Data Brokers and the Sale of Data on U.S. Military Personnel

Risks to Privacy, Safety, and National Security

By Justin Sherman, Hayley Barton, Aden Klein, Brady Kruse, and Anushka Srinivasan

November 2023

—

Overview: The data brokerage ecosystem is a multi-billion-dollar industry comprised of companies gathering, inferring, aggregating, and then selling, licensing, and sharing data on Americans as well as providing technological services based on that data. After previously discovering that data brokers were advertising data about current and former U.S. military personnel, this study sought to understand (a) what kinds of data that data brokers were gathering and selling about military servicemembers and (b) the risk that a foreign actor, such as a foreign adversary government, could acquire the data to undermine U.S. national security. This study involved scraping hundreds of data broker websites to look for terms like "military" and "veteran," contacting U.S. data brokers from a U.S. domain to inquire about and purchase data on the U.S. military, and contacting U.S. data brokers from a *.asia* domain to inquire about and purchase the same. It concludes with a discussion of the risks to U.S. military servicemembers and U.S. national security, paired with policy recommendations for the federal government to address the risks at hand.

Table of Contents:

Authors:
Justin Sherman is a senior fellow at Duke University's Sanford School of Public Policy and leads its data brokerage research project.
Hayley Barton is a Master of Public Policy (MPP) and Master of Business Administration (MBA) student at Duke University and a former research assistant on Duke's data brokerage research project.
Aden Klein is a senior at Duke University and a research assistant on Duke's data brokerage research project.
Brady Kruse is an MPP student at Duke University and a research assistant on Duke's data brokerage research project.
Anushka Srinivasan is a sophomore at Duke University and a former research assistant on Duke's data brokerage research project.

Acknowledgements: The authors would like to thank the multiple reviewers for their comments earlier versions of this study. The authors would also like to especially thank Professor David Hoffman, the Principal Investigator (PI) for the grant, and Professor Ken Rogerson for their support throughout the duration of this work.

Funding Statement: This research was sponsored by the United States Military Academy (USMA) and was accomplished under Cooperative Agreement Number W911NF-22-2-0099. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Military Academy or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints of the final, public research products published by Duke University for Government purposes notwithstanding any copyright notation herein.

Executive Summary

This report describes the process and results of a 12-month-long study into data brokers and the sale of data on U.S. military servicemembers and veterans. The study was designed to explore two central questions:

1. What kinds of data are data brokers currently gathering and selling related to U.S. military servicemembers and veterans?
2. What is the risk that a foreign adversary could exploit the data brokerage ecosystem to access this data on U.S. military servicemembers and use it in harmful ways?

The key findings of this study are summarized below and come from all four phases of the project: Phase 1: Scraping Data Broker Websites; Phases 2 and 3: Buying Servicemembers' Data; and Phase 4: Analysis of Purchased Data.

Major Takeaways:

- It is not difficult to obtain sensitive data about active-duty members of the military, their families, and veterans, including non-public, individually identified, and sensitive data, such as health data, financial data, and information about religious practices. The team bought this and other data from U.S. data brokers via a *.org* and a *.asia* domain for as low as $0.12 per record. Location data is also available, though the team did not purchase it.
- Data broker methods of determining the identity of customers are inconsistent and evidence a lack of industry best-practices.
- Currently, these inconsistent practices are highly unregulated by the U.S. government.
- The inconsistencies of controls when purchasing sensitive, non-public, individually identified data about active-duty members of the military and veterans extends to situations in which data brokers are selling to customers who are outside of the United States.
- Access to this data could be used by foreign and malicious actors to target active-duty military personnel, veterans, and their families and acquaintances for profiling, blackmail, targeting with information campaigns, and more.

Phase 1: Scraping Data Broker Websites:

- Data brokers are advertising the fact that they hold and can sell data on current and former members of the U.S. military, ranging from data that is aggregated (e.g., the number of people in a ZIP code with a given characteristic) to data that is clearly identified and linked to specific individuals.
- We found a total of 7,728 hits for the word "military" and 6,776 hits for the word "veteran" across 533 data brokers' websites, built from the Vermont and California state data broker registries. These mentions ranged from data brokers advertising data about veterans (e.g., "veterans that own a motorcycle," "military readers") to one data broker noting its ability to find a deceased veteran's "claim or discharge number" by searching death records.

- Several data broker websites advertise data on military families, with dataset titles such as “Military Families Mailing List” and “Hard Core Military Families.”

Phases 2 and 3: Buying Servicemembers’ Data:

- We contacted 12 U.S. data brokers about purchasing information on U.S. military servicemembers and veterans. These 12 data brokers were selected based on our landscape analysis and knowledge of the industry and are referred to within this report as Broker 1, Broker 2, and so on. We ultimately purchased data from three brokers based on the type of data and variables offered. We conducted all research in compliance with Duke University’s research ethics processes.
- As part of their sales process, multiple data brokers sent us lists of hundreds of identifiable demographic variables that we could select from within their consumer databases. This is a relatively standard practice for data brokers but underscores that a prospective buyer could purchase many different data points on individuals, selected from a large menu of options.
- We found a lack of robust controls when asking some data brokers about buying data on the U.S. military and when actually purchasing data from some data brokers, such as identity verification, background checks, or detective controls to ascertain our intended uses for the purchased data. For example, Broker 4 told us that it would have to verify our identity before selling us data on the military unless we paid by wire instead of credit card.[1] We then paid by wire, and Broker 4 provided us with the data we requested on members of the U.S. military without asking about or verifying our identity.
- Some other brokers did appear to have some controls in place. Two brokers refused to sell to us based on our lack of a website and the fact that we were not a “verified” company. One location data broker said it would not sell geolocation data around “sensitive” locations, including military sites, but could provide geolocation data on many other areas of the country.
- Broker 5 asked, when we inquired about purchasing data, if we intended to make the data public, publish research on the data, or provide investors or policymakers with the data. We decided not to purchase data from this broker.
- Brokers 10 and 11 requested that we sign nondisclosure agreements (NDAs) covering our interactions with the broker, details on the data available for purchase, and any purchased datasets. We did not sign any NDAs and did not purchase from these brokers.
- For several of the brokers, as noted above, the controls in place were primarily focused on requiring confidentiality around the data purchasing itself and to make certain the customer was a company.

Phase 4: Analysis of Purchased Data:

- All datasets that we purchased included individual, personally identifiable information on military personnel in the United States. None of these datasets were anonymized nor aggregated, even when providing sensitive information (such as net worth, religion, or health) and without verifying the purchaser’s identity. (“Sensitive”

[1] All wire transfers were facilitated through Duke University and indicated as such upon receipt.

is used here in an analytic sense and not drawn from a specific law or regulation.) Anyone with a few hundred dollars can obtain the same type of data that we did and use it for any purpose, good or bad.

- The purchased data cost us between $0.12 to $0.32 per U.S. military servicemember when buying between 4,951 and 15,000 identifiable records at a time. Based on advertising from other brokers, identifiable datasets pertaining to the U.S. military can be purchased for as little as $0.01 per military servicemember for much larger purchases.
- Using a U.S. domain, we purchased three datasets from Broker 3. The first was individually identified contact data on 5,000 active-duty military personnel; the second was individually identified contact data on 5,000 friends and family members of military personnel; and the third was individually identified contact data on 15,000 military personnel plus 15 checkboxes indicating ailments and health conditions. The first two contact datasets cost $0.20 per military servicemember with name, address, email, and specific branch and/or agency (active-duty only), while the health dataset cost around $0.22 per military servicemember. They collectively included military personnel living in all 50 states.
- Using a U.S. domain, we purchased one dataset from Broker 4 containing individually identified information on 5,000 active-duty military personnel, including those servicemembers' names, home addresses, email addresses, and wireless phone numbers. The dataset cost $0.125 per military servicemember.
- Using a U.S. domain, we purchased one dataset from Broker 6 containing individually identified information on every active-duty military servicemember in their records, geofenced to Washington, DC, Maryland, and Virginia.[2] For those 4,951 military personnel, the dataset included their name, home address, email address, political affiliation, gender, age, income, net worth, credit rating, occupation, presence of children in the home (yes/no), marital status, homeowner/renter status, home value, and religion. The dataset cost was $0.213 per military servicemember.
- Using a *.asia* domain name, email address, and Singaporean IP address, we purchased one dataset from Broker 3 containing individually identified information on 5,000 active-duty military servicemembers and veterans, geofenced to Washington, DC, Maryland, and Virginia. The dataset included their name, home address, email address, and cell phone number. The dataset cost $0.32 per military servicemember or veteran. The broker demonstrated no restrictions on its ability to sell to a *.asia* domain data on active-duty military servicemembers and veterans.
- Using a *.asia* domain name, email address, and Singaporean IP address, we purchased one dataset from Broker 4 containing individually identified information on 5,000 total active-duty military servicemembers and veterans. The dataset included their name, home address, wireless phone number, email, age, sex, marital status, homeowner status, estimated home value, interest in charitable donations, interest

[2] A geofence is a virtual perimeter drawn around a physical location. For the purposes of this study, we refer to "geofencing" in instances where we requested and purchased data for a more limited geographic region than the full United States (for example, only data for specific ZIP codes, states, or military bases). Since we are focused on national security concerns, we geofenced to states or locations that have prominent military installations. Some location data brokers that we contacted but did not purchase from offered even more granular geofencing to specific buildings, addresses, shapes, or other locations, including tracking foot traffic.

in current affairs / politics, and interest in gambling / casinos. The dataset cost $0.12 per military servicemember or veteran. The broker implied an internal restriction on selling financial data to unverified customers, but nonetheless sold other identifiable and potentially sensitive data on active-duty military servicemembers and veterans to a *.asia* domain without verification.

- Using a *.asia* domain name, email address, and Singaporean IP address, we purchased one dataset from Broker 6 containing individually identified information on active-duty military servicemembers and veterans, geofenced to Fort Bragg, North Carolina; Fort AP Hill, Virginia; and Quantico, Virginia; in addition to generally geofencing Washington, DC, Maryland, and Virginia. The dataset included 5,048 individually identified servicemembers, along with their contact information and demographic data. For those military personnel, the dataset included their name, home address, email address, gender, age, net worth, levels of education, occupation, numbers of children, ages of children, sexes of children, marital status, homeowner/renter status, ethnicity, language, religion, and credit rating. The dataset cost $0.25 per military servicemember. We did not observe any controls that differentiated the location of the purchasing entity.

Introduction

Data brokers are companies that collect, aggregate, and infer data about individuals for the purposes of selling and sharing it. The "data brokerage ecosystem" is the multi-billion-dollar industry of these companies, thousands of which are based in the United States. According to our team's definition, the data brokerage ecosystem includes companies that gather data on their own customers and then sell it, such as mobile apps that collect and then surreptitiously sell users' geolocation data. The data brokerage ecosystem also includes third-party companies that have no direct relationship with consumers but still sell data on them. The ecosystem spans large, publicly traded companies (such as Experian and Oracle) and smaller, privately incorporated data brokers. It also includes companies that sell datasets with individuals' names and companies that provide services based on data (such as identity verification) but do not provide the underlying data to buyers.

Some of these companies may be familiar to consumers. For example, Oracle is a well-known technology company that also brokers data. The three major credit reporting agencies—Equifax, Experian, and TransUnion—are also in the business of brokering data. However, many other data brokers, such as Acxiom and Verisk, make hundreds of millions of dollars or more each year brokering data on millions of people but are likely unfamiliar to consumers. There are also companies with which consumers directly interact but may not realize are selling their data. For example, it was uncovered in 2020-2023 that a family safety app, a Muslim prayer app, and a major telehealth company were all selling or sharing consumers' information with third parties.

Collectively, the U.S. data brokerage industry gathers data on virtually every American. Our team's previous research has found data brokers marketing the fact that they collect data on hundreds of millions of individuals in the United States. This includes companies gathering and selling data on individuals' demographic information (e.g., race, ethnicity, religion, sexual orientation), political preferences and beliefs, lifestyle behaviors, home addresses and GPS locations, economic and financial situations, and health and mental health conditions.[1] Companies gather and derive this information from many sources, including by scraping public records, embedding code into mobile apps, and paying companies to sell data on their own customers.

Once data brokers have this data, many build packages of data on specific groups of people. These packages focus on individuals with shared characteristics, ranging from datasets on heavy coffee drinkers or avid podcast listeners to datasets on students, first responders, and elderly Americans. In our team's first report in August 2021, we identified multiple data brokers advertising the fact that they had data related to U.S. military servicemembers and veterans.[2] Upon finding this information many months ago, we began considering the extent to which data brokerage had implications for U.S. national security.

We arrived at two central questions:

1. What kinds of data are data brokers currently gathering and selling related to U.S. military servicemembers and veterans?
2. What is the risk that a foreign adversary could exploit the data brokerage ecosystem to access this data on U.S. military servicemembers and use it in harmful ways?

After further reviewing the existing literature, it was clear that some research had been conducted on the issue (described in detail below), but that our central questions remain unaddressed. In 2021, we submitted a proposal to the U.S. Military Academy at West Point, in response to a solicitation for grant proposals, to tackle these two central questions. Our proposed effort had three main prongs:

1. Develop a better understanding of the data that data brokers gather and sell on current and former U.S. military personnel;
2. Purchase data from U.S. data brokers on U.S. military personnel via a U.S. domain to understand the kinds of data available and the process for acquiring it; and
3. Purchase data from U.S. data brokers on U.S. military personnel via a non-U.S. domain to understand the kinds of data available, the process for acquiring it, and the extent to which that process might differ when buying data via a non-U.S. domain.

This report first provides background on data brokers and U.S. national security, as well as the U.S. regulatory gaps around data brokerage broadly. Second, it describes the risks to U.S. national security that could arise from foreign and malign actors accessing and exploiting brokered data on U.S. military personnel. Third, it describes the study's methodology, including the university research ethics processes with which the team complied. Fourth, it describes the key findings from the research. Finally, it concludes with a discussion of policy recommendations for the U.S. federal government to address the risks associated with data brokerage and the sale of data on former and active-duty U.S. military personnel.

Background

Most early research and reporting on data brokers focused on privacy-related harms to American consumers.[3] The links between data brokerage and national security have recently attracted more attention, particularly with respect to the Chinese government's collection of data on populations outside of China, including on U.S. citizens.[4] However, as described in this section, significant research gaps remain when it comes to the collection and sale of data on U.S. military servicemembers and veterans.

Existing Research on Data Brokers and National Security

While some existing research focuses on data brokers and military servicemembers, there is much more research to be done. Most of the previous research on data brokers and national security focuses on data about all U.S. persons, rather than focusing on servicemembers as we do in this report. Research in both categories is described here.

National security concerns over data brokers were reignited in 2020 during the debate in the U.S. over the banning of TikTok. Experts demonstrated concern over the large extent to which the Chinese government could otherwise acquire personal data, outside of collection through apps like TikTok and WeChat, and the risks associated with this data collection. University of Virginia professor and *Trafficking Data* author Aynne Kokas noted that social media platforms "do present significant and specific security risks," and that "data exfiltration by Chinese firms from the United States is…pervasive."[5]

However, social media companies are only one method of data collection for foreign adversaries. Data brokers offer similar data for sale on the open marketplace. Yale Law School senior fellow and China expert Samm Sacks has repeatedly stated that under current law, U.S. companies can sell or provide data to third-party data brokers, which in turn could simply provide data to foreign and malign actors.[6] Caitlin Chin at the Center for Strategic & International Studies (CSIS) think tank has written that the data brokerage industry "is generally unforthcoming about the specific identities or locations of their clients, so the full extent of data brokerage partnerships with foreign governments is currently unknown."[7]

Former Federal Trade Commission (FTC) lawyer Whitney Merrill has said of TikTok that "China could buy similar mobile data from data brokers or ad networks" and that "most ad networks are collecting the same, if not worse, information" as what is collected by TikTok.[8] A NATO StratCom report notes that brokers frequently provide data to buyers "without significant screening" and that malicious actors could obtain data from data brokers by exploiting the companies' often "insufficient cybersecurity practices."[9] Two Singapore-based national security analysts wrote in 2020 that "the amount of data stored with data brokers opens them to be prime targets of cyberattacks from state or non-state actors."[10]

The wide availability of this data on consumers represents a pressing national security issue. Sen. Ron Wyden (D-Ore.) has repeatedly spoken about these issues, stating in June 2022 that "right now it's perfectly legal for a company in China to buy huge databases of sensitive information from data brokers about the movements or health records of millions of Americans, and then share that information with the Chinese government. That's a huge problem for our country's security."[11] Sen. Cynthia Lummis (R-Wyo.) also said at the time that "allowing foreign adversaries unrestricted access to Americans' private, sensitive data places U.S. companies at a competitive disadvantage and threatens our national security."[12]

Research from the Modern War Institute, the Army Cyber Institute at the United States Military Academy (which provided the grant that funded this research project), and elsewhere has highlighted that malicious actors can use personal data to define target audiences, push content to those audiences, and attempt to influence their thinking on particular issues as well as physically target individuals.[13] An article in *Lawfare* details how location data purchased from data brokers could be used for malicious activities, utilizing the data to track down specific individuals, even if the data is supposedly "anonymized."[14] A 2022 article in *The Intercept* describes how data brokers use phone data to geolocate individual people; one broker claimed, for demonstrative purposes, that it could do so for people who visited the NSA and CIA headquarters.[15] The statement was not corroborated in that specific case but is plausible.

Jonathan Panikoff, the former director of the U.S. government's Investment Security Group, which oversees the intelligence community's work on the Committee on Foreign Investment in the U.S. (CFIUS), has written that "the lack of federal regulation related to commercial data brokers, which today can and do legally collect and resell the data of millions of Americans, is a glaring gap that needs to be filled immediately."[16] He explained:

> A ban on TikTok, for example, would do nothing to prevent data brokers from aggregating the same consumer data from other apps and re-selling it to commercial entities, including those in China.[17]

Data brokers and the collection, use, and sale of data on military personnel have received less, albeit some, attention in this research and analysis. The aforementioned NATO StratCom report defines five avenues of compromising national security stemming from data held by data brokers: "personnel, equipment, information, facilities, and activities."[18] For example, it said, data from brokers can be used to "identify [and potentially blackmail] personnel," collect information about equipment capabilities, or track troop movements.[19] Another NATO StratCom report lists four potentially disruptive activities to national security: manipulation, impersonation, doxing, and revealing sensitive data.[20] It (positively) cites the example of an open-source investigative outlet's use of open-source data to track Russian activity as an example of what could (negatively) happen to other countries.[21]

Kirsten Hazelrig at nonprofit defense contractor MITRE has written that there is "ample open-source evidence" that this data can be used to:

- Target influential individuals for blackmail and coercion;
- Physically map and target sensitive sites, security measures, high-risk personnel, and operations;
- Create near real-time situational awareness of U.S. soft targets; and
- Target offensive cyber operations and network exploitation.[22]

There has been research generally related to data brokers and potential national security risks as well as the risk of foreign governments using data about government and military personnel to target them with ads, track them, blackmail them, and more. However, there has not been substantial research published specifically addressing the potential harm from data brokers and the collection, sale, and use of data related to U.S. military personnel.

Importantly, several questions remained unanswered after reviewing the literature:

1. What kinds of data are data brokers currently gathering and selling related to U.S. military servicemembers, veterans, and their families and acquaintances? Is some of this data in fact aggregated insights, rather than the underlying raw data? Is some of this data individually identified and clearly linked to specific servicemembers?
2. What is the risk that a foreign adversary could exploit the data brokerage ecosystem to access this data on U.S. military servicemembers? Are data brokers gathering and selling data on servicemembers that is not already publicly accessible or otherwise available? In other words, what could be the value-add to a foreign adversary in potentially accessing this brokered data?

The Data Brokerage Regulatory Gap

There are considerable gaps in the regulation of the data brokerage ecosystem. While some laws apply to data brokerage (e.g., around credit reporting), they do not cover all uses of that kind of data by all kinds of companies, organizations, and individuals. Other uses of data, such as the brokering of geolocation information, are largely 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, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 applies only to certain covered health entities, like hospitals and primary healthcare providers. Those entities are subject to privacy and cybersecurity controls around their collection, use, and sharing of individuals' personal health information. However, HIPAA does not apply to numerous mobile health apps, social media companies, online advertisers, data brokers, and many other kinds of corporate actors that have no business relationship with a HIPAA-covered entity. These organizations outside the narrow scope of HIPAA are therefore free to legally gather, buy, package, sell, and share Americans' individually identified and health-related data—and they do, such as whether people have prescriptions for antidepressants or whether they are believed to be pregnant.

The Family Educational Rights and Privacy Act (FERPA) of 1974 is another example of a narrow privacy law. FERPA governs covered educational institutions' use and disclosure of students' data. However, 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) of 1998, for its part, places 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 about minors. The Fair Credit Reporting Act (FCRA) of 1970 regulates credit reporting agencies' use of data but is not comprehensive in regulating the use of credit-related and financial data by other companies.

Most recently, the Daniel Arendt Judicial Privacy and Security Act, passed in December 2022 as part of the National Defense Authorization Act (NDAA) for FY2023,[23] tackles some issues associated with public records, "people search websites," and the publication of federal judges' personally identifiable information. The legislation was introduced after a violent individual obtained information online in July 2020 about New Jersey federal judge Esther Salas—and then went to her home, shot her husband, and shot and killed her 20-year-old son Daniel.[24] The Act permits federal judges to, in some cases, attempt to prevent some data brokers from selling their personal information and to limit the publication of some of their personal information on some websites.

At the state level, California and Vermont have data broker registry laws, which require some third-party data brokers to annually submit information to the state—such as name and business email address—to be published in a public, online registry.[25] Beyond the registry requirement, these two state laws do not place controls on the collection, aggregation, inference, and sale of data itself. In September 2023, California passed the Delete Act, which builds on the state's third-party data broker registry law and California's consumer state privacy laws—the 2018 California Consumer Privacy Act (CCPA) and the 2020 California Privacy Rights Act (CPRA)—to establish a centralized website through which California consumers have a one-stop-shop to request that the registered third-party data brokers not sell their "personal information" (as defined under the CCPA/CPRA).[26]

One relevant executive branch policy covers one mechanism of the collection and use of data related to military personnel in a very limited capacity. In 2018, the Department of Defense prohibited the use of GPS functions in deployed locations after fitness app Strava publicly shared a map "showing where users jog, bike and exercise."[27] Strava's publication of the map "inadvertently highlight[ed] the locations of U.S. military facilities in some of the most dangerous spots in the world" and had potential consequences for "international aid workers, intelligence operatives and millions of other people in many countries."[28] The data also revealed the location of some military facilities and spy outposts that were secret.[29] On top of that, several independent organizations proved that one can identify individuals by name based on this information.

While the Defense Department's policy around fitness wearables did not address data brokerage per se, it was designed to mitigate a risk associated with the unregulated sharing of data related to military personnel—deemed to pose a considerable risk to both individuals and missions. (Strava first responded to the news by saying that military servicemembers should "opt out" of their data being gathered and that it would work with governments to understand where there are sensitive areas that could be mapped.[30] The company subsequently restricted access to its heatmap, including by limiting its view to registered Strava users.[31])

The Challenge of "Anonymization"

Some data brokers frequently argue that some of the data sold is "anonymized." However, "anonymization" is not a technically meaningful term as used by many of these companies. There are indeed statistical techniques that can be used to provide more protection to individuals' data, such as differential privacy.[32] It is possible that some data brokers implement these kinds of controls that provide more masking of individuals' identities in datasets. But 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.[33]

Decades of computer science research show that combining datasets together, using statistical techniques, and employing other linkages can often unmask the individuals behind a data point.[34] One recent study, for example, found that with only 15 specific demographic attributes, it would be possible to "re-identify" 99.98% of Americans in a dataset.[35] Moreover, claims of "anonymization" obscure the fact that many data brokers are selling datasets that *do* include individuals' names and other clearly identifiable attributes. This fits within a data brokerage business model that is predicated on enabling organizations and individuals to use data to target, profile, and track people.

Risks to National Security

There is insufficient policy attention to the risks data brokerage poses to U.S. national security.

As is the focus of this study, there is a risk that foreign and malign actors could acquire personal data concerning U.S. military servicemembers and veterans through multiple methods. This includes buying data from brokers directly or through front organizations, hacking into data brokers' servers, or compromising the servers of data brokers' clients that have acquired data. Foreign governments' intelligence services could potentially use the acquired information against members of the U.S. military. Where foreign governments are concerned, this exploitation could range from learning sensitive information about, blackmailing, and then coercing military personnel to outing servicemembers' sexual orientations, releasing information that damages servicemembers' reputations, stalking and tailing personnel, or microtargeting personnel with particular messages.

In short, an industry that builds and sells detailed profiles on Americans could be exploited by hostile actors to target military servicemembers and veterans, as a subset of the U.S. population. Many veterans often still know currently classified information, even if they are no longer active-duty members of the military.

The Risk of Foreign and Malign Actors Exploiting Brokered Data on Military Personnel

The data brokerage ecosystem poses risks to national security by compiling large, detailed datasets on U.S. military personnel and subsequently selling that data on the open market. Data brokers advertise datasets containing information that can be used to identify members of the U.S. military and other politically sensitive targets, including detailed information on servicemembers' medical conditions, financial situation and credit score, political affiliation and religious identity, gender and sexuality, address and contact information, children and families, hobbies such as gambling or international travel, and other sensitive information, along with detailed geolocation data that can be used to identify military locations and movements. A few elements of this data may already be in the public domain, such as the names of U.S. military servicemembers (which are largely unclassified) or detailed, identified voting records linked to military housing on installations. However, sensitive information such as individual servicemembers' health conditions and financial information is not typically publicly available nor easily obtainable in aggregate.

Foreign and malign actors with access to these datasets could uncover information about high-level targets, such as military servicemembers, that could be used for coercion, reputational damage, and blackmail. For instance, data related to income level, credit score, marital status, sexual orientation, mental health conditions, sexual health conditions, gambling, and servicemembers' families is on the open market for sale and could be used for these purposes. This data could also be used to expose military members of the U.S. intelligence community through location information about visits to sensitive facilities. In a similar vein, foreign and malign actors could use detailed profiles and inferences to microtarget members of the military or government officials to spread disinformation or radicalize subgroups. Much attention has been paid to questions surrounding social media platforms, foreign governments (e.g., the Chinese and Russian governments), and known cases and risks of running targeted advertisements to U.S. persons. Yet, the research community and policymakers have paid little attention to how the large data packages compiled and sold by data brokers, including those that encompass military personnel and political data, could be exploited by foreign states and malign actors.

Location data presents its own unique risks. Foreign and malign actors could use location datasets to stalk or track high-profile military or political targets.[36] These movements could reveal sensitive locations—such as visits to a place of worship, a gambling venue, a health clinic, or a gay bar—which again could be used for profiling, coercion, blackmail, or other purposes. They could also imply other, reputationally damaging lifestyle characteristics, such as infidelity. The ability to target specific individuals in a large dataset is not just hypothetical. Using a location dataset they received, *New York Times* reporters "followed military officials with security clearances as they drove home at night" and "tracked law enforcement officers as they took their kids to school."[37] While the reporters were not seeking to cause harm with the dataset they acquired, a malicious actor could similarly track military officials and directly target them. A similar threat became concrete when a nonprofit purchased location data, seemingly originating from apps such as the gay dating app Grindr, and sifted through the dataset to identify a specific closeted priest, follow his movements, and out him.[38] A similar process could be used to target high-profile members of the military or any other individual that a foreign actor would be interested in targeting.

Aggregated insights from location data could also be valuable and damaging to U.S. national security. The aforementioned Strava incident from 2018 made available location data from U.S. military bases and undisclosed intelligence sites in various countries.[39] Location datasets could also be used to estimate military population or troop buildup in specific areas around the world or even identify areas of off-base congregation to target.[3] It is possible that this kind of information could also enable foreign intelligence organizations to identify when a targeted person was using tradecraft to avoid detection. For instance, a person who has stated they are headed to one location and instead visits another could be identified as an intelligence operative or someone working in another sensitive national security area.

[3] While this study focuses on the national security implications of the data brokerage ecosystem, there are other risks to military servicemembers from the sale of this data, such as criminals using data to scam veterans or military families.

Methodology

With these national security risks in mind, we designed a study around our two guiding questions:

1. What kinds of data are data brokers currently gathering and selling related to U.S. military servicemembers and veterans?
2. What is the risk that a foreign adversary could exploit the data brokerage ecosystem to access this data on U.S. military servicemembers and use it in harmful ways?

The research proceeded in three phases, which were designed to collectively answer these two questions and learn more about the process by which data brokers might sell data on military servicemembers and veterans. The team proceeded to:

1. Scrape hundreds of data brokers' websites, looking for key terms such as "military" and "veteran";
2. Set up a U.S. domain and email, contact a select group of U.S. data brokers asking to buy data on U.S. military personnel, and then attempt to buy data on U.S. military personnel; and
3. Set up a *.asia* domain and email, contact the same group of U.S. data brokers from which we purchased data from with the U.S. domain, and then attempt to buy the same data on U.S. military personnel.

The website scraping was designed to increase the team's understanding of data brokers' marketing of data on military personnel and to identify possible data brokers to contact about the sale of data on servicemembers and veterans. The U.S. domain component was designed to help the team understand what kinds of data are being collected and sold on U.S. military personnel, as well as the process by which that data could be purchased from an unverified domain. Similarly, the *.asia* domain component was designed to help the team understand the process by which data could be purchased on U.S. military personnel through a *.asia* domain.

Following Duke University guidelines, the project submitted an Institutional Review Board (IRB) protocol, and Duke's Office of Research Support (ORS) determined that this research is not human subjects research and therefore did not require that the project file a full application with the Duke IRB. As researchers, we sought to disclose as little as possible about our identity to data brokers during the purchasing process, until asked. During phone calls, we provided only our first names and did not immediately mention any affiliation with Duke University, until and unless asked, instead broadly (and truthfully, non-deceptively) stating that we were researchers performing market research on consumer data and U.S. military personnel. We did this because we did not pursue university approval to use deception—and because part of our research was to determine the degree to which data brokers would investigate us as their customers, which they could uncover by asking more information, looking at our payment information, or checking our *.asia* website (discussed more below).

The team also implemented internal technical controls limiting access to the purchased data to only these authors and a few select research supervisors and staff.

Phase 1 Methodology: Scraping Data Broker Websites

First, we built a computer program to automatically scrape information from websites. The program was built to download all publicly accessible pages of a website, including the HTML code that renders the website. It then scanned those webpages and their HTML code for any mentions of an inputted word or phrase. If that word or phrase was mentioned anywhere on the website, the scraping program would output that mention into an Excel file along with the surrounding context of its usage.

To create a list of data broker websites, we drew from the Vermont and California state data broker registries. Under their applicable laws, Vermont and California require companies defined as "data brokers" to provide the respective state governments with information such as their company names, updated company website URLs, and company contact information, which is then published to a public registry. Combining both datasets gave us a list of 561 websites that fell within the Vermont and/or California definition of a "data broker." Out of the initial list of websites, we discovered that a few had provided URLs and contact information that were outdated or otherwise inaccurate. Some URLs could be repaired by removing extra characters or editing the URL to the homepage page of the domain, but other entries had to be removed from the study. Ultimately, we ran the website scraping program on 533 websites. It took approximately one month to download the websites and search through them.

We created a list of key terms that the scraping program would search for and count, spanning potential consumer groups of interest, especially military personnel. The military terms were included because that was the focus of this study. Other terms were included because Duke's data brokerage research team has other workstreams focused on those particular populations, such as elderly Americans and people with Alzheimer's. Further, the list was merely intended to help uncover mentions of military personnel on data brokers' websites (alongside some other groups), not to be a comprehensive or comparative assessment of every demographic that might be discussed on brokers' websites. Our key term list was:

- DoD
- military
- veteran
- sweepstakes
- pregnancy
- pregnant
- alzheimers
- elder
- naive
- department of defense
- government employee
- mental illness
- active duty
- astrology
- military base
- abortion
- ovulation
- opportunity seekers
- reproductive healthcare

Phase 2 Methodology: Buying Servicemembers' Data via U.S. Domain

After completing the website scraping and background analysis, we set up computer, phone, and email domain infrastructure to facilitate contacting data brokers and purchasing data.

Identifiable data can be collected online through IP addresses and persistent device identifiers. Thus, we avoided contacting data brokers with our personal laptops, phones, and email addresses. Instead, we opted to use devices with minimized attribution to us, virtual private networks (VPNs) to limit the traceability of the source of our network traffic, and email addresses registered to a unique domain.

We then used this infrastructure and U.S. domain name to reach out to 12 data brokers. These 12 brokers were selected based on our web scraping results (from the first study phase), the team's prior knowledge of certain brokers, and research into brokers advertising data related to U.S. persons (of which military personnel could be a subset) or related to military personnel specifically. We then purchased datasets from three of these data brokers, discussed in the subsequent findings.

These data brokers are described throughout the rest of the paper as Broker 1, Broker 2, and so on. The project leadership has decided to anonymize the names of the data brokerage companies due to statements made by some of those companies that they deem their sales process as confidential. The project leadership does not necessarily agree that those sales processes are due any confidentiality.

Computers for Data Purchasing

For much of our online activity, we relied on VPNs and Duke-owned Chromebooks that had recently been restored to factory default settings. We used the Chromebooks to send emails to data brokers and conduct phone calls using a Google Voice number. Before engaging in online activity, we activated a VPN to mask our Duke University IP address. For the U.S.-based communications, we used a Chicago IP address. To the best of our knowledge, emails and voice calls to data brokers appeared to originate from the VPN IP address. Our precautions were necessary; one broker displayed on-screen our location and IP address every time we logged into or visited its website.

To activate the Chromebooks, we briefly used a personal device enabled with a VPN to create a new Gmail account. However, all email contact with data brokers was done through a unique domain name created specifically for the purposes of this project, as described below. Each time, before accessing the email account, we activated the Chicago-based VPN on the Duke-owned and recently wiped Chromebook.

Phones for Data Purchasing

We also needed a valid phone number in order to activate our new U.S. domain. In most cases, phone numbers are easily identifiable to a specific person, so we decided beforehand to acquire prepaid phones to mask our identity. We purchased several simple "feature phones," non-smartphones that are generally not equipped to collect extensive telemetry or consumer data. We purchased these phones from multiple electronic stores near Duke University, using cash to prevent the creation of an electronic purchase record. We also purchased prepaid phone plans and SIM cards that did not require an existing phone number to activate.

When visiting the stores, we took considerable precaution to prevent data collection on ourselves. Team members wore medical masks, hats, and clothing that covered any tattoos or identifiable marks. While in the store, we did not say our names or mention our research project. These precautions may initially seem extreme, but the electronic stores feature multiple test units of smart devices that collect and identify people based on video and audio input, such as Amazon Echo or Google Nest, and it is unknown if brokers receive data from these types of devices. Furthermore, we did not bring our phones and other mobile devices with us, preventing mobile tracking data from creating a record of our time at the stores.

Despite using prepaid SIM cards and feature phones, we took further precautions to prevent data collection. When activating the feature phones, we chose an area code near Chicago. The phones were kept off unless being actively used and were never connected to a wireless network. To create an additional layer of privacy, we did not actually contact the data brokers using the feature phones. We feared that, despite having a Chicago area code, the phones would rely on cell phone towers near Duke University. Instead, we used the phones to activate Google Voice numbers, also created with a Chicago area code. All phone calls with data brokers took place online via Google Voice and, to the best of our knowledge, did not create a cellular phone record.

Some data brokers asked us to have a Zoom or Google Meet call instead of a traditional phone call, which was answered from a Duke University-owned Chromebook connected to a VPN and with the video camera off. However, one broker recorded the Zoom call without our consent (we could tell based on the recording announcement and presence of the recording icon in the Zoom call), potentially creating a traceable record.

Creation of U.S. Domain

To facilitate email contact with U.S. data brokers, we purchased a new U.S. domain. We feared that brokers would be suspicious of a simple Gmail account; therefore, we purchased the U.S. domain "datamarketresearch.org" and set up an intentionally vague, associated email address. All U.S. domain email contact with data brokers was done through this email address while logged into the Chicago-based VPN from recently wiped Chromebooks.

Phase 3 Methodology: Buying Servicemembers' Data via *.asia* Domain

For contacting U.S. data brokers from a Singaporean IP address, we purchased the domain "dataanalytics.asia." We again created an intentionally vague email address using this domain, which was then used to contact the brokers. Each time, before accessing the email account, we activated the Singapore-based VPN on another one of the Chromebooks.

The use of the *.asia* domain did not require either a feature phone or a Google Voice number. All contact for the *.asia* portion of the research was done over email or via Google Meet / Zoom calls scheduled by the brokers. Our team was physically present in the United States for all of those contacts. During phone calls, we provided only our first names and did not immediately mention any affiliation with Duke University, until and unless asked, instead broadly (and truthfully, non-deceptively) stating that we were researchers performing market research on consumer data and U.S. military personnel. (As mentioned below, we stated clearly on the *.asia* domain website that we are affiliated with Duke University, although we have no evidence that the brokers we contacted looked at that website.)

Before engaging in foreign communication, we submitted the aforementioned IRB protocol, after which Duke's Office of Research Support determined that this research was not human subjects research and did not require that the project file a full application with the Duke University IRB. Then, to aid our research efforts, we created a simple, professional website on the "dataanalytics.asia" domain. The website was primarily composed of placeholder text, used broad language to describe a general data research group, explicitly stated that we are affiliated with Duke University, and included a disclaimer that the site was currently under construction. To prevent data collection, we programmed the website rather than use a website-building service. The website was also hosted on an Amazon web server located in Singapore, giving it a Singaporean IP address.

Once this was all configured, we contacted the three brokers that had sold data to us via a U.S. domain, along with one geolocation broker, using a *.asia* domain name and a website hosted on a Singaporean IP address. Our team selected Singapore in our initial grant proposal because of its tech industry and important geopolitical position between the U.S. and China. All of the brokers responded to our requests. We purchased datasets from three brokers and cut off communication with one broker after they requested that we sign an NDA.

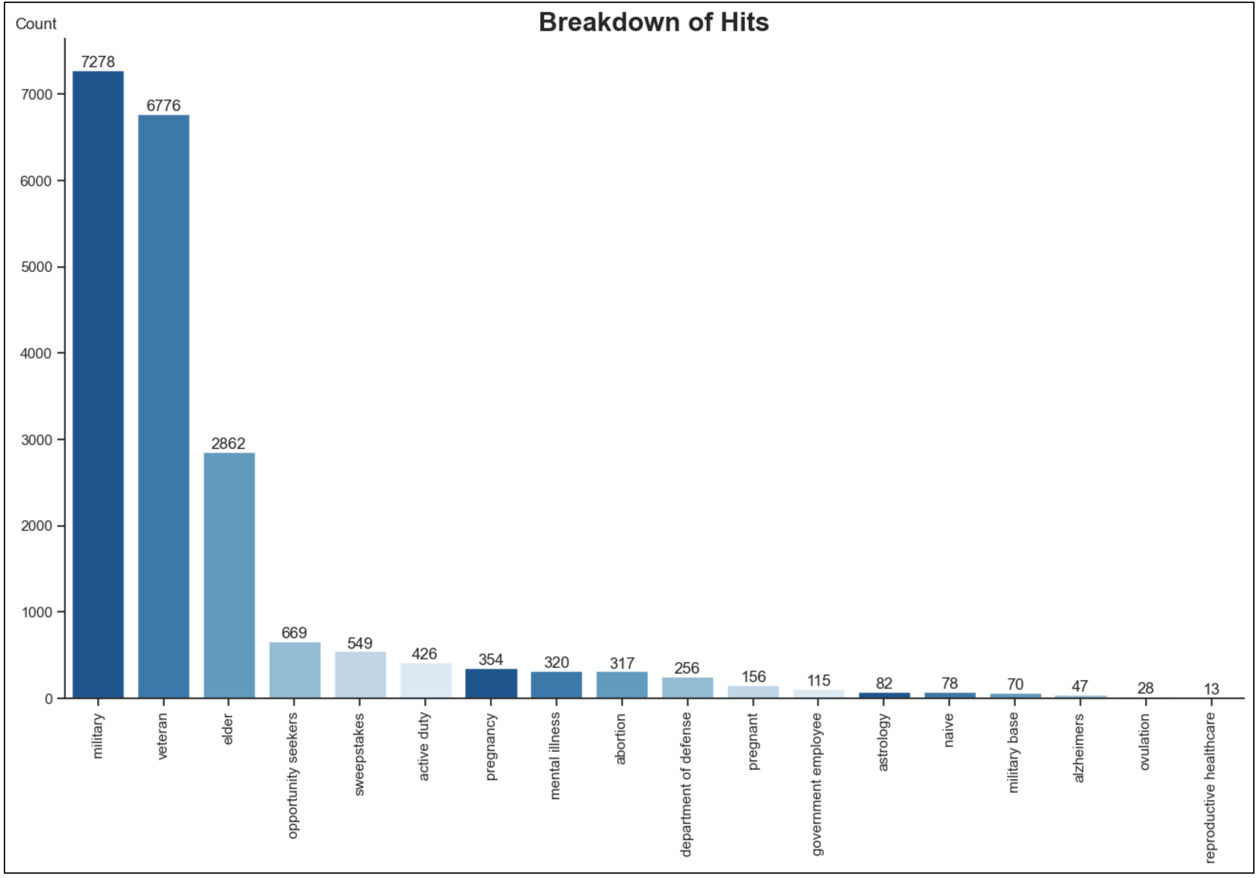
Phase 1 Results: Scraping Data Broker Websites

For the first phase of our project, we scraped hundreds of data brokers' websites in order to assess their advertisement of data about military servicemembers and identify brokers from which we might potentially purchase data.

Military Data

Terms related to the military were by far the most common out of our list of terms that we found on data brokers' websites—"military" had 7,278 hits and "veteran" had 6,776 hits across the collective 533 data broker websites. Figure 1 provides an overall count of each key term, including others that are military-related, such as "active duty," and others that are not military-related, such as "elder," but which reflect other Duke data brokerage research project workstreams examining data brokers and the harms and risks to individuals. It is important to note that this is not an exhaustive list of key terms; rather, it is the list of terms that we decided would be most useful to our project and to the other data brokerage research projects ongoing at Duke.

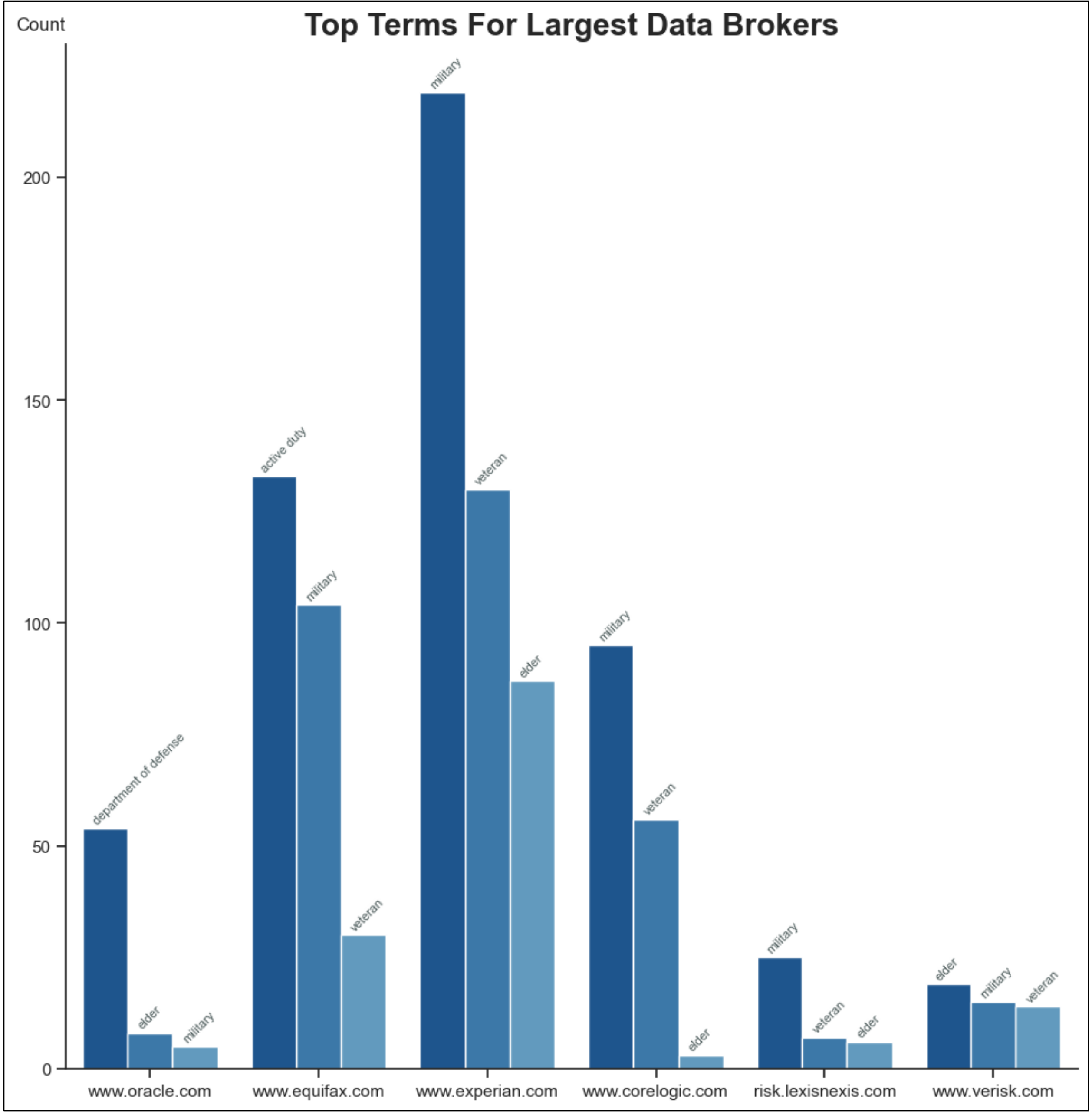
Figure 1: Searches of 533 Data Broker Websites — Key Term Hits



Many data brokers advertise data related to U.S. military servicemembers, often supplemented by characteristic, habit, and/or demographic data. For example, many brokers advertise datasets pertaining to U.S. military servicemembers' hobbies, careers, and interests. These include datasets with such titles as "Veterans that own a motorcycle," "military readers," and "Veteran Owned Construction Companies." One broker advertises data on veterans who are "responsive…to one or a variety of causes" (the causes were not specified), claiming that "[g]iving back and helping others is something that is in their heart & soul…[y]ou can already see them opening up their wallet for your offer today." In addition, the broker advertises other demographic information related to those individuals, including "their branch of service, income, age & even gender." One people search website—a type of data broker that scrapes public records and makes them available for search and sale online—describes its ability to find a deceased "Veteran's claim or discharge number" by searching death records. Overall, data on military personnel is widespread online and could easily be discovered via a search engine.

In Figure 2, we filter down our results to only include the top three most frequently discovered terms across six major U.S. data brokers' websites—those of Oracle, Equifax, Experian, CoreLogic, LexisNexis, and Verisk. These companies were selected because they are some of the most prominent data brokers in the United States, as discussed in a previous report from the Duke data broker team, *Data Brokers and Sensitive Data on U.S. Individuals* (2021).[40] Terms like "veteran," "department of defense," "active duty," and "military" are often found on these companies' sites. This chart also includes other terms like "elder" that are not the focus of this study but were included in the analysis because they relate to other data brokerage research project workstreams; they are described below only because they appeared in the data about most-used terms.

Figure 2: Searches of Six Major Data Broker Websites — Key Term Hits



Brokers also advertise their ability to provide up-to-date records. One broker advertises "18,000,000 verified military veterans representing the largest veterans marketing list on the market! And the veteran leads come with a money-refund quality guarantee. All veteran mailing list names and addresses are updated – verified on a monthly basis." Many other brokers include similar claims concerning the authenticity of the data they sell on military servicemembers and other populations, although it is unclear how they are supposedly verifying their data. In some instances, descriptions of military personnel were seemingly used in reference to both active-duty servicemembers and veterans, the latter of which may still know currently classified information and therefore be at risk of blackmail—but of course do not have a current role in the military, unlike active-duty personnel.

Many brokers advertise data not only on veterans or military personnel, but on related people and businesses. One data broker advertises data on companies near military bases: "Communities that are heavily connected to the nearby military base. Every restaurant, hotel, gas station, household, church, etc [sic] has a connection to the military base." Another data broker advertises a list entitled "Federal Civilian and Military Purchasing Officers Mailing List." Some data brokers advertise data on military families as well, such as "Military Families Mailing List" and "Hard Core Military Families."

Several data brokers advertise their ability to quickly organize data into different packages based on the prospective buyer's requests. One data broker states that its data can be organized by "branch of service, geographic location, age, sex, ethnic origin and income." Many advertised mailing lists are state-specific, such as one dataset titled "Active Military Personnel in Alabama."

Phase 2 Results: Buying Servicemembers' Data via U.S. Domain

As part of this study, we went through the data purchasing process ourselves to better understand what data on military servicemembers is available for purchase on the open market, and what guardrails (or lack thereof) are in place around its sale and use. This section describes our findings from purchasing data via a U.S. domain on members of the U.S. military and military families.

Figure 3: Contact and Sales Process for Contacted Data Brokers

Contacted Data Broker	Contact and Sales Process
Broker 1	We attempted to purchase data directly through the website without contacting a representative; after accidentally using a credit card flagged for fraud for unrelated reasons, the company rejected our purchase (did not send data) and stopped responding to communications.
Broker 2	Broker said via email it only works "with verified companies and contacts in order to prevent fraud"; we stopped communicating.
Broker 3*	We reached out via email; broker wanted sample marketing materials, to which we said we would not contact anyone in the data and did not have any; we then bought data.
Broker 4*	We reached out via email; broker initially asked to verify identity over phone but provided us with option to skip identity verification by purchasing via wire; we then paid by wire, skipped the referenced identity verification process, and received the data.
HBroker 5	Broker required a phone call before moving to the purchasing phase; we had a call, on which representative expressed concern about lack of (i) website linked to our domain and (ii) "existence of a legitimate organizational entity" for our group; we then stopped communicating; broker also required a marketing sample.
Broker 6*	We purchased data via email; we inputted a university mailing address on purchase order but were not asked to provide any other information.
Broker 7	We reached out via email; we did not proceed since advertised data seemed potentially similar and broker asked us to fill out a W9 form.
Broker 8	Broker required a phone call before moving to purchasing phase; we had a call, and it seemed open to using location data to track military personnel; after the call, it stopped replying to follow-up emails and calls on our inquiry (unclear why).
Broker 9	Broker required a phone call before moving to purchasing phase; we had a call, and broker said it said it was not able to directly identify active-duty military personnel for us, nor provide us with location data around military bases.
Broker 10	Broker required a phone call before moving to purchasing phase; we had a call; we did not proceed since advertised data seemed potentially similar

	and broker required us to sign an NDA before providing a data dictionary of data options.
Broker 11	Broker required a phone call before moving to purchasing phase; did not have call since advertised data seemed similar and broker required us to sign an NDA before speaking.
Broker 12	Broker required a phone call before moving to purchasing phase; did not have call since advertised data seemed similar

* Broker from which we ultimately purchased data.

Verification Process

Overall, we encountered inconsistent identity or data usage verification throughout our data purchasing process. Five of the 12 data brokers that we contacted from the U.S. domain, including the three brokers that we purchased data from (Brokers 3, 4, and 6), did not have any verification process apart from asking us to provide payment information on a form. (It is unclear to what extent this served an identity verification function as opposed to mere payment processing.) Broker 4 initially asked to verify our identity over the phone, but then said we could skip its identity verification process if we paid with a wire transfer instead of credit card. We then paid via wire instead of credit card, and the data broker provided us with the data we requested on members of the U.S. military without asking about or verifying our identity. This suggests that, at least for some data brokers, verification processes are more about ensuring that they get paid than they are about identifying purchasers and understanding the potential risks associated with that purchaser acquiring and using data. The lack of verification process is supported by Broker 4's entry with the Vermont data broker registry, which indicates that there is "no" credentialing process.

We had a similar experience with Broker 1, when we accidentally used a credit card that had been flagged for unrelated fraud to complete a purchase. After the card was rejected, the broker stopped communicating with us despite our repeated attempts to follow up. Broker 1 never attempted to verify our identity but cut off business when our payment did not go through.

Around half of the data brokers that we contacted had some sort of verification process, ranging from a required phone call (on the less in-depth end) to verification of what they called our "company's" status (on the more in-depth end). Six of the 12 data brokers required us to have a phone call with them before moving on to the purchasing phase (Brokers 5, 8, 9, 10, 11, and 12), which are described in more detail below. These phone calls seemed to serve a dual purpose of the broker conducting some verification alongside providing sales and marketing information about their services and datasets.

Two brokers did explicitly push back against selling to us based on our lack of a website and the fact that we were not a "verified" company. Broker 2 wrote in an email, "We're a B2B company only working with verified companies and contacts in order to prevent fraud." After our phone call with Broker 5, the company representative voiced concern about our lack of (i) a website linked to our domain and (ii) the "existence of a legitimate organizational entity" for our group, writing that "You'll need to make it absolutely clear that what you plan to do with the data isn't illegal or nefarious." The Broker 3 and Broker 5 representatives also asked us to share specific examples of our "marketing materials" or a mail piece that would be sent to individuals in the dataset, which they wanted to review before providing us with the data. These identity verification and data use provisions were interesting to see from some data brokers. Nonetheless, we were able to do business with other brokers selling similar data that did not require a verification process. For Broker 3, from which we did purchase a dataset, we communicated via email that we were not planning to contact any of the individuals in the datasets we purchased. Broker 3 then did not require us to submit a mail piece.

In summary, the apparent controls in use by data brokers ranged from one location data broker refusing to sell geolocation data on "sensitive" areas (including military sites), to brokers appearing to follow simple check-the-box phone scripts, to brokers asking for more information about prospective customers before proceeding with a buy process, to brokers not performing identity verification functions before proceeding with the purchase data.

Phone Calls

Six of the 12 brokers that we contacted via the U.S. domain required us to have a phone call before moving on to the purchasing phase (Brokers 5, 8, 9, 10, 11, and 12), and of that group, we initiated phone calls with Brokers 5, 8, 9, and 10. We opted not to have calls with Brokers 11 and 12 because their data seemed comparable to that of other brokers we contacted, and Broker 11 required a nondisclosure agreement in order to share any more information about its data. (We did not receive the NDA and stopped responding after they asked for information in order to produce the agreement.)[4] The phone calls with brokers 5, 8, 9, and 10 were helpful for learning about the data brokers' respective data collection processes (as the brokers described them) and the data available for purchase (as the brokers described it), although we did not end up purchasing data from any of the brokers with which we had phone calls.

We chose to have phone calls with Brokers 8 and 9 based on the location data services described on their websites. The information from these phone calls was particularly instructive because the data brokers differed significantly in their sales pitches and their approaches to data sales.

Broker 9 communicated that it did not sell location data around "sensitive" locations including military bases, schools, government buildings, hospitals, and abortion clinics, although it was unable to clarify how wide the unavailability radius was for these locations or whether it still collected this information (even if allegedly not selling it). Broker 9 explained that it could overlay other data with location data, including demographics, "supply and demand," per person spending, healthcare data, social media, and more based

[4] The data broker asked us for the following information in order to generate the NDA: legal name of organization, organization type, state of incorporation, billing address, name/title/email of signatory, billing contact and email, and any authorized end users.

on the home location of a given device (these terms were used by the broker and it was not clear what some, such as "supply and demand," mean). Broker 9 also recorded the call; although the automated Zoom voice said "recording in progress" out loud, the people on the other end of the call did not explicitly ask us if they could record the call before they began doing so without our consent. As previously mentioned, we could tell that we were being recorded based on the recording announcement and presence of the recording icon in the Zoom call. We ended up not proceeding with a purchase from Broker 9 because it said it was not able to directly identify active-duty military personnel for us, nor provide us with location data around military bases.

Broker 8 was more open to the idea of using location data to track military personnel. In fact, a company representative explained on our call that the company worked entirely on data related to national security, defense, and police, typically through contracts rather than directly with government agencies. Broker 8 explained that it would be easy to deduce where individuals lived, including on a military base, based on the home location of a device.

Both Brokers 8 and 9 acknowledged using mobile device location data from software development kits (SDKs) built into apps, but claimed that data was only collected after an "opt-in" process from the user and that the data was "anonymized" with a device ID rather than the name of an individual. (See the above discussion in the previous section of "anonymization.")

We also chose to have phone calls with Brokers 5 and 10, which specialized in contact lists and demographic data rather than location data. We elected not to proceed with a purchase from Broker 10 because it asked us to sign an NDA before it would provide us with a data dictionary that sounded quite similar to other data dictionaries we had received from brokers without an NDA or verification process.

After our phone discussions with Brokers 5 and 8, we were interested in purchasing data, but both brokers indicated that they did not want to continue doing business with us. Broker 5 seemed suspicious of our identity based on our lack of a website and "existence of a legitimate organizational entity," while Broker 8 cut off all contact by neglecting to reply to several follow-up emails and phone calls that we sent and made after the initial call.

NDAs and Data Sharing Restrictions

Most data brokers that we contacted, including the three that we purchased data from (Brokers 3, 4, and 6), did not ask us to sign an NDA and had very few restrictions on how purchased data could be used or shared. Another data broker asked for more specific details about how the data would be used, and two others said that they were unable to share more information with us without an NDA, as described below.

Broker 5 asked for specific details about who the data or subsequent research would be shared with before pursuing a sale, asking, "Is your intent to make the underlying data open to the public or shared with stakeholders (investors or policymakers) other than the purchaser?" and "Will this research be published?" It also asked for sample marketing materials or content that we would be sending to or sharing with the individuals encompassed in the dataset we were looking to purchase. (Note that Broker 3, which we did purchase from, also asked us for a mail piece when purchasing health data. However, we were not planning to contact—and did not contact—any of the individuals in any of the datasets we purchased, and we told Broker 3 as much. It then did not require us to submit any such materials.)

Brokers 10 and 11 were unable to share more specific information about the data they sell, including a data dictionary, without our team signing an NDA. A representative from Broker 11 explained: "Without appropriate background context and an NDA in place it's a bit difficult for me to share information." Again, we were able to purchase some data from other data brokers about active-duty military personnel, and the brokers we purchased from did have confidentiality terms in their sale conditions or purchase orders, specifically prohibiting activities like resale of the data. We could not compare the exact list of available data fields from Brokers 10 and 11 with those of the brokers we purchased from; our phone call with Broker 10 indicated that the fields were similar in nature to those that we had seen from other brokers, but it is not entirely clear.

Data Cost

The datasets purchased with the U.S. domain from Brokers 3, 4, and 6—all of which were clearly identifiable and included names and contact information—ranged in cost from $0.125 to $0.22 per military servicemember, depending on the provider and the set of variables included.

We also received cost estimates from several other brokers from which we did not end up purchasing data. Ironically, Broker 1's quote included a $25 "Privacy Fee," but it was not clear what the purpose of the fee was or what it covered. For Broker 7, the quoted cost per record increased depending on which variables we wanted to purchase. Broker 7 offered to sell us the names, email addresses, and phone numbers of 3,980 active-duty military personnel in DC, MD, and VA for around $0.21 each. After we asked about other information that could be purchased on those servicemembers, the broker told us that adding additional variables such as birth date, gender, income, political donations, foreign investments, and number of children would bring the cost up to around $0.31 per record. See Broker 7's cost breakdown for additional variables below:

- Age: +$5/thousand
- Birth Date: +$5/thousand
- Gender: +$5/thousand
- Income: +$10/thousand
- Occupation: +$15/thousand
- Political Donations: +$5/thousand
- Foreign Investments: +$5/thousand
- Casino/Gambling: +$5/thousand
- Credit Rating: +$15/thousand
- Net Worth: +$10/thousand
- Family Size: +$5/thousand
- Number of Children: +$5/thousand

The lowest quote that we received via the U.S. domain for a dataset with contact information on military personnel was over the phone from Broker 10, at only $0.05 per servicemember record with email. Broker 10 said it would cost us $0.08 per military servicemember to also include those individuals' cell phone numbers, with a minimum $1,000 order. We ended up not purchasing from this broker due to its more restrictive NDA requirements.

Payment Methods

The three brokers that we purchased data from all provided a wide variety of payment options to make payment as easy as possible. Broker 3 allowed us to pay by wire/ACH transfer, physical check, or credit card. Broker 4 initially asked us to pay by credit card and have a phone call to verify identity, but then said we could skip this step if we paid with a wire transfer instead of credit card, which we did. Broker 6 provided the options of credit card, PayPal payment, or bank transfer, and we chose to pay by credit card.

When making payments, all brokers required a name and billing address for a purchase invoice, so we provided the name of someone from our team and the address of Duke's Sanford School of Public Policy, without explicitly indicating a Duke affiliation. Only Broker 4 asked us to confirm that we provided a Duke University billing address, but still sold data to us via wire transfer (and, again, let us skip its identity verification process in doing so).

Accuracy Claims Made by Data Brokers

Of the data brokers that we contacted, Broker 3 made the most concrete claims about accuracy, claiming that it could "provide a minimum of 90% accuracy on our data and any records with inaccurate info exceeding that would be replaced by us." When asked about the process of verifying the data itself, Broker 3 shared that it verifies the emails "using professional email verification services such as Zerobounce, Data Validation, Bounceless, Etc." and removes any invalid ones before sending the data. However, it did not seem like any of the other data fields (names, demographic information, etc.) were subject to a verification process.

Other brokers made more vague claims about the level of accuracy and verification within their datasets. Broker 6 said that if datasets were not to the purchaser's 100% satisfaction, that it would fix the data. It did not elaborate further.

Data Delivery and Storage Methods

All of the brokers that we purchased data from (Brokers 3, 4, and 6) initially provided data via an online file transfer protocol (FTP), meaning they digitally uploaded and transferred the purchased dataset to us directly, either in *.xlsx* or *.csv* format. Interestingly, Broker 6 originally provided an Excel file through one such FTP, but when we noticed that several variables were missing from the dataset compared to what we had purchased, Broker 6 subsequently updated the file and simply sent it via email attachment, rather than reuploading the file to the FTP. Providing data via email rather than a secure FTP opens up avenues for the data to be intercepted or accessed by an unintended party. Broker 4 incorporated its file transfer system into its own online data portal, which also included tools to search for additional datasets to purchase.

We downloaded purchased data only to the Duke University-owned Chromebooks via VPN, then uploaded it to a secure Duke OneDrive folder. The datasets were only accessible by these authors on the data brokerage research team and a few select research supervisors and staff. We performed all analysis of identifiable data in Excel within the secure OneDrive folder, and only aggregate tabulations were downloaded locally for chart-making or other analysis purposes. Once research was completed, all data files were deleted from the Chromebooks.

Efforts Toward Continuous Business Relationship

The three data brokers from which we purchased data expressed interest in continuing a business relationship with us. Broker 3 expressed its thanks "for giving us yet another opportunity to work with you" and said that "we look forward to a continued business relationship." Broker 4 followed up at least three times throughout the purchase process and after we had already purchased data to ask if we wanted to buy any more data, including more data on U.S. military servicemembers. Broker 6 was less persistent, but provided a $150 voucher toward a future purchase after our original dataset was missing a variable that we had requested.

Six of the data brokers appeared to subscribe our domestic purchasing email to their email lists, as evidenced by our continued receipt of regular newsletters from brokers 3, 4, 5, 9, 10, and 12, including weekly updates from some of those six brokers advertising their products.

Phase 3 Results: Buying Servicemembers' Data via *.asia* Domain

Once we had purchased five datasets from three data brokers via the U.S. domain, we contacted those same three data brokers again through the *.asia* domain (along with one additional location data broker) to attempt to buy the same data. We were able to purchase data from Brokers 3, 4, and 6 via the U.S. domain with minimal verification and no required NDA. To investigate if this process changes when data brokers interact with purchasers with websites outside the U.S., we attempted to purchase another dataset using the *.asia* domain and Singaporean IP address. We also made a second attempt to purchase data from Broker 8, the location data broker, since it had abruptly ended communication with us from the U.S. domain. Ultimately, we were able to purchase similar data from Brokers 3, 4, and 6 via the *.asia* domain as to the U.S. domain with negligible differences in the purchasing process.

Verification Processes

Despite our use of a *.asia* domain, we experienced a similarly minimal verification process as we did when purchasing data using the U.S. domain. Only one of the four brokers contacted expressed concern over who we were.

Broker 3 required multiple contact attempts before responding to our email. Once a sales representative responded to our request, we were able to purchase data with very little difficulty. Broker 3 posed no substantial questions about our identity and appeared to have no barriers to selling to our *.asia* domain. At one point, Broker 3 requested to schedule a phone call; we declined, and it proceeded to sell us data anyway.

Interestingly, Broker 3 asked for a sample mail piece from the U.S. domain, but did not from the *.asia* domain. In both cases, we did not provide a sample mail piece and were sold data. Broker 3 also offered more data fields for purchase when we reached out via our U.S. domain than it did when we reached out via our *.asia* domain. Whereas data purchased from the U.S. domain included fields related to health conditions or specific military branches, Broker 3 only offered identifiable contact information to the *.asia* domain. It is unclear if this is due to internal restrictions or because we worked with a different sales representative.

Broker 4 was the only broker that imposed a restriction on the sale of some data fields to unverified customers. The sales representative from Broker 4 stated that they were "having a hard time internally as you're asking for a lot of outputs that we can only do in certain situations and we can't see your site or vet your company and it's a *.asia* domain." Nonetheless, the broker did not attempt to schedule a phone call with us, as it had attempted to do when we contacted from the U.S. domain. Broker 4 asked what we intended to do with the data (for example, "direct cold emails, direct postal mail, [or] online matching[?]") but accepted a general answer that we did not plan to contact people directly and were instead performing market research. Later, Broker 4, in an attempt to entice us to purchase more data, asked us "[y]ou're not marketing to the data so why not just take 100k [records] at once?" Broker 4 sold us a dataset that included all of our requested fields except "financial [fields] like income or summarized credit."

Broker 6 never questioned our identity, attempted a phone call, or asked what we planned to do with the data, allowing us to purchase the dataset quickly and easily.

Similar to our experience via the U.S. domain, Broker 8 required a phone call and an NDA before selling data. We did not attempt to purchase data from Broker 8 because of the NDA requirement.

Phone Calls

We had only two phone calls through the Singaporean domain, both with Broker 8. Broker 8 scheduled only one phone call when we contacted it from the U.S. domain before ceasing communication.

During the phone calls, Broker 8, a geolocation broker, told us that the company offers two types of services—raw geolocation data and data enhancement services. Broker 8 stated in the phone call that the raw geolocation data draws on a massive database of GPS coordinates tied to an "anonymous" mobile ID. The database ranges from three years ago to five days before the current day and is composed of GPS coordinates collected via SDKs when an individual opens a partnered application on their phone. Broker 8's data enhancement services, on the other hand, are only available to customers who already possess datasets. Broker 8 claimed that, if we were to provide them with a labeled dataset, such as one that we had purchased from another broker, Broker 8 would add additional fields such as habits or interests to the data, inferred from the GPS data, but would not provide GPS data. This implies that, despite collecting data via mobile IDs, Broker 8 is internally able to match IDs to names.

NDAs and Data Sharing Restrictions

After the phone calls, Broker 8 stated that additional information on pricing and purchasing would require an NDA, at which point we ceased communication. Brokers 3, 4, and 6 did not require an NDA and had minimal data sharing restrictions before sending us our purchased data, equivalent to the U.S. domain.

Data Cost

The pricing schemes were overall similar to those offered to the U.S. domain. Broker 3 was the most expensive data broker, charging us $0.32 per individual for data on 5,000 servicemembers and veterans for a total price of $1,600. Broker 3 did not send us its pricing scheme, instead providing a quote.

Broker 4 sold us data for $0.12 per individual, charging $600 for a dataset of 5,000 servicemembers and veterans. Broker 4 sent us the below pricing scheme for datasets involving validated consumer emails. Broker 4 never told us a consistent pricing scheme for additional data fields or smaller purchases. Broker 4 also asked if we needed the emails "cleaned and validated" but did not state if this service came at an additional cost.

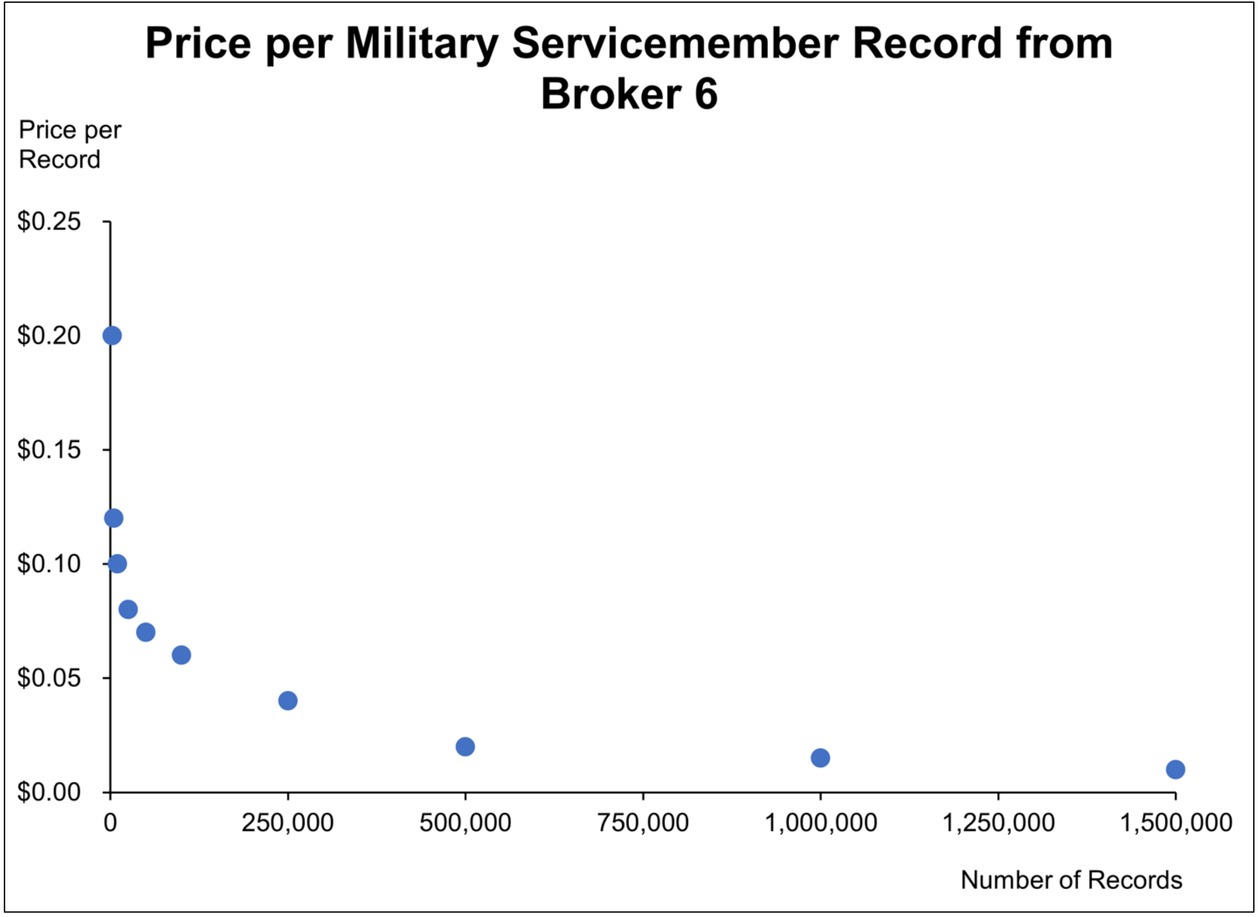
Broker 6 sold us data for $0.25 per servicemember or veteran, for a total cost of $1,262 for 5,048 individuals. Broker 6 said that additional fields could be added for $0.005 per field. Broker 6 also included the below pricing schemes, seemingly for basic "leads" (names, emails, and addresses of individuals). At large quantities, Broker 6 offered us leads for as low a cost as $0.01 per individual.

Figure 4: Price per Military Servicemember Record from Broker 6 (Table)

Number of Servicemembers / Veteran	Price per Servicemember / Veteran
2,500	$0.20
5,000	$0.12
10,000	$0.10
25,000	$0.08
50,000	$0.07
100,000	$0.06
250,000	$0.04
500,000	$0.02
1,000,000	$0.015
1,500,000+	$0.01

This price per record distribution is visualized in Figure 5 below.

Figure 5: Price per Military Servicemember Record from Broker 6 (Visualized)



Payment Methods

Both Brokers 3 and 4 accepted payment via wire. We initially asked Broker 6 to allow us to pay by wire, which it agreed to do. However, we later asked to pay by credit card due to payment processing issues on our end. Broker 6 sent us a PayPal invoice over email and supplied the data within two days of purchase. Duke's mailing address was listed on all wire forms.

Accuracy Claims Made by Data Brokers

Similar to what we heard from brokers contacted from the U.S. domain, the brokers contacted from the *.asia* domain made claims about "validation" and "verification." Broker 3 referred to its datasets as "verified records" in its emails. Broker 4 alluded to a validation and cleaning of email addresses when discussing price plans. Broker 6 provided a thorough email that advertised its accuracy and validation services as "spectacular details" about the data. Broker 6 claimed that "each and every lead is verified and vetted to match your exact criteria." Broker 6's validation procedures include "hard bounce verification cleanse," removal of IDs that cause "transmission issues," "spam trap removal," "removal of honey pots, black boxes, deep pulls, ghost accounts, and traps," and "full legal compliance." The email detailing validation procedures had several grammatical errors and technical jargon that was not clearly understandable, such as “complainer database scrub.”

Brokers 3, 4, and 6 implied that we purchased only a small subset of their database on military servicemembers and veterans. Broker 3 claimed to have 1,502,394 records about military servicemembers and veterans that included only email and 573,498 records that included email, phone, and address. Broker 4 claimed to have 587,595 “lead prospects” on active-duty servicemembers and 1,997,878 “lead prospects” on veterans. Broker 6 claimed to have emails for 107,372 active-duty servicemembers and 4,251,203 veterans.

Data Delivery and Storage Methods

Brokers 3, 4, and 6 delivered the data through FTP via a download URL to a secure platform. Datasets were all either *.csv* files or Excel files, and Broker 6 provided an accompanying “decoder” that defined variables used in the dataset.

Phase 4 Results: Analysis of Purchased Data

Types of Data Available on Military Personnel

The 12 data brokers that we contacted offered two primary types of datasets on military personnel: (i) lists of identifiable contact information and demographic details; and (ii) more granular location data. Due to cost, availability, and verification concerns, all eight datasets that we purchased fell into the first category. Some data brokers offered pre-made lists targeting certain populations (for example, "Military Personnel and Officials"), while others designed lists for us based on our exact specifications including location, military affiliation, and availability of certain demographic variables.

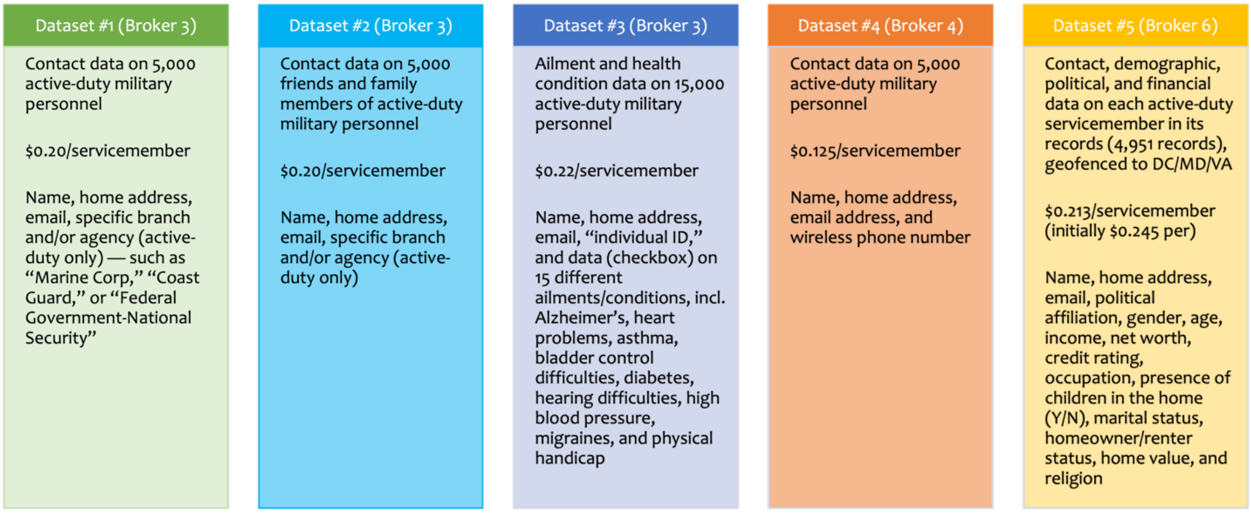
As part of their sales process, multiple data brokers sent us lists of hundreds of demographic variables that we could select from within their consumer database. Broker 5 sent us an Excel file with around 740 available variables for purchase within its "core" consumer database. Broker 7 emailed us a 31-page brochure detailing around 400 variables available for purchase, including everything from "Gun Enthusiast," "Assimilation Code" (English speaking, bilingual, or unassimilated), "Ethnic Code," whether they had air conditioning, mortgage amount, what type of cooking and travel they were interested in, "Smoker/Tobacco," which causes they donated to, age and gender of children, which credit card they have, and much more, allowing a purchaser to build very targeted profiles. Broker 4 likewise sent an Excel file of 475 variables available for purchase, including "Casino," "Sweepstakes / Contests," ethnicity, occupation, clothing size, political donations, "Charitable Donor," "Christian Families," among many others. Finally, Broker 6 sent a list of 169 data fields which included similar variables, including demographic information such as language, religion, and ethnicity.

Some brokers also provided us with aggregated data before making a purchase. For example, when we expressed interest in filtering list data by geography, Broker 6 provided a breakdown of aggregated totals and other information by U.S. state (this breakdown was provided when we reached out with both the U.S. and *.asia* domain). Broker 4 provided a total number of records that it owns for veterans and active-duty military personnel—1,997,878 and 587,595 respectively—as well as how many records have emails or phone numbers associated with them.

From the U.S. domain, we purchased 34,951 identified, personal contact records from three data brokers for a total cost of $6,931.69 (on average, under $0.20 per record). From the *.asia* domain, we purchased 15,048 identified, personal contact records for a total cost of $3,362 (on average, approximately $0.22 per record). The following sections provide more specifics on the datasets and demographic details purchased from each broker.

The data purchased from data brokers via the U.S. domain—as well as the cost and scope of the data—is captured in Figure 6.

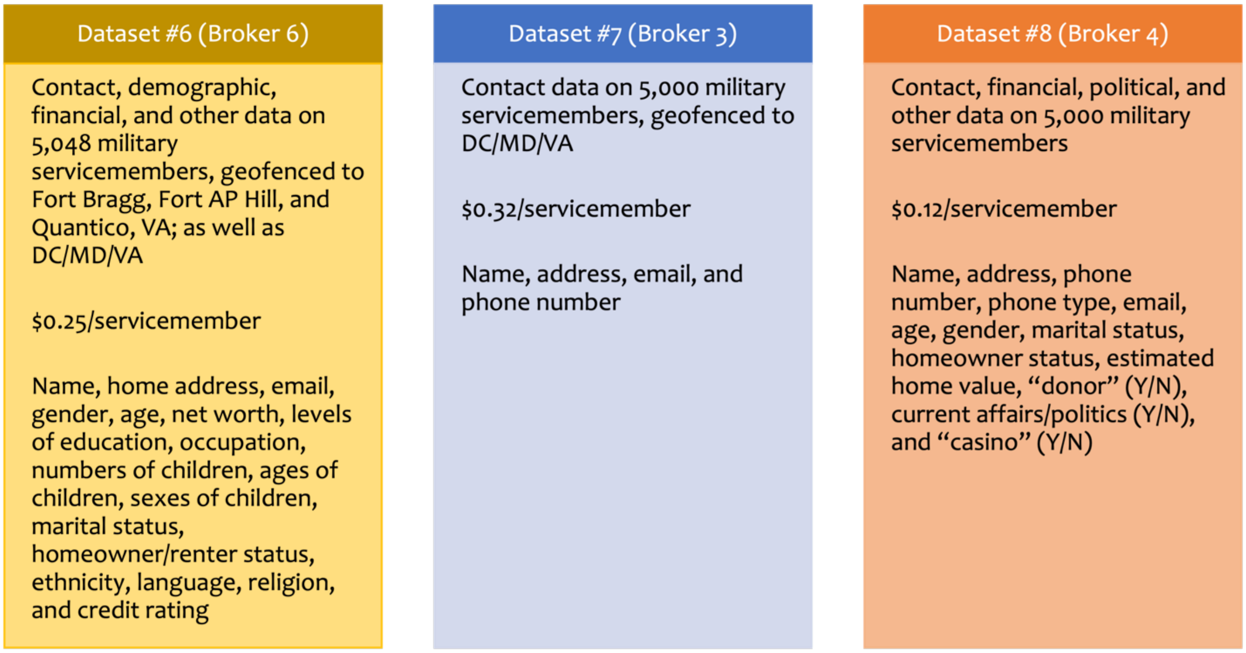
Figure 6: Data Purchased from Data Brokers via U.S. Domain



Dataset #1 (Broker 3)	Dataset #2 (Broker 3)	Dataset #3 (Broker 3)	Dataset #4 (Broker 4)	Dataset #5 (Broker 6)
Contact data on 5,000 active-duty military personnel	Contact data on 5,000 friends and family members of active-duty military personnel	Ailment and health condition data on 15,000 active-duty military personnel	Contact data on 5,000 active-duty military personnel	Contact, demographic, political, and financial data on each active-duty servicemember in its records (4,951 records), geofenced to DC/MD/VA
$0.20/servicemember	$0.20/servicemember	$0.22/servicemember	$0.125/servicemember	$0.213/servicemember (initially $0.245 per)
Name, home address, email, specific branch and/or agency (active-duty only) — such as "Marine Corp," "Coast Guard," or "Federal Government-National Security"	Name, home address, email, specific branch and/or agency (active-duty only)	Name, home address, email, "individual ID," and data (checkbox) on 15 different ailments/conditions, incl. Alzheimer's, heart problems, asthma, bladder control difficulties, diabetes, hearing difficulties, high blood pressure, migraines, and physical handicap	Name, home address, email address, and wireless phone number	Name, home address, email, political affiliation, gender, age, income, net worth, credit rating, occupation, presence of children in the home (Y/N), marital status, homeowner/renter status, home value, and religion

Figure 7 below captures the data purchased from U.S. data brokers, as well as the cost and scope, via the *.asia* domain.

Figure 7: Data Purchased from Data Brokers via *.asia* Domain



Dataset #6 (Broker 6)	Dataset #7 (Broker 3)	Dataset #8 (Broker 4)
Contact, demographic, financial, and other data on 5,048 military servicemembers, geofenced to Fort Bragg, Fort AP Hill, and Quantico, VA; as well as DC/MD/VA	Contact data on 5,000 military servicemembers, geofenced to DC/MD/VA	Contact, financial, political, and other data on 5,000 military servicemembers
$0.25/servicemember	$0.32/servicemember	$0.12/servicemember
Name, home address, email, gender, age, net worth, levels of education, occupation, numbers of children, ages of children, sexes of children, marital status, homeowner/renter status, ethnicity, language, religion, and credit rating	Name, address, email, and phone number	Name, address, phone number, phone type, email, age, gender, marital status, homeowner status, estimated home value, "donor" (Y/N), current affairs/politics (Y/N), and "casino" (Y/N)

The team was able to purchase similar, individually identified data through the *.asia* domain, including datasets geofenced to Fort Bragg; Fort AP Hill; Quantico, Virginia (home to a Marine Corps base, among other facilities); and Washington, DC, Maryland, and Virginia.

Data Purchased from Broker 3

We purchased three datasets from Broker 3 from the U.S. domain and one from the *.asia* domain. The first U.S. dataset contained 5,000 identifiable contact records for active-duty military personnel, the second U.S. dataset included similar information for 5,000 individuals within veterans' households, the third U.S. dataset contained 15,000 identifiable records with health information and ailments, and the fourth *.asia* dataset contained 5,000 identifiable contact records seemingly containing both active-duty military personnel and veterans.

The first and second dataset included the name, address, and email of each individual at a rate of $0.20 per individual. The active-duty military dataset also included the specific branch and/or agency of each individual. Combined, the datasets included individuals living in all 50 U.S. states. Maryland and Virginia were most represented at 906 and 898 records, respectively.

The active-duty military dataset included the specific branch and/or agency that an individual works in, labeled as their "company." The dataset included 339 unique branches or agencies, with varying levels of detail. Some records listed only a branch such as "US Coast Guard" or "US Marine Corp," while others included more specific agencies such as "US Navy Dept Naval Shipyard," "Pentagon Force Protection Agcy [sic]," or "Defense Advanced Research." The dataset also included a few non-military agencies, such as the Transportation Security Administration (TSA) and the General Services Administration (GSA).

Broker 3 sorted the personnel into four categories: "Recruiting-US Armed Forces," "Federal Government-National Security," "State Government-National Security," and "Military Bases." How the broker groups individuals into these categories is unknown.

Finally, Broker 3 included a "Title" column in the active-duty military personnel dataset with 39 unique labels. The title column groups individuals into a particular field of work, such as IT, HR, or "Operations." The title columns can also provide insight into the rank of the individual, applying titles such as "Exec Officer," "Executive," and "Manager."

The health dataset from Broker 3 included identifiable medical ailments for 15,000 servicemembers including first name, last name, suffix, address, email, function, and "individual ID" (which seems to be some form of persistent identifier), along with checkboxes for 15 categories of ailments: Allergies, Alzheimer's, Angina/heart problems, Arthritis/rheumatism, Asthma, Bladder control difficulties, Diabetes, Emphysema, Frequent headaches, Hearing difficulties, High blood pressure, High cholesterol, Migraines, Osteoporosis, and Physical handicap. The cost of this dataset was $3,250 for 15,000 servicemembers, less than $0.22 per individual.

Within the dataset, the number of ailments for a given servicemember ranged from zero ailments (possessed by around 86% of servicemembers) to all 15 ailments (possessed by one individual). The average number of ailments per servicemember was 0.3723. The frequency of each ailment within the dataset can be found in the below table:

Figure 8: Breakdown of Identified Health Data Records Provided by Broker 3 in Dataset 3

Ailment	Number of Records	Ailment (cont.)	Number of Records (cont.)
Allergies	1,543	Diabetes	237
Asthma	572	Angina/heart problems	126
Migraines	561	Bladder control difficulties	101
High blood pressure	552	Physical handicap	70
Frequent headaches	541	Osteoporosis	63
High cholesterol	495	Emphysema	37
Arthritis/rheumatism	436	Alzheimer's	11
Hearing difficulties	240		

We were able to purchase this sensitive, identifiable health information from a data broker, with very little verification, at a cost of less than $0.22 per servicemember.

The dataset sold to the *.asia* domain contained only individually identifiable contact information, such as address, phone number, and email, geofenced to specific states.

Data Purchased from Broker 4

We were able to purchase individually identifiable military contact records from Broker 4 with both the U.S. and *.asia* domain, at an even more affordable price point. From the U.S. domain, we purchased 5,000 records for active-duty military including name, address, email, and wireless phone for a total cost of $625, or $0.125 per servicemember. From the *.asia* domain, we were able to purchase 5,000 records containing both active-duty servicemembers and veterans for a total cost of $600, or $0.12 per individual. Interestingly, this second purchase via the *.asia* domain had more fields than the purchase from the U.S. domain. In addition to the fields purchased from the U.S. domain, we purchased age, sex, marital status, home ownership status, estimated home value, interest in charitable donations, interest in current affairs / politics, and a field called "casino."

The datasets from Broker 4 included records from all 50 states and the District of Columbia. When we asked Broker 4 how the sample was selected, we learned that the sample is "based on an nth select across the entire universe available. That way you get evenly distributed quantities throughout the entire universe, versus all the records in 1 single state for example." Compared with the addresses in Broker 3's dataset, which are much more evenly distributed geographically, Broker 4's addresses are heavily skewed toward a smaller number of states, especially Texas, Florida, California, and Virginia.

Data Purchased from Broker 6

We purchased data from Broker 6 using both the U.S. and the *.asia* domains.

Our U.S.-purchased dataset from Broker 6 contained not only identifiable contact information for active-duty military servicemembers, but also rich and detailed demographic information. We also limited our focus geographically for this dataset and elected to purchase the data of all 4,951 servicemembers available in the DC/MD/VA region. Although Broker 6 advertised hundreds of potential demographic variables on its website, we elected to receive the following variables in our purchase: name, address, email, political affiliation, gender, age, income, net worth, occupation, presence of children (yes/no), marital status, homeowner/renter status, home value, religion, and credit rating. We received all of this information at a total cost of around $0.213 per servicemember, or $1,056.69 for all 4,951 servicemembers' information.

Our analysis of the identifiable information purchased from Broker 6 via the U.S. domain revealed some interesting trends about military personnel in the DC/MD/VA region. 72% of the individuals in the dataset were affiliated with the Republican party, while 22% were affiliated with the Democratic party and 6% were Independent. 51% of individuals in the dataset were female while 45% were male, with the rest unknown or missing. 88% of individuals were between the age of 25 and 55. The distribution of certain characteristics in our dataset varies meaningfully from the servicemember population as a whole, potentially indicating that the data was collected from a particular segment of servicemembers rather than randomly from the servicemember population as a whole. For example, government and survey data indicates that only 34% of servicemembers identify as Republican (33% Independent), women account for about 17%, and over 80% are under age 45.[41]

We initially attempted to purchase health data from Broker 6 via the U.S. domain as well, but we were told that Broker 6 was "internally prohibited" from providing this type of data for armed services, public safety, or government contacts. As we learned from our conversation with Broker 9 on the provision of location data around "sensitive locations," it seems that some brokers do have more internal restrictions when it comes to certain populations or types of data. Broker 6's refusal to provide us with location data about the armed services underscored this fact.

Our *.asia*-purchased dataset from Broker 6 also included sensitive and identifiable demographic information on 5,048 active military servicemembers and veterans, geofenced to U.S. military bases and Washington, DC, Virginia, and Maryland. We had no trouble purchasing this data using a *.asia* domain. The dataset included the following variables in addition to names, emails, and addresses: gender, age, net worth, education, occupation, number of children, age of children, sex of children, marital status, home ownership status, home value, "money seeker," ethnicity, language, religion, and credit rating. The names of children were not provided. Interestingly, the only data fields that the broker did sell to the U.S. domain but not the *.asia* domain were political affiliation and income, but it is unclear if this is due to the foreign domain or interacting with different sales representatives.

Several of the records delivered to the *.asia* domain were located in or near U.S. military installations. Twenty-eight records had an address listed in Fort Bragg, North Carolina, and 52 in Fayetteville, North Carolina. Sixteen records were listed in Fort Belvoir, Virginia. One record was listed as Andrews Air Force Base, Maryland, and five in Quantico, Virginia. Interestingly, the dataset had two fields related to ethnicity. One was a broad ethnic group ("Hispanic," "All African American Ethnic Groups," "Western European," etc.) while the other was much more specific ("English," "Scotch," "Estonian," etc.). Similarly, the dataset had two fields related to occupation—one broad and one more specific. The general occupation category included labels such as "Craftsman / Blue Collar," and "Military," while the detailed occupation category had job titles such as "Homemaker," "Armed Forces," or "Air Traffic Control." Both categories were over 50% military occupations. The data also included several fields that contained large amounts of null values. The religion field was 52% empty. The records that were complete were overwhelmingly Protestant, along with a much smaller number of "Buddhist," "Islamic," or "Shinto" records. The "Money Seekers" field was indicated for 359 out of the 5,048 records.

The number of children field was very dense at 77% complete. Occasionally, the two fields appeared to contradict each other. For example, the number of children field might have indicated two children total for a given record, but the ages / sex of children fields indicated three children: a positive indicator on each of "Age 06–10 Male," "Age 06–10 Female," and "Age 16–17 Female." It is unclear if the data was inaccurate or only meant to be an approximation—a data broker may be unsure of a child's sex or exact age. Nonetheless, these fields represent very sensitive information that was sold to us via the *.asia* domain.

Purchased Records Were Neither Anonymized nor Aggregated

All datasets that we purchased included individual, personally identifiable information on military personnel in the United States. None of these datasets were anonymized nor aggregated, even when providing sensitive information (such as net worth, religion, or health) and without verifying the purchaser's identity. Legally, anyone with a few hundred dollars can obtain the same type of data that we did and use it for any purpose, harmful or otherwise.

Direct and Inferential Data Gathering

Some purchased variables included an inferential component, such as the marital status variable provided by Broker 6. In our 4,951-servicemember dataset from Broker 6, there were 2,823 individuals identified as "Single," with an additional eight "Inferred Single," and 2,085 individuals were identified as "Married," with an additional seven "Inferred Married."

When we asked Broker 6 how the "inferred" status was calculated, it answered that inferential variables were calculated from a "confidence rate" based on the number of unique data points available during the compilation and aggregation process. Simply put, it seems as if the broker was correlating other data points to a marriage variable, and then using those data points to predict marital status for certain individuals. In our case, the "Married" or "Single" designations had higher reliability than the "inferred" ones because they were based on more "multi-point data sources" to verify against. However, Broker 6 assured us that the "inferred" status was still very reliable.

Where Did the Data Brokers Get this Data?

Data brokers shared with us that they acquire data from a wide variety of sources and platforms in order to create their lists. These statements are summarized in the below figure. We did not attempt to verify their accuracy and are providing the level of detail given to us by the brokers.

Figure 9: Data Sources for Contacted Data Brokers (Based on Their Statements)

Contacted Data Broker	**Data Sources**
Broker 1	Did not ask
Broker 2	Did not ask
Broker 3*	Medical records; government records; surveys; healthcare directories
Broker 4*	Active military occupational data; on-base housing information; Department of Veterans Affairs mortgage data; nonprofits serving military and veteran causes; public records; utility and new phone connection records; quote forms; order forms; sweepstake forms; partnerships with list providers and others
Broker 5	Working with commissaries on military approved buyers
Broker 6*	Partnerships with over 900+ sources, including data gathered from public records, social media accounts, online purchase records, public tax documents, credit reports, national clearinghouse records, and phone/email/postal surveys; call center compilation live feeds
Broker 7	Did not ask
Broker 8	Almost entirely from SDKs
Broker 9	Partners on the app store; SDKs
Broker 10	Voter data; data from commercial sources
Broker 11	Did not ask
Broker 12	Did not ask

* Broker from which we ultimately purchased data.

According to an email from Broker 4, it receives contact list data both from official sources such as active military occupational data, on-base housing information, and Department of Veterans Affairs mortgage data, as well as nonprofit organizations that serve military and veterans causes. On its website, Broker 4 acknowledges additional sources, including public records, utility and new phone connection records, quote forms, order forms, and sweepstake forms, plus partnerships with other list providers and affiliates.

Broker 6 referenced similar sources for its contact lists and demographic data, involving partnerships with “over 900+” sources including public records, social media accounts, online purchase records, public tax documents, credit reports, national clearinghouse records, and phone, email, and postal mail surveys. It also collects data from “call center compilation live feeds.”

With regard to location data, both Brokers 8 and 9 acknowledged using mobile device location data from software development kits (SDKs) built into apps. For health data, Broker 3 referenced its “highly authentic sources” including medical records, government records, surveys, and healthcare directories.

When asked by the research team, five of the 12 brokers, including the three from which we purchased data, claimed that their data was only collected after an “opt-in” process from the user. The exact mechanism for the “opt-in” was unclear, including the degree to which the individual in fact knew this data collection and subsequent use was occurring. Consumers’ use of an app or platform is often construed as acceptance of the privacy policy associated with that app or platform, even when the consumers have not actually read the policy, clicked on a specific button, or affirmatively ticked a box to acknowledge acceptance.

A data broker buying data from an app or platform with a privacy policy that has been passively accepted by a consumer via the consumer’s use of the app or platform may represent that the consumer “opted in,” as a broker would have it, to the disclosure of their data in accordance with that policy. Individuals are frequently unaware of how their data is being used, even when they “opt in” by passively accepting a privacy policy associated with an app or other platform that they use.

Privacy Policies and Terms of Service

Throughout the data buying process, we collected the privacy policies and terms of service for brokers when we could either find the documents online or were given access through conducting business with the broker. We reviewed privacy policies and/or terms of service for Brokers (or the parent companies of Brokers) 1, 2, 3, 4, 5, 6, and 8.

Most policies draw a distinction between data subjects (individuals who appear in datasets sold by brokers) and customers (visitors to the broker’s website or purchasers of data from the broker). Many policies state that data on customers/website visitors, such as IP address or email information, will be collected for limited purposes, such as marketing of the broker’s services, business communication, and website analytics (e.g., how long a user spends on the broker’s site).

Conclusion and Policy Recommendations

The data brokerage ecosystem gathers and sells data on U.S. military personnel, including sensitive, individually identified, and non-public information about active-duty servicemembers' finances, health conditions, political beliefs, children, and religions. Such activities focused on military personnel sit within the broader, multi-billion-dollar data brokerage ecosystem that gathers and sells data on virtually every single American.

This study set out to evaluate what kinds of data that data brokers gather and sell about U.S. military servicemembers and veterans—and the risk that a foreign actor could acquire this data in order to inflict harm on the U.S. military and U.S. national security. Our ability to purchase sensitive, individually identified, non-public information about military personnel with almost no vetting, including from a *.asia* domain, for as low as $0.12 per record, underscores the substantial risk. Meaningful policy action is needed to address this ecosystem and mitigate national security risks facing the United States.

Foreign governments have historically sought data about American persons and organizations for espionage, election interference, and other purposes. Their interest in the U.S. military in particular is high, and they could obtain such data through the data brokerage ecosystem, either by purchasing it legally or by hacking into the databases of brokers or their customers. Sensitive data on members of the U.S. military and the broader U.S. national security community, including location, financial situation, medical conditions, political affiliation, and religion, could be used for purposes ranging from profiling, scamming, blackmail, and coercion to outing, reputation-damaging, stalking and tailing, microtargeting, and conducting other analyses on members of the national security community.

Policymakers should consider the following steps:

Congress should pass a comprehensive U.S. privacy law, with strong controls on the data brokerage ecosystem. The most effective step to prevent harms from data brokerage for all Americans would be a strong, comprehensive privacy law. For example, the American Data Privacy and Protection Act, introduced in 2022 in the 117th Congress (not yet reintroduced in the current Congress), includes provisions to generally prohibit companies from transferring individuals' personal data without their affirmative express consent and to establish a centralized registry through which consumers can opt out of the sale of some of their data by some third-party data brokers.[42] It also includes requirements for companies to implement security practices to protect and secure personal data against unauthorized access. Such provisions could introduce new controls around the collection and use of personal data about Americans, and in doing so encompassing members of the U.S. military and their families.

A comprehensive privacy law should also govern the use of public records and publicly available information, which can be sources of potential harm depending on the context, and especially when aggregated by data brokers. For example, people search websites' scraping of property filings, voting registries, and other government records enables people search data brokers to aggregate data, build profiles linked to individuals, and publish them online for search and sale—including about members of the U.S. military. In a previous study, one of the authors of this paper was able to find profiles for sale on people search websites that appeared to correspond to the home addresses, contact information, family information, and other data of senior U.S. military figures.[43]

Several states have passed privacy laws, including California, Colorado, Connecticut, Iowa, Virginia, and Utah. For example, the California Consumer Privacy Act[44] gives California residents the right to delete, opt out of the sale of, and correct their data. However, the opt-out mechanism requires action by the consumer, who often is unaware their data is being collected. Finally, a piecemeal state-by-state approach allows data brokers to continue collecting data on residents of states that have not yet passed privacy laws. A preemptive federal privacy law would instead prevent the collection and sale of consumer data without informed and clear consent across the country and for all U.S. persons, though states would still be and are permitted to pass their own more restrictive legislation. They could also pass legislation protecting certain categories of data, akin to the Washington My Health, My Data Act[45] or the Illinois Biometric Information Privacy Act.[46]

Congress and the executive branch should supplement a privacy law with national security-focused data controls. In addition to a comprehensive privacy law, Congress and the executive branch should also consider more targeted controls focused specifically on the risks associated with data brokerage and U.S. national security. For example, data brokers could be prohibited from collecting and selling identifiable data related to government employees and active-duty servicemembers for the purpose of sale to third parties, with a possible exception for circumstances in which it is deemed to be highly necessary, low risk, and well-defined. Congress and the executive branch should also consider how some particularly sensitive types of data, such as location data, are at risk of compromise when widely collected and aggregated in the first place.

The Defense Department should assess the risks from data brokerage in its contracts. In addition to legislative action, the Department of Defense should conduct an internal contractual data flow assessment of how it receives, transfers, and releases personal data about civilian employees and uniformed members of the military. The full list of data sources drawn on by data brokers is unclear but may include more official sources such as active military occupational data or privatized on-base housing information. Such an assessment may reveal opportunities to curtail the flow of sensitive military information to data brokers while still allowing for data flows that are necessary to military functions.

The Department of Defense could also implement controls in its contracting requirements. For example, the Department of Defense could reserve the right to restrict a contractor's sale of any data, related to the contract or otherwise, to external entities throughout the contract period *and* restrict the future sale of data to entities that was obtained due to the contract. The Department of Defense could also create and mandate the use of screening protocols that must be implemented by the contractor for the sale of any data, related to the contract or otherwise, that verifies that the data purchaser is a legitimate and non-nefarious business entity. The Department of Defense could also restrict a contractor's collection of data on current or former U.S. military personnel throughout the duration of the contract.

Regulatory agencies should pursue new policies, enforcement actions, and rulemaking where applicable concerning the privacy, cybersecurity, financial opportunity, and other risks associated with the data broker industry and the sale of data about military servicemembers. The Federal Trade Commission is currently undergoing rulemaking on commercial surveillance and data security. Drawing on its authorities under Section 5 of the FTC Act—to enforce against unfair or deceptive business acts or practices—the FTC could implement rules preventing the re-identification of data unless the individual explicitly provided fully informed consent. Furthermore, the Consumer Financial Protection Bureau (CFPB) could aid in the research into data brokers by asking data brokers for information about their practices to shed further light on places where regulation may be needed. Specifically, the CFPB could request information related to data streams, or the specific avenues in which brokers get their data; the prevalence and development of SDKs; how data brokers infer data points; how data brokers re-identify seemingly anonymous data; and how data on military servicemembers is collected and sold, especially credit- and financial-related data.

Congress should provide more funding to regulatory agencies to enforce any new laws or regulations related to the data broker industry and national security. In order to be able to investigate and enforce any resulting regulations, agencies such as the FTC and the CFPB would need sufficient resources. The FTC, for instance, lacks the resources to sufficiently carry out privacy investigations and enforcement actions, and new laws or regulations would only increase the demand for additional personnel.[47]

The data brokerage industry is a multi-billion-dollar ecosystem. It touches everything from consumer reporting agencies to small geolocation data brokers; to mobile apps selling their users' information; to medium-sized enterprises offering advertising, profiling, and other services related to individuals' data. Within that ecosystem is a large amount of data gathered, packaged, inferred, and sold about members of the U.S. military. This study's findings about the availability of individually identified, non-public, sensitive data about military servicemembers—and the variability of (and in some cases lack of) controls around the data's collection, aggregation, and sale—suggests that there are a number of risks to U.S. national security that have gone unaddressed in current law, policy, regulation, and technology. Until there are changes in the way this data is gathered, shared, analyzed, licensed, and sold, these risks will persist.

Endnotes

[1] 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://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf; 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/; Alistair Simmons, *Data Brokers and the Sale of Students' Data* (Durham: Duke University Sanford School of Public Policy, July 2023), https://techpolicy.sanford.duke.edu/data-brokers-and-the-sale-of-students-data/.

[2] Sherman, *Data Brokers and Sensitive Data on U.S. Individuals*.

[3] See, for example, the Senate Commerce Committee's 2013 investigative report into data brokerage and the FTC's 2014 report on data brokers: 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 2013. https://www.commerce.senate.gov/services/files/0d2b3642-6221-4888-a631-08f2f255b577; U.S. Federal Trade Commission. *Data Brokers: A Call for Transparency and Accountability: A Report of the Federal Trade Commission*. Washington, D.C.: Federal Trade Commission, May 2014. https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014.

[4] See, e.g., Trevor Logan, "The United States Has a Data Broker Problem," Foundation for Defense of Democracies, May 13, 2021, https://www.fdd.org/analysis/2021/05/13/the-united-states-has-a-data-broker-problem/.

[5] Aynne Kokas, "China already has your data. Trump's TikTok and WeChat bans can't stop that," *The Washington Post*, August 11, 2020, https://www.washingtonpost.com/outlook/2020/08/11/tiktok-wechat-bans-ineffective/.

[6] Kevin Roose, "Don't Ban TikTok. Make an Example of It," *The New York Times*, July 26, 2020, https://www.nytimes.com/2020/07/26/technology/tiktok-china-ban-model.html.

[7] Caitlin Chin, "U.S. Digital Privacy Troubles Do Not Start or End with TikTok," Center for Strategic & International Studies, October 6, 2022, https://www.csis.org/analysis/us-digital-privacy-troubles-do-not-start-or-end-tiktok.

[8] Kevin Collier, "TikTok a privacy threat? Sure, but so are most of your smartphone apps," NBC, July 13, 2020, https://www.nbcnews.com/tech/security/tiktok-privacy-threat-sure-so-are-most-your-smartphone-apps-n1233625.

[9] Henrik Twetman and Gundars Bergmanis-Korats, *Data Brokers and Security: Risks and Vulnerabilities Related to Commercially Available Data* (Riga: NATO Strategic Communications Centre of Excellence, January 2021), https://stratcomcoe.org/publications/data-brokers-and-security/17, 14, 20.

[10] Dymples Leong and Teo Yi-Ling, "Data Brokers: A Weak Link in National Security," *The Diplomat*, August 21, 2020, https://thediplomat.com/2020/08/data-brokers-a-weak-link-in-national-security/.

[11] Office of Senator Ron Wyden, "Wyden, Lummis, Whitehouse, Rubio and Hagerty Introduce Bipartisan Legislation to Protect Americans' Private Data from Hostile Foreign Governments," wyden.senate.gov, June 23, 2022, https://www.wyden.senate.gov/news/press-releases/wyden-lummis-whitehouse-rubio-and-hagerty-introduce-bipartisan-legislation-to-protect-americans-private-data-from-hostile-foreign-governments.

[12] Ibid.

[13] Lauren Mannix, "Personal Data Exploitation and Social Media Manipulation as a Security Threat for NATO Nations and Democratic Societies," *Journal of Military and Strategic Studies* 22, no. 1 (October 2022), https://jmss.org/article/view/76243; Joe Littell, Maggie Smith, and Nick Starck, "The Devil is in the Data: Publicly Available Information and the Risks to Force Protection and Readiness," Modern War Institute at West Point, September 20, 2022, https://mwi.usma.edu/the-devil-is-in-the-data-publicly-available-information-and-the-risks-to-force-protection-and-readiness/; Jessica Dawson and Brandon Pugh, "Ukraine conflict heightens US military's data privacy vulnerabilities," *Defense News*, April 14, 2022, https://www.defensenews.com/opinion/2022/04/14/ukraine-conflict-heightens-us-militarys-data-privacy-vulnerabilities/.

[14] Michael Kans, "Data Brokers and National Security," *Lawfare*, April 29, 2021, https://www.lawfareblog.com/data-brokers-and-national-security.

[15] Sam Biddle and Jack Poulson, "American Phone-Tracking Firm Demo'd Surveillance Powers by Spying on CIA and NSA," *The Intercept*, April 22, 2022, https://theintercept.com/2022/04/22/anomaly-six-phone-tracking-zignal-surveillance-cia-nsa/.

[16] Jonathan Panikoff, "Banning TikTok alone will not solve the problem of US data security," Atlantic Council, March 31, 2023, https://www.atlanticcouncil.org/blogs/new-atlanticist/banning-tiktok-alone-will-not-solve-the-problem-of-us-data-security/.

[17] Ibid.

[18] Twetman and Bergmanis-Korats, *Data Brokers and Security*, 23.

[19] Ibid., 25.

[20] Rolf Fredheim et al., *The Current Digital Arena and its Risks to Serving Military Personnel* (Riga: NATO Strategic Communications Centre of Excellence, February 2019), https://stratcomcoe.org/publications/the-current-digital-arena-and-its-risks-to-serving-military-personnel/102, 8.

[21] Ibid.

[22] Kirsten Hazelrig, "Surveillance Technologies Are Imbedded into the Fabric of Modern Life—The Intelligence Community Must Respond," MITRE, January 9, 2023, https://www.mitre.org/news-insights/publication/surveillance-technologies-are-imbedded-intelligence-community-must-respond, 2.

[23] Joey Fox, "Congress passes Daniel Anderl judicial security act via defense bill," *New Jersey Globe*, December 16, 2022, https://newjerseyglobe.com/congress/congress-passes-daniel-anderl-judicial-security-act-via-defense-bill/

[24] Suzanne Smalley, "Brokers' sales of U.S. military personnel data overseas stir national security fears," *Cyberscoop*, April 20, 2022, https://cyberscoop.com/data-brokers-national-security-risk/; 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; Nina Totenberg, "Judge Esther Salas Remembers the Night of Assailant's Attack on Her Family," NPR, November 19, 2020, https://www.npr.org/2020/11/19/936783691/judge-esther-salas-remembers-the-night-of-assailants-attack-on-her-family.

[25] California Civil Code § 1798.99.80. https://casetext.com/statute/california-codes/california-civil-code/division-3-obligations/part-4-obligations-arising-from-particular-transactions/title-18148-data-broker-registration/section-17989980; Vermont Statute 9 V.S.A. § 2430. https://legislature.vermont.gov/statutes/section/09/062/02430.

[26] California Delete Act. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB362.

[27] Liz Sly, "U.S. military reviewing its rules after fitness trackers exposed sensitive data," *The Washington Post*, January 29, 2018, https://www.washingtonpost.com/world/the-us-military-reviews-its-rules-as-new-details-of-us-soldiers-and-bases-emerge/2018/01/29/6310d518-050f-11e8-aa61-f3391373867e_story.html.

[28] Ibid.

[29] Alex Hern, "Fitness tracking app Strava gives away location of secret US army bases," *The Guardian*, January 28, 2018, https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

[30] Alex Hern, "Strava suggests military users 'opt out' of heatmap as row deepens," *The Guardian*, January 29, 2018, https://www.theguardian.com/technology/2018/jan/29/strava-secret-army-base-locations-heatmap-public-users-military-ban.

[31] David Ingram, "Exclusive: Fitness app Strava overhauls map that revealed military positions," Reuters, March 13, 2018, https://www.reuters.com/article/us-strava-privacy-exclusive/exclusive-fitness-app-strava-overhauls-map-that-revealed-military-positions-idUSKCN1GP1WE.

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

[33] See, e.g., Bennett Cyphers, "How Law Enforcement Around the Country Buys Cell Phone Location Data Wholesale," Electronic Frontier Foundation, August 31, 2022, https://www.eff.org/deeplinks/2022/08/how-law-enforcement-around-country-buys-cell-phone-location-data-wholesale.

[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.

[36] Kim Lyons, "Congress investigating how data broker sells smartphone tracking info to law enforcement," *The Verge*, June 25, 2020, https://www.theverge.com/2020/6/25/21303190/congress-data-smartphone-tracking-fbi-security-privacy; 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.

[37] Thompson and Warzel, "Twelve Million Phones."

[38] 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/.

[39] Jeremy Hsu, "The Strava Heat Map and the End of Secrets," *WIRED*, January 29, 2018, https://www.wired.com/story/strava-heat-map-military-bases-fitness-trackers-privacy/.

[40] Sherman, *Data Brokers and Sensitive Data on U.S. Individuals*.

[41] Frank Newport, "Military Veterans of All Ages Tend to Be More Republican," Gallup, May 25, 2009, https://news.gallup.com/poll/118684/military-veterans-ages-tend-republican.aspx; "Department of Defense Releases Annual Demographics Report — Upward Trend in Number of Women Serving Continues," U.S. Department of Defense, December 14, 2022, https://www.defense.gov/News/Releases/Release/Article/3246268/department-of-defense-releases-annual-demographics-report-upward-trend-in-numbe/; "Demographics of the U.S. Military," Council on Foreign Relations, July 13, 2020, https://www.cfr.org/backgrounder/demographics-us-military.

[42] H.R.8152. American Data Privacy and Protection Act. June 21, 2022. https://www.congress.gov/bill/117th-congress/house-bill/8152/text

[43] Sherman, *Data Brokers and Sensitive Data on U.S. Individuals*.

[44] California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa.

[45] My Health, My Data Act. http://app.leg.wa.gov/billsummary?BillNumber=1155&Year=2023.

[46] Illinois Biometric Information Privacy Act. https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57.

[47] U.S. Federal Trade Commission. *FTC Report to Congress on Privacy and Security*. Washington, D.C.: Federal Trade Commission, September 2021. https://www.ftc.gov/system/files/documents/reports/ftc-report-congress-privacy-security/report_to_congress_on_privacy_and_data_security_2021.pdf. 7; Testimony of Ashkan Soltani. Hearing before the Senate Committee on Commerce, Science, and Transportation on "Protecting Consumer Privacy." https://www.commerce.senate.gov/services/files/5771F646-244C-4E39-8844-D0AEE1940E00. 3-5; Justin Sherman, "The Key to Protecting Privacy Is Locked in an Underfunded Government Agency," *Slate*, July 14, 2023, https://slate.com/technology/2023/07/federal-trade-commission-funding-privacy.html.