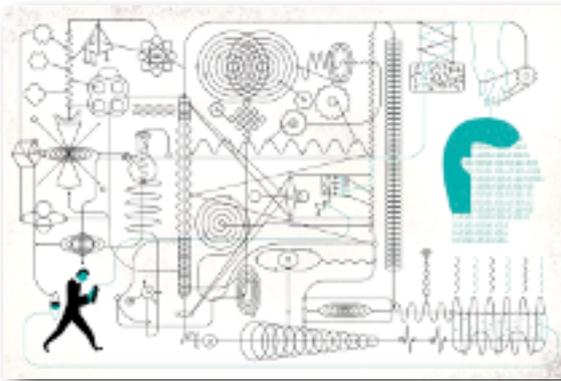




by [Jeremy B. Merrill](#)
April 30, 5 a.m. EDT

How to Wrestle Your Data From Data Brokers, Silicon Valley — and *Cambridge Analytica*

It can be done but it's not easy.



Cambridge Analytica thinks that I'm a "Very Unlikely Republican." Another political data firm, *ALC Digital*, has concluded I'm a "Socially Conservative," Republican, "Boomer Voter."

In fact, I'm a 27-year-old millennial with no set party allegiance.

For all the fanfare, the burgeoning field of mining our personal data remains an inexact art.

One thing is certain: My personal data, and likely yours, is in more hands than ever. Tech firms, data brokers and political consultants build profiles of what they know — [or think they can reasonably guess](#) — about your purchasing habits, personality, hobbies and even what political issues you care about.

You can find out what those companies know about you but be prepared to be stubborn. Very stubborn. To demonstrate how this works, we've chosen a couple of representative companies from three major categories: data brokers, big tech firms and political data consultants.

Few of them make it easy. Some will show you on their websites, others will make you ask for your digital profile via the U.S. mail. And then there's *Cambridge Analytica*, the controversial Trump campaign vendor that has come under intense fire in light of a report¹ in the British newspaper *The Observer* and in *The New York Times* that the company used improperly obtained data from *Facebook* to help build voter profiles.

To find out what the chaps at the British data firm have on you, you're going to need both stamps and a "cheque."

Once you see your data, you'll have a much better understanding of how this shadowy corner of the new economy works. You'll see what seemingly personal information they know about you ... and you'll probably have some hypotheses about where this data is coming from.

You'll also probably see some predictions about who you are that are hilariously wrong.

¹<https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>

And if you do obtain your data from any of these companies, please let us know your thoughts at politicaldata@propublica.org. We won't share or publish what you say (unless you tell us that's it's OK).

Cambridge Analytica and Other Political Consultants

Making statistically informed guesses about Americans' political beliefs and pet issues is a common business these days, with dozens of firms selling data to candidates and issue groups about the purported leanings of individual American voters.

Few of these firms have to give your data. But *Cambridge Analytica* is required to do so by an obscure European rule.

▣ Cambridge Analytica:

Around the time of the 2016 election, Paul-Olivier Dehaye, a Belgian mathematician and founder of a website that helps people exercise their data protection rights called [PersonalData.IO](https://personaldata.io), approached me with an idea for a story. He flagged some of *Cambridge Analytica's* claims about the power of its "psychographic" targeting capabilities and suggested that I demand my data from them.

So I sent off a request, following Dehaye's coaching, and citing the *UK Data Protection Act 1998*, the British implementation of a little-known European Union data-protection law that grants individuals (even Americans) the rights to see the data European companies compile about individuals.

It worked. I got back a spreadsheet of data about me. But it took months, cost ten pounds — and I had to give them a photo ID and two utility bills. Presumably they didn't want my personal data falling into the wrong hands.

▣ How You Can Request Your Data From Cambridge Analytica:

⌘ Visit Cambridge Analytica's website here and fill out this web form².

⌘ After you submit the form, the page will immediately request that you email to data.compliance@cambridgeanalytica.org a photo ID and two copies of your utility bills or bank statements, to prove your identity. This page will also include the company's bank account details.

⌘ Find a way to send them 10 GBP. You can try wiring this from your bank, though it may cost you an additional \$25 or so — or ask a friend in the UK to go to their bank and get a cashier's check. Your American bank probably won't let you write a GBP-denominated check. Two services I tried, *Xoom* and *TransferWise*, weren't able to do it.

⌘ Eventually, *Cambridge Analytica* will email you a small *Excel* spreadsheet of information and a letter. You might have to wait a few weeks. Celeste LeCompte, *ProPublica's* vice president of business development, requested her data on March 27 and still hasn't received it.

² <https://datarequests.cambridgeanalytica.org/>

Because the company is based in the United Kingdom, it had no choice but to fulfill my request. In recent weeks, the firm has come under intense fire after *The New York Times* and the British paper *The Observer* disclosed that it had used improperly obtained data from Facebook to build profiles of American voters. Facebook told me that data about me was likely transmitted to Cambridge Analytica because a person with whom I am “friends” on the social network had taken the now-infamous “This Is Your Digital Life” quiz. For what it’s worth, my data shows no sign of anything derived from Facebook.

What You Might Get Back From Cambridge Analytica:

Cambridge Analytica had generated 13 data points about my views: 10 political issues, ranked by importance; two guesses at my partisan leanings (one blank); and a guess at whether I would turn out in the 2016 general election.

They told me that the lower the rank, the higher the predicted importance of the issue to me.

| | |
|--|--------------------------|
| National Debt Importance Rank [1-10] | 6 |
| Gun Rights Importance Rank [1-10] | 4 |
| Traditional Social and Moral Values Importance Rank [1-10] | 8 |
| Environment Importance Rank [1-10] | 9 |
| Education Importance Rank [1-10] | 5 |
| National Security Importance Rank [1-10] | 1 |
| Immigration Importance Rank [1-10] | 10 |
| Socially Progressive Civil Rights Importance Rank [1-10] | 7 |
| Jobs and Economy Importance Rank [1-10] | 2 |
| Healthcare Importance Rank [1-10] | 3 |
| Registered Partisanship | Very Unlikely Republican |
| Unregistered Partisanship | |
| 2016 General Election Turnout Propensity | Low |
| 2016 General Election Turnout Propensity | Low |
| Unregistered Partisanship | Very Unlikely Republican |

Alongside that data labeled “models” were two other types of data that are run-of-the-mill and widely used by political consultants. One sheet of “core data” — that is, personal info, sliced and diced a few different ways, perhaps to be used more easily as parameters for a statistical model. It included my address, my electoral district, the census tract I live in and my date of birth.

The spreadsheet included a few rows of “election returns” — previous elections in New York State in which I had voted. (Intriguingly, Cambridge Analytica missed that I had voted in 2015’s *snoozefest* of a vote-for-five-of-these-five judicial election.

It also didn’t know about elections in which I had voted in North Carolina, where I lived before I lived in New York.)

ALC Digital

ALC Digital is another data broker, which says that its info is “audiences are built from multi-sourced, verified information about an individual.” Their data is distributed via Oracle Data Cloud, a service that lets advertisers target specific audience of people — like, perhaps, people who are Boomer Voters and also Republicans.

The firm brags in an Oracle document posted online³ about how hard it is to avoid their data collection efforts, saying,

“It has no cookies to erase and can’t be ‘cleared.’ ALC Real World Data is rooted in reality, and doesn’t rely on inferences or faulty models.”

□ How You Can Request Your Data From ALC Digital:

Here’s how to find the predictions about your political beliefs data in *Oracle Data Cloud*:

⌘ Visit <http://www.bluekai.com/registry/>. If you use an ad blocker, there may not be much to see here.

⌘ Click on the Partner Segments tab.

⌘ Scroll on through until you find *ALC Digital*.

⌘ You may have to scroll for a while before you find it.

⌘ And not everyone appears to have data from *ALC Digital*, so don’t be shocked if you can’t find it. If you don’t, there may be other fascinating companies with data about who you are in your *Oracle* file.

□ What You Might Get Back From ALC Digital:

When I downloaded the data last year, it said I was “*Socially Conservative*,” “*Boomer Voter*” — as well as a female voter and a tax reform supporter.

Recently, when I checked my data, those categories had disappeared entirely from my data. I had nothing from *ALC Digital*.

ALC Digital is not required to release this data. It is disclosed via the *Oracle Data Cloud*. Fran Green, the company’s president, said that *Aristotle*, a longtime political data company,

“provides us with consumer data that populates these audiences.”

She also said that

“we do not claim to know people’s ‘beliefs.’”

Big Tech

Big tech firms like *Google* and *Facebook* tend to make their money by selling ads, so they build extensive profiles of their users’ interests and activities. They also depend on their users’ goodwill to keep us voluntarily giving them our locations, our browsing histories and plain ol’ lists of our friends

³ <http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf>

and interests. (So far, these popular companies have not faced much regulation.) All three make it easy to download the data that they keep on you.

Firms like *Google* and *Facebook* firms don't sell your data — because it's their competitive advantage. *Google's* privacy page⁴ screams in 72 point type:

"We do not sell your personal information to anyone."

As websites that we visit frequently, they sell access to our attention, so companies that want to reach you in particular can do so with these companies' sites or other sites that feature their ads.

Facebook

How You Can Request Your Data From *Facebook*:

You of course have to have a *Facebook* account and be logged in:

- ⌘ Visit <https://www.facebook.com/settings> on your computer.
- ⌘ Click the "Download a copy of your Facebook data" link.
- ⌘ On the next page, click "Start My Archive."
- ⌘ Enter your password, then click "Start My Archive" again.
- ⌘ You'll get an email immediately, and another one saying "Your Facebook download is ready" when your data is ready to be downloaded. You'll get a notification on *Facebook*, too. Mine took just a few minutes.
- ⌘ Once you get that email, click the link, then click *Download Archive*. Then reenter your password, which will start a zip file downloading..
- ⌘ Unzip the folder; depending on your computer's operating system, this might be called uncompressing or "expanding." You'll get a folder called something like "facebook-jeremybmerrill," but, of course, with your username instead of mine.
- ⌘ Open the folder and double-click "index.htm" to open it in your web browser.

What You Might Get Back From *Facebook*

Facebook designed its archive to first show you your profile information. That's all information you typed into *Facebook* and that you probably intended to be shared with your friends. It's no surprise that *Facebook* knows what city I live in or what my AIM screen name was — I told *Facebook* those things so that my friends would know.

⁴ https://privacy.google.com/intl/en/how-ads-work.html?utm_source=google&utm_medium=ad-settings&utm_campaign=inbound-site-link

But it's a bit of a surprise that they decided to feature a list of my ex-girlfriends — what they blandly termed “*Previous Relationships*” — so prominently.

As you dig deeper in your archive, you'll find more information that you gave *Facebook*, but that you might not have expected the social network to keep hold of for years: if you're me, that's the *Nickelback* concert I apparently *RSVPed* to, posts about switching high schools and instant messages from my freshman year in college.

But finally, you'll find the creepier information: what *Facebook* knows about you that you *didn't* tell it, on the “*Ads*” page. You'll find “*Ads Topics*” that *Facebook* decided you were interested in, like *Housing*, *ESPN* or the town of *Ellijay, Georgia*. And, you'll find a list of advertisers who have obtained your contact information and uploaded it to *Facebook*, as part of a so-called *Custom Audience* of specific people to whom they want to show their ads.

You'll find more of that creepy information on your *Ads Preferences* page⁵. Despite Mark Zuckerberg telling Rep. Jerry McNerney, D-Calif., in a hearing earlier this month that “*all of your information is included in your 'download your information,'*” my archive didn't include that list of ad categories that can be used to target ads to me. (Some other types of information aren't included in the download, like other people's posts you've liked. Those are listed here⁶, along with where to find them — which, for most, is in your *Activity Log*⁷.)

This area may include *Facebook's* guesses about who you are, boiled down from some of your activities. Most Americans' will have a guess about their politics⁸ — *Facebook* says I'm a “*moderate*” about *US Politics* — and some will have a guess about so-called “*multicultural affinity*,”⁹ which *Facebook* insists is not a guess about your ethnicity, but rather what sorts of content “*you are interested in or will respond well to.*” For instance, *Facebook* recently added that I have a “*Multicultural Affinity: African American.*” (I'm white — though, because *Facebook's* definition of “*multicultural affinity*” is so strange, it's hard to tell if this is an error on *Facebook's* part.)

Facebook also doesn't include your browsing history — the subject of back-and-forths between Mark Zuckerberg and several members of Congress — it says it keeps that just long enough to boil it down into those “*Ad Topics.*”

For people without *Facebook* accounts, *Facebook* says to email datarequests@support.facebook.com or fill out an online form¹⁰ to download what *Facebook* knows about you. One puzzle here is how *Facebook* gathers data on people whose identities it may not know. It may know that a person using a phone from Atlanta, Georgia, has accessed a *Facebook* site and that the same person was last week in Austin, Texas, and before that Cincinnati, but it may not know that that person is me.

⁵ <https://www.facebook.com/ads/preferences>

⁶ https://www.facebook.com/help/405183566203254?helpref=page_content

⁷ <https://www.facebook.com/me/allactivity>

⁸ <https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html>

⁹ <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>

¹⁰ <https://www.facebook.com/help/contact/180237885820953>

It's in principle difficult for the company to give the data it collects about logged-out users if it doesn't know exactly who they are.

□ Google

Like *Facebook*, *Google* will give you a zip archive of your data. *Google's* can be much bigger, because you might have stored gigabytes of files in *Google Drive* or years of emails in *Gmail*.

But like *Facebook*, *Google* does not provide its guesses about your interests, which it uses to target ads. Those guesses are available elsewhere.

□ How You Can Request Your Data From Google:

⌘ Visit <https://takeout.google.com/settings/takeout/> to use *Google's* cutely named *Takeout service*.

⌘ You'll have to pick which data you want to download and examine. You should definitely select *My Activity*, *Location History* and *Searches*. You may not want to download gigabytes of emails, if you use *Gmail*, since that uses a lot of space and may take a while. (That's also information you shouldn't be surprised that *Google* keeps — you left it with *Gmail* so that you could use *Google's* search expertise to hold on to your emails.)

⌘ *Google* will present you with a few options for how to get your archive. The defaults are fine.

⌘ Within a few hours, you should get an email with the subject "Your *Google* data archive is ready." Click *Download Archive* and log in again. That should start the download of a file named something like "takeout-20180412T193535.zip."

⌘ Unzip the folder; depending on your computer's operating system, this might be called "uncompressing" or "expanding."

⌘ You'll get a folder called *Takeout*. Open the file inside it called "index.html" in your web browser to explore your archive.

□ What You Might Get Back From Google:

Once you open the *index.html* file, you'll see icons for the data you chose in step 2. Try exploring "Ads" under "My Activity" — you'll see a list of times you saw *Google Ads*, including on apps on your phone.

Google also includes your search history, under "Searches" — in my case, going back to 2013. *Google* knows what I had forgotten: I *Googled* a bunch of dinosaurs around Valentine's Day that year ... And it's not just web searches: the *Sound Search history* reminded me that at some point, I used that service to identify *Natalie Imbruglia's* song "Torn."

Android phone users might want to check the "Android" folder: *Google* keeps a list of each app you've used on your phone.

Most of the data contained here are records of ways you've directly interacted with *Google* — and the company really does use the those to improve how their services work for me. I'm glad to see my searches auto-completed, for instance.

But the company also creates data about you: Visit the company's *Ads Settings page*¹¹ to see some of the "topics" *Google* guesses you're interested in, and which it uses to personalize the ads you see. Those topics are fairly general — it knows I'm interested in "Politics" — but the company says it has more granular classifications that it doesn't include on the list. Those more granular, hidden classifications are on various topics, from sports to vacations to politics, where *Google* does generate a guess whether some people are politically "left-leaning" or "right-leaning."

Data Brokers

Here's who really does sell your data. Data brokers like the credit reporting agency *Experian* and a firm named *Epsilon*.

These sometimes-shady firms are middlemen who buy your data from tracking firms, survey marketers and retailers, slice and dice the data into "segments," then sell those on to advertisers.

Experian

Experian is best known as a credit reporting firm, but your credit cards aren't all they keep track of. They told me that they "firmly believe people should be made aware of how their data is being used" — so if you print and mail them a form, they'll tell you what data they have on you.

"Educated consumers," they said, "are better equipped to be effective, successful participants in a world that increasingly relies on the exchange of information to efficiently deliver the products and services consumers demand."

How You Can Request Your Data From Experian:

⌘ Visit *Experian's Marketing Data Request site* and print the *Marketing Data Report Request form*.

⌘ Print a copy of your ID and proof of address.

⌘ Mail it all to *Experian at Experian Marketing Services*
PO Box 40
Allen, TX 75013

⌘ Wait for them to mail you something back.

⌘ What You Might Get Back From *Experian*: Expect to wait a while. I've been waiting almost a month.

¹¹ <https://adssettings.google.com/authenticated>

They also come up with a guess about your political views that's integrated with *Facebook* — our *Facebook Political Ad Collector project*¹² has found that many political candidates use *Experian's* data to target their *Facebook* ads¹³ to likely supporters.

You should hope to find a guess about your political views that'd be useful to those candidates — as well as categories derived from your purchasing data.

Experian told me they generate the data they have about you from a long list of sources, including public records and “historical catalog purchase information” — as well as calculating it from predictive models.

□ Epsilon

□ How You Can Request Your Data From *Epsilon*:

⌘ Visit *Epsilon's* Marketing Data Summary Request¹⁴ form.

⌘ After entering your name and address, *Epsilon* will answer some of those identity-verification questions that quiz you about your old addresses and cars. If your identity can't be verified with those, *Epsilon* will ask you to mail in a form.

⌘ Wait for *Epsilon* to mail you your data; it took about a week for me.

□ What You Might Get Back From *Epsilon*:

Epsilon has information on “demographics” and “lifestyle interests” — at the household level. It also includes a list of “household purchases.”

It also has data that political candidates use to target their *Facebook* ads, including Randy Bryce, a Wisconsin Democrat who's seeking his party's nomination to run for retiring Speaker Paul Ryan's seat, and Rep. Tulsi Gabbard, D-Hawaii.

In my case, *Epsilon* knows I buy clothes, books and home office supplies, among other things — but isn't any more specific. They didn't tell me what political beliefs they believe I hold. The company didn't respond to a request for comment.

□ Oracle

Oracle's Data Cloud aggregates data about you from *Oracle*, but also so-called third party data from other companies.

□ How You Can Request Your Data From *Oracle*:

¹² <https://www.propublica.org/article/help-us-monitor-political-ads-online>

¹³ <https://projects.propublica.org/facebook-ads/?targets=%5B%7B%22target%22%3A%22Agency%22%7D%5D&lang=en-US>

¹⁴ <https://us.epsilon.com/marketing-data-summary-request>

⌘ Visit <http://www.bluekai.com/registry/>. If you use an ad blocker, there may not be much to see here.

⌘ Explore each tab, from “Basic Info” to “Hobbies & Interests” and “Partner Segments.”

⌘ Not fun scrolling through all those pages? I have 84 pages of four pieces of data each.

⌘ You can't search. All the text is actually images of text. Oracle declined to say why it chose to make their site so hard to use.

□ What You Might Get Back From Oracle:

My *Oracle* profile includes nearly 1500 data points, covering all aspects of my life, from my age to my car to how old my children are to whether I buy eggs. These profiles can even say if you're likely to dress your pet in a costume for Halloween. But many of them are off-base or contradictory.

Many companies in *Oracle*'s data, besides *ALC Digital*, offer guesses about my political views: Data from one company uploaded by *AcquireWeb* says that my political affiliations are as a Democrat and an Independent ... but also that I'm a “Mild Republican.” Another company, an *Oracle* subsidiary called *AddThis*, says that I'm a “Liberal.” *Cuebiq*, which calls itself a “location intelligence” company, says I'm in a subset of “Democrats” called “Liberal Professions.”

If an advertiser wants to show an ad to *Spring Break Enthusiasts*, *Oracle* can enable that. I'm apparently a *Spring Break Enthusiast*. Do I buy eggs? I sure do. Data on *Oracle*'s site associated with *AcquireWeb* says I'm a cat owner ...

But it also “knows” I'm a dog owner, which I'm not.

Al Gadbut, the CEO of *AcquireWeb*, explained that the guesses associated with his company weren't based on my personal data, but rather the tendencies of people in my geographical area — hence the seemingly contradictory political guesses. He said his firm doesn't generate the data, but rather uploaded it on behalf of other companies. *Cuebiq*'s guess was a “probabilistic inference” they drew from location data submitted to them by some app on my phone. Valentina Marastoni-Bieser, *Cuebiq*'s senior vice president of marketing, wouldn't tell me which app it was, though.

Data for sale here includes a long list what TV shows I — supposedly — watch.

But it's not all wrong. *AddThis* can tell that I'm “Young & Hip.”

Takeaways:

The above list is just a sampling of the firms that collect your data and try to draw conclusions about who you are — not just sites you visit like *Facebook* and controversial firms like *Cambridge Analytica*.

You can make some guesses as to where this data comes from — especially the more granular consumer data from *Oracle*. For each data point, it's worth considering: Who'd be in a position to sell a list of what TV shows I watch, or, at least, a list of what TV shows people demographically like me watch? Who'd be in a position to sell a list of what groceries I, or people similar to me in my area,

buy? Some of those companies — companies who you're likely paying, and for whom the internet adage that "if you're not paying, you're the product" doesn't hold — are likely selling data about you without your knowledge. Other data points, like the location data used by *Cuebiq*, can come from any number of apps or websites, so it may be difficult to figure out exactly which one has passed it on.

Companies like *Google* and *Facebook* often say that they'll let you "correct" the data that they hold on you — tacitly acknowledging that they sometimes get it wrong. But if receiving relevant ads is not important to you, they'll let you opt-out entirely — or, presumably, "correct" your data to something false.

An upcoming European Union rule called the *General Data Protection Regulation* portends a dramatic change to how data is collected and used on the web — if only for Europeans. No such law seems likely to be passed in the U.S. in the near future.

More on Protecting Your Data

- **So What the Hell Is Doxxing?** - What doxxing really is, plus advice on how to protect yourself from unwanted exposure of personal and private information online.