additional tools that deliver metrics that are useful while respecting the privacy of our users is an essential part of user experience. The tools below strive to connect necessary data collection with an understanding of the types of concerns people might have about how much analytics data we collect.

Matomo (formerly Piwik)
Matomo is an open-source web analytics service that provides comprehensive analytics with options for cloud hosting or self-hosting.

Countly
Countly is an enterprise grade analytics and marketing platform for mobile, web, desktop and IoT applications committed to giving you control of your analytics data. 100% data ownership, 100% control.

SimpleAnalytics
Streamlined analytics from a small, privacy-conscious team that only collects the fundamental metrics: page views, referrers, top pages, and screen sizes.

Open Web Analytics
An open source web analytics software that can be installed and run according to your specifications. Fairly mature and stable releases are available and would be useful for people looking to control all aspects of their analytics program.

Server Logs (Access and Error logs)
"Out of the Box" or raw log files can provide wide-ranging analytics. In many cases, these logs provide the foundational data for analytics programs and can be mined and visualized with web analytics metrics in mind. Tools like Webalizer and GoAccess are some good places to start.

---

[1] forceSSL and anonymizeIP are described more fully in the Analytics.js Field Reference: https://developers.google.com/analytics/devguides/collection/analyticsjs/field-reference

[2] More information: https://support.google.com/analytics/answer/1011397?hl=en

[3] More information: https://developers.google.com/analytics/devguides/collection/analyticsjs/display-features#disable-advertising-features

[4] More information: https://support.google.com/analytics/answer/7667196?hl=en

[5] More information: https://www.blastam.com/blog/5-actionable-steps-gdpr-compliance-google-analytics

[6] More information: https://developers.google.com/analytics/devguides/collection/analyticsjs/user-opt-out

STAFF SKILLS AND COMPETENCIES

There are a number of ways for people to engage in learning around the topics of privacy and web analytics. Many of these discussions focus on the practices an individual can take to safeguard their data or make themselves invisible. We are approaching this conversation with a more collective, organizational viewpoint. References in this section are outlining broad privacy certifications and requirements in the context of skills or certifications that might be brought into your organization to aid in responsible analytics practices.

Core Privacy Concepts

- Information security management and governance, including frameworks, controls, cryptography and identity, and access management (IAM).

Understanding privacy vulnerabilities

- Network and remote server vulnerabilities, web browser vulnerabilities, data breach liabilities for both university-owned and vendor-built systems

Auditing data

- Developing information lifecycle plans, data identification and classification systems, data flow diagrams, data retention and deletion

Preparing data

- De-identification and pseudonymization of Personally Identifiable Information

Multiple certifications and learning opportunities exist for building up these privacy skills. The International Association of Privacy Professionals runs a number of training opportunities related to general privacy and best practices for data stewardship with an eye toward protecting users is a good place to start. If you are interested in some more theoretical reading on how to process and visualize raw information with a deep view of the biases and implications of this kind of work, check out *Thinking, Fast and Slow* by Daniel Kahneman.

PRIVACY POLICIES

A privacy policy helps create a bridge of communication and trust between a website and its user community[7]. For libraries--traditional places of trust--the privacy policy is an especially important instrument for transparency and trust-building[8]. A positive user experience can result from a straightforward, aesthetically-pleasing policy that speaks directly to the user in everyday terms[9]. A library might also consider implementing a cookie notice that (1) alerts users to the privacy policy and (2) provides an opt-out mechanism for analytics tracking.

[7] Bansal, Gaurav, Fatemeh 'Mariam' Zahedi, and David Gefen. 2015. "The Role of Privacy Assurance Mechanisms in Building Trust and the Moderating Role of Privacy Concern." European Journal of Information Systems 24 (6): 624–44. https://doi.org/10.1057/ejis.2014.41.

[8] Arcand, Manon, Anne Vincent, Jacques Nantel, and Mathieu Arles-Dufour. 2007. "The Impact of Reading a Web Site's Privacy Statement on Perceived Control over Privacy and Perceived Trust." Online Information Review 31 (5): 661–81. https://doi.org/10.1108/14684520710832342.

[9] Waldman, Ari Ezra. 2018. "A Statistical Analysis of Privacy Policy Design." Notre Dame Law Review 93. http://ndlawreview.org/2018/04/a-statistical-analysis-of-privacy-policy-design/.

Five-Point Plan for Privacy-Aware Analytics

Ultimately, data is useful for evidenced-based decision making. For libraries, we can provide value-added contributions by helping our organizations implement analytics with more thoughtfulness and critical-awareness of privacy and users. With this in mind, we offer the following set of indicators to help refine analytics implementation and privacy practices. These indicators can be viewed as a checklist that people can use to assess analytics programs for alignment with our privacy-conscious lens.

Privacy Indicator 1: Collect only the data needed for your use case.

Privacy Indicator 2: Support analytics tools that allow retention and downloading of your own data in open formats.

Privacy Indicator 3: Support analytics tools that allow the setting of a data retention strategy and enable the complete removal of data.

Privacy Indicator 4: Implement analytics tools that allow for deidentification and/or pseudonymization, and the removal of personally identifiable information (PII).

Privacy Indicator 5: Implement analytics tools that have support for emerging international privacy standards (e.g., General Data Protection Regulation).

PART 2: SOCIAL — COMMUNICATION STRATEGIES FOR PRIVACY AND ANALYTICS

In addition to the technical aspects of configuring web analytics, developing robust privacy policies, and the like, taking action to further privacy requires engaging in dialogue with a variety of stakeholders. This section presents some strategies for handing some of the questions and (perhaps) objections that may emerge.

CHOOSE YOUR MINDSET

For those who are trained as librarians, prioritizing privacy is a part of the mindset of our field. It is a plank in our code of ethics. We have historically taken specific (e.g., deleting circulation records) and public (e.g., response to the PATRIOT Act) positions to advocate for and protect privacy. It can be frustrating to encounter those who have a different system of ethics or who are unaware of the concerns that seem obvious to librarians and other privacy-oriented professions.

Recognizing that we do not all have the same background and training is a useful strategy for choosing the mindset with which we approach our conversations. Assume that others do have an ethical framework from which they approach their work and that if they have agreed to the conversation they are seeking to do so responsibly.

LISTEN CAREFULLY

Building off the mindset of assumed good intentions, be interested in learning about the ways others see issues related to privacy and what principles underscore their decision-making and priorities. If they raise objections to your views, be certain to take time to understand their specific objections and not assume you know what they mean. Before responding, be clear on what you are responding to. To the best of your ability, be as generous as possible in assessing the question or objection. It can be difficult, admittedly, to listen if you are having an intense emotional reaction so try for meta-awareness of not only your thoughts but also your affective state and how they are impacting your participation in a dialogue.

BE PREPARED

When it is your turn to speak or share, it is perhaps obvious that it is useful to be prepared. Think in advance about some of the issues that are likely to be raised and how you might respond.

For example, one might hear "People don't care about privacy today—why should we?" Though library ethics are part of the answer, it may be useful to acknowledge that people do regularly agree to give away data about themselves. This is a fact. But, what's more important is the interpretation of this reality. It isn't that people don't care about privacy. Instead, research shows that they do care about privacy but are resigned currently to the reality that their privacy is being taken from them (see for example, ***Sharing Data for Deals? More Like Watching It Go With a Sigh*** - https://www.nytimes.com/2018/12/24/business/media/data-sharing-deals-privacy.html).

Someone also might ask "How can we gather the data that we need for strategic decision-making and service development (e.g., personalization) if we focus on privacy?" Protecting privacy does not have to mean that no data is collected but rather that data collection is minimized, well-managed, and informed by best practices in user research. Focusing on transparency and consent will improve privacy even though data is still being collected. Lisa Janicke Hinchliffe has provided guidance on this topic in P***rivacy in User Research - Can You?*** (https://scholarlykitchen.sspnet.org/2018/09/05/privacy-in-user-research-can-you/).

FIND COMMON GROUND WITH A LONG-VIEW

Protecting privacy is critical for institutional health. Doing something to better protect privacy is better than doing nothing and leaving the status quo. Look for things that all parties can agree upon to move forward towards privacy protection. Opt-in is an improvement on opt-out; opt-out is an improvement on no choice at all. Look for areas of agreement and move forward with those. Any given conversation is likely not the last conversation but rather a stage in a discussion over time.

REFERENCES

RELATED ORGANIZATIONS

- Data Justice Lab
- Open Privacy Research Lab
- Library Freedom Project
- International Association of Privacy Professionals

PRIVACY

- Blum, Dan. 2014. "Privacy by Design and the Online Library Environment." Information Standards Quarterly 26(3). https://www.niso.org/sites/default/files/stories/2017-08/FE_Blum_Privacy_by_Design_isqv26no3.pdf
- Cavoukian, Ann. 2011. "Privacy by Design: The 7 Foundational Principles." https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf
- Igo, Sarah E. 2018. *The Known Citizen: A History of Privacy in Modern America*. Cambridge, Massachusetts London, England: Harvard University Press.
- Lee, Una, and Dann Toliver. 2017. "Building Consentful Tech." http://www.lib.montana.edu/privacy-forum/Building-Consentful-Tech-Zine.pdf.
- Selinger, Evan. 2018. "Stop Saying Privacy Is Dead." Medium (blog). October 11, 2018. https://medium.com/s/story/stop-saying-privacy-is-dead-513dda573071.
- Wittkower, Dylan Eric. 2016. "Lurkers, Creepers, and Virtuous Interactivity: From Property Rights to Consent and Care as a Conceptual Basis for Privacy Concerns and Information Ethics." First Monday 21 (10). http://firstmonday.org/ojs/index.php/fm/article/view/6948.

LIBRARY VALUES AND ETHICAL PRACTICES

- Asher, Andrew, Kristin Briney, Gabriel J. Gardner, Lisa Janicke Hinchliffe, Bethany Nowviskie, Dorothea Salo, and Yasmeen Shorish. 2018. "Ethics in Research Use of Library Patron Data: Glossary and Explainer." https://osf.io/xfkz6/.
- Molls, Emma. 2018. "Of Vendors & Values." https://doi.org/10.6084/m9.figshare.7416278.v1.
- Newman, Bobbi, and Bonnie Tijerina, eds. 2017. Protecting Patron Privacy: A LITA Guide. Lanham, Maryland: Rowman & Littlefield Publishers. https://rowman.com/isbn/9781442269699/protecting-patron-privacy-a-lita-guide.
- Shorish, Yasmeen. 2018. "Valuing Privacy in a Surveillance Society." presented at the Data, Libraries and Justice ENY/ACRL Conference, Colgate University. https://osf.io/hwbfa/.
- Yoose, Becky. 2017. "Balancing Privacy and Strategic Planning Needs: A Case Study in De-Identification of Patron Data." Journal of Intellectual Freedom & Privacy 2 (1): 15–22. https://doi.org/10.5860/jifp.v2i1.6250.

CRITICAL DATA COLLECTION AND SURVEILLANCE

- Browne, Simone. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham: Duke University Press Books.
- Gilliard, Chris. 2018. "Friction-Free Racism." Real Life, October 15, 2018. https://reallifemag.com/friction-free-racism/.
- Kahneman, Daniel. Thinking, Fast and Slow. 2015. Print. http://www.worldcat.org/oclc/917473664
- "Managing Google Analytics with an Eye on Library Privacy." Stanford University Libraries. 2019.
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: PublicAffairs.

AUTHORSHIP AND ACKNOWLEDGEMENTS

This Action Handbook is co-authored by Scott W. H. Young, Jason Clark, Sara Mannheimer, and Lisa Janicke Hinchliffe. This publication is a part of *A National Forum on Web Privacy and Web Analytics*, and is generously supported by the Institute of Museum and Library Services.

MONTANA STATE UNIVERSITY LIBRARY

INSTITUTE of Museum and Library SERVICES

NATIONAL WEB PRIVACY FORUM 2018 – 2019