

**Data Mining:
How Companies Now Know Everything About You
By Joel Stein, *TIME* March 10, 2011**

Three hours after I gave my name and e-mail address to Michael Fertik, the CEO of Reputation.com, he called me back and read my Social Security number to me. "We had it a couple of hours ago," he said. "I was just too busy to call."

In the past few months, I have been told many more-interesting facts about myself than my Social Security number. I've gathered a bit of the vast amount of data that's being collected both online and off by companies in stealth — taken from the websites I look at, the stuff I buy, my Facebook photos, my warranty cards, my customer-reward cards, the songs I listen to online, surveys I was guilted into filling out and magazines I subscribe to.

Google's Ads Preferences believes I'm a guy interested in politics, Asian food, perfume, celebrity gossip, animated movies and crime but who doesn't care about "books & literature" or "people & society." (So not true.) Yahoo! has me down as a 36-to-45-year-old male who uses a Mac computer and likes hockey, rap, rock, parenting, recipes, clothes and beauty products; it also thinks I live in New York, even though I moved to Los Angeles more than six years ago. Alliance Data, an enormous data-marketing firm in Texas, knows that I'm a 39-year-old college-educated Jewish male who takes in at least \$125,000 a year, makes most of his purchases online and spends an average of only \$25 per item. Specifically, it knows that on Jan. 24, 2004, I spent \$46 on "low-ticket gifts and merchandise" and that on Oct. 10, 2010, I spent \$180 on intimate apparel. It knows about more than 100 purchases in between. Alliance also knows I owe \$854,000 on a house built in 1939 that — get this — it thinks has stucco walls. They're mostly wood siding with a little stucco on the bottom! Idiots.

EXelate, a Manhattan company that acts as an exchange for the buying and selling of people's data, thinks I have a high net worth and dig green living and travel within the U.S. BlueKai, one of eXelate's competitors in Bellevue, Wash., believes I'm a "collegiate-minded" senior executive with a high net

worth who rents sports cars (note to Time Inc. accounting: it's wrong unless the Toyota Yaris is a sports car). At one point BlueKai also believed, probably based on my \$180 splurge for my wife Cassandra on HerRoom.com, that I was an 18-to-19-year-old woman.

RapLeaf, a data-mining company that was recently banned by Facebook because it mined people's user IDs, has me down as a 35-to-44-year-old married male with a graduate degree living in L.A. But RapLeaf thinks I have no kids, work as a medical professional and drive a truck. RapLeaf clearly does not read my column in TIME.

Intellidyn, a company that buys and sells data, searched its file on me, which says I'm a writer at Time Inc. and a "highly assimilated" Jew. It knows that Cassandra and I like gardening, fashion, home decorating and exercise, though in my case the word like means "am forced to be involved in." We are pretty unlikely to buy car insurance by mail but extremely likely to go on a European river cruise, despite the fact that we are totally not going to go on a European river cruise. There are tons of other companies I could have called to learn more about myself, but in a result no one could have predicted, I got bored.

Each of these pieces of information (and misinformation) about me is sold for about two-fifths of a cent to advertisers, which then deliver me an Internet ad, send me a catalog or mail me a credit-card offer. This data is collected in lots of ways, such as tracking devices (like cookies) on websites that allow a company to identify you as you travel around the Web and apps you download on your cell that look at your contact list and location. You know how everything has seemed free for the past few years? It wasn't. It's just that no one told you that instead of using money, you were paying with your personal information.

The Creep Factor

There is now an enormous multibillion-dollar industry based on the collection and sale of this personal and behavioral data, an industry that Senator John Kerry, chair of the Subcommittee on Communications, Technology and the Internet, is hoping to rein in. Kerry is about to introduce a bill that would require companies to make sure all the stuff they know about you is secured

from hackers and to let you inspect everything they have on you, correct any mistakes and opt out of being tracked. He is doing this because, he argues, "There's no code of conduct. There's no standard. There's nothing that safeguards privacy and establishes rules of the road."

At Senate hearings on privacy beginning March 16, the Federal Trade Commission (FTC) will be weighing in on how to protect consumers. It has already issued a report that calls upon the major browsers to come up with a do-not-track mechanism that allows people to choose not to have their information collected by companies they aren't directly doing business with. Under any such plan, it would likely still be O.K. for Amazon to remember your past orders and make purchase suggestions or for American Express to figure your card was stolen because a recent purchase doesn't fit your precise buying patterns. But it wouldn't be cool if they gave another company that information without your permission.

Taking your information without asking and then profiting from it isn't new: it's the idea behind the phone book, junk mail and telemarketing. Worrying about it is just as old: in 1890, Louis Brandeis argued that printing a photograph without the subject's permission inflicts "mental pain and distress, far greater than could be inflicted by mere bodily harm." Once again, new technology is making us weigh what we're sacrificing in privacy against what we're gaining in instant access to information. Some facts about you were always public — the price of your home, some divorce papers, your criminal records, your political donations — but they were held in different buildings, accessible only by those who filled out annoying forms; now they can be clicked on. Other information was not possible to compile pre-Internet because it would have required sending a person to follow each of us around the mall, listen to our conversations and watch what we read in the newspaper. Now all of those activities happen online — and can be tracked instantaneously.

Part of the problem people have with data mining is that it seems so creepy. Right after I e-mailed a friend in Texas that I might be coming to town, a suggestion for a restaurant in Houston popped up as a one-line all-text ad above my Gmail inbox. But it's not a barbecue-pit master stalking me, which would indeed be creepy; it's an algorithm designed to give me more useful, specific ads. And while that doesn't sound like all that good a deal in exchange

for my private data, if it means that I get to learn when the next Paul Thomas Anderson movie is coming out, when Wilco is playing near my house and when Tom Colicchio is opening a restaurant close by, maybe that's not such a bad return.

Since targeted ads are so much more effective than nontargeted ones, websites can charge much more for them. This is why — compared with the old banners and pop-ups — online ads have become smaller and less invasive, and why websites have been able to provide better content and still be free. Besides, the fact that I'm going to Houston is bundled with the information that 999 other people are Houston-bound and is auctioned by a computer; no actual person looks at my name or my Houston-boundness. Advertisers are interested only in tiny chunks of information about my behavior, not my whole profile, which is one of the reasons M. Ryan Calo, a Stanford Law School professor who is director of the school's Consumer Privacy Project, argues that data mining does no actual damage.

"We have this feeling of being dogged that's uncomfortable," Calo says, "but the risk of privacy harm isn't necessarily harmful. Let's get serious and talk about what harm really is." The real problem with data mining, Calo and others believe, arises when the data is wrong. "It's one thing to see bad ads because of bad information about you. It's another thing if you're not getting a credit card or a job because of bad information," says Justin Brookman, the former chief of the Internet bureau of the New York attorney general's office, who is now the director of the Center for Democracy and Technology, a nonprofit group in Washington.

Russell Glass, the CEO of Bizo — which mines the fact that people are business executives and sells that info to hundreds of advertisers such as American Express, Monster.com, Citibank, Sprint and Google — says the newness of his industry is what scares people. "It's the monster-under-the-bed syndrome," Glass says. "People are afraid of what they really don't understand. They don't understand that companies like us have no idea who they are. And we really don't give a s — -. I just want a little information that will help me sell you an ad." Not many people, he notes, seem to be creeped out by all the junk mail they still get from direct-marketing campaigns, which buy the same information from data-mining companies. "I have a 2-year-old

daughter who is getting mail at my home address," he says. "That freaks me out."

Why That Ad Is Following You

Junk mail is a familiar evil that's barely changed over the decades. Data mining and the advertising it supports get more refined every month. The latest trick to freak people out is retargeting — when you look at an item in an online store and then an ad for that item follows you around to other sites.

Last year, Zappos was the most prominent company in the U.S. to go all out in behavioral retargeting. And people got pissed off. One of the company's mistakes was running ads too frequently and coming off as an annoying, persistent salesman. "We took that brick-and-mortar pet peeve and implied it online," says Darrin Shamo, Zappos' director of direct marketing. Shamo learned, the hard way, that people get upset when their computer shows lingerie ads, even if they had been recently shopping for G-strings, since people share computers and use them in front of their kids. He also learned that ads that reveal potential Christmas gifts are bad for business.

Since then, Zappos has been experimenting with new ads that people will see no more than five times and for no longer than eight days. Zappos has also dumbbed the ads down, showing items that aren't the ones you considered buying but are sort of close, which people greatly prefer. And much like Amazon's "Customers who bought 1984 also bought Brave New World"—style recommendation engine, the new ads tell people what Zappos knows about them and how they got that information ("a company called Criteo helps Zappos to create these kinds of personalized ads"). It also tells them how they can opt out of seeing them ("Some people prefer rainbows. And others prefer unicorns. If you prefer not to see personalized ads, we totally get it").

If that calms the angry 15% of the people who saw these ads, Zappos will stick with them. Otherwise, it plans on quitting the retargeting business. Shamo thinks he'll just need to wait until the newness wears off and people are used to ads tailored for them. "Sometimes things don't move as fast as you think," he says.

They're not even moving that much faster with the generation that grew up with the Internet. While young people expect more of their data to be mined and used, that doesn't mean they don't care about privacy. "In my research, I found that teenagers live with this underlying anxiety of not knowing the rules of who can look at their information on the Internet. They think schools look at it, they think the government looks at it, they think colleges can look at it, they think employers can look at it, they think Facebook can see everything," says Sherry Turkle, a professor at MIT who is the director of the Initiative on Technology and Self and the author of *Alone Together: Why We Expect More from Technology and Less From Each Other*. "It's the opposite of the mental state I grew up in. My grandmother took me down to the mailbox in Brooklyn every morning, and she would say, 'It's a federal offense for anyone to look at your mail. That's what makes this country great.' In the old country they'd open your mail, and that's how they knew about you."

Data mining, Turkle argues, is a panopticon: the circular prison invented by 18th century philosopher Jeremy Bentham where you can't tell if you're being observed, so you assume that you always are. "The practical concern is loss of control and loss of identity," says Marc Rotenberg, executive director of the Electronic Privacy Information Center. "It's a little abstract, but that's part of what's taking place."

The Facebook and Google Troves

Our identities, however, were never completely within our control: our friends keep letters we've forgotten writing, our enemies tell stories about us we remember differently, our yearbook photos are in way too many people's houses. Opting out of all those interactions is opting out of society. Which is why Facebook is such a confusing privacy hub point. Many data-mining companies made this argument to me: How can I complain about having my Houston trip data-mined when I'm posting photos of myself with a giant mullet and a gold chain on Facebook and writing columns about how I want a second kid and my wife doesn't? Because, unlike when my data is secretly mined, I get to control what I share. Even narcissists want privacy. "It's the difference between sharing and tracking," says Bret Taylor, Facebook's chief technology officer.

To get into the Facebook office in Palo Alto, Calif., I have to sign a piece of physical paper: a Single-Party Non-Disclosure Agreement, which legally prevents me from writing the last paragraph. But your privacy on Facebook — that's up to you. You choose what to share and what circle of friends gets to see it, and you can untag yourself from any photos of you that other people put up. However, from a miner's point of view, Facebook has the most valuable trove of data ever assembled: not only have you told it everything you like, but it also knows what your friends like, which is an amazing predictor of what you'll like.

Facebook doesn't sell any of your data, partly because it doesn't have to — 23.1% of all online ads not on search engines, video or e-mail run on Facebook. But data-mining companies are "scraping" all your personal data that's not set to private and selling it to any outside party that's interested. So that information is being bought and sold unless you squeeze your Facebook privacy settings tight, which keeps you from a lot of the social interaction that drew you to the site in the first place.

The only company that might have an even better dossier on you than Facebook is Google. In a conference room on the Google campus, I sit through a long privacy-policy PowerPoint presentation. Summary: Google cares! Specifically, Google keeps the data it has about you from various parts of its company separate. One category is the personally identifiable account data it can attach to your name, age, gender, e-mail address and ZIP code when you signed up for services like Gmail, YouTube, Blogger, Picasa, iGoogle, Google Voice or Calendar. The other is log data associated with your computer, which it "anonymizes" after nine months: your search history, Chrome browser data, Google Maps requests and all the info its myriad data trackers and ad agencies (DoubleClick, AdSense, AdMob) collect when you're on other sites and Android phone apps. You can change your settings on the former at Google Dashboard and the latter at Google Ads Preferences — where you can opt out of having your data mined or change the company's guesses about what you're into.

Nicole Wong, deputy general counsel at Google, says the company created these tools to try to reassure people who have no idea how all this information is being collected and used. "When I go to TIME.com as a user, I think only TIME.com is collecting my data. What I don't realize is that for every ad on

that page, a company is also dropping a code and collecting my data. It's a black box — and we've tried to open up the box. Sometimes you're not even sure who the advertisers are. It's just a bunch of jumping monkeys or something." Google really does want to protect your privacy, but it's got issues. First, it's profit-driven and it's huge. But those aren't the main reasons privacy advocates get so upset about Google. They get upset because the company's guiding philosophy conflicts with the notion of privacy. As the PowerPoint says right up top: "Google's mission: to organize the world's information and make it universally accessible and useful." Which is awesome, except for the fact that my information is part of the world's information.

Tracking the Trackers

To see just what information is being gathered about me, I downloaded Ghostery, a browser extension that lets you watch the watchers watching you. Each time you go to a new website, up pops a little bubble that lists all the data trackers checking you out. This is what I discovered: the very few companies that actually charge you for services tend not to data mine much. When you visit TIME.com, several dozen tracking companies, with names such as Eyeblaster, Bluestreak, DoubleClick and Factor TG, could be collecting data at any given time.

If you're reading this in print as a subscriber, TIME has probably "rented" your name and address many times to various companies for a one-time use. This is also true if you subscribe to Vanity Fair, Cosmopolitan or just about any other publication.

This being America, I don't have to wait for the government to give me an opt-out option; I can pay for one right now. Michael Fertik, the CEO and founder of Reputation.com, who nabbed my Social Security number, will do it for me for just \$8.25 a month. His company will also, for a lot more money, make Google searches of your name come up with more flattering results — because when everyone is famous, everyone needs a public relations department. Fertik, who clerked for the chief judge of the Sixth Circuit after graduating from Harvard Law School, believes that if data mining isn't regulated, everyone will soon be assigned scores for attractiveness and a social-prowess index and a complainer index, so companies can avoid serving

you — just as you now have a credit score that they can easily check before deciding to do business with you. "What happens when those data sets are used for life transactions: health insurance, employment, dating and education? It's inevitable that all of these decisions will be made based on machine conclusions. Your FICO score is already an all-but-decisional fact about you. ABD, dude! All but decisional," says Fertik.

Even if I were to use the services of Reputation.com, there's still all the public information about me that I can't suppress. Last year, thousands of people sent their friends a Facebook message telling them to opt out of being listed on Spokeo.com, which they described as the creepiest paparazzo of all, giving out your age, profession, address and a photo of your house. Spokeo, a tiny company in Pasadena, Calif., is run by 28-year-old Stanford grad Harrison Tang. He was surprised at the outcry. "Some people don't know what Google Street View is, so they think this is magic," Tang says of the photos of people's homes that his site shows. The info on Spokeo isn't even all that revealing — he purposely leaves off criminal records and previous marriages — but Tang thinks society is still learning about data mining and will soon become inured to it. "Back in the 1990s, if you said, 'I'm going to put pictures on the Internet for everyone to see,' it would have been hard to believe. Now everyone does it. The Internet is becoming more and more open. This world will become more connected, and the distance between you and me will be a lot closer. If everybody is a walled garden, there won't be an Internet."

I deeply believe that, but it's still too easy to find our gardens. Your political donations, home value and address have always been public, but you used to have to actually go to all these different places — courthouses, libraries, property-tax assessors' offices — and request documents. "You were private by default and public by effort. Nowadays, you're public by default and private by effort," says Lee Tien, a senior staff attorney for the Electronic Frontier Foundation, an advocacy group for digital rights. "There are all sorts of inferences that can be made about you from the websites you visit, what you buy, who you talk to. What if your employer had access to information about you that shows you have a particular kind of health condition or a woman is pregnant or thinking about it?" Tien worries that political dissidents in other countries, battered women and other groups that need anonymity are vulnerable to data mining. At the very least, he argues, we're responsible to

protect special groups, just as Google Street View allows users to request that a particular location, like an abused-women's shelter, not be photographed.

Other democratic countries have taken much stronger stands than the U.S. has on regulating data mining. Google Street View has been banned by the Czech Republic. Germany — after protests and much debate — decided at the end of last year to allow it but to let people request that their houses not be shown, which nearly 250,000 people had done as of last November. E.U. Justice Commissioner Viviane Reding is about to present a proposal to allow people to correct and erase information about themselves on the Web. "Everyone should have the right to be forgotten," she says. "Due to their painful history in the 20th century, Europeans are naturally more sensitive to the collection and use of their data by public authorities."

After 9/11, not many Americans protested when concerns about security seemed to trump privacy. Now that privacy issues are being pushed in Congress, companies are making last-ditch efforts to become more transparent. New tools released in February for Firefox and Google Chrome browsers let users block data collecting, though Firefox and Chrome depend on the data miners to respect the users' request, which won't stop unscrupulous companies. In addition to the new browser options, an increasing number of ads have a little i (an Advertising Option Icon), which you can click on to find out exactly which companies are tracking you and what they do. The technology behind the icon is managed by Evidon, the company that provides the Ghostery download. Evidon has gotten more than 500 data-collecting companies to provide their info.

It takes a lot of work to find out about this tiny little i and even more to click on it and read the information. But it also took people a while to learn what the recycling symbol meant. And reading the info behind the i icon isn't necessarily the point, says Evidon CEO Scott Meyer, who used to be CEO of About.com and managed the New York Times' website. "Do I look at nutritional labeling? No. But would I buy a food product that didn't have one? Absolutely not. I would be really concerned. It's accountability."

FTC chairman Jon Leibowitz has been pleased by how effective he's been at using the threat of legislation to scare companies into taking action and dropping their excuse that they don't know anything about you personally, just

data associated with your computer. "We used to have a distinction 10 years ago between personally identifiable information and non-PII. Now those distinctions have broken down." In November, Leibowitz hired Edward Felten, the Princeton computer-science professor famous for uncovering weaknesses in electronic-voting machines and digital-music protection, to serve as the FTC's chief technologist for the next year. Felten has found that the online-advertising industry is as eager as the government is for improved privacy protections. "There's a lot of fear that holds people back from doing things they would otherwise do online. This is part of the cost of privacy uncertainty. People are a little wary of trying out some new site or service if they're worried about giving their information," Felten says.

He's right: oddly, the more I learned about data mining, the less concerned I was. Sure, I was surprised that all these companies are actually keeping permanent files on me. But I don't think they will do anything with them that does me any harm. There should be protections for vulnerable groups, and a government-enforced opt-out mechanism would be great for accountability. But I'm pretty sure that, like me, most people won't use that option. Of the people who actually find the Ads Preferences page — and these must be people pretty into privacy — only 1 in 8 asks to opt out of being tracked. The rest, apparently, just like to read privacy rules.

We're quickly figuring out how to navigate our trail of data — don't say anything private on a Facebook wall, keep your secrets out of e-mail, use cash for illicit purchases. The vast majority of it, though, is worthless to us and a pretty good exchange for frequent-flier miles, better search results, a fast system to qualify for credit, finding out if our babysitter has a criminal record and ads we find more useful than annoying. Especially because no human being ever reads your files. As I learned by trying to find out all my data, we're not all that interesting.

— With reporting by Eben Harrell / London