Protecting Privacy on the Web

A Study of HTTPS and Google Analytics Implementation in Academic Library Websites

Kenning Arlitsch, Dean of the Library
Scott W.H. Young, UX & Assessment Librarian

CNI Fall Membership Meeting
December 11, 2018
Presentation Outline

- THIRD-PARTY TRACKING
- WEB ANALYTICS
- LIBRARY PRIVACY PRINCIPLES
- NEW RESEARCH AND RESULTS
- RECOMMENDATIONS FOR PRACTICE

DISCUSSION PROMPTS
Third-Party Tracking
Third-Party Tracking

- Web analytics services track user search/browse behavior
- Information is sometimes passed to other vendors
- Tracking is often performed without fully informed user consent
The practice of third-party tracking on websites is widespread, and has only increased in prevalence, variety, and complexity over time.

Libraries are increasingly partnering with third-party vendors who deploy analytics and tracking software that may not align with our values and commitments to privacy.
Mitigation Techniques

- Use a different service—consider moving away from Google Analytics
- IP anonymization
- Opt-out mechanisms
- Secure HTTP connections (HTTPS)
  - Encrypts communication
  - Assures users that site is the one they’ve requested
  - Websites should automatically redirect HTTP requests to HTTPS
- Google has penalized non-HTTPS sites since 2014
Web Analytics
Benefits
• Free to use ($)
• Easy to install
• Extensive/sophisticated
• Helpful graphs
• Can provide useful insights

Costs
• Pass user data to Google
• May not align with our values
• Inaccuracies
• Can produce dubious insights
Libraries collect lots of data through third-party services like Google Analytics, but we lack an understanding of the technology and an appreciation for its privacy costs.
Library Profession Privacy Principles
“Libraries, publishers and software providers have a shared obligation to foster a digital environment that respects library users’ privacy as they search, discover and use those resources and services.”

-NISO Privacy Principles
“Library and information services should reject electronic surveillance and any type of illegitimate monitoring or collection of users’ personal data or information behavior...”

- Privacy Statement
“...the right to privacy is the right to open inquiry without having the subject of one’s interest examined or scrutinized by others.”

-Library Bill of Rights
Coalition for Networked Information

“Libraries collecting data using Google Analytics are realizing they may be violating the ALA Library Bill of Rights...this is but one example of how easily convenient web-based service offerings can come with unexpected consequences.”

-Privacy in the Age of Analytics
97% of librarians surveyed agree or strongly agree that libraries should never share personal information and circulation or a court order.

• Privacy has long been a concern of libraries.

• Given the extent of third-party tracking, it is exceedingly difficult to implement analytics trackers like Google Analytics without compromising the privacy for users that libraries have championed.
Showing our value  ↔  Living our values
New Research and Results

https://doi.org/10.1108/OIR-02-2018-0056
RQ1: Do libraries implement HTTPS, with proper redirect practices?

RQ2: Do libraries that use Google Analytics implement the available privacy protection measures?
Research Question 1

Do libraries implement HTTPS, with proper redirect practices?

- Do they protect privacy with a secure connection (via HTTPS) between the user’s browser and the library’s website?
- Use a permanent redirect to enforce the use of secure connections?
- Redirect secure page requests to a non-secure version of the page in violation of recommended practice?
Research Question 2

Do libraries that use Google Analytics implement the available privacy protection measures?

- Does the library use Google Analytics?
- Does it protect user privacy via a secure connection between the library website and Google’s servers?
- Does it obfuscate individual user tracking using Google’s IP Anonymization feature?
Research Methodology and Method

• Webometrics methodology
  • Subset of Informetrics
  • Initially focused on statistical analyses of word and phrase frequencies
  • Modern definitions include “quantitative aspects of the construction and use of information resources, structures and technologies on the Web”
    • (Bjorneborn and Ingwersen, 2004).

• Covert Observation Research method
  • Observes behavior without revealing identity or presence of the researcher.
  • Our observed subjects are information structures on publicly hosted machines:
    • Presence or lack of HTTPS and Google Tracking Code
Study Population

279 US and international academic libraries (16 countries)

Membership in one or more of following:
  • Association of Research Libraries
  • OCLC Research Library Partnership
  • Digital Library Federation
HTTPS Implementation for Academic Library Home Pages (n=279)

- 173 (62%) HTTPS Implemented
- 106 (38%) No HTTPS
Redirect implementation for academic library websites with HTTPS (n=173)

- Redirect to HTTPS
- No Redirect to HTTPS

- Redirect to HTTPS: 56 (32%)
- No Redirect to HTTPS: 117 (68%)
Implementation of Google Analytics and/or Google Tag Manager on Academic Library Homepages (n=279)

- Google Analytics and/or Google Tag Manager Implemented
  - 34 (12%)
- Neither Google Analytics or Google Tag Manager Implemented
  - 245 (88%)
Privacy Protection Features for Academic Library Homepages with Google Analytics and/or Google Tag Manager (n=245)

- No Google Privacy Protection: 3 (1%)
- Library-to-Google HTTPS: 34 (14%)
- Google IP Anonymization: 208 (85%)
Recommendations for Practice

- **Implement HTTPS**: Ensure redirection, even when request is for insecure (HTTP) site.
- **Implement IP Anonymization for Google Analytics**: Capability is built in to GA, but rarely turned on.
- **Provide user education related to online privacy**.
- **Obtain informed consent from users**.
- **Conduct risk/benefit analysis when using third-party service providers**.
Research funded by IMLS

Grant proposal:
“Measuring Up: Assessing Accuracy of Reported Use and Impact of Digital Repositories” - scholarworks.montana.edu/xmlui/handle/1/8924

Publication:
https://doi.org/10.1108/OIR-02-2018-0056
Discussion

- Why have so few libraries implemented available privacy tools?
  - Awareness?
  - Lack of technical expertise?
  - Other barriers?
- Which methods for achieving privacy are the most promising or effective?
- What are our knowledge gaps?
- What are our most significant pressures and challenges in this area?
- How can we reshape prevailing norms that encourage privacy incursions?