Data Brokerage and Threats to U.S. Privacy and Security

Written Testimony

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

U.S. 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

Chair Warren, Ranking Member Cassidy, and distinguished members of the Subcommittee, I appreciate the opportunity to testify about privacy issues facing American citizens.

I am a fellow at Duke University's Sanford School of Public Policy, where I lead a research project focused on the data brokerage ecosystem. We study the virtually unregulated industry and practice of data brokerage—the collection, aggregation, analysis, buying, selling, and sharing of data—and its impacts on civil rights, national security, and democracy. I am also affiliated with the Atlantic Council and with American University Washington College of Law, where I work on cybersecurity, internet policy, and geopolitical issues.

Data brokerage is a threat to civil rights, to U.S. national security, and to democracy. The entire data brokerage ecosystem—from companies whose entire business model is data brokerage, to the thousands of other advertisers, technology giants, and companies that also 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 must not wait to resolve the debate over such a law to regulate the data brokerage industry.

There are three steps Congress should take now:

- Strictly control the sale of data collected by data brokers to foreign companies, citizens, and governments;
- Strictly control the sale of data in sensitive categories, like genetic and health information and location data; and
- Stop data brokers from circumventing those controls by "inferring" data.

The Data Brokerage Problem

Today, and for several decades, thousands of companies have surreptitiously collected data from public and private sources about each and every American. Often, these companies will use tools to "infer" additional data about each American. These companies then repackage and resell that data on the open market, with very few controls. This is the data brokerage ecosystem.

Data brokerage is a virtually unregulated practice in the United States (except for two, limited state laws and some narrowly targeted federal regulations discussed below). Brokered data is used to target consumers, marginalized communities, veterans, military servicemembers, government employees, first responders, students, and children. Too often, this targeting is exploitative.

- Military personnel: Data brokers advertise data about millions of U.S. military personnel. Criminals have acquired this data to run educational scams against veterans because of federal military benefits.[1] Foreign governments could acquire this data to profile military personnel, track them 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, in the future, there is no need for the Chinese government or any other foreign intelligence agency to even hack many U.S. government databases when the data can be legally purchased from American data brokers, who problematically appear to do very little customer vetting.
- Survivors of domestic 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.[2] 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.
- Individuals with mental health conditions: Data brokers advertise data on millions of Americans' mental health conditions. Companies can legally purchase this data from other firms, circumventing existing health privacy laws, and use it to exploit consumers. Criminals could acquire this data to run scams against senior citizens with Alzheimer's and dementia.[3] Foreign governments could even acquire this data for intelligence purposes. Once again, there is little evidence data brokers conduct robust customer screening.

Our research at Duke University has found data brokers widely and publicly advertising data regarding millions of Americans' sensitive demographic information, political preferences and beliefs, and whereabouts and real-time locations, as well as data on first responders, government

[1] Tariq Habash and Mike Saunders, "The Predatory Underworld of Companies that Target Veterans for a Buck," Student Borrower Protection Center, February 1, 2019, https://protectborrowers.org/the-predatory-underworld-of-companies-that-target-veterans-for-a-buck/.
[2] 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 December 2, 2021, https://nnedv.org/mdocs-posts/people-searches-data-brokers/.
[3] Criminals have already used broker data to facilitate elder scams. See, e.g., U.S. Department of Justice, "List Brokerage Firm Pleads Guilty To Facilitating Elder Fraud Schemes," Justice.gov, September 28, 2020, https://www.justice.gov/opa/pr/list-brokerage-firm-pleads-guilty-facilitating-elder-fraud-schemes.

employees, and current and former members of the U.S. military.[4] Data brokers gather your race, ethnicity, religion, gender, sexual orientation, and income level; major life events like pregnancy and divorce; medical information like drug prescriptions and mental illness; your real-time smartphone location; details on your family members and friends; where you like to travel, what you search online, what doctor's office you visit, and which political figures and organizations you support. All of this is aggregated, analyzed, and packaged into datasets for sale with such titles as "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."[5] All of this information is typically collected without any consumer notice or consumer consent.

Hundreds of data brokers make selling this data their entire business model, and thousands more companies, from small businesses to technology giants, buy, sell, and share data as part of this ecosystem. The entities using this data include banks, credit agencies, insurance firms, internet service providers, predatory loan companies, online advertisers, U.S. law enforcement and security agencies, and perpetrators of domestic violence—not to mention the foreign governments, criminals, terrorist organizations, and violent individuals that could potentially acquire the data. There are single data brokers alone that advertise thousands of individual data points on billions of people around the world. Large brokers also spend millions of dollars lobbying against strong U.S. federal privacy legislation that would undercut their business models.[6]

The harms are well-documented. Scammers have acquired data to run educational scams against veterans, military servicemembers, and their families.[7] Abusive individuals have used people search websites—where data brokers scrape public records and publish Americans' addresses and other information on the internet—to hunt down and stalk, intimidate, harass, and even murder individuals trying to escape them.[8] Financial firms have used brokered data to market products to consumers that "limit or obscure their access to loans, credit, and financial services."[9] GPS location data companies have secretly tracked citizens attending protests and demonstrations and identified their ages, genders, ethnicities, and other sensitive demographic characteristics—all of which they can legally sell.[10] Health insurance companies have aggregated millions of Americans' medical diagnosis, test, prescription, and socioeconomic data—as well as sensitive demographic information like race, education level, net worth, and family structure—to market their products

[4] 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/.
[5] 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.
[6] 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.
[7] Habash and Saunders, "The Predatory Underworld of Companies that Target Veterans for a Buck."
[8] Sherman, *Data Brokers and Sensitive Data on U.S. Individuals*.
[9] Testimony of Pam Dixon before the U.S. Senate Committee on Banking, Housing, and Urban Affairs, "Data Brokers, Privacy, and the Fair Credit Reporting Act," June 11, 2019, https://www.banking.senate.gov/imo/media/doc/Dixon%20Testimony%206-11-19.pdf, 1.
[10] 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/.

and, possibly, calculate how much they can charge consumers.[11] 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.[12] The data law enforcement and other customers use may not even be updated, complete, or accurate.[13] The list of known harms goes on. And with all this data, companies can easily identify individuals by name.

The potential harms are also numerous. Domestic extremists could acquire real-time GPS location data to target politicians at home. Foreign governments 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.[14] 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. Not to mention, threat actors can simply hack into the data brokers, online advertising firms, and other entities housing this highly sensitive data.

Companies can collect this data on Americans:

- directly, whether those individuals know it or not;
- indirectly, by purchasing or licensing the data or by plugging into data sources like online advertising networks or third-party software development kits (SDKs); and
- by running algorithms to predict (what they often call "infer") sensitive information about individuals, from income level to sexual orientation.

Based on our research at Duke University, the companies selling this data on the open market conduct varying degrees of know-your-customer due diligence: some appear to conduct some due diligence before initiating a data purchase agreement, some appear to conduct a little due diligence, and some appear to conduct none at all. For those that appear to conduct some due diligence, it is unclear how comprehensive that vetting is in practice. Further, based on the copious evidence of data brokerage-linked harms (from domestic violence to consumer exploitation), there is very little to suggest data brokers implement controls to prevent harmful uses of their data once sold. Data brokers may also require clients to sign nondisclosure agreements preventing them from identifying where they obtained U.S. citizens' data.

As part of talking about the power of Big Tech, the dangers of modern surveillance, and data threats to Americans' civil rights, U.S. national security, and democracy, we must focus on this entire data brokerage ecosystem.

---

[11] One such company has alleged it does not use this data for pricing. Marshall Allen, "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates," *Pro Publica*, July 17, 2018, https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

[12] 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.

[13] 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.

[14] 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.

The Regulatory Gap

Data brokerage is a virtually unregulated practice. While there are some narrow controls around the collection, aggregation, buying, selling, and sharing of certain types of data—such as with the Health Insurance Portability and Accountability Act (HIPAA) and covered health providers, or with the Family Educational Rights and Privacy Act (FERPA) and covered educational institutions—these regulations are very limited and easily circumventable. It is remarkably easy to collect, aggregate, analyze, buy, sell, and share data on Americans, even millions at a time, without running into any legal barriers, regulatory requirements, or mandatory disclosures.

Two state laws mention data brokers: one in California and one in Vermont.[15] However, these laws are limited and insufficient to prevent the harms identified for four main reasons—and, therefore, this Committee should lead and enact legislation to regulate the data brokerage ecosystem.

First, both state laws focus merely on disclosure. They do not put meaningful restrictions on data collection, aggregation, or analysis or on the buying, selling, and sharing of data by companies classified as "data brokers." Instead, they focus on requiring those companies to register with the state, after which basic company information (e.g., the company's name) is published in a registry on the respective state government's website.[16] The Vermont law also imposes a few basic technical requirements to protect the security of what it describes as individuals' "personally identifiable information," though this is aimed at preventing data breaches instead of putting controls on data sales.[17]

Second, these laws define data brokers (generally) as only those companies buying and selling data on people with whom they do not have a direct business relationship. This definition excludes every single company that buys, sells, and shares data on its own customers from coverage under a "data broker" law. In practice, if this definition were paired with substantive controls, much of the data brokerage ecosystem would escape regulation. This definition is also insufficient because some firms occupy gray areas vis-à-vis these laws: for example, Oracle has registered as a data broker in both states, but it appears to buy and sell data it did not directly collect from consumers—as well as data it may collect directly but through subsidiaries. The same could be argued with respect to online advertisers, which frequently have direct interactions with consumers but often in ways consumers do not recognize or understand.[18]

Third, even with the given definitions of data brokers, the two state laws do not target the underlying ecosystem—the collecting, aggregating, analyzing, buying, selling, and sharing of Americans' data. The practice of buying and selling data with virtually no restrictions enables consumer exploitation, civil rights abuses, and direct threats to U.S. national security, but it is not meaningfully controlled by these laws. And even with a legal focus on specific "data broker" entities, many firms that engage in data brokerage are not captured in the laws, due to last-minute definitional changes obtained by industry lobbyists prior to state-level enactment: companies that

[15] These are, respectively, California Civil Code § 1798.99.80 and Vermont Statute 9 V.S.A. § 2430.
[16] The California registry can be found at: https://oag.ca.gov/data-brokers. The Vermont registry can be found at: https://bizfilings.vermont.gov/online/DatabrokerInquire/DataBrokerSearch.
[17] Vermont Statute 9 V.S.A. § 2447. Data broker duty to protect information; standards; technical requirements.
[18] For more on how these laws provide lessons for writing a federal privacy legislation, see: 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.

buy and sell data on their direct customers; third-party code providers that plug into apps and websites to collect data on unwitting individuals; companies that run real-time bidding networks for online advertisements, where dozens of companies get access to data on consumers whom they could target with paid ad access.

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 which 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.[19] Basing laws too much on this distinction 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 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.

The Congressional Response

Congress has an opportunity to regulate the data brokerage ecosystem, protecting Americans' civil rights, U.S. national security, and democracy in the process. While a strong, comprehensive consumer privacy law is important, Congress must not wait to resolve the debate on such a law to regulate the data brokerage industry.

There are three steps Congress can take now:

Strictly control the sale of data collected by data brokers 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 a risk 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, and military personnel. 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 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.

[19] 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.

Strictly control the sale of data in sensitive categories, like genetic and health information and location data. Congress should also 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. 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 controls.

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' spending histories, WiFi connection histories, phone call logs, and other information to derive the data that is supposed to be controlled in the first place, without *technically* "collecting" GPS location itself. Congress should stop data brokers from circumventing controls 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.