5-8-2015

Who Does the Internet Think You Are? Three Tools That Teach Students How They Are Actively Profiled Online, All the Time

Robin Camille Davis
CUNY John Jay College

Who does the Internet think you are?

Three tools that teach students how they are actively profiled online, all the time

Robin Camille Davis
Emerging Technologies & Distance Services Librarian
John Jay College of Criminal Justice
Download slides (PDF, 3MB)

Presentation for the LACUNY Institute at John Jay College of Criminal Justice on May 8, 2015. The theme of the Institute was "Privacy and Surveillance: Library Advocacy for the 21st Century."

Presentation script

Intro

I've been incorporating a small amount of critical digital literacy* into library class sessions I lead, and I've recently begun teaching a workshop called "See through the internet" that touches on digital literacy and privacy topics. In this presentation, I'm going to focus the question, *Who does the internet think you are?* This is aimed at librarians who want simple tools to demonstrate to students and to each other one facet of the internet that is often invisible.

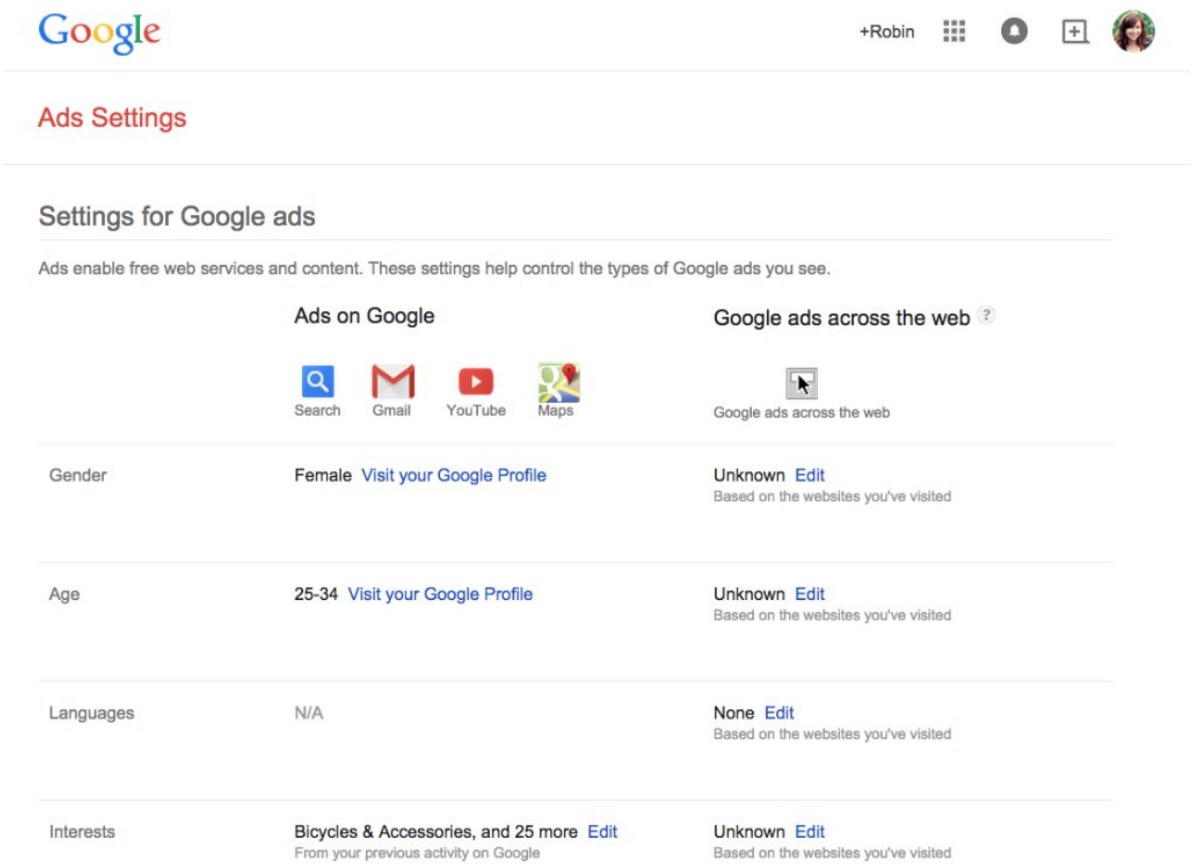
**More about critical digital literacy in libraries: CRTCLDGTL (reading group) and #critlib (live discussion every Tuesday night)*

Your internet experience is tuned to you, but not by you. The internet has a fragmented but specific idea about you. When you use your own computer or device, or log into your accounts — websites, ads, search results, and sometimes even product prices are tailored to you specifically — but how? A vast collection of data describes you to a number of organizations who use this information to shape the internet you experience. It shapes why you see different google results from your friend, it dictates why you see certain ads and not others, and it may even dictate the prices of products or services you buy online. This data about you is not easy for you to see, but we have some tools to bring this activity into the foreground.

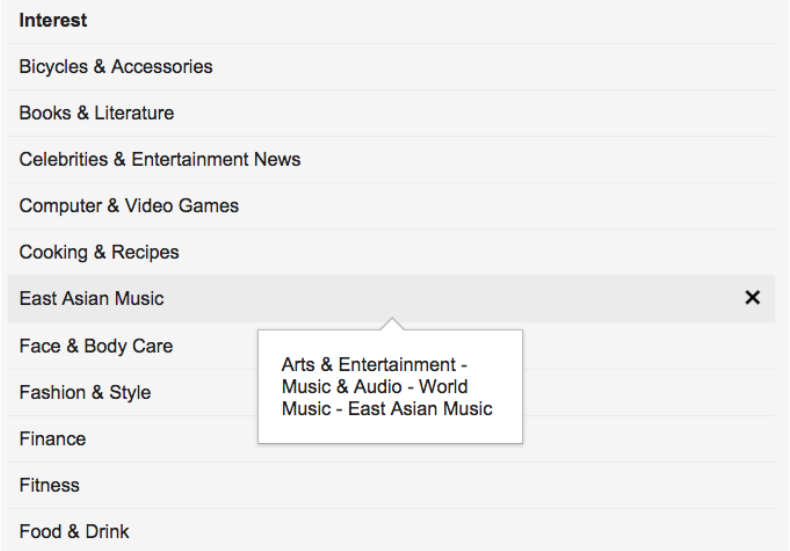
We'll be looking at 3 tools today to help you and your students understand more about what information is tracked and how. We'll focus mainly on how advertising companies glean information about you, since ads take up a lot of the visual real estate of the web and are often a major part of websites' revenue. We'll start with the simplest tool on the list.

Google ad settings: see how a major advertiser profiles you

Google, as we know, is known mainly for search, but it is also one of the world's biggest advertisers, and in a tepid spirit of transparency, Google makes its generalized ad profile about you available for you to see. Here's mine.



This is from my own profile, when I'm logged into Google. So what do we see? Because I once enthusiastically set up a Google+ profile, Google knows my gender and age pretty well because I volunteered that information. When demonstrating this page, what's most fun to look at is what Google thinks I am interested in. Bicycles and accessories? Yup, that's my jam! Let's take a look at Interests (by clicking Edit) to see what else I'm into.



Books & Literature, sure, Finance, I guess, Fitness… These are pretty broad interests, but they can get weirdly specific in some cases, like East Asian Music. Google lists interests on my ad profile based on what I search for, what search results I click on, keywords in my email, and what sites I visit, if those sites use Google's ad service. For instance, because I really like Javanese gamelan music, I've watched a lot of videos of gamelan performances. So that's why Google thinks I'm interested in East Asian Music. And based on that, some ads will show up for me, and some won't. Google's system might extrapolate that a person who likes East Asian Music would be more inclined to click on an ad for yoga than for zumba when she searches for exercise classes, for instance.

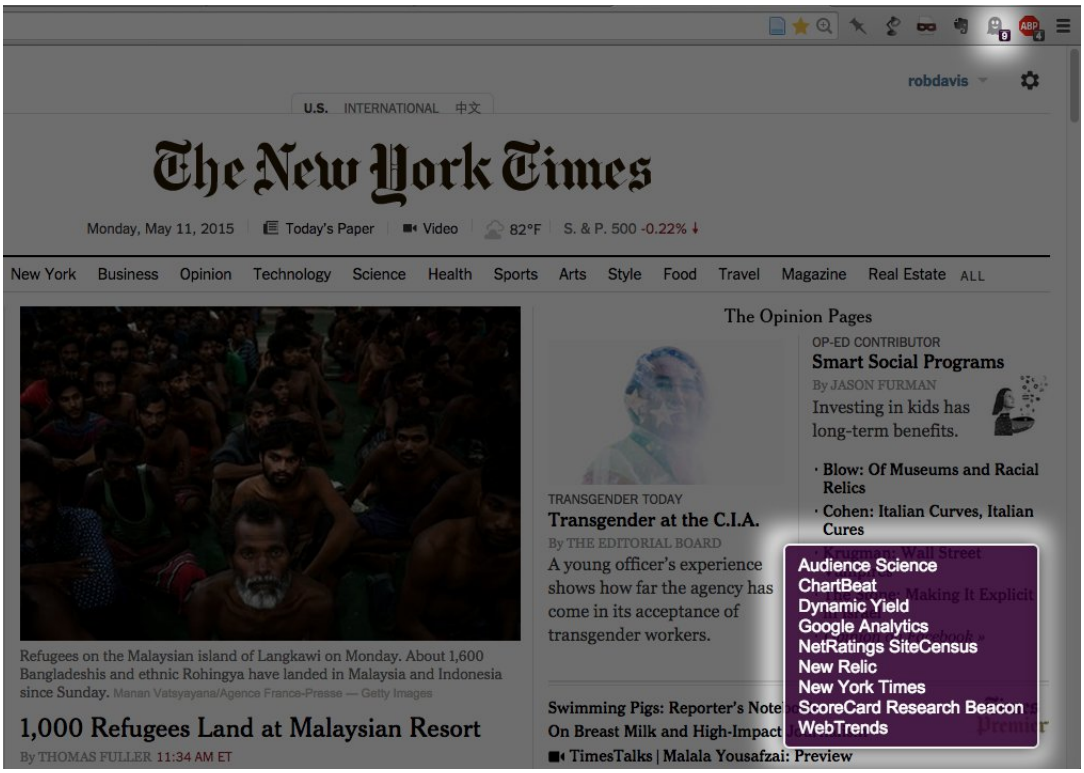
If you're demoing with students, have someone else log into their Google account and compare. Why is this interesting? This is a simplified and human readable example of what information advertisers can infer about you, even when you aren't consciously volunteering information. To see the ad profile at work, search Google for various keywords and compare your screens to each other to see if the ads differ.

Ghostery & Privacy Badger: use plugins to see what's being tracked

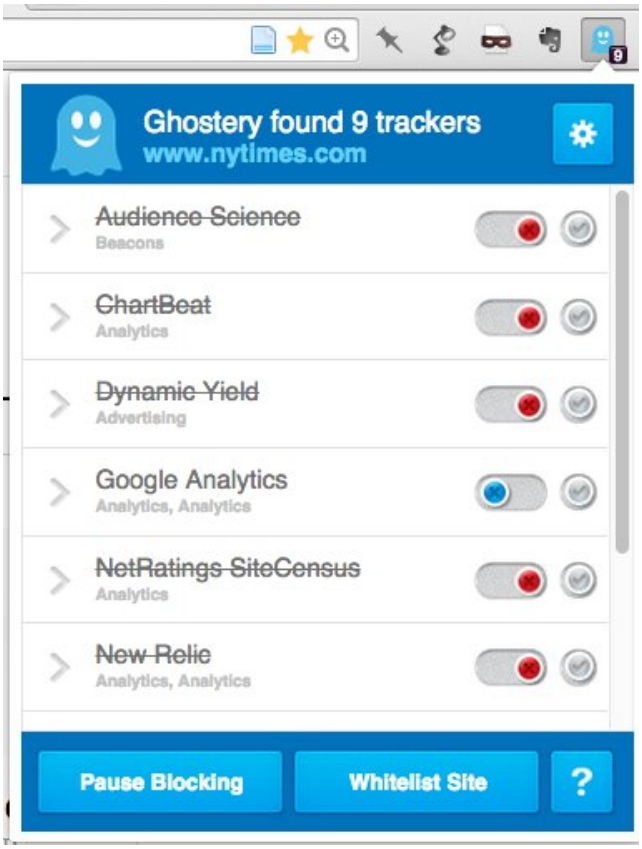
So, now we've gotten a small glimpse at a few kinds of information that an advertiser is interested in keeping about you. This information used by targeted advertisers can paint a very accurate picture about your buying habits, browsing history, and other information that may indicate whether you'd click on Ad A or Ad B.

So, how can targeted advertising do this? By tracking what you see on the web using **cookies**. To me, cookies are delicious and never nefarious, so I think it's a mistake that someone named these tiny tracking devices "cookies." So I'm calling them trackers instead, to drive my point home.

It's hard or even impossible to tell *who's* tracking you,*where* these trackers are embedded, and *what*information they're collecting. To lift the veil, I use a tool called Ghostery, simply as a way to see the trackers on a given website. This is a browser plugin for all major browsers.



You'll see that in Chrome, a little ghost icon with a number will tell me how many trackers are on each web page I visit. For the *New York Times* website at this time, that's 9 (but the number can change after refreshing the page, based on which ads are displayed). In addition, a little window will pop up at the bottom to list out the names of the companies whose trackers are on the page. Somehow, to me, seeing the names makes it feel more real that there are little tiny trackers all over the web.

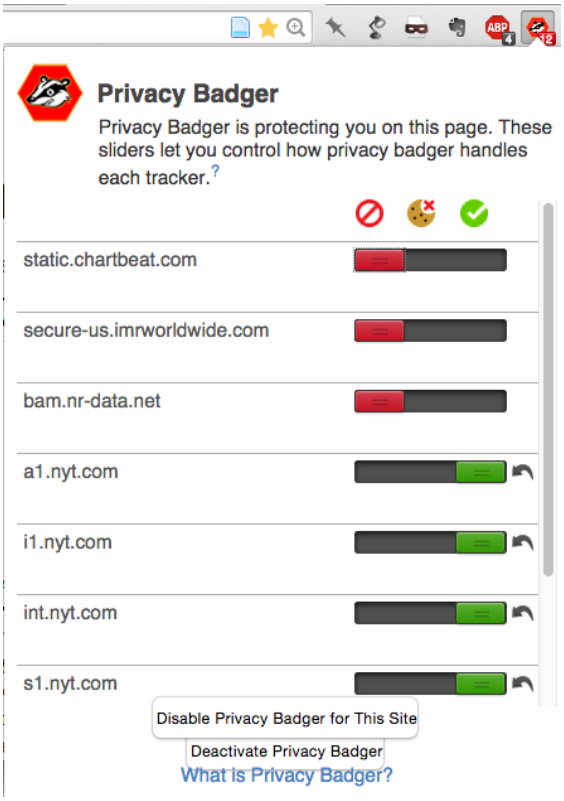


Clicking on the little ghost icon will open up a window with more information. Not only can I see the names of the trackers are on the page, but which category Ghostery believes they fall under. And you can see it's not only advertising — trackers are useful for analytics, aka page visit stats, which we probably all use. They're also useful for social media, like those "Tweet this" buttons.

The last thing to point out are those red and blue toggle switches. They allow you to disable the trackers entirely, which is a powerful thing. Not only can you finally see which companies are tracking you and why, but you are now empowered to block them, one by one.

For simply displaying trackers, their categories, and the option to block them, Ghostery is the most user-friendly tool I've used, and I continue to use it and demonstrate it regularly. But a heads up: Ghosterymakes money in several ways, one of which is by asking users to opt in to "donate" their data. Ghostery uses this data to help advertising companies figure out how not to get blocked. As far as I know, this is opt-in, not opt-out.

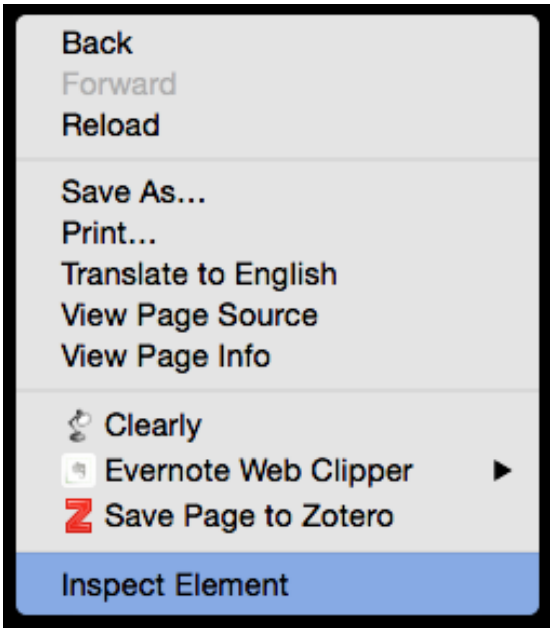
But if you're looking for a privacy tool that maybe doesn't deal at all with advertisers, one that's made by the internet good guys, you might want to look at Privacy Badger, made by the EFF.



Privacy Badger has many of the same features as Ghostery, but with the added feature of logging not only who's tracking you, but how much (red: a lot; yellow: some, green: not much). And if the Badger sees that the tracker is on many different sites, it automatically disables it for you. So it's much more aggressive than Ghostery. The only downside is that sometimes it disables relatively benign trackers, like those used for library database functionality, so it requires some manual pruning. But again, it is made by the good guys.

Use one of these plugins for a week, and see how your view of the Internet changes. These trackers are constantly running, constantly tracking you for dozens of organizations. Some websites use dozens (or even hundreds) of trackers; some only use a few, or none at all. Try it out on your institution's website. Try it out on library databases. I tried it out on LexisNexis, and was surprised to see a couple of advertising trackers there. Why would those be there? (Seriously, I'm asking.)

Inspect Element: use your browser tools to see through the internet



The last tool I want to point out is your desktop browser. It's an option called Inspect Element.

This is a feature in Chrome, Firefox, and Safari, though you have to enable Developer Tools to be able to use it. I use Inspect Element mainly as diagnostic tool for figuring out what's wrong with a website. But it's also useful to see what's going on when you load a website.

Right click anywhere on a webpage and hit **Inspect Element**. My demo screenshots are from Chrome, by the way.

Now you have a split window that can show you the raw code of the webpage; every file that page loads, like images and javascript code, in a timeline with their status under Network; the contents of each file, under Sources; and under Resources, I can see the actual cookies.

Name	Value	Domain	Path	Expires / Max-Age
DYID	-828662282760163595	.dynamicyield.com	/	2017-04-30T18:03:33....
DYSES	c3bfd564fab1feafffa04d5d684d582d	.dynamicyield.com	/	Session
_dy_ses_load_seq	9085%3A1431367413393	.srv.dynamicyield.com	/	2015-06-10T18:03:33....
_dycst	m.c.	.dynamicyield.com	/	2015-06-10T18:03:35....
_dyexps	13056%7C176743%3A%3A0%3A143136741...	.dynamicyield.com	/	2015-06-10T18:03:33....
_dyprd		.srv.dynamicyield.com	/	2015-06-10T18:03:35....
_dyprdobj		.srv.dynamicyield.com	/	2015-06-10T18:03:35....

Some of the cookies from "Dynamic Yield" used on nytimes.com

Okay. This is it. These are the trackers we've been talking about. That's what they look like. They're tiny two-part text files. They've got names and values. The cookies are categorized by source when you toggled open the "Cookies" menu, so you can see how many cookies each tracking company is using on your computer. In addition, if you want to know what exactly the cookie does, you can search its name in Cookiepedia.

So what's *inside* these cookies? What do they mean? What's that gibberish, like*c3bfd564fab1feafffa04d5d684d582d*? It's not very human-readable. I'll turn to a nice explanatory video that explains the content of these cookies:

So, that gibberish is the unique identifier that the tracker keeps on your computer to tag you as a specific user. When you visit a website that uses that ad company, the website will connect that unique ID and pass it on to the ad company. Then the ad company consults its database to find this unique ID, adds any more info it just got, like what page you're on, and it looks at all the information it's gathered about this unique ID over time. It then decides which ad is most appropriate, and those are the ads you see on a webpage. This happens in a split second, when you load the webpage. Again, trackers aren't always used by ad companies. It could be a more benign case: simply telling Twitter who you are so you're logged in so you don't have to enter your password again, or telling an analytics company that the webpage just got another unique visitor.

Students get a lot of info about keeping their private information private online. They may have heard about clearing out their cookies all the time without ever having*seen* what a cookie looks like. Since we're training information-literate students, we want them to understand what exactly we mean when we say "cookie" — it's our responsibility to turn that metaphor into something real.

So, to wrap up, we've looked at three tools that drill down how our online behavior is tracked by many, many organizations. These tools are useful not just for us, but for demonstrating to students, too. We saw one advertiser's general profile of us through Google's ad settings; we saw how Ghostery and Privacy Badger tell us the names of the organizations tracking us; and we saw how browser tools can help us see the actual cookies that track us. Go forth and show your students how they are actively profiled online, all the time!

Questions? robincamilledavis@gmail.com or @robincamille

Presentation & slides available online at robincamille.com/presentations/lacuny15