

USER TRACKING ON ACADEMIC PUBLISHER PLATFORMS

CNI Spring 2019 Membership Meeting

Cody Hanson (@codyh)

University of Minnesota Libraries

codyhanson@umn.edu

Thanks

Michael Berkowski, Sunshine Carter,
Angela Galvan, Jason Griffey, Eric Hellmann,
Lisa Hinchliffe, David Lacy, Matthew Reidsma,
Dorothea Salo, Franklin Sayre, Nancy Sims,
Elizabeth Temple

My thanks do not imply
their endorsement.

<https://www.codyh.com/writing/tracking.html>

Key findings

- The articles most frequently accessed by UMN patrons include code on their publisher pages that is designed to identify users and link their identity to the pages they visit.
- These tools derive user identity in part through metadata that is not currently governed by our typical definition of Personally Identifiable Information.
- I do not believe that it is currently possible to ensure that the use of electronic library resources can be private.

BACKGROUND

Dec. 2018 CNI Fall Membership Meeting

- [Protecting Privacy on the Web: A Study of HTTPS and Google Analytics Implementation in Academic Library Websites](#) – Kenning Arlitsch and Scott W. H. Young of Montana State University
- [Evaluating and Closing Privacy Gaps for Online Library Services](#) – Micah Altman of MIT, Lisa Janicke Hinchliffe of the University of Illinois, and Katie Zimmerman of MIT
- [RA21: Resource Access for the 21st Century – Pilot Results and New Recommended Practices](#) - Todd Carpenter of the National Information Standards Organization (NISO), Jean Shipman of Elsevier, Ralph Youngen of the American Chemical Society.

PUBLISHERS DON'T
NEED RA21 TO
IDENTIFY USERS.

THE STUDY

January and February, 2019
University of Minnesota Libraries Resources

Can an analysis of the source code of publisher platform pages provide evidence of if and how publishers can identify library users?

Can an analysis of the source code of publisher platform pages provide evidence of if and how publishers can identify library users?

Yes.

100 most-frequently accessed articles at UMN

- The 100 most frequently-accessed DOIs through the UMN Libraries' proxy server
- Recorded mid-2016 through 2018
- 15 different publisher platforms represented

15 platforms

- One article from each platform
- Resolved DOI through doi.org from an on-campus IP
- Captured a complete archive of the page, including first- and third-party assets and code
- Read source code, to the best of my ability
- Analyzed the live page with Ghostery

UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

M LIBRARIES

Search for books,

Hours & locations

 **Walter Library**
Today: 8:00 a.m.–5:00 p.m.
📍 East Bank

 **Wilson Library**
Today: 7:00 a.m.–Midnight

GHOSTERY

Simple View Detailed View  

1

www.lib.umn.edu

Trackers Blocked: 0 Page Load: 0.65 secs

Trust Site

Restrict Site

Pause Ghostery 

  

Block All

Collapse All

 **Site Analytics**
1 TRACKER

 Google Analytics

List View

Third-party code

- On average, each platform had 18 third-party assets loaded on their article page
- Median was 10
- One had none
- One had over 100
- A total of 139 distinct third-parties across the 15 platforms

Significance of third-party code

JavaScript can access:

- Page address
- Page contents
- User actions on the page
- Browser info
- User IP address
- Contents of existing browser cookies set against the same domain

JavaScript can also load additional JavaScript from other sources.

Significance of third-party code

By loading third-party JavaScript, publisher platforms are sharing the content of user research inquiries with third parties, along with information that can be used to specifically identify the user.

Facebook

Four of fifteen publisher platform pages included Facebook code.

On sites with Facebook code, we can assume that the identity of users with a Facebook cookie in their browser is being combined with information about the publisher page they're visiting and stored by Facebook.

On sites with Facebook code, we can assume that the fragmentary identity metadata of users **without** a Facebook cookie in their browser is being combined with information about the publisher page they're visiting and stored by Facebook.

Google

Fourteen of fifteen publisher platform pages include Google code.

On sites with Google code, we can assume that the identity of users with a Google cookie in their browser is being combined with information about the publisher page they're visiting and stored by Google.

I assume the same holds true for users without a Google cookie.

Browser fingerprinting

A technique to generate a unique identifier for a user about whom little else is known using web browser metadata, e.g.:

- User Agent - identifies browser, OS, and versions thereof
- Screen Size and Color Depth - identifies the specific monitor hardware
- System Fonts - the list of fonts available to the browser
- Browser Plugins – installed extensions, such as ad blockers
- Are Cookies Enabled?
- Do Not Track Enabled?

Information drawn in large part from
EFF's [Panopticlick](#) tool.

Within our dataset of several hundred thousand visitors tested in the past 45 days, only **one in 103433.5 browsers have the same fingerprint as yours.**

AUDIENCE TOOLS

How browser metadata becomes PII

[Identity](#)[Solutions](#) ▾[Industries](#)[Clients](#)[Partners](#)[Resources](#) ▾[Blog](#) ▾[Let's Connect](#)[Resources & Tools](#)

Target the Right Customers Using the Right Data

The more you know about your customers the more impactful your marketing can be. And in today's connected world where consumers move rapidly across devices and touchpoints, having the wrong data will not only produce a bad experience, it can put you and your organization at risk. It's time to stop guessing and start knowing with accurate and verified customer identity data.

Neustar's Customer Identity File gives you access to the most accurate and complete name, address and phone data available. Over 150 million USPS Delivery Point Validation (DPV®) households - compiled, verified, and enhanced with 450+ fields of demographic, behavioral, financial, property, segmentation and geographic attributes. Whether you are sitting on mountains of customer data or looking to build your own repository from scratch, Neustar gives you the power of true consumer intelligence.

Benefits

UNCOVER GREATER INSIGHTS

Obtain a complete consumer profile of all U.S. households with their names, addresses, phone numbers, demographics, and more! Gain a more complete understanding of your existing customers to deliver a better experience for retention.

Neustar

At least four of the fifteen platform pages included Neustar code.

“Identity is not static. It is dynamic. Only Neustar’s OneID system has holistic identity resolution, corroborated as often as every fifteen minutes, with eleven billion daily updates from multiple sources.”

Trait Storage



Audience Traits

CRM

DMA

Demographic



Gender

Platform

Products

Purchases

Social

Subscription

Campaigns

Channels

Company - LF

Customer Onsite

Option A

Trait ID

Name

Description

74624

Age 20 - 29

Age range trait

74625

Age 30 - 39

Age range trait

74627

Age 40 - 49

Age range trait

4056508

Age 50+

Age range trait

74621

Income - \$100,000 to 124,999

Income range trait

74622

Income - \$125,000 to 149,999

Income range trait

74623

Income - \$150,000+

Income range trait

Adobe Audience Manager

At least six of the fifteen publisher platforms included Adobe Audience Manager code.

Can “...turn fragmented data, from any channel or device, into meaningful audiences that you can act on right away.”

Can be used to “Deliver offers only to users when they are logged in, or based on previous log in activity.”

Can be enriched with data from Acxiom, which boasts “Comprehensive consumer data on approximately 250 million U.S. addressable consumers...”

Oracle Marketing Cloud

At least four of fifteen publisher platform pages included Oracle Marketing Cloud code.

Advertises its ability to “connect with an individual customer across all channels and devices” using the Oracle ID Graph.

The Oracle ID Graph “ingests massive amounts of IDs across cookies, login, HH [household], email, and mobile ad IDs...The Oracle ID Graph can reach over 90% of people online in the US and in markets that matter internationally...”

AddThis

Legal Policies

[Terms of Service](#)

[Comment Policy](#)

[Mobile SDK License Agreement](#)

[Extension and Bookmarklet](#)

[Terms of Service](#)

[Pixel Partners](#)

[Partners](#)

Cookie & Pixel Partners

When a website visitor visits a Publisher Site, Oracle and Oracle Partners to set cookies and fire pixels to collect AddThis Data to enable the synchronization of internal unique identifiers between AddThis and our third party partners and to facilitate online behavioral advertising as described in detail in the [Privacy Policy](#).

The “Opt-Out” button available at the left of this page will opt the website visitor out of data collection by these third party partners. Please see the “[What are your choices?](#)” section in the Privacy Policy for more information.

AddThis

Eleven of fifteen publisher platform pages included AddThis code.

AddThis gathers and shares user activity and identity information with a network of over 40 ad networks and data brokers. Among these partners:

- Neustar
- Adobe
- Oracle
- Google

I have now mentioned only six of the 139 different third-parties with assets on these 15 publisher platforms.

1plusX, Adometry, Adbrain, AdGear, **Adobe**, Affectv, AOL One, AppNexus, Branch, Cardlytics, Centro, Collective, DataXu, Drawbridge, Dstillery, Ebay, eXelate, Exponential, **Google**, GroupM/Xaxis, LiveRamp, LocalResponse, Lotame, MediaHead, MediaMath, Modern Advertising, Nano Interactive, **Neustar**, Oracle (BlueKai, Datalogix), RadiumOne, Resonate, RocketFuel, Samsung, Semasio, Tapad, The Trade Desk, TubeMogul, Turn, Videology, VisualDNA, Vizury, Weborama, X+1, Yahoo-DataX

1plusX, Adometry, Adbrain, AdGear, **Adobe**, Affectv, AOL One, AppNexus, Branch, Cardlytics, Centro, Collective, DataXu, Drawbridge, Dstillery, **Ebay**, eXelate, Exponential, **Google**, GroupM/Xaxis, LiveRamp, LocalResponse, Lotame, MediaHead, MediaMath, Modern Advertising, Nano Interactive, **Neustar**, Oracle (BlueKai, Datalogix), RadiumOne, Resonate, RocketFuel, **Samsung**, Semasio, Tapad, The Trade Desk, TubeMogul, Turn, Videology, VisualDNA, Vizury, Weborama, X+1, Yahoo-DataX

I DO NOT BELIEVE IT IS
POSSIBLE FOR USE OF
LICENSED RESOURCES
TO BE PRIVATE.

ALA Patron Bill of Rights Article VII

"All people, regardless of origin, age, background, or views, possess a right to privacy and confidentiality in their library use. Libraries should advocate for, educate about, and protect people's privacy, safeguarding all library use data, including personally identifiable information."

<http://www.ala.org/advocacy/intfreedom/librarybill>

Stanford, et. al

“As research libraries, we do not sell patron data. We do not share it. We object to, and reject, subscription agreements that silently expose it to third-party interests, whether they be commercial or governmental.”

<https://library.stanford.edu/using/special-policies/statement-patron-privacy-and-database-access>

Model license language

- [Evaluating and Closing Privacy Gaps for Online Library Services](#) – Micah Altman of MIT, Lisa Janicke Hinchliffe of the University of Illinois, and Katie Zimmerman of MIT

<https://www.codyh.com/writing/tracking.html>

@codyh
codyhanson@umn.edu