Data Brokerage, the Sale of Individuals' Data, and Risks to Americans' Privacy, Personal Safety, and National Security

Written Testimony

Justin Sherman
Senior Fellow and Research Lead, Data Brokerage Project
Duke University Sanford School of Public Policy

U.S. House Committee on Energy and Commerce

Subcommittee on Oversight and Investigations

Hearing on "Who is Selling Your Data: A Critical Examination of the Role of Data Brokers in the Digital Economy"

April 19, 2023

Chair Griffith, Vice Chair Lesko, Ranking Member Castor, and distinguished members of the Subcommittee, I appreciate the opportunity to testify about data brokers and threats to Americans' privacy, to personal safety, and to national security.

I am a senior fellow at Duke University's Sanford School of Public Policy, where I lead our research project on the data brokerage ecosystem. We study the virtually unregulated industry and practice of data brokerage—the collection, inference, aggregation, analysis, buying, selling, and sharing of data on individuals—and its impacts on civil rights, consumer privacy, personal safety, and national security. I am also the founder and CEO of Global Cyber Strategies, a Washington, DC-based research and advisory firm, and a nonresident fellow at the Atlantic Council, where I work on cybersecurity, privacy, internet policy, and geopolitical issues.

Data brokerage is a threat to Americans' civil rights, consumers' privacy and well-being, and U.S. national security. The entire data brokerage ecosystem—from companies whose entire business model is data brokerage, to the thousands of other apps, advertisers, tech giants, and companies that collect, buy, sell, and share Americans' personal data—profits from unregulated surveillance of every American, particularly the most vulnerable. While I support a strong, comprehensive consumer privacy law, Congress should act now to regulate the data brokerage ecosystem.

There are three steps Congress should take now:

1. Strictly control the sale of data to foreign companies, citizens, and governments;
2. Ban the sale of data completely in some sensitive categories, such as with health and location data, and strictly control the sale of data in other categories;
3. Stop data brokers from circumventing those controls by "inferring" data.

In this written testimony, I describe:

- The data brokerage problem facing the United States, including the landscape of companies involved and our team's research into the data sold on Americans on the open market;[1]
- Data brokers' data streams, including the three main ways data brokers get data, the broken notion of consumers "consenting" to the sale of their data, and why "anonymization" is not a technically meaningful term but a marketing term, frequently used by data brokers to falsely imply data without names cannot be linked back to individuals;
- The harms and risks to Americans' privacy, civil rights, personal safety, and well-being and to U.S. national security—ranging from stalking and gendered violence, to the murder of a U.S. federal judge's son, to scams of elderly Americans and people with Alzheimer's, to the considerable risks to military servicemembers and U.S. national security;
- How data brokerage is a virtually unregulated practice and how the definition of data brokerage should encompass both third parties as well as first parties that collect and sell data on their own customers; and
- The steps Congress can take now to protect Americans' privacy and personal safety, as well as U.S. national security, from the risks posed by the data brokerage ecosystem.

To adequately address the power of Big Tech, the dangers of modern surveillance, and data threats to Americans' privacy and civil rights, U.S. national security, and democracy, we must focus on this entire data brokerage ecosystem.

The Data Brokerage Problem

The U.S. has a data brokerage problem. Today, and for the past several decades, hundreds and thousands of companies have surreptitiously collected data from public and private sources about each and every American. Oftentimes, they will use tools and techniques to "infer," or predict, additional data about Americans. These companies then repackage and resell that data on the open market, with very few controls. This is the data brokerage ecosystem, worth billions of dollars in the United States, and composed of everything from large, publicly traded companies like Experian and Oracle to smaller data brokers that hide from the public eye—and companies that quietly sell data on their customers on the side, just to make an additional profit.

Data brokerage is a virtually unregulated practice in the United States (except for a few, limited state laws and some narrowly targeted federal regulations discussed below). Brokered data is widely available; purchasable at low cost; often sold by brokers with little to no vetting; and can be used to profile, track, and target consumers, including people in marginalized communities, veterans, military servicemembers, government employees, first responders, elderly Americans, people with Alzheimer's, students, and teenagers. The customers for this data range from banks and other financial institutions, insurance and health insurance firms, companies doing market research and running advertisements, and law enforcement agencies buying data without warrants to predatory loan companies, criminal scammers, abusive and violent individuals, and, potentially,

[1] I would also refer the Subcommittee and readers to my previous testimony on the data brokerage ecosystem: Justin Sherman, "Data Brokerage and Threats to U.S. Privacy and Security," Written Testimony before the Senate Committee on Finance: Subcommittee on Fiscal Responsibility and Economic Growth, Hearing on "Promoting Competition, Growth, and Privacy Protection in the Technology Sector," December 7, 2021. https://www.finance.senate.gov/imo/media/doc/Written%20Testimony%20-%20Justin%20Sherman.pdf.

foreign actors. Because the data brokers selling this data appear to do very little customer vetting, the data they sell is often accessible to nearly anyone.

- Military servicemembers: Data brokers gather, package, and advertise highly sensitive data on current and former members of the U.S. military, which poses privacy and safety risks to servicemembers. In 2018, the Federal Trade Commission (FTC) successfully filed an injunction against a "lead generation" data broker that created websites falsely advertised to prospective servicemembers as military recruitment sites (such as *army.com*, *armyreserves.com*, *armyenlist.com*, *navyenlist.com*, and *marinesenlist.com*), when in reality the company was not doing military recruitment but collecting data on those prospective servicemembers and quietly selling it to post-secondary schools.[2] Tens of thousands of potential recruits visited the sites each month and were asked to input data such as their names, home addresses, email addresses, phone numbers, dates of birth, and educational histories.[3] In other cases, data brokers do not necessarily have to use that level of deception but can collect data legally, due to a lack of regulation. For instance, in 2007, the *New York Times* uncovered a mass-fraud scheme targeting World War II veterans, among others, such as a 92-year-old Army veteran whose name a list broker put in a dataset and sold to telemarketing criminals.[4] Many other data brokers advertise data points on hundreds of thousands of current and former U.S. military personnel—as well as their families and the presence of children in the home.[5] Foreign governments could acquire this data to profile, track, and target military personnel and their families and otherwise undermine U.S. national security. The Chinese government's 2015 hack of the Office of Personnel Management was one of the most damaging data breaches the federal government has suffered—yet there is no need for the Chinese government or any other foreign intelligence agency to even hack many U.S. databases when so much data can be legally purchased from U.S. data brokers, which appear to do very little customer vetting.
- Survivors of domestic and gendered violence: Data brokers known as "people search websites" aggregate millions of Americans' public records and make them available for search and sale online. Abusive individuals have used this data—including highly sensitive information on individuals' addresses, whereabouts, property filings, contact details, and family members—to hunt down and stalk, harass, intimidate, and even murder other individuals, predominantly women and members of the LGBTQ+ community.[6] The

[2] Complaint for Civil Penalties, Permanent Injunction, and Other Equitable Relief. *United States of America v. Sunkey Publishing, Inc.* (2018). https://www.ftc.gov/system/files/documents/cases/sunkey_filed_complaint.pdf. See also: Stipulated Final Order for Permanent Injunction and Civil Penalty Judgment. *United States of America v. Sunkey Publishing, Inc.* (2018). https://www.ftc.gov/system/files/documents/cases/sunkey_proposed_order.pdf.
[3] Ibid., 8. Also see, on lead generation: U.S. Federal Trade Commission. *"Follow the Lead" Workshop*. Washington, D.C.: Federal Trade Commission, September 2016. https://www.ftc.gov/system/files/documents/reports/staff-perspective-follow-lead/staff_perspective_follow_the_lead_workshop.pdf.
[4] Charles Duhigg, "Bilking the Elderly, With a Corporate Assist," *The New York Times*, May 20, 2007, https://www.nytimes.com/2007/05/20/business/20tele.html.
[5] See, e.g., Justin Sherman, "Data Brokers Are Advertising Data on U.S. Military Personnel," *Lawfare*, August 23, 2021, https://www.lawfareblog.com/data-brokers-are-advertising-data-us-military-personnel.
[6] This goes back decades. See, e.g., Supreme Court of New Hampshire. *Helen Remsburg, Administratrix of the Estate of Amy Lynn Boyer, v. Docusearch, Inc., d/b/a Docusearch.Com & a* (2003). Also see: National Network to End Domestic Violence, "People Searches & Data Brokers," last accessed April 10, 2023, https://nnedv.org/mdocs-posts/people-searches-data-brokers/; Kaveh Waddell, "How FamilyTreeNow Makes Stalking Easy," *The Atlantic*, January 17, 2017, https://www.theatlantic.com/technology/archive/2017/01/the-webs-many-search-engines-for-

Subcommittee's 2006 hearing on data brokers highlighted this very problem and risk.[7] Still today, though, there is little in U.S. law stopping data brokers from collecting, publishing, and selling this data on victims and survivors of intimate partner violence. Although these brokers often point to opt-out functionalities on their websites, they still expose Americans' data by default, usually without those Americans knowing—and for those who do attempt to remove their information from people search websites, the process is often onerous and ineffective.[8] Further, many states and cities have implemented "address confidentiality" programs, which allow individuals to hide their addresses from public records. These public programs are an important step towards mitigating the risks of stalking and gendered violence. But they are still woefully insufficient, because they leave the underlying practice of selling Americans' home addresses and other data unregulated—and often require women and other individuals to meet an unreasonably high burden of proof (e.g., providing evidence to the state or local government that stalking has already occurred) in order to protect the privacy and physical safety of themselves and their families.

- Elderly Americans and people with Alzheimer's: Data brokers sell data on elderly Americans and people with Alzheimer's, dementia, and other brain health conditions. In 2020 and 2021, the Department of Justice charged three data brokers—Epsilon, Macromark, and KBM—with conspiracy to commit mail and wire fraud for knowingly selling, for roughly a decade each, lists of vulnerable Americans, including elderly Americans and people with Alzheimer's, to criminal scammers.[9] The criminal scammers then used that brokered data to steal millions of dollars from these people.[10] Each data broker sold this data knowingly, on some of our country's most vulnerable people, because they profited off the sale of the data. For instance, as the Justice Department described in its court filing against Epsilon, "the Employees were familiar with the clients' practices, as well as their deceptive solicitations"; "worked to develop and increase business with clients engaged in fraud despite receiving notice that those and similar clients had been arrested, charged with crimes, convicted, and otherwise were subject to law enforcement actions"; and "engaged in this conduct, in part, to benefit Epsilon, to enrich themselves through sales-based compensation, and to enable the fraudulent clients to solicit new customers."[11] As Alistair Simmons, a student researcher on our team, noticed when reviewing these

---

your-personal-information/513323/; Adi Robertson, "Senators ask FTC to fight stalkers exploiting people search sites," *The Verge*, March 4, 2021, https://www.theverge.com/2021/3/4/22313613/ftc-senator-letter-stalking-abuse-data-broker-people-search-sites.

[7] Committee on Energy and Commerce: Subcommittee on Oversight and Investigations, Hearing on "Internet Data Brokers: Who Has Access to Your Private Records?," June 21, June 22, and September 29, 2006, https://www.govinfo.gov/content/pkg/CHRG-109hhrg31363/pdf/CHRG-109hhrg31363.pdf, 6, 11.

[8] For example, people search websites may comply with a request to remove an individual's profile only for the website to repopulate that person's profile a day later—or to remove only one webpage with that person's information while leaving their information accessible via other profiles gathered, packaged, and sold on that person's family. See, e.g., Mara Hvistendahl, "I Tried to Get My Name off People-Search Sites. It Was Nearly Impossible," Consumer Reports, August 20, 2020, https://www.consumerreports.org/personal-information/i-tried-to-get-my-name-off-peoplesearch-sites-it-was-nearly--a0741114794/.

[9] *United States of America v. Epsilon Data Management, LLC* (2021). https://www.justice.gov/opa/press-release/file/1360881/download; *United States of America v. Macromark, Inc.* (2020). https://www.justice.gov/civil/case/file/1326376/download; *United States of America v. KBM Group, LLC* (2021). https://www.justice.gov/civil/case/file/1404091/download.

[10] Ibid.

[11] *United States of America v. Epsilon Data Management, LLC* (2021). 4.

documents, the data brokers took note of vulnerable Americans who were scammed and then identified them for additional, future scamming—as we put it, "recycling victims' information to target them again."[12] Despite the Justice Department's actions against these three companies,[13] we have already found additional data brokers advertising lists of elderly Americans and people who are suffering from Alzheimer's. Scammers have already stolen millions of dollars from vulnerable consumers using this kind of brokered data, and the risks persist without any new laws or regulations in place.

- Americans with mental health conditions: Data brokers collect and sell data on Americans' mental health conditions, including data on people suffering from depression and anxiety and the prescriptions they take. Recently, we published a report by Joanne Kim, a former student researcher on our team, detailing a number of data brokers advertising both aggregated and individually linked data on Americans with depression, attention disorder, insomnia, anxiety, ADHD, bipolar disorder, panic disorder, Post-Traumatic Stress Disorder (PTSD), and more.[14] In some cases, the advertised datasets also included data on individuals' races, ethnicities, ages, genders, zip codes, religions, number of children in the home, marital statuses, net worth, credit scores, dates of birth, or whether the person was a single parent.[15] She also found a data broker advertising data on Americans with cancer and on people who had suffered strokes.[16] Many other data brokers advertise this information, which is highly invasive. It also concerns people who already face numerous barriers and significant stigma in accessing care. This data could be used in predatory marketing activities or abused by scammers. But gaps in the Health Insurance Portability and Accountability Act (HIPAA) mean that numerous apps and websites outside the scope of HIPAA's narrow "covered entities" are entirely free to legally collect, aggregate, and sell, license, and share Americans' health information on the open market.[17]

These examples of data collection, inference, aggregation, and sale are extremely invasive and pose enormous risks to vulnerable Americans, whether former veterans seeking benefits, survivors of gendered violence seeking to protect themselves and their families, or older Americans suffering from Alzheimer's who are targeted by predatory scammers. Yet, these examples hardly stand alone in the kinds of data collected and sold by data brokers seeking to make a profit.

---

[12] Alistair Simmons and Justin Sherman, "Data Brokers, Elder Fraud, and Justice Department Investigations," *Lawfare*, July 25, 2022, https://www.lawfareblog.com/data-brokers-elder-fraud-and-justice-department-investigations.

[13] The companies were also required to implement a variety of compliance and control measures around their data practices, including filing reports to the Justice Department, but there is no external oversight of the three data brokers' practices beyond those reports (which are also not available to the public).

[14] Joanne Kim, *Data Brokers and the Sale of Americans' Mental Health Data* (Durham: Duke University Sanford School of Public Policy, February 2023), https://techpolicy.sanford.duke.edu/data-brokers-and-the-sale-of-americans-mental-health-data/.

[15] Ibid., 4-5.

[16] Ibid., 5.

[17] See, e.g., Kirk J. Nahra, "Moving Toward a New Health Care Privacy Paradigm," Wiley Rein LLP, November 2014, https://www.healthit.gov/sites/default/files/facas/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf; Alexis Guadarrama, "Mind the Gap: Addressing Gaps in HIPAA Coverage in the Mobile Health Apps Industry," *Houston Law Review* 55, no. 4 (2018), https://houstonlawreview.org/article/3876-mind-the-gap-addressing-gaps-in-hipaa-coverage-in-the-mobile-health-apps-industry; Justin Sherman, "GoodRx, Health Data Brokerage, and the Limits of HIPAA," *Lawfare*, March 6, 2023, https://www.lawfareblog.com/goodrx-health-data-brokerage-and-limits-hipaa.

Our research at Duke University has found data brokers advertising sensitive data on hundreds of millions of Americans, including data on individuals' demographic characteristics, political preferences and beliefs, home addresses and geolocations, and health and mental health conditions.[18] The data brokerage industry collects, infers, and sells data on your race, religion, gender, sexual orientation, marital status, income level, credit rating, children, home address, geolocation, political preferences, health conditions, mental health conditions, device usage, and much, much more. Some of this brokered data is more "aggregated," such as providing population-level statistics without supplying buyers with underlying data. (Although, the data brokers are often still collecting the raw data and just choosing to not provide it to buyers.) Other times, this brokered data is linked to individuals by name or by another persistent identifier, such as a mobile advertising ID. Both of these cases raise privacy questions about the protection of Americans' data—and related issues like civil rights, personal safety, and national security.

Data brokers also sell data on Americans in packages, or what many call "marketing segments." Brokers' pitch is that buyers can then use the compiled and pre-packaged datasets to profile or target individuals. We have found data brokers advertising packages of data for sale on the open market on students, teenagers, active-duty U.S. military personnel, veterans, U.S. government employees, elderly Americans, people with Alzheimer's, adults with cancer, individuals suffering from depression, and more. Research from the World Privacy Forum in 2013 found a data broker advertising the packaged home addresses of police officers and advertising a dataset that it described as a "Rape Sufferers List," for just $0.079 per name.[19] Other, real dataset titles include "Rural and Barely Making It," "Ethnic Second-City Strugglers," "Retiring on Empty: Singles," "Tough Start: Young Single Parents," "Credit Crunched: City Families," "viewership-gay," "African American," "Jewish," "working class," "unlikely voters," and "seeking medical care."[20] The packaging of this data creates additional privacy risks to individuals because the brokers have already gone through the work of compiling and sorting the data—and making it ready to use for profiling, tracking, or targeting. Information can also be gleaned from datasets even if that information is not formally included in the dataset as an explicit characteristic.

There are additionally cases in which data brokers do not provide buyers with the underlying datasets per se but instead offer services based on the underlying data they have gathered and inferred. For instance, a data broker might offer identity verification services, wherein an

---

[18] See, e.g., Justin Sherman, *Data Brokers and Sensitive Data on U.S. Individuals* (Durham: Duke University Sanford School of Public Policy, August 2021), https://sites.sanford.duke.edu/techpolicy/report-data-brokers-and-sensitive-data-on-u-s-individuals/; Kim, *Data Brokers and the Sale of Americans' Mental Health Data*.

[19] Pam Dixon, Written Testimony before the Senate Committee on Commerce, Science, and Transportation, Hearing on "What Information Do Data Brokers Have on Consumers, and How Do They Use It?," December 18, 2013, http://www.worldprivacyforum.org/wp-content/uploads/2013/12/WPF_PamDixon_CongressionalTestimony_DataBrokers_2013_fs.pdf.

[20] U.S. Senate Committee on Commerce, Science, and Transportation. *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes*. Washington, D.C.: Senate Committee on Commerce, Science, and Transportation, December 18, 2013. https://www.commerce.senate.gov/services/files/0d2b3642-6221-4888-a631-08f2f255b577. ii; U.S. Federal Trade Commission. *A Look At What ISPs Know About You: Examining the Privacy Practices of Six Major Internet Service Providers*. Washington, D.C.: Federal Trade Commission, October 21, 2021. https://www.ftc.gov/system/files/documents/reports/look-what-isps-know-about-you-examining-privacy-practices-six-major-internet-service-providers/p195402_isp_6b_staff_report.pdf. 22.

individual can submit a query to the data broker for verification (e.g., matching a person's name to a last known address) but will not receive access to the entire database held by the broker.

In compliance with our university research ethics processes, our Duke data brokerage team has also purchased data from data brokers to understand what data these companies actually sell; the level of identifiability of the data; the controls data brokers may or may not impose and enforce around the sale and use of the data; the costs of different kinds and sizes of datasets; and the risks to individuals, communities, and society (among others). This data is often disturbingly low-cost. In a forthcoming study on military data, for example, we purchased individually identified data on members of the U.S. military for as cheap as $0.125 per servicemember. Buyers with more funding, such as law enforcement agencies, health insurance companies, and potentially foreign actors, can spend hundreds of thousands of dollars a year on data broker subscription services, where brokers provide licensed access to enormous databases.

We have found little evidence that data brokers implement robust controls, such as know-your-customer identity verification controls, technical controls to limit buyers' misuse of purchased data, and detective controls to ensure that purchasers are not lying about their intended data use cases. For instance, in our team's February 2023 report by Joanne Kim on the sale of mental health data, a data broker that she contacted told her that mental health data was too sensitive to discuss without first conducting a background check—but then continued sending her "data samples" anyway, providing the researcher with a subset of a real dataset on Americans' mental health conditions to entice her to purchase the full dataset.[21] In the aforementioned Justice Department action against the three data brokers selling data on vulnerable Americans to scammers, there was just one documented case where a data broker (KBM) had a control in place to prevent this kind of insidious data sale and use.[22] One of KBM's employees determined that a prospective buyer was a criminal scammer and blocked the sale of data, but other employees overrode the block.[23] The data broker then sold data on Americans to the scammer anyway.[24]

The only controls we have seen relatively consistently in our team's research are data brokers (i) ensuring the buyer's credit card will work and (ii) often requiring buyers to sign nondisclosure agreements (NDAs) to limit their disclosure of the fact that they purchased data on U.S. individuals, which data broker they purchased data from, how much the data cost, and/or other information. This increases the opacity of the data brokerage ecosystem and makes it harder for consumers to understand that their data is being collected and sold on the open market. It likewise makes it harder for legislators and regulators to understand how these companies profit off the sale of Americans' data. In our case, NDAs can prevent researchers from publishing the full details of their findings and can limit their ability to inform academic, public, and policy discourse about data, privacy, and national security risks. NDAs may not just enter the picture when individuals buy data from brokers. We have interacted with multiple data brokers that have asked us to sign NDAs before they even speak with us about the data they sell. Rather than vetting customers, data brokers clearly seek to use these controls to keep their activities quiet and uninterrupted.

[21] Kim, *Data Brokers and the Sale of Americans' Mental Health Data*.
[22] *United States of America v. KBM Group, LLC* (2021). B-6-B-7.
[23] Ibid.
[24] Ibid., B-7.

On top of conducting this research through data purchasing, we have also spoken with several data brokers about their data sale processes. We have not found any data brokers willing to provide documentation to support their claims that they place controls around the sale of data on Americans, from health data to location data to data on elderly Americans, teenagers, and military servicemembers. This remains an ongoing point of inquiry for our team. Again, however, the publicly known instances of harm associated with data brokerage raise the question of whether most brokers have any controls in place at all (or if there are controls on paper, if they are meaningfully enforced). In many cases, as with the sale of Americans' health data on the open market, or the sale of data on teenagers[25] or U.S. military servicemembers, it is also worth considering whether the introduction of controls would appropriately mitigate the risks of surreptitiously collecting, inferring, aggregating, and selling the data to begin with.

Data Brokers' Data Streams, "Consent," and "Anonymization"

There are, in my own classification, generally three main ways that data brokers acquire information on Americans—so they can subsequently package and sell that data:

1. *Directly*, including data brokers buying up companies and services that gather data directly (such as apps and websites) and paying app developers to include the data broker's software development kit (SDK) in the developer's app, after which the developer can just let the app run while the broker "sits" within the app and siphons data on users;
2. *Indirectly*, including data brokers scraping public records (e.g., property records, voting records, etc.), gathering data from real-time bidding networks for online ads, and paying app developers to transmit data to data brokers via server-to-server transfers, once the app developers have collected information on app users and stored it on their own servers; and
3. *"Inference,"* or prediction—data brokers using algorithms and other techniques to make predictions about individuals' characteristics, such as by using purchase information and home ZIP code to predict household income, location data from smartphones to predict religion (e.g., visits to churches, mosques, synagogues), or app installations on a phone to predict sexual orientation (e.g., the presence of LGBTQ+ dating apps).

Many data brokers, when asked about their data practices, will claim that Americans "consent" to the packaging and sale of their data. In particular, brokers will often point out that many apps, websites, and other companies collecting data will include clauses in their privacy policies and terms of service that refer to the possibility of that first-party collector sharing data on consumers. This is a bad-faith and patently ridiculous argument. Most consumers do not read privacy policies.

Among others, a 2019 Pew Research Center survey found that 81% of Americans agree to privacy policies at least monthly, but that only 9% of Americans say they always read a privacy policy

[25] Take, for instance, the case of the NJ-based data broker ALC Inc. that was discovered in 2019, by the *Philadelphia Inquirer*, to be advertising data on 1.2 million students aged 14-17, "including their names, addresses, high schools, and hobbies" as well as data on their "parents' names, household incomes, and ethnicity," among others. ALC, in filing its entry in the Vermont data broker registry, had previously asserted that it "has no knowledge nor do we allow the collection or use of data on any persons under the age of 18." See: Christian Hetrick, "N.J. data broker tried to sell personal info on a million kids but didn't tell state officials," *Philadelphia Inquirer*, March 19, 2019, https://www.inquirer.com/business/technology/alc-princeton-data-broker-personal-info-million-kids-vermont-law-20190319.html.

before agreeing to a company's terms and conditions.[26] A 2021 survey by Security.org found that 37% of people skim the documents, 35% don't read them at all, and 16% search for and read a few key parts of the documents;[27] only 11% say they fully read privacy policies before agreeing.[28] The information asymmetry facing consumers is also huge: a 2008 study calculated that if consumers wanted to read the privacy policies for the services they use, it would take each person an average of 244 hours a year.[29] As the authors put it, "the national opportunity cost for just the time to read policies is on the order of $781 billion" (and that was in 2008).[30] The *New York Times*, to give another example, examined 150 companies' privacy policies in 2019 and found that they were difficult to read (an "incomprehensible disaster," the article's title said), with many even more complex than the texts that doctors, lawyers, and other professionals must understand in their jobs.[31] I spend much of my time researching data privacy issues and data brokerage, and I do not have time to read every single line in every single privacy policy of every single application I use.

Further, this idea of "consent" is not fully informed. Most consumers do not understand and cannot be reasonably expected to understand how the data brokerage ecosystem operates. There are many companies gathering data on consumers through websites, mobile apps, and other products and services and then—with very little or no transparency—selling and sharing the data with other actors. An individual downloading a weather app with a built-in GPS feature has no reasonable expectation the app might share their location data with a data broker which then sells it to advertisers or to law enforcement. (The FTC took an enforcement action in this vein in 2013 against flashlight app Brightest Flashlight Free, which indicated to users that location data would only be used internally but in reality shared and sold the data with third parties.[32]) Similarly, teenagers using a mental health app have no reasonable expectation that the app could, in many cases, entirely legally gather and sell data on teenagers suffering from depression or anxiety. Moreover, even if consumers did understand how the data brokerage ecosystem operates, that is distinct from fully understanding its harms. The "inference" means of acquiring data on Americans underscores this further: data brokers may attain information like religion, income level, or sexual orientation that consumers scarcely, if ever, input into a form but which data brokers are predicting and then selling, with serious risks to consumers.

The term "consent" also suggests it is freely given, but this is not the case with data brokers gathering and selling Americans' data. Even if consumers did fully understand what was

---

[26] Brooke Auxier et. al, "Americans' attitudes and experiences with privacy policies and laws," Pew Research Center, November 15, 2019, https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/.
[27] Eric Griffith, "Everyone Wants Data Privacy, But No One Reads Privacy Agreements," *PC Magazine*, April 19, 2021, https://www.pcmag.com/news/everyone-wants-data-privacy-but-no-one-reads-privacy-agreements.
[28] Ibid.
[29] Aleecia M. McDonald and Lorrie Faith Cranor, "The Cost of Reading Privacy Policies," *I/S: A Journal of Law and Policy for the Information Society* (2008), https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/, 17.
[30] Ibid., 2.
[31] Kevin Litman-Navarro, "We Read 150 Privacy Policies. They Were an Incomprehensible Disaster," *The New York Times*, June 12, 2019, https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.
[32] Cecilia Kang, "Flashlight app kept users in the dark about sharing location data: FTC," *The Washington Post*, December 5, 2013, https://www.washingtonpost.com/business/technology/flashlight-app-kept-users-in-the-dark-about-sharing-location-data-ftc/2013/12/05/1be26fa6-5dc7-11e3-be07-006c776266ed_story.html.

happening in the data brokerage ecosystem and how it could or does harm them, the focus on individuals distracts from the systemic problems at play. There are an immense amount of information and financial asymmetries stacked against consumers impacted by data brokerage. People are regularly forced to interact with data brokers, whether to get a new credit card, put in a deposit for an apartment, or apply for a loan; many insurance companies buy data on consumers, too. Whether or not individuals "consent" to data brokerage is not a question limited to merely using an app that has a privacy policy somewhere if their not-consenting means they cannot access housing, money, employment opportunities, and other essentials.

Another frequent data broker argument is that the data sold is "anonymized." This, too, is a bad-faith argument. Anonymization is not a technically meaningful term; it is a marketing term. There are indeed statistical techniques that can be used to provide more protection to individuals' data in datasets, such as with differential privacy.[33] These are important measures, and it is vital to invest in the research and development of privacy-preserving techniques for individual data points and datasets, for cases in which we as a society have collectively deemed a certain data use to be appropriate. However, the claim that removing a name from a dataset makes the data "anonymized," which is often what data brokers and other companies refer to when using this term, falsely implies that it cannot be relinked to an individual. Decades of computer science research decisively show otherwise.[34] One recent study found that with only 15 specific demographic attributes, it would be possible to "re-identify" 99.98% of Americans in a dataset.[35] This reality makes the "anonymization" argument itself misleading and wrong.

Moreover, claims of "anonymization" obscure the fact that many data brokers are selling datasets that *do* include individuals' names. We ourselves have purchased some of this data, in compliance with our university research ethics processes. It would be all too easy for an individual or organization with malicious intent and not bound by university research ethics to use this data in harmful ways. For example, we have purchased individually identified datasets from data brokers, which included consumers' names, contact information, and home addresses, where the broker (i) did not verify our identity and (ii) asked once via email if we planned to contact individuals in the dataset but did not follow up or perform any kind of verification that our reply of "no" was indeed

[33] See, e.g., "Differential Privacy," Harvard University Privacy Tools Project, accessed April 15, 2023, https://privacytools.seas.harvard.edu/differential-privacy.

[34] See, e.g., Latanya Sweeney, "Weaving Technology and Policy Together to Maintain Confidentiality," *Journal of Law, Medicine & Ethics* 25, nos. 2 & 3 (1997): 98-110, https://dataprivacylab.org/dataprivacy/projects/law/law1.html; Latanya Sweeney, "Simple Demographics Often Identify People Uniquely," Carnegie Mellon University, 2000, https://dataprivacylab.org/projects/identifiability/index.html; Michael Barbaro and Tom Zeller Jr., "A Face Is Exposed for AOL Searcher No. 4417749," *The New York Times*, August 9, 2006, https://www.nytimes.com/2006/08/09/technology/09aol.html; Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets," University of Texas-Austin, 2008, https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf; Marie Douriez et al., "Anonymizing NYC Taxi Data: Does It Matter?" *2016 IEEE International Conference on Data Science and Advanced Analytics* (2016), https://ieeexplore.ieee.org/document/7796899; Yongqi Dong et al., "Revealing New York taxi drivers' operation patterns focusing on the revenue aspect," *2016 12th World Congress on Intelligent Control and Automation* (2016), https://ieeexplore.ieee.org/document/7578771.

[35] Luc Rocher, Julien M. Hendrickx, and Yves-Alexandre de Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature Communications* 10, no. 3069 (2019), https://www.nature.com/articles/s41467-019-10933-3.

true. Broadly, it is also worth pointing out that a core part of the data brokerage business model is allowing people and companies to profile, track, and/or target individuals. The notion that data brokers would not want the ability to identify specific Americans within datasets defies a core of their business pitch in the first place.

The Harms of Data Brokerage

Data brokers profit from collecting, inferring, aggregating, analyzing, buying, selling, and sharing data on Americans. This has contributed to and resulted in a number of known harms.

For decades, "people search websites" have enabled domestic and intimate partner violence by scraping public records and posting the data in them for search and sale online. Abusive individuals have bought or obtained information on people to hunt down and stalk, harass, intimidate, harm, and even murder other people, largely impacting women and members of the LGBTQ+ community. In 2020, an individual purchased information online about New Jersey federal judge Esther Salas and her family and then went to her home, shot her husband, and shot and killed her 20-year-old son.[36] Criminal scammers have bought information from data brokers—in some cases, where the brokers are fully aware their clients are scammers—to steal from elderly Americans, people with Alzheimer's and other cognitive health issues, and other vulnerable individuals.[37] For example, on top of the aforementioned cases, in 2020 the Justice Department charged multiple lead brokers as part of an indictment of a $300 million nationwide telemarketing fraud scheme; the lead brokers "bought and sold lead lists of victim-consumers to fraudulent magazine sales companies," where "many of the consumers on this list were elderly and susceptible to fraudulent and deceptive sales tactics" and were sold away for as little as $10 or $15 per name.[38]

Health insurance companies have purchased data from data brokers—including data on race, education level, marital status, net worth, social media posts, payments of bills, and more—to profile consumers and predict the costs of providing healthcare to those people.[39] Scammers have bought payday loan applicants' financial information, which at least one data broker illegally sold, to steal millions of dollars from those people.[40] Financial firms have used brokered data to market products to consumers that "limit or obscure their access to loans, credit, and financial services."[41] GPS location data companies have secretly tracked citizens attending protests and demonstrations

[36] Esther Salas, "My Son Was Killed Because I'm a Federal Judge," *The New York Times*, December 8, 2020, https://www.nytimes.com/2020/12/08/opinion/esther-salas-murder-federal-judges.html.
[37] See, e.g., the aforementioned cases: *United States of America v. Epsilon Data Management, LLC* (2021); *United States of America v. Macromark, Inc.* (2020); *United States of America v. KBM Group, LLC* (2021).
[38] U.S. Department of Justice, "Sixty Defendants Charged in $300 Million Nationwide Telemarketing Fraud Scheme," justice.gov, October 28, 2020, https://www.justice.gov/usao-mn/pr/sixty-defendants-charged-300-million-nationwide-telemarketing-fraud-scheme.
[39] Marshall Allen, "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates," NPR, July 17, 2018, https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.
[40] U.S. Federal Trade Commission, "FTC Charges Data Brokers with Helping Scammers Take More Than $7 Million from Consumers' Accounts," FTC.gov, August 12, 2015, https://www.ftc.gov/news-events/news/press-releases/2015/08/ftc-charges-data-brokers-helping-scammer-take-more-7-million-consumers-accounts.
[41] Pam Dixon, "Data Brokers, Privacy, and the Fair Credit Reporting Act." Testimony before the U.S. Senate Committee on Banking, Housing, and Urban Affairs. June 11, 2019. https://www.banking.senate.gov/imo/media/doc/Dixon%20Testimony%206-11-19.pdf, 1.

and identified their ages, genders, ethnicities, and other sensitive demographic characteristics—all of which they can legally sell.[42]

It's not just companies; individuals can also use brokered location data to inflict harm on other, specific people and populations of Americans. In 2019, the *New York Times* obtained a dataset from a location data company containing over 50 billion location pings from more than 12 million Americans' phones.[43] The journalists were able to identify individuals visiting the Playboy Mansion overnight and people traveling to celebrities' estates—all the way to "military officials with security clearances as they drove home at night" and "law enforcement officers as they took their kids to school."[44] They were able to infer additional data from the location dataset, including signs of failing marriages, indications of drug addiction, and visits to psychological facilities.[45] While the journalists were not doing this work for the purposes of harming the individuals, it would be all too easy for an individual with malicious intentions to similarly track Americans through brokered datasets. In another case, a nonprofit organization purchased location data covering 2018-2021 that originated from multiple gay dating and hookup apps, such as Grindr, and sifted through that location data to identify, track, and out a closeted priest.[46] Location data is so sensitive because (i) it allows an individual or organization to follow a specific person, (ii) physical movements over time are highly unique to individuals, and (iii) location data enables an individual or organization to derive additional information about a person based on their movements.

Law enforcement and security agencies have purchased data broker data on U.S. citizens, ranging from home utility data to real-time locations, without warrants, public disclosure, and robust oversight.[47] This harms Americans and society through law enforcement effectively buying its way around important, democratic controls on government overreach and warrantless monitoring of U.S. persons. The data law enforcement and other customers use may also not even be updated,

[42] Zak Doffman, "Black Lives Matter: U.S. Protesters Tracked By Secretive Phone Location Technology," *Forbes*, June 26, 2020, https://www.forbes.com/sites/zakdoffman/2020/06/26/secretive-phone-tracking-company-publishes-location-data-on-black-lives-matter-protesters/.
[43] Stuart A. Thompson and Charlie Warzel, "Twelve Million Phones, One Dataset, Zero Privacy," *The New York Times*, December 19, 2019, https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html.
[44] Ibid.
[45] Ibid.
[46] Michelle Boorstein and Heather Kelly, "Catholic group spent millions on app data that tracked gay priests," *The Washington Post*, March 9, 2023, https://www.washingtonpost.com/dc-md-va/2023/03/09/catholics-gay-priests-grindr-data-bishops/.
[47] See, e.g., Nina Wang et al., *American Dragnet: Data-Driven Deportation in the 21st Century* (Washington, D.C.: Georgetown Law Center on Privacy & Technology, May 2022), https://americandragnet.org; Sharon Bradford Franklin, Greg Nojeim, and Dhanaraj Thakur, *Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers* (Washington, D.C.: Center for Democracy & Technology, December 2021), https://cdt.org/insights/report-legal-loopholes-and-data-for-dollars-how-law-enforcement-and-intelligence-agencies-are-buying-your-data-from-brokers/; Drew Harwell, "ICE investigators used a private utility database covering millions to pursue immigration violations," *The Washington Post*, February 26, 2021, https://www.washingtonpost.com/technology/2021/02/26/ice-private-utility-data/; Joseph Cox, "How an ICE Contractor Tracks Phones Around the World," *VICE*, December 3, 2020, https://www.vice.com/en/article/epdpdm/ice-dhs-fbi-location-data-venntel-apps.

complete, or accurate.[48] Law enforcement could then be carrying out investigations or performing other actions based on inaccurate information, which furthers the harm already inflicted on Americans and society. The list of known harms associated with the virtually unregulated data brokerage ecosystem goes on. Research, investigative reporting, criminal prosecutions, oversight investigations, and regulatory actions continue to further expose harmful uses of this data.

The potential harms are also numerous. Domestic extremists could acquire real-time GPS location data or home address information to target politicians and judges at home. Foreign governments, such as the Chinese government, could acquire Americans' data to run disinformation campaigns, uncover spies, blackmail U.S. government employees, and conduct other kinds of intelligence and military operations. Criminal organizations will continue purchasing this data to run scams and phishing campaigns.[49] Individuals will continue using address, whereabout, and GPS data to stalk and commit violence against fellow citizens. Companies will continue buying data on consumers and then make decisions and target advertisements based on sensitive demographic characteristics like race, ethnicity, gender, sexual orientation, religion, income level, family structure, political affiliation, and immigration status. Employers could also learn sensitive data about their own employees, ranging from health conditions to political affiliation to lifestyle information, that workers have not freely and knowingly chosen to disclose. Not to mention, threat actors can simply hack into the data brokers, online advertising firms, and other entities housing this highly sensitive data that has already been pre-packaged. When data brokers fail to protect and secure their datasets, exemplified by data broker Equifax's inadequate security leading to a major hack by the Chinese military in 2017, it further creates risks to U.S. individuals and to society broadly.[50]

The harms and risks affect every American—of every income level, region, and political stripe—and they fall hardest on the most vulnerable U.S. individuals and populations.

The Regulatory Gap

Data brokerage is mostly unregulated. The few privacy laws the U.S. has enacted are focused on how some entities in a few select industries or sectors use specific kinds of data.

For example, HIPAA applies only to certain covered health entities, like hospitals and primary healthcare providers, and does not apply to mobile health apps, social media companies, online advertisers, data brokers, and many other kinds of corporate actors that are neither covered entities nor business associates of those entities. These organizations outside the narrow scope of HIPAA are therefore free to legally gather, buy, package, sell, and share Americans' health-related data—

[48] See, e.g., United States District Court. Central District of California. *Gerardo Gonzalez et al. vs. Immigration and Customs Enforcement et al.* (2019). https://www.courthousenews.com/wp-content/uploads/2019/09/Gonzalez.v.ICE_.detainer.final_.order_.9.27.pdf.

[49] See, e.g., U.S. Federal Trade Commission, "FTC Charges Data Brokers with Helping Scammer Take More Than $7 Million from Consumers' Accounts," FTC.gov, August 12, 2015, https://www.ftc.gov/news-events/press-releases/2015/08/ftc-charges-data-brokers-helping-scammer-take-more-7-million.

[50] U.S. Department of Justice, "Chinese Military Personnel Charged with Computer Fraud, Economic Espionage, and Wire Fraud for Hacking into Credit Reporting Agency Equifax," justice.gov, February 10, 2020, https://www.justice.gov/opa/pr/chinese-military-personnel-charged-computer-fraud-economic-espionage-and-wire-fraud-hacking; Justin Sherman, "Data brokers and data breaches," Duke University Sanford School of Public Policy, September 27, 2022, https://techpolicy.sanford.duke.edu/blogroll/data-brokers-and-data-breaches/.

and they do, such as whether people have prescriptions for antidepressants, are suffering from cancer, or are believed to be pregnant.[51] (Some companies also falsely hide under the cover of HIPAA to suggest a degree of data privacy protection: in the FTC's recent enforcement action against GoodRx, for example, the telehealth and prescription drug company was falsely portraying itself as covered by HIPAA (it is not), displaying a HIPAA certification seal on its website (which is not real), and quietly sharing millions of consumers' health data, including health conditions and prescription information, with third parties.[52]) The Family Educational Rights and Privacy Act (FERPA) is another example. FERPA governs covered educational institutions' use and disclosure of students' data—but its narrow scope allows many other actors, including those brokering data, to sell information about students with virtually no restrictions. The Children's Online Privacy Protection Act (COPPA), for its part, places important protections around the collection and use of data from children under 13, but it does not regulate the collection and use of data on teenagers, including minors under 18. It also does not clearly prevent the sale of all data on minors.

At the state level, the California and Vermont data broker laws do not place any restrictions on the collection, aggregation, analysis, packaging, buying, selling, and sharing of consumer data. They are registry laws, meaning they primarily force companies that fit a narrow definition of a "data broker" to submit some information to the state. (Even within that limited scope, registry information sometimes appears outdated, duplicated, or with broken links.) Some state privacy bills, like a data broker registry bill in Delaware and the Michigan state legislature's new privacy bill, would expand on the narrow definition of data broker used in the California and Vermont laws.[53] However, they also do not place strong controls on data brokerage. Other state privacy laws, such as the Virginia and Colorado laws, allow consumers to opt out of the sale of their information but provide numerous exceptions for companies, including for data that does not match the definition of "personal data" and for "publicly available information."[54] The laws do not place strong controls on the business of data brokerage itself, and their do-not-sell provisions place an unreasonable burden on consumers to try to marginally address systemic surveillance practices.

Both the California and Vermont registry laws are a step towards more visibility into data brokerage and represent the work of many privacy experts and activists, working against the lobbying forces of data brokers which pushed to weaken the legislation before its passage. Nonetheless, there is much more to be done.

---

[51] Kim, *Data Brokers and the Sale of Americans' Mental Health Data*; Shoshana Wodinsky and Kyle Barr, "These Companies Know When You're Pregnant—And They're Not Keeping It Secret," *Gizmodo*, July 30, 2022, https://gizmodo.com/data-brokers-selling-pregnancy-roe-v-wade-abortion-1849148426; Alfred Ng, "Data brokers resist pressure to stop collecting info on pregnant people," *Politico*, August 1, 2022, https://www.politico.com/news/2022/08/01/data-information-pregnant-people-00048988.

[52] U.S. Federal Trade Commission, "FTC Enforcement Action to Bar GoodRx from Sharing Consumers' Sensitive Health Info for Advertising," ftc.gov, February 1, 2023, https://www.ftc.gov/news-events/news/press-releases/2023/02/ftc-enforcement-action-bar-goodrx-sharing-consumers-sensitive-health-info-advertising.

[53] Delaware H.B. 262. https://legis.delaware.gov/BillDetail/79022; Michigan S.B. 1182. https://www.legislature.mi.gov/(S(yiquhvromywxgybpxwrgpa4u))/mileg.aspx?page=GetObject&objectname=2022-SB-1182.

[54] Virginia Title 59.1 Chapter 53. Consumer Data Protection Act. https://law.lis.virginia.gov/vacode/title59.1/chapter53/; Colorado S.B. 21-190. Colorado Privacy Act. https://leg.colorado.gov/sites/default/files/2021a_190_signed.pdf.

There is a related question of defining "data brokers." The California and Vermont registry laws generally define data brokers as companies that sell information about individuals with whom they have no direct business relationship.[55] In other words, a company that only sells information about its own customers is not considered a data broker under the California and Vermont registry laws. Federal and state bills around data brokerage frequently reflect this distinction.[56]

Our research program takes a broader, and in some ways more intuitive, view: if a company brokers data, it is engaged in data brokerage. One possible definition of data brokerage is the collection, aggregation, analysis, buying, selling, and sharing of data, irrespective of the company's relationship with the consumer whose data is being sold or monetized. The logic is that while "first-party" and "third-party" is an important distinction, companies in both categories broker Americans' data. Similarly, there are some companies that make data brokerage their entire business model. Other companies engage in data brokerage even though it is not their primary means of making money. Although the two types of entities make different percentages of their revenue from brokering data, both categories of companies are in the business of profiting off the collection, buying, and selling (or, for instance, licensing) of consumers' data. All told, the companies engaged in data brokerage range from companies publicly advertising themselves as data brokers, to companies that broker data but call themselves advertisers or marketing businesses, to mobile applications that brand themselves as providing a particular product (like a weather app or family safety app) while quietly selling data on their users to make a profit.

Our understanding of data brokerage aligns with the FTC's previous analyses of data brokers. In the Commission's 2012 report, the FTC noted a possible distinction between "(1) entities that maintain data for marketing purposes; (2) entities subject to the FCRA [Fair Credit Reporting Act]; and (3) entities that may maintain data for other, non-marketing purposes that fall outside of the FCRA."[57] In its 2014 report, it defined data brokers as "companies that collect consumers' personal information and resell or share that information with others."[58] Again, companies that sell data on their own customers were still considered data brokers. This view also aligns with the Consumer Financial Protection Bureau (CFPB)'s 2023 request for information on data brokers and other business practices involving the collection and sale of consumers' information. It stated that "data

[55] California Civil Code Title 1.81.48. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.48.&part=4.&chapter=&article=; Vermont Statute 9 V.S.A. § 2430. https://legislature.vermont.gov/statutes/section/09/062/02430.

[56] See, e.g., Justin Sherman, "Examining State Bills on Data Brokers," *Lawfare*, May 31, 2022, https://www.lawfareblog.com/examining-state-bills-data-brokers; Justin Sherman, "Federal Privacy Rules Must Get 'Data Broker' Definitions Right," *Lawfare*, April 8, 2021, https://www.lawfareblog.com/federal-privacy-rules-must-get-data-broker-definitions-right.

[57] U.S. Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers* (Washington, D.C.: Federal Trade Commission, March 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf, 65.

[58] U.S. Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability* (Washington, D.C.: Federal Trade Commission, May 2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf,i.

brokers encompass actors such as first-party data brokers that interact with consumers directly, as well as third-party data brokers with whom the consumer does not have a direct relationship."[59]

By failing to account for the spectrum of companies brokering data, state and federal privacy legislation and debate have excluded some data brokers from the legal discussion and from regulation. For example, *The Markup* uncovered in December 2021 that "family safety app" Life360 was secretly selling precise location data on its parent and child users—and in 2020 made almost 20 percent of its revenue from this brokerage.[60] Despite the fact that Life360 was selling data on its own users, it would not have to register as a data broker under the California and Vermont registry laws. As a result, policy and legal regulations that use a narrow definition of data brokers will not comprehensively address the spectrum of data brokerage activities ongoing today. Companies involved in data brokerage in some capacity have a strong financial interest in limiting the scope of legal, regulatory, and policy activities vis-à-vis "data brokers" for precisely this reason. This is especially important when it comes to health and location data, as mobile apps that collect health and location data on their own users and then surreptitiously sell it are a primary way in which that kind of information (which consumers have no reasonable expectation is sold and shared) becomes available for purchase on the open market. Without including that vector of data transmission in regulation, policymakers would not fully tackle the risks facing Americans.

Large data brokers also spend millions of dollars lobbying against strong U.S. federal privacy legislation that would undercut their business models.[61]

Lastly, these laws rely on the notion that some data is clearly personally identifiable while other data is not. There is a difference between data with an individual's name attached and data that does not have a name attached, but that line is increasingly blurring. The sheer volume of data that exists on any given American—including for sale on the open market—means individuals, companies, and government agencies can easily combine datasets together to unmask or "reidentify" the person behind a piece of information. For instance, researchers unmasked supposedly anonymized ride data for New York City taxi drivers and could then calculate drivers' incomes.[62] The earlier discussion of the misleading term "anonymization" speaks to this point as well. Basing laws too much on the distinction between personally identifiable and not personally identifiable does not recognize the complicated reality, where simply removing a name or Social Security Number from a dataset does not meaningfully protect individuals' privacy. This distinction can also allow companies to circumvent the narrow legal restrictions that do protect

[59] Consumer Financial Protection Bureau. *Request for Information Regarding Data Brokers and Other Business Practices Involving the Collection and Sale of Consumer Information*. Washington, D.C.: Consumer Financial Protection Bureau, March 2023. https://www.federalregister.gov/documents/2023/03/21/2023-05670/request-for-information-regarding-data-brokers-and-other-business-practices-involving-the-collection.
[60] Jon Keegan and Alfred Ng, "The Popular Family Safety App Life360 Is Selling Precise Location Data on Its Tens of Millions of Users," *The Markup*, December 6, 2021, https://themarkup.org/privacy/2021/12/06/the-popular-family-safety-app-life360-is-selling-precise-location-data-on-its-tens-of-millions-of-user.
[61] Alfred Ng and Maddy Varner, "The Little-Known Data Broker Industry Is Spending Big Bucks Lobbying Congress," *The Markup*, April 1, 2021, https://themarkup.org/privacy/2021/04/01/the-little-known-data-broker-industry-is-spending-big-bucks-lobbying-congress.
[62] Marie Douriez et al., "Anonymizing NYC Taxi Data: Does It Matter?" *2016 IEEE International Conference on Data Science and Advanced Analytics*, October 2016, https://ieeexplore.ieee.org/document/7796899.

individuals' data, because they can buy, sell, and share Americans' information without a name attached and simply acquire other identifying data or perform their own reidentification separately.

Steps Congress Can Take Now

Data brokerage is a virtually unregulated industry in the United States. The U.S. needs heavy restrictions on data brokerage to properly prevent and mitigate harms and reduce the many risks to consumers and to society—from the persistent harms to consumers' privacy and safety to the considerable risks to national security. Allowing companies to collect, analyze, aggregate, and sell, share, or otherwise monetize Americans' data without regulation is only inviting harms to occur.

Playing whack-a-mole, or pursuing enforcement against companies one-by-one on a case-by-case basis, will not suffice, for at least two main reasons. The first is that data brokerage is a multi-billion-dollar industry in the United States, and companies have incredibly strong incentives to continue selling data. There have been several documented cases in which data brokers that are knowingly selling data to criminal scammers, even data on vulnerable Americans such as elderly people with Alzheimer's or a 92-year-old Army veteran, have simply kept doing so even after their criminal clients get caught—or have even assisted those buyers with evading law enforcement scrutiny.[63] Our Duke data brokerage research team has purchased highly sensitive, individually identified data on Americans from data brokers, with no vetting, only for them to follow up pitching us on additional datasets to purchase. Going after a single bad actor is not going to disrupt the strong market incentives to sell data—or stop what every other data broker is doing.

The second reason whack-a-mole is insufficient is that in many cases, once a regulatory agency looks to step in, the harm has already occurred. This is not to say regulatory enforcement is unimportant; of course, any strong regulatory regime requires both robust rules and robust enforcement. The FTC, for example, is doing important work in this area, long supported by both Democratic and Republican Commissioners, to crack down on some data brokers' unfair or deceptive trade practices. But an ex post-facto regulatory action cannot undo the harm of an abusive individual buying a woman's home address online and then stalking and assaulting her. An ex post-facto regulatory action cannot undo the ruining of someone's life because they were outed with data related to their lifestyle and sexual activity. An ex post-facto regulatory action cannot undo the damage a scammer inflicts on elderly Americans and people with Alzheimer's because the scammer legally bought their contact information from a data broker and stole their savings. An ex post-facto regulatory action will not undo the harm of a foreign nation-state actor exploiting the widespread availability of data on hundreds of millions of U.S. citizens. While the FTC plays a key role, and it should continue doing so, the best, most comprehensive solutions will also depend on federal privacy laws focused on this ecosystem.

As my colleague Professor David Hoffman has written, "data privacy need not stifle innovation, but it must protect the public from harm."[64]

---

[63] See, e.g., *United States of America v. Epsilon Data Management, LLC* (2021); *United States of America v. Macromark, Inc.* (2020); *United States of America v. KBM Group, LLC* (2021); Duhigg, "Bilking the Elderly, With a Corporate Assist."

[64] David Hoffman, "Rein In Data Brokers," *The New York Times*, July 15, 2019, https://www.nytimes.com/2019/07/15/opinion/intel-data-brokers.html.

There are three steps Congress should take now:

1. *Strictly control the sale of data to foreign companies, citizens, and governments.* Currently, there is virtually nothing in U.S. law preventing American companies from selling citizens' personal data—from real-time GPS locations and health information to data on military personnel and government employees—to foreign entities, including those entities which pose risks to U.S. national security. As a result, it is far too easy for a foreign government to set up a front company through which it can simply buy highly sensitive data on millions of Americans, including members of Congress, federal government employees, military personnel, and their families. In response, Congress should develop a set of strict controls on data brokers' sales of data to foreign companies, citizens, and governments—weighing outright prohibitions in some cases (e.g., on selling data on government employees and military personnel) and conditional restrictions in others (e.g., banning sale to a particular end user determined, through a robust and transparent security review process, to have requisite links to a foreign military or intelligence organization). As more and more U.S. citizen data is available for sale on the open market, this set of restrictions would better protect national security and also protect against exploitation of American consumers by foreign corporations. Congress should also consider areas in which outright bans on the sale of certain types of sensitive data would best protect national security (see item #2).
2. *Ban the sale of data completely in some sensitive categories, such as with health and location data, and strictly control the sale of data in other categories.* Congress should consider banning the sale of certain categories of data altogether. While many kinds of data can be used in harmful ways, some categories are arguably more sensitive than others. For instance, individuals' genetic information is highly sensitive. The U.S. intelligence community has made clear that the Chinese government is highly interested in gathering U.S. persons' genetic data.[65] Health data broadly is also sensitive: both Democrats and Republicans agreed nearly three decades ago, with the passage of HIPAA, that Americans' health data should be protected, as it is intimate to one's body and mind, highly traceable to an individual, and highly susceptible to abuse. Yet, Americans' health data, including individually identified health data, is constantly gathered and sold as part of the data brokerage ecosystem. Location data is also a very dangerous kind of data. With GPS data, law enforcement agencies operating without adequate oversight as well as foreign intelligence organizations, terrorist groups, criminals, and violent individuals could acquire this data to follow people around as they visit bars, restaurants, medical centers, divorce attorneys, police stations, religious buildings, military bases, listed and unlisted government facilities, their relatives' homes, and their children's schools. Based on tracking U.S. citizens as they walk, travel, shop, sit, and sleep, organizations and individuals intent on doing harm can also derive other sensitive information about Americans' health, income, lifestyle, and more. Congress should develop a list of sensitive data categories that each correspond to bans on sale or other strong controls.

[65] See, e.g., U.S. National Counterintelligence and Security Center. *China's Collection of Genomic and Other Healthcare Data from America: Risks to Privacy and U.S. Economic and National Security*. Washington, D.C.: National Counterintelligence and Security Center, February 2021. https://www.dni.gov/files/NCSC/documents/SafeguardingOurFuture/NCSC_China_Genomics_Fact_Sheet_2021revision20210203.pdf.

3. *Stop data brokers from circumventing those controls by "inferring" data.* If data brokers are prevented from collecting, aggregating, buying, selling, and sharing certain kinds of data and/or selling it to and sharing it with certain entities, they may still get data using their third vector—analyzing data and making "inferences" from it. For instance, if data brokers were prohibited specifically from buying and selling Americans' GPS location histories, a company could still, in line with current practice, mine individuals' purchase information, Wi-Fi connection histories, Bluetooth device links, and other information to derive the data that is supposed to be controlled in the first place, without technically "collecting" GPS location itself. I know of at least one case, to give another example, where a data broker advertising data on Americans with depression was claiming that it was not advertising data on depressed Americans per se, but that it was using information about people taking anti-depressants to provide companies with data on people interested in anti-depressant prescriptions. Congress should stop data brokers from circumventing regulatory controls and making these kinds of nonsensical semantic distinctions by implementing additional prohibitions around "inferring" categories of sensitive information about individuals. This will tackle the third main way data brokers currently get their data—and prevent companies from circumventing controls to keep exploiting Americans.

The data brokerage ecosystem perpetuates and enables civil rights abuses, consumer exploitation, and threats to U.S. national security and democracy. It operates with virtually no regulation. Rather than waiting to resolve the debate over a strong, comprehensive consumer privacy law—which is also sorely needed—Congress can and should act now to regulate data brokerage.