RADIO

LOUISE MATSAKIS BUSINESS 02.15.19 07:00 AM

THE WIRED GUIDE TO YOUR PERSONAL DATA (AND WHO IS USING IT)

ON THE INTERNET, the personal data users give away for free is transformed into a precious commodity. The puppy photos people upload train machines to be smarter. The questions they ask Google uncover humanity's deepest prejudices. And their location histories tell investors which stores attract the most shoppers. Even seemingly benign activities, like staying in and watching a movie, generate mountains of information, treasure to be scooped up later by businesses of all kinds.

Personal data is often compared to oil—it powers today's most profitable corporations, just like fossil fuels energized those of the past. But the consumers it's extracted from often know little about how much of their information is collected, who gets to look at it, and what it's worth. Every day, hundreds of companies you may not even know exist gather facts about you, some more intimate than others. That information may then flow to academic researchers, hackers, law enforcement, and foreign nations—as well as plenty of companies trying to sell you stuff.

What Constitutes "Personal Data"?

The internet might seem like one big privacy nightmare, but don’t throw your smartphone out the window just yet. “Personal data” is a pretty vague umbrella term, and it helps to unpack exactly what it means. Health records, social security numbers, and banking details make up the most sensitive information stored online. Social media posts, location data, and search-engine queries may also be revealing but are also typically monetized in a way that, say, your credit card number is not. Other kinds of data collection fall into separate categories—ones that may surprise you. Did you know some companies are analyzing the unique way you tap and fumble with your smartphone?

All this information is collected on a wide spectrum of consent: Sometimes the data is forked over knowingly, while in other scenarios users might not understand they’re giving up anything at all. Often, it’s clear *something* is being collected, but the specifics are hidden from view or buried in hard-to-parse terms-of-service agreements.

Consider what happens when someone sends a vial of saliva to 23andme. The person knows they're sharing their DNA with a genomics company, but they may not realize it will be resold to pharmaceutical firms. Many apps use your location to serve up custom advertisements, but they don't necessarily make it clear that a hedge fund may also buy that location data to analyze which retail stores you frequent. Anyone who has witnessed the same shoe advertisement follow them around the web knows they're being tracked, but fewer people likely understand that companies may be recording not just their clicks but also the exact movements of their mouse.

In each of these scenarios, the user received something in return for allowing a corporation to monetize their data. They got to learn about their genetic ancestry, use a mobile app, or browse the latest footwear trends from the comfort of their computer. This is the same sort of bargain Facebook and Google offer. Their core products, including Instagram, Messenger, Gmail, and Google Maps, don't cost money. You pay with your personal data, which is used to target you with ads.

Who Buys, Sells, and Barters My Personal Data?

The trade-off between the data you give and the services you get may or may not be worth it, but another breed of business amasses, analyzes, and sells your information without giving you anything at all: data brokers. These firms compile info from publicly available sources like property records, marriage licenses, and court cases. They may also gather your medical records, browsing history, social media connections, and online purchases. Depending on where you live, data brokers might even purchase your information from the Department of Motor Vehicles. Don't have a driver's license? Retail stores sell info to data brokers, too.

The information data brokers collect may be inaccurate or out of date. Still, it can be incredibly valuable to corporations, marketers, investors, and individuals. In fact, American companies alone are estimated to have spent over $19 billion in 2018 acquiring and analyzing consumer data, according to the Interactive Advertising Bureau.

Data brokers are also valuable resources for abusers and stalkers. Doxing, the practice of publicly releasing someone's personal information without their consent, is often made possible because of data brokers. While you can delete your Facebook account relatively easily, getting these firms to remove your information is time-consuming, complicated, and sometimes impossible. In fact, the process is so burdensome that you can pay a service to do it on your behalf.

Amassing and selling your data like this is perfectly legal. While some states, including California and Vermont, have recently moved to put more restrictions on data brokers, they remain largely unregulated. The Fair Credit Reporting Act dictates how information collected for credit, employment, and insurance reasons may be used, but some data brokers have been caught skirting the law. In 2012 the "person lookup" site Spokeo settled with the FTC for $800,000 over charges that it violated the FCRA by advertising its products for purposes like job background checks. And data brokers that market themselves as being more akin to digital phone books don't have to abide by the regulation in the first place.

There are also few laws governing how social media companies may collect data about their users. In the United States, no modern federal privacy regulation exists, and the government can even legally request digital data held by companies without a warrant in many circumstances (though the Supreme Court recently expanded Fourth Amendment protections to a narrow type of location data).

The good news is, the information you share online does contribute to the global store of useful knowledge: Researchers from a number of academic disciplines study social media posts and other user-generated data to learn more about humanity. In his book, *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are,* Seth Stephens-Davidowitz argues there are many scenarios where humans are more honest with sites like Google than they are on traditional surveys. For example, he says, fewer than 20 percent of people admit they watch porn, but there are more Google searches for "porn" than "weather."

Personal data is also used by artificial intelligence researchers to train their automated programs. Every day, users around the globe upload billions of photos, videos, text posts, and audio clips to sites like YouTube, Facebook, Instagram, and Twitter. That media is then fed to machine learning algorithms, so they can learn to “see” what’s in a photograph or automatically determine whether a post violates Facebook’s hate-speech policy. Your selfies are literally making the robots smarter. Congratulations.

The History of Personal Data Collection

Humans have used technological devices to collect and process data about the world for thousands of years. Greek scientists developed the “first computer,” a complex gear system called the Antikythera mechanism, to trace astrological patterns as far back as 150 BC. Two millennia later, in the late 1880s, Herman Hollerith invented the tabulating machine, a punch card device that helped process data from the 1890 United States Census. Hollerith created a company to market his invention that later merged into what is now IBM.

By the 1960s, the US government was using powerful mainframe computers to store and process an enormous amount of data on nearly every American. Corporations also used the machines to analyze sensitive information including consumer purchasing habits. There were no laws dictating what kind of data they could collect. Worries over supercharged surveillance soon emerged, especially after the publication of Vance Packard’s 1964 book, *The Naked Society*, which argued that technological change was causing the unprecedented erosion of privacy.

THE TRACKERS TRACKING YOU

Online trackers can be divided into two main categories: same-site and cross-site. The former are mostly benign, while the latter are more invasive. A quick taxonomy:

- **Traditional Cookies**
  Facebook, Google, and other companies use these extremely popular cross-site trackers to follow users from website to website. They work by depositing a piece of code into the browser, which users then

unwittingly carry with them as they surf the web.

- **Super Cookies**
  Supercharged cookies can be difficult or impossible to clear from your browser. They were most famously used by Verizon, which had to pay a $1.35 million fine to the FCC as a result of the practice.

- **Fingerprinters**
  These cross-site trackers follow users by creating a unique profile of their device. They collect things like the person's IP address, their screen resolution, and what type of computer they have.

- **Identity trackers**
  Instead of using a cookie, these rare trackers follow people using personally identifiable information, such as their email address. They collect this data by hiding on login pages where people enter their credentials.

- **Session cookies**
  Some trackers are good! These helpful same-site scripts keep you logged in to websites and remember what's in your shopping cart—often even if you close your browser window.

- **Session replay scripts**
  Some same-site scripts can be incredibly invasive. These record everything you do on a website, such as which products you clicked on and sometimes even the password you entered.

The next year, President Lyndon Johnson's administration proposed merging hundreds of federal databases into one centralized National Data Bank. Congress, concerned about possible surveillance, pushed back and organized a Special Subcommittee on the Invasion of Privacy. Lawmakers worried the data bank, which would "pool statistics on millions of Americans," could "possibly violate their secret lives," *The New York Times* reported at the time. The project was never realized. Instead, Congress passed a series of laws governing the use of personal data, including the Fair Credit Reporting Act in 1970 and the Privacy Act in 1974. The regulations mandated transparency but did nothing to prevent the government and corporations from collecting information in the *first* place, argues technology historian Margaret O'Mara.

Toward the end of the 1960s, some scholars, including MIT political scientist Ithiel de Sola Pool, predicted that new computer technologies would continue to facilitate even more invasive personal data collection. The reality they envisioned began to take shape in the mid-1990s, when many Americans started using the internet. By the time most everyone was online, though, one of the first privacy battles over digital data brokers had already been fought: In 1990, Lotus Corporation and the credit bureau Equifax teamed up to create Lotus MarketPlace: Households, a CD-ROM marketing product that was advertised to contain names, income ranges, addresses, and other information about more than 120 million Americans. It quickly caused an uproar among privacy advocates on digital forums like Usenet; over 30,000 people contacted Lotus to opt out of the database. It was ultimately canceled before it was even released. But the scandal didn't stop other companies from creating massive data sets of consumer information in the future.

Several years later, ads began permeating the web. In the beginning, online advertising remained largely anonymous. While you may have seen ads for skiing if you looked up winter sports, websites couldn't connect you to your real identity. (HotWired.com, the online version of WIRED, was the first website to run a banner ad in 1994, as part of a campaign for AT&T.) Then, in 1999, digital ad giant DoubleClick ignited a privacy scandal when it tried to de-anonymize its ads by merging with the enormous data broker Abacus Direct.

Privacy groups argued that DoubleClick could have used personal information collected by the data broker to target ads based on people's real names. They petitioned the Federal Trade Commission, arguing that the practice would amount to unlawful tracking. As a result, DoubleClick sold the firm at a loss in 2006, and the Network Advertising Initiative was created, a trade group that developed standards for online advertising, including requiring companies to notify users when their personal data is being collected.

But privacy advocates' concerns eventually came true. In 2008, Google officially acquired DoubleClick, and in 2016 it revised its privacy policy to permit personally-identifiable web tracking. Before then, Google kept its DoubleClick browsing data separate from personal information it collected from services like Gmail. Today, Google and Facebook can target ads based on your name—exactly what people feared DoubleClick would do two decades ago. And that's not all: Because most people carry tracking devices in their pockets in the form of smartphones, these companies, and many others, can also follow us wherever we go.

The Future of Personal Data Collection

Personal information is currently collected primarily through screens, when people use computers and smartphones. The coming years will bring the widespread adoption of new data-guzzling devices, like smart speakers, censor-embedded clothing, and wearable health monitors. Even those who refrain from using these devices will likely have their data gathered, by things like facial recognition-enabled surveillance cameras installed on street corners. In many ways, this future has already begun: Taylor Swift fans have had their face data collected, and Amazon Echos are listening in on millions of homes.

We haven't decided, though, how to navigate this new data-filled reality. Should colleges be permitted to digitally track their teenage applicants? Do we really want health insurance companies monitoring our Instagram posts? Governments, artists, academics, and citizens will think about these questions and plenty more.

And as scientists push the boundaries of what's possible with artificial intelligence, we will also need to learn to make sense of personal data that isn't even *real*, at least in that it didn't come from humans. For example, algorithms are already generating "fake" data for other algorithms to train on. So-called deepfake technology allows propagandists and hoaxers to leverage social media photos to make videos depicting events that never happened. AI can now create millions of synthetic faces that don't belong to anyone, altering the meaning of stolen identity. This fraudulent data could further distort social media and other parts of the internet. Imagine trying to discern whether a Tinder match or the person you followed on Instagram actually exists.

Whether data is fabricated by computers or created by real people, one of the biggest concerns will be how it is analyzed. It matters not just what information is collected but also what inferences and predictions are made based upon it. Personal data is used by algorithms to make incredibly important decisions, like whether someone should maintain their health care benefits, or be released on bail. Those decisions can easily be biased, and researchers and companies like Google are now working to make algorithms more transparent and fair.

Tech companies are also beginning to acknowledge that personal data collection needs to be regulated. Microsoft has called for the federal regulation of facial recognition, while Apple CEO Tim Cook has argued that the FTC should step in and create a clearinghouse where all data brokers need to register. But not all of Big Tech's declarations may be in good faith. In the summer of 2018, California passed a strict privacy law that will go into effect on January 1, 2020, unless a federal law supersedes it. Companies like Amazon, Apple, Facebook, and Google are now pushing for Congress to pass new, less stringent privacy legislation in 2019 before the California law kicks in. Even in a divided Congress, lawmakers could come together around privacy—scrutinizing Big Tech has become an important issue for both sides.

Some companies and researchers argue it's not enough for the government to simply protect personal data; consumers need to own their information and be compensated when it's used. Social networks like Minds and Steemit have experimented with rewarding users with cryptocurrency when they share content or spend time using their platforms. Other companies will pay you in exchange for sharing data—your banking transactions, for instance—with them. But allowing people to take back ownership likely wouldn't solve every privacy issue posed by personal data collection. It might also be the wrong way to frame the issue: Instead, perhaps, less collection should be permitted in the first place, forcing companies to move away from the targeted-advertising business model altogether.

Before we can figure out the future of personal data collection, we need to learn more about its present. The cascade of privacy scandals that have come to light in recent years—from Cambridge Analytica to Google's shady location tracking practices—have demonstrated that users still don't know all the ways their information is being sold, traded, and shared. Until consumers actually understand the ecosystem they've unwittingly become a part of, we won't be able to grapple with it in the first place.

Learn More

- **The Privacy Battle to Save Google From Itself**
  Google's sprawling privacy apparatus includes thousands of employees and billions of dollars in cumulative investment. But the company is still an advertising behemoth and fundamentally makes money by monetizing the personal data it collects from users. Yet Google has also played a leadership role in creating industry standards for transparency and data protection. More than a dozen privacy employees at Google spoke to WIRED about how they make sense of the paradox of their work, insisting that there's no internal pressure to compromise privacy protections to make a larger profit.

- **Few Rules Govern Police Use of Facial-Recognition Technology**
  One of the most sensitive pieces of personal data you possess isn't hidden

at all: It's your face. The issue has become contentious for civil rights activists, and Amazon in particular has faced backlash—even from its own employees—over use of the technology, especially for law enforcement purposes. With the exception of two states however, few laws regulating the use of facial recognition exist.

- **Carriers Swore They'd Stop Selling Location Data. Will They Ever?** In 2018, US phone carriers promised to stop selling customer location data after journalists discovered it had ended up in the hands of questionable third parties. Not even a year later, the same carriers were caught doing it again. The question now is how the Federal Communications Commission will handle the issue. The agency has the authority to make it illegal for carriers to sell this kind of information, but so far it hasn't said whether the law should apply to location data. In the meantime, consumers are left to take Verizon, Sprint, T-Mobile, and AT&T's promises at face value.
- **I Sold My Data for Crypto. Here's How Much I Made** A new wave of companies is peddling an alluring message: Users should own their own data and get a cut of its value, instead of allowing it to be monetized by advertising companies and data brokers for free. Sign up for one of these apps and the buyers will contact you directly, offering cryptocurrency tokens in exchange for information like your bank transactions, medical history, or the fluctuations of your smart thermostat.
- **Get Ready for a Privacy Law Showdown in 2019** Companies like Amazon, Apple, Facebook, and Google are pushing hard for federal digital privacy legislation in 2019, and not quite out of the goodness of their hearts. Last summer, California's state legislature passed a groundbreaking privacy bill that is set to go into effect on January 1, 2020. Tech giants are now racing to supersede the law with more industry-friendly federal legislation. Even though Congress is divided politically, it looks like a deal could be reached. Reigning in Big Tech has become a bipartisan issue.

- **Your Smartphone Choice Could Determine Whether You Get a Loan**
  In Europe, some lenders are using passive signals, like what kind of phone you have, to determine whether you should qualify for a loan. Research from the National Bureau of Economic Research suggests those indicators can predict consumer behavior as accurately as traditional credit scores. But these factors aren't necessarily ones consumers are aware of or know to change.

- **The Wired Guide to Data Breaches**
  There's no such thing as perfect security, and it's impossible to safeguard against every potential data breach. But how worried should users be when they find out their personal information was leaked or stolen? To answer that question, it helps to know a little about the history of data breaches. Armed with context, consumers can determine whether they need to take extra precautions after a security incident happens.

- **What Does a Fair Algorithm Actually Look Like?**
  Lawmakers largely haven't decided what rights citizens should have when it comes to transparency in algorithmic decision-making. There isn't a "right to explanation" for how a machine came to a conclusion about your life. Some researchers are conceptualizing what such a right should might look like in the future.

Thanks to Ghostery, Mozilla, the Electronic Frontier Foundation, and Seth Stephens-Davidowitz for their help in creating this guide.

Last updated February 13, 2019.

Enjoyed this deep dive? Check out more WIRED Guides.

RELATED VIDEO

CULTURE

Are You the Most Average Person in America?

Do you get 8.7 hours of sleep every night? Spend a daily total of 30 minutes in the bathroom? Use Facebook for 40 minutes (and Tinder for 77 minutes) each day? If so, you just might be the most average person in the U.S.

#WIRED GUIDE #DATA #GOOGLE #PRIVACY

VIEW COMMENTS

MORE BUSINESS

GOING PRO

To Compete With Google, OpenAI Seeks Investors–and Profits

TOM SIMONITE

HYPERTEXT

30 Years On, Reports of the Web's Death Are Exaggerated

KLINT FINLEY

WIRED OPINION

The Huawei Case Is Part of a New US-China Cold War Over Tech

ZACHARY KARABELL

WIRED Q&A

Why the US Needs to Enforce Antitrust Laws

KLINT FINLEY

SXSW

This Big Facebook Critic Fears Tech's Business Model

ANDREA VALDEZ

POLITICS

The Last Place Big Tech Wants to Be Is on the Defense

ZACHARY KARABELL

GET OUR NEWSLETTER

WIRED's biggest stories delivered to your inbox.

Enter your email

→ SUBMIT

FOLLOW US ON PINTEREST

See what's inspiring us.

→ FOLLOW

SUBSCRIBE
ADVERTISE
SITE MAP
PRESS CENTER
FAQ
ACCESSIBILITY HELP
CUSTOMER CARE
CONTACT US
SECUREDROP
COUPONS
NEWSLETTER
WIRED STAFF
JOBS
RSS