Tracking the Surveillance and Information Practices of Data Brokers: A Report

Rahul Kanwal & Kevin Walby

July 4, 2024

CAIJ
Centre for Access to Information and Justice

THE UNIVERSITY OF WINNIPEG

Tracking the Surveillance and Information Practices of Data Brokers: A Report

Recommended citation: Kanwal, Rahul and Kevin Walby. 2024. *Tracking the Surveillance and Information Practices of Data Brokers*. Winnipeg, MB: Centre for Access to Information and Justice.

Report design: Alex Luscombe.

Cover photo: ©Kevin Ku/Unsplash.

Please direct inquiries to:
Centre for Access to Information and Justice
University of Winnipeg
Department of Criminal Justice
Centennial Hall, 3rd Floor
515 Portage Avenue
Winnipeg, Manitoba
Canada R3B 2E9

www.uwinnipeg.ca/caij

Contents

About the Authors

Rahul Kanwal is a researcher working with the CAIJ who is completing his BS Honours in Applied Computer Science at the University of Winnipeg. His academic pursuits are focused on topics within the domains of Data analytics, Machine Learning and Algorithms.

Dr. Kevin Walby is Associate Professor of Criminal Justice and Director of the Centre for Access to Information and Justice at the University of Winnipeg, Canada. Kevin can be reached at caijuwinnipeg@gmail.com.

About the CAIJ

The Centre for Access to Information and Justice (CAIJ) at the University of Winnipeg aims to be an international hub for public interest research on matters of freedom of information (FOI) and access to justice in Canada and beyond. Through local and international collaborative projects, the CAIJ promotes a multi-disciplinary and critical approach to research and policy engagement. The CAIJ investigates government practices, tracks general trends in FOI and access to justice, as well as charts national and regional variations in these practices. The CAIJ advances theoretical, empirical, and policy-oriented studies of FOI and access to justice in the form of workshops, reports, articles, and books produced by its members.

The CAIJ's mission and goals include:

- Advancing knowledge of FOI and access to justice practices through multi-disciplinary and critical collaborative research projects;
- Organizing knowledge mobilization and research-driven working groups, workshops, seminars, training, and conferences on FOI and access to justice;
- Providing a welcoming and enabling context for students and visiting scholars working in the areas of FOI and access to justice in Canada and beyond;
- Engaging in outreach with a community and public interest focus.

For more information, please visit the Centre's website.

Report Summary

Data brokering is a multibillion dollar industry comprised of thousands of companies that specialize in collecting and analyzing consumer data. The data brokering industry has expanded quickly in the last two decades as a result of developments in artificial intelligence and data science. These entities collect information from a combination of public records, publicly available information, and non-public, proprietary sources (Rostow, 2017). Data brokers have different origins and business models, and there is variation in how value is extracted from data (Reviglio, 2022). As defined by the US Federal Trade Commission, data brokers "are companies that collect personal information about consumers from a variety of public and non-public sources and resell the information to other companies" (Federal Trade Commission, 2012). The European Data Protection Supervisor (EDPS) defines data brokers as entities that collect personal information about consumers and sell that information to other organisations (Rieke et al., 2016). Retail companies rely on information collected by data brokers to create targeted advertisements to boost sales. Political campaign teams use data broker insights to make predictions about voter behaviour. Public health officials (during the COVID-19 pandemic) leverage data broker information to generate intelligence about people's movements and the implications for the spread of contagion. Beyond the private sector, data brokers partner with the Social Security Administration and Departments of Justice, Homeland Security, and State (US Government Accountability Office, 2006) (Crain, 2018). Government also uses data brokers to execute background checks and credit monitoring.

In this systematic review, we examine existing literature on data brokers and related issues such as methods of surveillance, privacy concerns, and government regulations. Following PRISMA guidelines, we conducted a literature search using ProQuest, Web of Science, Google Scholar, IEEE Xplore, and JSTOR databases considering all articles published until May 31, 2023. We selected studies containing the words "data brokering", "data brokers" and "data brokers surveillance" for our analysis. Of a total of 135 articles located, 110 articles met the necessary conditions for further review. The findings were categorized into three main sections: surveillance practices of data brokers and privacy concerns, government initiatives and legislation, and suggested changes or remedies. Overall, this report explores the information-gathering and surveillance practices of data brokers, as well as the laws affecting them and recommendations for modifying the current framework. This knowledge presented here will be valuable for thinking about future research on data brokers and marginalization.

Introduction

From communicating with friends and family, buying a pair of shoes, using coupons to buy groceries, driving to work, to going for a jog, nearly every aspect of our lives in the 21st century produces a digital trace. Technologies are embedded in the most intimate and mundane parts of our lives (West, 2019). Tracking these movements and activities, some companies have specialized in collecting and analyzing these data for decades. For example, the practice of segmenting consumers for marketing purposes dates back to at least the 1970s, when a company called Claritas pitched a "lifestyle segmentation system" that promised to help marketers gain insight into their customers' preferences (Rieke et al., 2016). Companies that collect, buy, and sell these data are often referred to as data brokers. Data brokers specialize in collecting and analyzing individual data and repackaging it for buyers. The global data broker industry is comprised of thousands of companies generating some US$200 billion in annual revenue (Crain, 2018). The data collected by data brokers is often collected without the consent or knowledge of the individuals involved, integrated and synthesized using advanced analytic tools, then sold to other data brokers and businesses for a variety of purposes (Anthes, 2014). Data brokers do not inform the users of the implications and the intended uses of the data being collected, and the consent to collect user data is simply implied with the general "accept-all" terms and conditions. Users are often unaware that their data is being exploited, bought and sold, and being used to manipulate their consumer interests. Data brokers also obtain and sell highly sensitive data on individuals pertaining to race, ethnicity, gender, sexual orientation, immigration status, income level, and political beliefs (like support for the NAACP or National LGBTQ Task Force in the United States) that can be used to undermine civil rights (Sherman, 2021).

There is little research on data brokers. Kim (2023) provides one of the only studies that has involved interviews with data brokers, and found that data brokers were highly suspicious regarding research inquiries. Kim argues "The unregulated collection, aggregation, sharing, and sale of data on individuals' mental health conditions puts vulnerable populations at greater risk of discrimination, social isolation and health complications" (pg. 17), suggesting that comprehensive law addressing data broker surveillance is long overdue (pg. 18). Crain (2018) argues that data brokers create an environment of "pervasive commercial surveillance" (pg. 89). Data brokers play a key role in surveillance capitalism and the commodification of personal data in the 21st century. Data brokers extend the surveillance capacity of financial institutions and corporations, repurposing our own behavioural data as targeted advertising that manufactures our desires. For Crain, the commodification of personal data is not a glitch in the system, rather "it is the system" (pg. 100). As we discuss below, some laws could feasibly regulate data brokers, but various factors have culminated to prevent this from happening at the time of writing. Data brokers and online advertisers have formed lobby groups and trade groups to deter regulation of any kind (pg. 95). According to Reviglio (2022), big data and big tech lobby groups have continued to fight against such legislation, often gutting sections of proposed bills.

The sheer amount of records that data brokers are accessing almost defies comprehension. In 2012, one company, Acxiom, was analyzing 50 trillion transactions a year in the USA (Roderick, 2014: 730). Roderick (2014) referred to Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Peekyou, Rapleaf, and Recorded Futures as the main data brokers (pg. 732). Thousands of data broker companies exist worldwide and their surveillance capacity only grows with each new technological development. Data brokers allow companies to target the consumption patterns of citizens and thus predict and even shape future consumer behaviour. The activities of data brokers also normalize consumer debt and consumerism. Data brokers are in the business of collecting and selling data on people. The data these brokers sell is commonly used to feed marketing as well as political campaigns (Venkatadri et al., 2019). Basic information about a person's age, gender, and location is estimated to be worth $0.0005 per person (Rostow, 2017), while more targeted commercial information-such as persons looking to purchase a car or a vacation is worth at least $0.0021 per person (Rostow, 2017). Data brokers could also play a role in political campaigns and cyber espionage (Reviglio, 2022). Data brokers often combine different datasets such as individual browsing histories, locations, and voter lists, to predict voting preferences through the use of advanced algorithms. This predictive data can be purchased by political parties to target specific areas with lower voter turnout for targeted campaign efforts. Furthermore, data brokers collect vast quantity of sensitive data from various social media platforms such as Facebook and TikTok on

individuals based in the US and Canada.

This data can be sold to companies, including foreign entities, without strict regulations governing its use or distribution, thus facilitating cyber espionage. One such example is the TikTok ban in the US due to concerns that it could be used for surveillance or espionage by China. Cyber policy scholar Samm Sacks argues that American companies can still sell data to data brokers, even after buying ownership of foreign-based apps (Roose, 2020). Data brokers even collect pharmaceutical data. For example, MS collects prescription and purchasing data from individual pharmacies, identifies physician trends in prescribing pharmaceuticals, and then profiles the physicians in an effort to assist pharmaceutical companies in marketing their drugs to those doctors (Palk & Muralidhar, 2017). Beyond the private sector, data brokers partner with the Social Security Administration and Departments of Justice, Homeland Security, and State (US Government Accountability Office, 2006) (Crain, 2018). Data brokers sell several forms of sensitive data, including communications, biometric, and license plate reader data to law enforcement and intelligence agencies. The practice is increasing, with multiple agencies spending upwards of tens of millions of dollars on multi-year contracts (Shenkman et al., 2022).

Finally, we discuss the legal landscape affecting data brokers. The practices of data brokers can be referred to as predatory (Kuempel 2016: 234) because they exploit our behavioural data to manipulate our consumer practices in the future. However, there are almost no regulations or laws for overseeing these organizations and their surveillance practices. In the United States, no generalized protection exists to shield consumers from the processing of their personal information by the private sector (Kuempel, 2016). Several industry-specific regulations such as HIPAA, FCRA, DDPA, and ECPA are present, but thus far are ineffective. Data brokers are addressed under a patchwork of regulatory frameworks that involve different types of specific information and uses, which is sometimes called the sectoral approach (Neally, 2019). Furthermore, data broker partnerships with various government law-enforcement agencies such as the US Department of Justice and Homeland Security make it more difficult to establish laws limiting data broker practices and holding them accountable.

Report Objectives

This report will thus provide an analysis of the following:

-collection and analytic strategies data brokers use to infer the sociodemographic attributes, opinions, and interests of digital citizens;

-sales of personal data that data brokers engage in;

-how data broker practices foster marginalization;

-the ways privacy and legal communities are responding to the activities of data brokers.

Methods

Search and Inclusion Criteria

Figure 1 provides a summary of the search and selection methodology used in this study. We followed the Preferred Reported Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines. All relevant articles containing the terms "data brokering," "data brokers," and "data brokers surveillance" were selected from the databases ProQuest, Web of Science, Google Scholar, heinOnline, and IEEE Xplore in May 2023. To broaden the search, we selected articles on big data, data privacy, and personal data pertaining to data brokers and were cited in the previously selected articles resulting in the addition of seven articles to the primary analysis. In addition to database searches, a manual search was carried out, looking through conference papers and journals that were relevant to data privacy and data brokering. Articles from institutions such as the Financial Times and Harvard Business Review were also included. Articles not written in English were also taken into consideration. Two articles were found, one with a Spanish abstract and the other with full text in Korean. However, none of them met our inclusion criteria. As a result, only English articles were selected in our final sample. Editorials, book reviews, and grey literature without proper referencing were also included in the primary analysis. Theses and dissertations were considered for primary analysis. However, only six met our inclusion criteria. We chose to exclude the other theses and dissertations. Government reports from organizations like the Federal Trade Commission and the Privacy Commissioner of Canada also met the requirements for inclusion and were included.

Twelve publications included the keyword "data brokering" but did not specifically address the data broker industry. Those publications were excluded from the study. Lastly, previously selected articles were screened for additional article selection on topics such as big data, data privacy, and personal data. This resulted in 12 additional references.

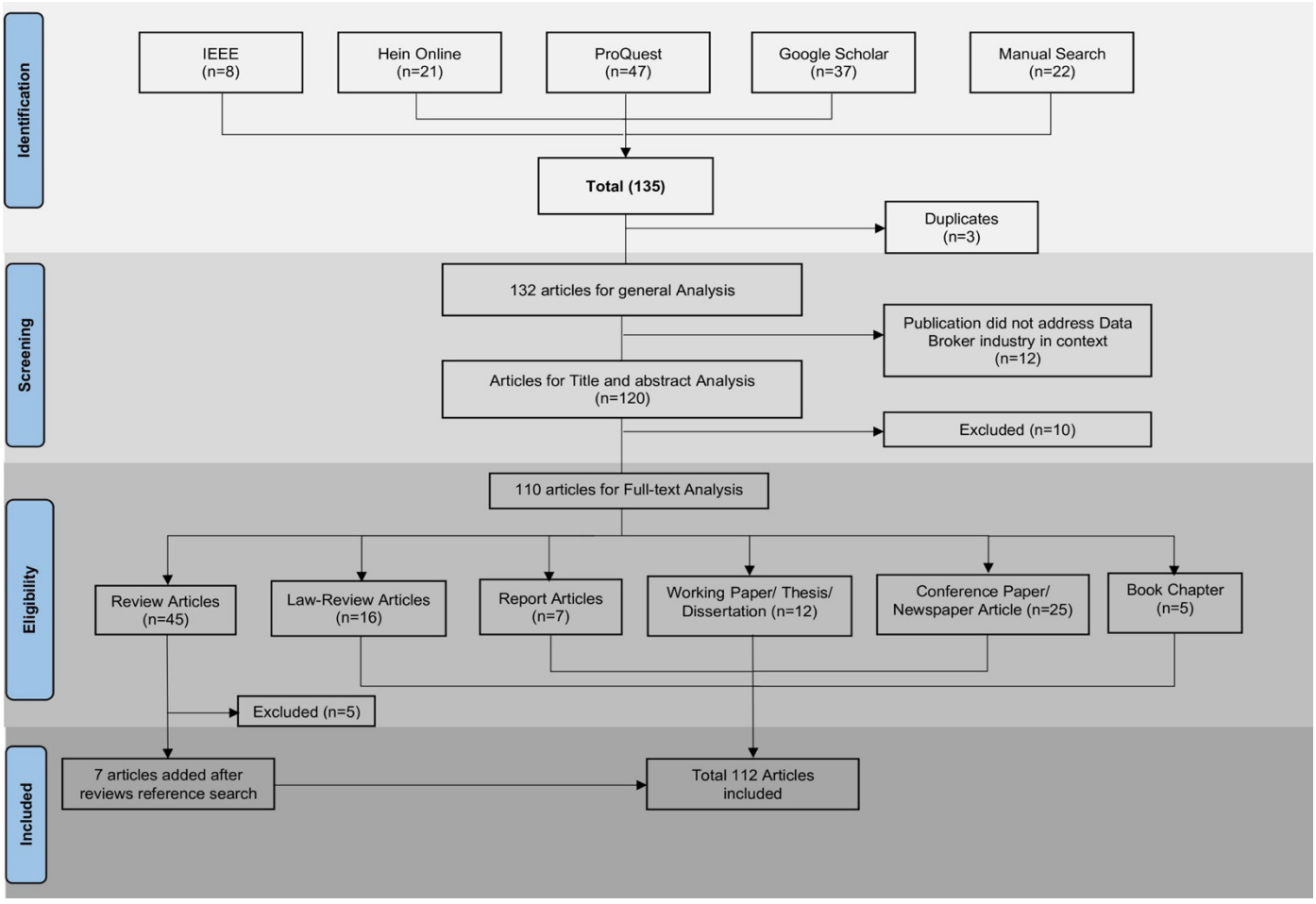

Figure 1. Flowchart depicting the selection of articles in accordance with PRISMA guidelines

Results

Following article selection, we extracted metadata such as type of publication and year of publication. We plotted the number of publications per year (Figure 2) and the type of publications (Figure 3). An Excel spreadsheet was used to organize the articles, comprising the first and last authors, type of publication, title, year of publication, and the link to the article.

We have divided the results into three categories: surveillance practices of data brokers and privacy concerns, government initiatives and legislation, and laws in effect.

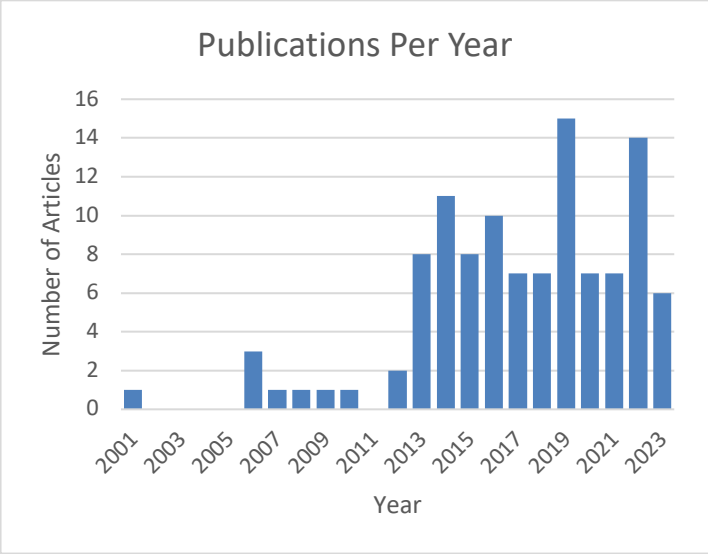

Figure 2. Publications per year of included articles

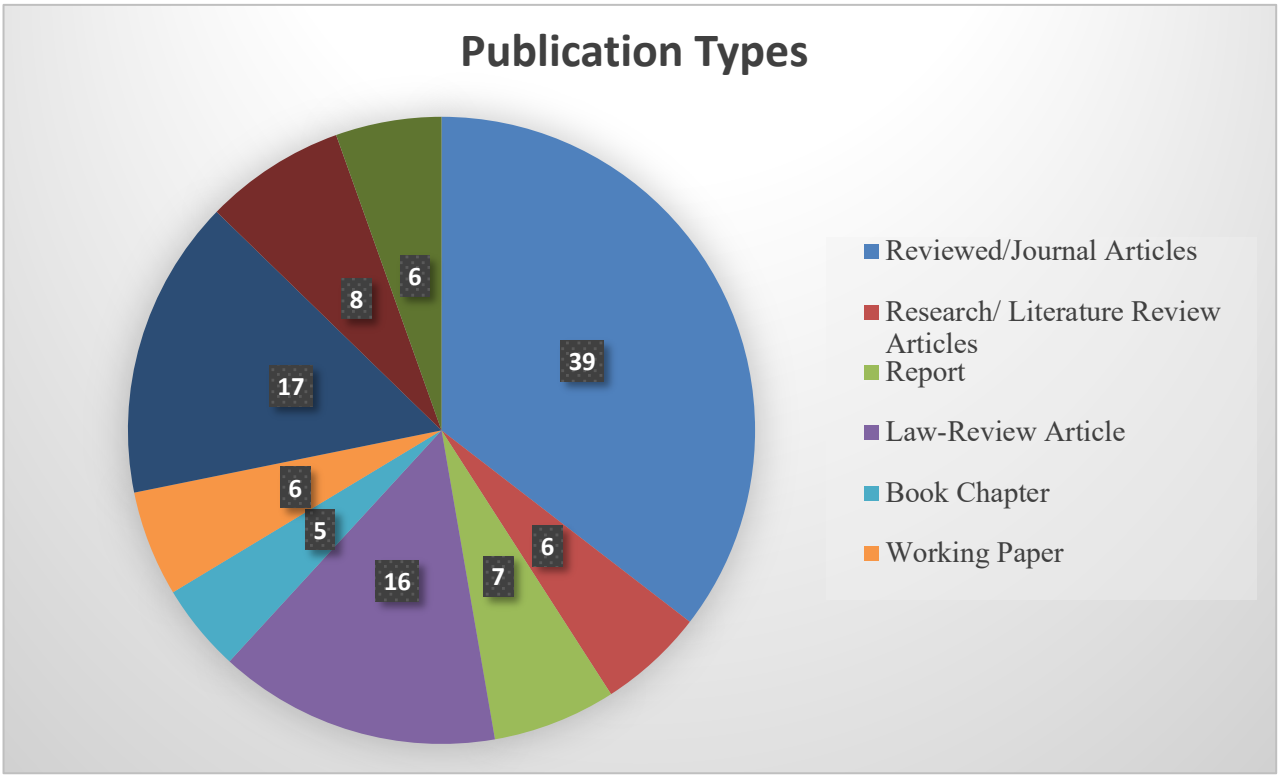

Figure 3. Publication types of included articles

Surveillance Practices of Data Brokers and Privacy Concerns

Data brokers tend to specialize in market niches with the aim of obtaining a competitive advantage (Baccaro, 2021). Based on the services data brokers offer, they can be grouped into three categories. These categories are risk-mitigation, people-search and marketing (also see Neally, 2019). Risk-mitigation primarily includes credit checks and background-information, while people-search includes lookup and people-search services. Marketing includes services such as real estate services, genealogy lookup services and services providing email and calling lists.

In the realm of risk mitigation, data brokers include both long-existing firms in credit scoring, geo demographics, and marketing sectors such as Experian or Claritas, as well as start-ups including those partnered with platform companies interested in leveraging their consumer data to develop personal loans and alternative credit scoring products (Zook, & Spangler, 2023).

People-search services are targeted at individuals seeking personal information about their acquaintances (Ruppert et al., 2017). Whenever a person attempts to search for themselves, a friend, a neighbor, or a business, they are attempting to access a data broker's 'people search' product to find personal information (Neally, 2019). These products typically consist of public knowledge and are used for personal financing or by governments and retailers (Neally, 2019). People search further includes services such as genealogy lookup services and services providing email and calling lists.

Marketing data brokers are primarily focused on targeted advertising and analytics. They offer to improve marketing strategies for both individuals and businesses. For example, Acxiom and Datalogix profile consumers for targeting purposes, collecting information such as demographics, socio-graphics, and purchasing behaviours. Data brokers including CoreLogic and eBureau sell detailed financial and property data analytics (Gu et al., 2021). Educational data brokers collect information from educational settings and sell it to commercial and non-commercial entities for use in for-profit activities, such as marketing and technological development (Arantes, 2023). Through analytics, marketing-focused data brokers advise their clients on consumer habits and preferences to improve product messaging, ad placements, and campaigns (Neally, 2019). Brokers such as Verisk (https://www.verisk.com/) and Interactive Data (https://ididata.com/) provide data analytics services and predictive data through their proprietary analytical software. Data brokers such as First Orion Corporation (https://firstorion.com/) provide calling solutions and Dataamerican.com (http://www.dataamerica.com/) supply mailing lists for potential customers, which can then be used for targeted telemarketing campaigns.

We prepared a list of registered data brokers in the US and Canada and identified 190 of them. We then categorized them into three categories based on the services they offer. These categories are risk-mitigation, people-search and marketing (also see Neally, 2019, and see below). Figure 4 represents the types of data brokers present in the US and Canada.

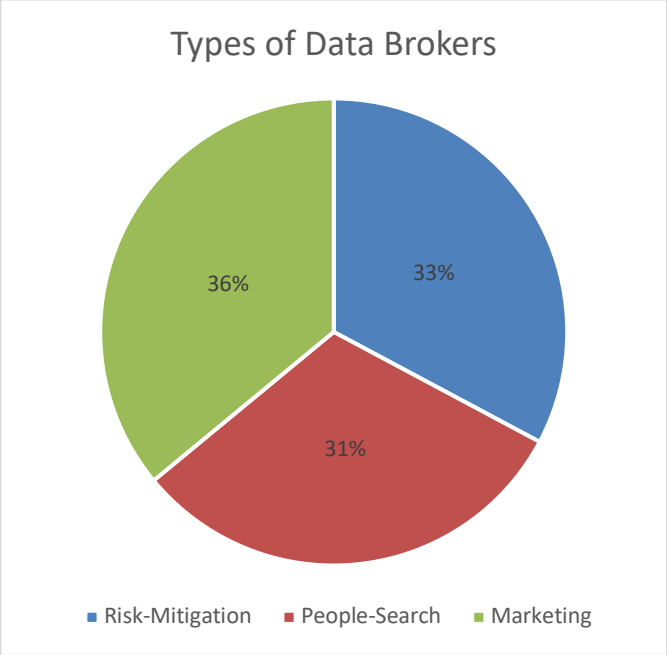

Figure 4. Types of Data Brokers

A data broker is unlikely to be the entity that initially collected the data that it subsequently makes available commercially (Shenkman et al., 2022). Data may pass through multiple providers before reaching a data broker (Shenkman et al., 2022). Data brokers use various methods for collecting data. Data is often gathered by using either direct or indirect

methods. Typical entry points for data acquirement are the filling in of forms, commercial transactions, internet searches, use of social network platforms, webmail, loyalty and discount programs including websites, retail, banks, drug stores and health plans, among many other interactions (Birckan et al., 2020). These include phishing attacks used by these companies, which are essentially fraudulent emails. When opening an account, signing up for a subscription, purchasing an item, browsing the internet, or using a phone, a person must, either explicitly or implicitly, agree to terms and conditions that allow the company to collect and use their information (Neally, 2019). Data brokers also collect data from commercial sources, such as retailers (Martin, 2020). This includes the point of sales locations at various retailers such as grocery stores and pharmacies, which require the user to log personally identifiable information to shop and make purchases. Data brokers also collect data from publicly available information through web scraping. Web scraping is a technique of extracting unstructured data from publicly available websites and transforming it into structured data that can be stored and analyzed in a database. This technique involves software agents, or web robots, mimicking human browsing interactions to access web pages and extracting relevant data using methods such as regular expressions, HTML parsing libraries, XPath expressions, and CSS selectors (Glez-Peña et al., 2014). However, many websites now require users to agree to terms and conditions that prohibit data scraping, potentially leading to legal consequences for violating these agreements (Luscombe, Dick, & Walby, 2022).

The census provides a vast amount of information, including geographic location cross-referenced with "ethnicity, age, education level, household makeup, income, occupations, and commute times" (Martin, 2020). To collect data from such public records, data brokers employ staff to gather this information through online databases, records requests, or simply sending researchers to county clerks' offices (Crain, 2018). Many companies and government entities sell their customer information as an additional revenue stream or exchange data as part of service agreements (Crain, 2018). Data brokers also purchase, license, or acquire data second-hand from companies that collect this information from their users (Sherman, 2021). People are becoming more dependent on modern technology like smartphones and smartwatches, putting consumer privacy at risk. These mobile devices provide rich new data about people including their location, the apps they use, and their contacts (Rieke et al., 2016). This includes not only smartphones, computers, tablets and smart watches, but also common digital devices such as water, gas, or light meters, which gather enough information to create accurate knowledge about consumers (Abad & Orón, 2016). The information can also be collected by mobile platform providers like Google and Apple, and by app developers and the data brokers that provide developers with analytics and advertising (Rieke et al., 2016). Data brokers may purchase data from these companies.

Data brokers also use indirect methods to gather information. Data brokers observe people's behaviour across many websites, making sophisticated use of browser cookies and other technologies (Rieke et al., 2016). Cookies were designed to enable websites to remember who web-page visitors are (Martin, 2020). Once companies realized they could use this technology to track consumer movement, "the third-party cookie was born" (Martin, 2020). Through cookies technologies, a website can track what content is viewed, how long it is viewed, connect the browsing activity to search history, and algorithmically develop a profile for marketing new products to the patron (Tsesis, 2014). Cookies and other tracking devices are often installed on computers without owners' knowledge and cached on computers (Tsesis, 2014). While users can usually opt out, most tend to accept cookies due to "information fatigue" since the user are confronted with an array of confusing options, such as accept-all, accept-some, accept only necessary, etc (Reviglio, 2022). Among the newest forms of tracking technology, still in its developmental stages, is "fingerprinting," which enables hosts to run JavaScript bench markers to circumvent conventional security methods, like using proxy servers to obfuscate identity or opting out of cookie placement (Tsesis, 2014). Fingerprinting tracks users "by collecting the properties of PCs, smartphones and tablets including their screen size, the software versions they're running and which plug- ins are installed" and is typically run in an effort to circumvent European and U.S. laws on the propagation of cookies (Tsesis, 2014: 107). Often data is extracted by smartphone apps through SDKs (Software Development Kit). Data brokers often provide this software to developers for free. SDKs are used to make apps faster at the cost of allowing data brokers to hoard data (see

Morrison, 2020; Reviglio, 2022). Data collection may also be sourced through web scraping and data crawling. Web scraping refers to automated programs known as bots that crawl (visit) web pages simulating human web surfing habits in order to collect specified bits of information from different websites. Data crawling uses similar techniques to retrieve information from any source (not necessarily limited to the web) (Reviglio, 2022). These techniques are legal and nearly impossible to avoid online. Data brokers may also use some network automatic capture tools such as Scrapy and Beautiful soup to obtain information, although many entities will set up technical barriers to prevent "automatic capture" tool information hunting (Nie & Han, 2019).

To prevent or limit third-party cookies, software solutions exist, often implemented as browser extensions like adblockers. And antivirus programs such as Malwarebytes (https://www.malwarebytes.com/solutions/free-ad-blocker) offer adblocking functionality. Adblockers block third-party cookies by intercepting the requests made by the browser and selectively blocking ads deemed risky or containing tracking scripts. The most familiar adblocker solutions are browser extensions such as Ghostery or Adblock Plus which suppress unnecessary requests to third-party advertisements and tracking servers such as Google Analytics and Adobe Analytics, limiting the risk of data leakage to these servers (Gervais, 2017). These blockers use filter lists containing rules based on URL patterns or other criteria to identify and block such requests. In a study of Ghostery extension with maximum protection level, it was found that Ghostery MaxProtection decreases the mean FPD node degree by approximately 80% compared to NoAdblocker (Gervais, 2017). FPD (first-party domain) node degree is the measure of engagement of third-party domains on first-party websites, revealing the extent of third-party cookie involvement or tracking on the primary websites. Adblockers such as Ghostery limit the ability of third-party domains to track users and collect personal information. Another study found only a small number of extensions effectively blocked the majority of stateful trackers (Merzdovnik, 2017). None of the analyzed extensions were able to block all fingerprinting services (Merzdovnik, 2017). Blockers decrease online tracking by blocking third-party cookies, but they do not offer a solution to the problem of online tracing. Users remain vulnerable to exposing personal data to data brokers through alternate means of data collection besides third-party cookies.

With so much information coming from so many sources, it is inevitable that errors arise in the digital dossiers that these data brokers compile. The errors can be difficult or impossible to correct (Anthes, 2014). Data brokers often exchange data even among themselves or purchase data from other companies collecting data (Baccaro, 2021). In a study published in 2013, the FTC reported one in five consumers had an error in one or more of their credit reports (Anthes, 2014).

Once information has been swept into the data broker marketplace, it becomes challenging to trace any given datum to its original source for any combination of the following reasons: (1) data brokers maintain that information sources and analytic processes are trade secrets, (2) information buyers and sellers are divorced from information collection by degrees of separation via complex markets, and (3) a significant portion of data brokers' information is computationally generated and has no "real" empirical source (Crain , 2018). Some of the data brokers received this information directly from the retailer, while others purchased it from other data brokers (Martin, 2020). Such sharing creates a large web of data exchanges, making it "virtually impossible for a consumer to determine the originator of a particular data element" (Martin, 2020). The anonymity of most data broker transactions has opened the door for nefarious groups to pose as legitimate businesses and obtain vital information about an individual – usually a Social Security number – and steal his or her identity (Brooks, 2001).

According to a 2020 study by the NATO Strategic Communications Centre of Excellence, there are over 5,000 data brokers worldwide, registering an industry of around $178 billion in revenue (Baccaro, 2021). Data brokers earn money from several sources, offering pre-packaged databases of information to potential buyers (Sherman, 2021). As noted above, products are sold in three broad categories: marketing, risk mitigation, and people search (Neally, 2019). Clients use risk mitigation products to verify customer information to prevent fraud (Neally, 2019). For example, when a person attempts to search for themselves, a friend, or a neighbour, they access a data broker's "people search" product to find personal

information. Marketing products are the most familiar products of data brokers and are used to create tailored messages for a client's consumers (Neally, 2019). Data brokers also generate revenues from brokerage fees and/or any value-added services regarding data analysis and data management (Oh et al., 2021). Multiple agencies are spending tens of millions of dollars on multi-year contracts seeking sensitive data, including location, communications, biometric, and license plate reader data, sold by data brokers to law enforcement and intelligence agencies (Shenkman et al., 2022). Many industries from health insurance to life insurance to banking to e-commerce purchase data from data brokers to run advertisements and target their services (Sherman, 2021). A 2018 investigation by ProPublica found that health insurers were purchasing data from data brokers (including data on individuals' race, marital status, education level, net worth, TV consumption, and bill payment history) to predict health costs (Sherman, 2021).

While personal data is sold for a variety of purposes, many of the brokers' customers use the information for targeted marketing. As per The Federal Trade Commission, a broker segments consumers into handy buckets with labels such as "urban scramble" (heavily populated with low-income Latinos and African Americans), "rural everlasting" (single men and women over the age of 66 with minimal education and modest net worth), and "married sophisticates" (upper- middle-class young adults with no children) (Anthes, 2014). More narrowly defined groups included "expectant parent," "diabetes interest," and "cholesterol focus" (Anthes, 2014). Customer profiling and segmentation aids in providing improved services by offering a personalized environment favourable to consumers; however, it also results in several negative ramifications (Mishra, 2021). Profiled data can be misused for discrimination against individuals.

Data brokers collect raw data and they also analyze it. Data brokers combine online and offline data by scraping public records to assemble data points that represent the attributes of millions of people (Mishra, 2021). Publicly available records such as home purchases, marriage, voter lists and so on overlap with other datasets to draw inferences (Mishra, 2021). One illustration of this is in a political setting. New methods of voter data collection and data analysis have improved and enriched traditional forms of political microtargeting like canvassing (Frederik et al., 2018). Political parties could target voters with tailored information that maximizes, or minimizes, voter engagement (Frederik et al., 2018), yet it jeopardizes people's fundamental right to vote.

Corporate manipulation of private information has increased exponentially with the development of algorithmic software used for gathering, packaging, and analytically harvesting data that consumers have provided, without remuneration, to merchants and social networks (Tsesis, 2014). For the extraction of knowledge from these data, statistical algorithms are used. Machine learning techniques are also being used, which can facilitate making the leap across informational and social contexts, generating unforeseen inferences (Birckan et al., 2020). Some data brokers also provide technologies or software to help their clients derive insights from their data (Ruppert et al., 2017). Moreover, data brokers claim to provide added value by aggregating discrete data sets through independently created algorithms, allowing third parties to develop a comprehensive picture of consumers and providing consumers with pertinent information (Palk & Muralidhar, 2017). Combining the browsing history with email addresses can provide a detailed picture of consumer preferences and enable targeted ads and offers (Gu et al., 2021).

Data brokers also claim to aid in risk mitigation services. The Federal Trade Commission's Report highlights how data brokers' products mitigate certain risks and aid in identity verification and prevention of fraud (Mishra, 2021). However, due to the lack of laws and regulations, most data brokers do not verify information for accuracy or inform individuals that their data is being categorized and sold (Neally, 2019). For example, a journalist obtained a copy of their information held by Oracle Data Cloud, and examined their data as provided by Acxiom's "About the Data" site, finding that more than 70% of their attributes from these sources were inaccurate (Venkatadri et al., 2019). Another journalist found their information from Acxiom was highly inaccurate, while yet another found nearly 50% of their personal information purchased for a $50 fee from an undisclosed company to be inaccurate (Venkatadri et al., 2019). The result is that most consumers have important, sometimes life-altering, decisions made for them based on inaccurate and easily

correctable information they never expected to be used against them (Neally, 2019). A real-life scenario is a person applying for a home loan. If the data provided by the data broker is inaccurate, the loan will not be approved. Various law-enforcement agencies such as the Department of Justice and Homeland Security partner with data brokers (Crain, 2018). When law enforcement relies on data broker search tools, even minor inaccuracies in the data could lead to dire consequences (Rieke et al., 2016). A simple case of mistaken identity could lead to a wrongful arrest, or could lead to officers using force against the wrong person (Rieke et al., 2016).

The practice of data brokerage is secretive, and there is often no means to appeal incorrect information (Wayne, 2012). While individuals can choose to opt out of data collection by each data broker, this requires a significant amount of time and provides little certainty that these data has been deleted (Reviglio, 2022). Furthermore, data brokers often do not provide clear guidance on how to opt out of data collection, correct captured inaccurate data, or delete data. The methods for opting out are often not well publicized, or are difficult to follow (Anthes, 2014). Effectively opting out from data broker data collection is time-consuming and does not leave certainty that this opt out includes all personal data. This leads to key questions such as whether or not users can ever exercise their right of erasure (Reviglio, 2022).

In response to growing concerns about the lack of control over personal data held by data brokers, several companies have emerged to provide removal services. These companies, such as Aura (https://www.aura.com/), Optery (https://www.optery.com/), Privacy Bee (https://privacybee.com/), and Incogni (https://surfshark.com/blog/introducing-incogni), specialize in assisting individuals with removing their data from data broker databases. For instance, Incogni streamlines the removal process by acting on a user's behalf. Users grant authorization to Incogni to manage data removal requests, after which the Incogni team contacts data brokers on behalf of the user. They facilitate the process to ensure the effective removal of the user's personal data from databases. Similarly, Optery offers a comprehensive solution for removing personal information from Google and over 335 other websites. Optery further asserts that their patented search technology enables them to locate and remove a greater number of customer profiles compared to their competitors. Although companies advertise the removal of personal information from numerous data brokers, they acknowledge that they cannot remove personal information from all data brokers. This is because once personal information has been packaged, sold and resold, it may live indefinitely on the servers operated by the data broker industry (Reviglio, 2022). If it is hacked, the profile joins the billions of other profiles being traded on the dark web. This represents another concerning feature of the data broker industry: endless profiling persistence (Reviglio, 2022). Moreover, data brokers often intentionally retain consumer data for indefinite periods of time. Furthermore, even if a person had a crime they had been convicted of successfully expunged from their record, there is no requirement for data brokers to remove that record unless the person seeks them out and asks that particular agency (Dudley, 2015). Therefore, to protect consumer privacy, strict regulations are needed. For Crain (2018), "failing to confront commodification and continuing down the current path will almost certainly represent one small step for privacy, one giant leap for commercial surveillance" (pg. 101).

Government Initiatives and Legislation

Fair information practice principles (FIPPs) are the most widely accepted privacy framework by governments globally. Fair information principles guide organizations in protecting personal information and restricting data collection. These principles address both the collection and use of data, recommending that data collection be minimized where feasible and that data be used for specific purposes. However, fair information principles leave much room for interpretation and varied application (Rieke et al., 2016).

In the United States, no generalized protection exists to shield consumers from the processing of their personal information by the private sector (Kuempel, 2016). Instead, there are several regulations including HIPAA, FCRA, DDPA, and ECPA, among others that provide industry-specific protection based on different kinds of information. Data brokers partner with US Department of Justice, Homeland Security, among others (US Accountability Office, 2006) (Crain, 2018). Numerous law enforcement authorities, including the public police, also use the data that data brokers have collected (Rieke et al., 2016). While investigating crimes, police

sometimes turn to data brokers to obtain information that they themselves cannot access (Dudley, 2015). They can avoid the Fourth Amendment and other legal issues by hiring civilians to investigate or report on previously collected data (Dudley, 2015). Indeed, enacting laws prohibiting data brokers is made more difficult because of data broker partnerships with several core government organizations across the US and Canada.

Laws in Effect

The Privacy Act of 1974 in the United States is one law in place to safeguard consumer privacy. The Privacy Act governs "the collection, maintenance, use, and dissemination of personal information by Federal agencies" (Palk & Muralidhar, 2017). Any information about an individual that is "linked to that individual by name or identifying particular" is protected from government release (Palk & Muralidhar, 2017). However, due to the digitization of public records and the increased online presence of people, access to personal information has become easier. The Privacy Act addressed these concerns by banning secret federal databases, allowing individuals to access and correct their own records, and prohibiting government from keeping databases of information on the First Amendment activities of individuals (McCain, 2009). However, the act has limited power to hold data brokers accountable. It only applies to the federal government and to private companies who are administering records for the government (Solove & Hoofnagle, 2005). Under the Privacy Act, a federal court cannot order a government agency to change its practices; it can only levy fines, or provide the data subject with access to their records, to amend inaccuracies in these documents. Thus, the Act imposes few or no privacy constraints on federal agencies, and no constraints at all on the commercial data brokers supplying the information (McCain, 2009). While The Privacy Act of 1974 was a step forward for consumer privacy, the act fails to cover many companies within the US, as most companies belong to the private sector and are not subject to this law (Vashey, 2020).

The internet serves as the main source of online data collected by data brokers. To enhance the safety, security, and dependability of the Internet, the Computer Fraud and Abuse Act (CFAA) provides authority to prosecute anyone who uses unethical methods to harm or steal online data. However, a primary issue is the possibility of a data breach going unnoticed at a company for several years. Furthermore, it is difficult to determine the extent of the data breach. The CFAA only applies to businesses facing a data security breach resulting from "the negligent design or manufacture of computer hardware, computer software, or firmware" and requires the plaintiffs to demonstrate they experienced more than $5,000 in damages (FairClough, 2016). For this reason, the legislation provides little protection for individuals whose personal information has been stolen by data brokers. Suing data brokers is not a feasible option, as courts require plaintiffs to demonstrate a particular harm before they recognize a privacy violation (Rostow, 2017) and it is difficult to convince the court of a privacy breach.

Due to the lack of laws and regulations governing the data broker industry, most brokers do not verify information for accuracy or inform individuals that their data is being categorized and sold (Neally, 2019). The Fair Credit Reporting Act (FCRA) is one such regulation that is meant to ensure the accuracy of the data captured by different entities. The law requires that entities collecting information for those involved in employment, credit, insurance and housing decisions must do so in a manner that ensures accuracy of the information (Brill, 2013). It verifies the accuracy and offers protection for personal credit information. The FCRA, however, does not regulate data brokers that collect and sell information that is not subject to legal restrictions or covered by government regulations. The only limit it imposes is on what brokers can sell to whom (Dudley, 2015). Roderick (2014) argues that The Fair Credit Reporting Act fails to account for the invasive logic of data broker surveillance. As Roderick (2014: 737) puts it, the state has not taken an active role in protecting the privacy of citizens when it comes to consumer data broker companies, especially when it comes to possible methods of regulating third-party data collection" (pg. 737).

Another act that regulates what information data brokers can collect is the Children's Online Privacy Protection Act ("COPPA"). COPPA protects children's privacy online by imposing the only limits on what may be collected from individuals. Children under 13 are considered too young to be able to judge what information is appropriate to share (Dudley, 2015). However, COPPA has several limits. First,

data collected from minors over the age of thirteen are not subject to the statute (Elvy, 2017). This leaves minors over the age of thirteen vulnerable at a critical stage in their lives when they need privacy protection the most. Second, the statute likely does not apply to data that adults supply about children, as its coverage is limited to information supplied directly by children (Elvy, 2017). COPPA is likely the only data collection specific statute that has strict compliance enforcement that cannot be easily circumvented by data brokers (Neally, 2019).

It is difficult to secure our personal information in the 21st century, when technology affects every aspect of our lives, including our health and medical conditions which are now being monitored by devices such as smartphones and smartwatches. The Health Information Portability and Accountability Act (HIPAA) governs how doctors and medical services must protect patient data (Rostow, 2017). All protected health information indexed to personally identifiable information, such as demographic information or data exposing patient background information, is protected by this privacy regulation (Dudley, 2015). The HIPAA extends to health care providers and their business associates. Brokers do not fall in the category of health care providers or business associates, so there is no burden for them created under HIPAA (Dudley, 2015). Companies that produce and maintain other technologies that might collect health data, such as wearables or social media platforms, are also often not covered by HIPAA (Kim, 2023). Although HIPPA requires the anonymization of data, this measure is not a reliable solution given that various de-anonymization algorithms can easily decode the dataset. More concerning is the practice of cross-referencing de-anonymized datasets with other data sets, which can yield harmful new insights about users. These insights can have catastrophic consequences for individuals. Anonymized health data, when in the process of being de-anonymized, may be cross-referenced with additional features such as browsing history and purchasing habits. This process could lead to manipulation of users through targeted marketing based on analyzed data, compelling them to purchase a specific drug that could prove harmful to their health.

The Gramm-Leach-Bliley Act (GBLA) regulates the personally identifiable information primarily present in financial institutions. There are guidelines set forth in the GLBA that prohibit any financial institution from providing non-public personal information with any non-affiliated third parties, thereby protecting consumer privacy (Neally, 2019). Non-affiliated third parties are entities not related to financial institutions through ownership or control, such as independent companies or service providers. A key element of the GLBA is that financial institutions (but not consumer data brokers) must provide notice of their privacy policies. Before disclosing any consumer's personal financial information to a non- affiliated third party (particularly for marketing purposes) they must offer an opportunity for the consumer to opt out (Roderick, 2014). Consumer data brokers remain outside the scope of the GLBA, and there are currently no laws requiring these companies to maintain the privacy of consumer data unless they use that data for credit, employment, insurance or housing (Roderick, 2014).

Social media platforms also serve as a potential database for data brokers to capture data from. The Stored Communications Act (SCA), which is a part of the 1986 Electronic Communications Privacy Act (ECPA), prohibits electronic communications services (ECS) and remote computing services (RCS) from disclosing digital communications to nongovernmental entities without the consent of the message's originator or recipient (Rostow, 2017). However, the SCA does not extend protections to metadata or communications once they are no longer in electronic storage, creating gaps in privacy coverage. Consequently, SCA does not cover social media posting or comments, and its language (enacted in 1986 as part of the Electronic Communications Privacy Act) is no longer accurate in today's technological environment (Rostow, 2017). Additionally, ECPA prohibits remote computing services and electronic communication services from sharing client information with any government agency. If those third parties are not RCS or ECS providers themselves, ECPA does not apply and does not prohibit them from selling or otherwise providing the information to the government (Shenkman et al., 2022). This ECPA loophole has allowed government agencies to purchase sensitive information from data brokers even though those agencies should have been required to obtain a warrant, a court order, or a subpoena under ECPA (Shenkman et al., 2022). The US Fourth Amendment requires the government to obtain a warrant to access information in which

individuals have a reasonable expectation of privacy. Statutes such as ECPA require the government to use legal processes to obtain certain types of data held by communications service providers (Shenkman et al., 2022).

Law enforcement and intelligence agencies purchase certain personal data about Americans from data brokers to evade Fourth Amendment safeguards as recognized by the Supreme Court (Shenkman et al., 2022). Data brokers further evade Fourth Amendment protections by exploiting the perceived "public" nature of the data they collect. While the Fourth Amendment safeguards against government overreach in searches and seizures, it does not always extend to information voluntarily shared with third parties or made publicly available. Data compiled by data brokers can contain detailed location history and personal profiles, which are often categorized as publicly available, falling outside the scope of Fourth Amendment safeguards. This situation raises a privacy concern, as individuals may be unaware of the extent to which their personal information is collected and analyzed. Thus, although the Fourth Amendment does not restrict the activities of data brokers (unless they are state actors in a particular context), government agencies that purchase location and other sensitive digital data without the cover of a warrant may technically be violating the Fourth Amendment (Shenkman et al., 2022).

The California Consumer Privacy Act (CCPA) of 2018 grants consumers four fundamental rights: (1) the right to know what data companies have collected about them; where it is sourced from; and how it is being used, sold, or disclosed; (2) the right to opt out of the sale or sharing of their data for business purposes, or the right for consumers under 16 years old not to have their information sold absent their or their parents' opt in; and (3) the right to sue companies that violate the law, (4) the right to have a business delete a consumer's personal information, with certain exceptions (Pardau, 2018). In addition to regulating the collection and sale of information, the law stipulates consumers' rights and data brokers' responsibilities regarding data deletion (Spivak, 2019). Furthermore, the CCPA not only dictates what data brokers must do with respect to compiling and selling data, it also establishes a strict reporting and disclosure scheme that aims to keep users informed of their rights (Spivak, 2019). The CCPA places the onus to enforce the law on state regulators rather than private citizens (Pardau, 2018). Therefore, it may help shield consumers from various implicitly accepted terms and conditions unethically obtained by data brokers.

Despite the breadth of the CCPA, it does not apply to all companies, corporations and all residents of California. It only applies to businesses that: (A) generate annual gross revenues in excess of twenty-five million dollars; (B) alone or in combination, annually buys, receives for the business's commercial purposes, sells, or shares for commercial purposes, alone or in combination, the personal information of 50,000 or more consumers, households, or devices; (C) derives 50 percent or more of its annual revenues from selling consumers' personal information (Spivak, 2019). In addition, it only applies to those who are permanent residents of California, i.e., those who are not persons simply passing through for business or leisure (Spivak, 2019). The CCPA also lists a number of "personal information" examples, including names, aliases, postal addresses, IP addresses, social security numbers, and "other similar identifiers," together with biometric information, geolocation data, "professional or employment-related information," and "education information". This definition, and the Act as a whole, "apply to the collection and sale of all personal information collected by a business from consumers," whether in electronic or paper form (Pardau, 2018). However, the CCPA excludes "publicly available information" from the definition of personal information. The current definition leaves room for varied interpretation, providing a loophole for data brokers to use.

The Vermont Data Broker Regulation (2018) has undertaken significant measures to protect consumer data, specifically data acquired by data brokers. In May of 2018, Vermont enacted the first law in the United States focused exclusively on regulating data brokers (Martin, 2020). The Act also required data brokers to register with the secretary of state by January 1, 2019, and pay a registration fee. It requires data brokers to disclose information about their practices, including whether the brokers provide a method for opting out, and if so, a description of that process (Martin, 2020). In addition to registering, data brokers must develop security measures in accordance with the law's standards (Martin, 2020). The Vermont Data Broker Regulation is a step towards more transparent data

procurement and operations of data brokers. Missing from this law is a mandate to allow users to opt out of data collection, a way to access or review what data is collected and sold about them, or a way to know how their data was obtained and who is buying it (Kraus, 2020).

The Federal Trade Commission (FTC) is an independent government agency that works to promote consumer rights and empower consumers through measures that shield them from the unethical behaviour of data brokers. The FTC's ability to enforce and protect data stems from four statutes: (1) Federal Credit Reporting Act (FCRA), (2) Gramm- Leach-Bliley Act (GLBA), (3) Health Insurance Portability and Accountability Act (HIPAA), and (4) Children's Online Privacy Protection Act (COPPA) (Neally, 2019). Without congressional authority to regulate data brokers, the FTC's powers are limited. One key provision that provides the FTC with authority is section 5 of the Federal Trade Commission Act (Martin, 2020). Section 5 prohibits "unfair or deceptive acts or practices in or affecting commerce." This provision provides the FTC with leeway in seeking out "blatantly" deceptive companies (Martin, 2020). Section 5 is a useful tool for the FTC, but the agency can only employ it when companies have misled their consumers (Martin, 2020). As a result, a company could "be vague about its commitment to privacy" to avoid a section 5 violation (Martin, 2020).

A 2014 Federal Trade Commission report, Data Brokers: A Call for Transparency and Accountability (FTC Report), made several legislative recommendations, which aimed to increase the transparency of data brokers. The key proposals included (1) creating a centralized internet portal in which data brokers identify themselves, (2) mandating disclosure requirements regarding data brokers' use of aggregated data and (3) increasing transparency regarding the sources of data brokers (Kuempel, 2016). The FTC recommended that the US Congress require data brokers to disclose the names or categories of their data sources on their websites so consumers can better remedy incorrect data or opt-out of its use (Kuempel, 2016).

In Canada, The Personal Information Protection and Electronic Documents Act (PIPEDA) regulates what type of information data brokers can collect. According to PIPEDA, information can only be gathered for legitimate reasons, and consent from the data subject is required before any use of personal information is made. Data brokers have low compliance with PIPEDA (Kim, 2006). First, some data brokers purport to be exempt from PIPEDA, because they claim not to collect, use, or disclose "personal information" (Kim, 2006). Second, brokers, who rarely collect information directly from the user, rely on data owners to obtain consent from the individuals affected (Kim, 2006). Third, when lists are sold or rented to a data user, brokers stipulate that the list will not be used other than for the intended purpose (Kim, 2006). This clause allows data brokers to avoid PIPEDA, meaning the law does not hold them accountable. PIPEDA has been criticized because of its remedial provisions, including the absence of meaningful sanctions or penalties for non-compliance (Davidson et al., 2022). Privacy commissioners in Canada (notably the Privacy Commissioner of Canada) have not been proactive in regulating data brokers and have done little to address these concerns.

Bill C-27, known as the Digital Charter Implementation Act, 2022, proposes amendments to the current privacy legislation in Canada. Bill C-27 contains three proposed Acts, which relate to consumer privacy, data protection, and AI systems (Arai, 2023). The proposed Acts are The Consumer Privacy Protection Act (CPPA), The Personal Information and Data Protection Tribunal Act (PIDPTA), and The Artificial Intelligence and Data Act (AIDA) (Arai, 2023). The CPPA aims to replace sections of the existing Personal Information Protection and Electronic Documents Act (PIPEDA). The CPPA introduces an updated framework governing the collection, use, and disclosure of personal information by private sector organizations engaging in commercial activities in Canada. Under the current Canadian privacy regime, consent for collecting personal information may be express or implied, with implied consent being sufficient under certain circumstances. However, CPPA would require a more stringent approach. It states that organizations must obtain "valid" consent from individuals, ensuring that information is provided in plain language that individuals would reasonably be expected to understand (Davidson et al., 2022). The Artificial Intelligence and Data Act (AIDA) is the federal government's first attempt to regulate artificial intelligence (Arai, 2023). AIDA lays out requirements for "persons responsible" for AI systems, including anonymizing data and conducting assessments for high-impact AI systems (Arai, 2023). However, it lacks specific

regulations, leaving many facets of AI regulation dependent on future developments. The Personal Information and Data Protection Tribunal Act (PIDPTA) establishes a specialized tribunal tasked with hearing appeals related to data protection and privacy issues, aiming to enhance enforcement of the Canadian Consumer Privacy Act (CPPA). Addressing a shortcoming of PIPEDA, through enforcement of CPPA, PIDPTA can introduce penalties up to the greater of $10 million or 3 per cent of an organization's gross global revenue in its financial year (Davidson et al., 2022). Moreover, organizations that contravene certain sections under the CPPA may be found guilty of an indictable offence and liable to a fine of up the greater of $25 million or 5 per cent of the organization's gross global revenue in its financial year (Davidson et al., 2022). Bill C-27 would be a step towards enhancing Canada's privacy regulations, although we have yet to see the law in action.

Suggested Remedies

The harms posed by data brokers are real. Individual identity theft is often regarded as a more serious issue than data breaches and data broker surveillance (Roderick 2014: 739), but this should not be the case. Below we suggest some remedies to reign in the surveillance powers of data brokers.

First, there should be restrictions on the purchasing of personally identifiable information. Currently, there is no legal regime that prevents brokers and other companies from sharing these data with other individuals and companies. A wide array of entities from political campaigns to antivirus companies buy and sell data with brokers. Political campaigns and parties should only have limited (if any) access to sensitive information as it can be exploited to manipulate voters and undermine their fundamental right to vote. For example, data on U.S. individuals shared and/or sold by U.S. data brokers could be used for activities that threaten elements of the U.S. democratic electoral system, such as foreign government micro-targeting of individuals with election disinformation intended to dissuade voter participation (e.g., as the Russian Internet Research Agency did to Black communities in 2016 (Sherman, 2021). Laws that restrict the purchase of these kinds of data from data brokers should be enacted.

Data brokers are always at risk of being hacked if they do not invest in cyber security, evidenced by the ChoicePoint data breach case in 2008. The primary cause of such breaches is data brokers allocating fewer resources for internal IT security and privacy protection. Given the sensitive data they have access to, it is necessary to have a government mandate dictating minimum privacy standards or measures a data broker should have in place to conduct business.

In addition, there should be an independent regulatory body dedicated to regulating data brokers. A useful analogy is the SEC (U.S. Securities and Exchange Commission) whose primary purpose is to enforce law against market manipulation. Rather than taking a "sectoral approach" to data collection which enacts a series of unconnected laws targeting specific markets (Kuempel, 2016), a more centralized approach is needed such as the European Union data protection framework which mandates the presence of a centralized data controller accountable for adhering to laws and regulations governing data privacy. The problem with data brokers is that they invade consumer rights to privacy and also subject consumers to indiscriminate, mass surveillance. Creating a centralized portal where data brokers must disclose information, mandating disclosure on data accumulated by data brokers and its usages, and other measures to increase transparency, are all possible (pg. 212). Self-regulation is not viable because data brokers have demonstrated little interest in regulating their conduct and their activities are already deeply embedded in circuits of capital. The European Union's (EU) approach to data and privacy regulation with initiatives such as the EU Data Directive shows some promise for regulating data brokers. The EU Data Directive also imposes some tougher penalties for lack of compliance (pg. 231).

There should also be a mandate ensuring transparency in data procurement and transfers. Data brokers often claim to work with 'anonymous' data, a claim that operates to shield unethical data transfers from oversight. Currently, data is exchanged between multiple data brokers, making accountability difficult. For practices of data procurement, a model similar to that of the EU Data Directive should be adopted. Under this model, data subjects are "entitled to know what personal data of theirs is being processed, the lawful basis of that processing, as well as whether or not their personal data is being processed by the controller or by a third-party processor" (Martin, 2020), and policy

should reflect this. In addition, data brokers should be prohibited from acquiring more data than necessary (i.e., data minimization), which can also help protect consumer data from the risk of breaches (Reviglio, 2022). In the current scenario, data brokers unethically acquire data through phishing, exploiting ambiguous user consent, and using covert tracking technologies like cookies and fingerprinting, often without explicit approval. Additionally, they engage in web scraping and data crawling, collecting information without transparent disclosure or user consent, casting the widest net possible.

Activities like theft or cheating yield quantifiable consequences, such as loss of life or financial impact. To prevent these, we can formulate policies based on the quantified implications. This often happens in the realm of criminal law. In contrast, the consequences of data collected by data brokers lack predictability and cannot be quantified. For example, data collection of user browsing habits can enable data brokers to predict voting preferences, posing a potential threat to individuals' fundamental right to vote. It is crucial to implement policies capable of foreseeing and quantifying the implications of collected data to prevent such cases. For instance, in this case, data on the browsing habits of users should be restricted as it can lead to the undermining of consumer's fundamental right to vote.

Currently, there are no international laws governing data brokers. While organizations such as the International Association of Privacy Professionals (IAPP), IARC Data Protection Policy, and UN Principles on Personal Data Protection and Privacy provide privacy best practices, data brokers are not legally required to follow them, as there are no laws enforcing compliance. To address this gap, intermediaries like web browsers, Internet Service Providers (ISPs), and web servers can also play a vital role in regulating data brokers. In an effort to implement a centralized international approach, ISP's or Web Browsers can restrict the collection of data-by-data brokers to safeguard consumer privacy. Given the challenges of implementing comprehensive international policy, it may be easier to implement a policy at the level of intermediaries. Google has taken a step forward to address this concern by limiting the collected third-party cookies, which serve as a significant source of collection for data brokers.

Finally, while data brokers use algorithms and predictive models to produce added value from the raw data sets they amass, the added value often leads to potential misuse of personal information such as in the case of political microtargeting and targeted commercials. This means there is a close connection to data science, making data science students ideal candidates for these roles. Similar to teaching Corporate Social Responsibility (CSR) to business students, the government or other authorities should develop a curriculum to educate students about the responsible use of data. This approach to data ethics education would help potential future data broker employees make informed and ethical decisions, thereby preventing unethical practices in the field.

Acknowledgement and Addendum

We acknowledge the Social Sciences and Humanities Research Council of Canada (SSHRC) and the Knowledge Synthesis Program (KSG) for funding this project. It is important to note that we have also attempted to contact data brokers for interviews so that they could add their perspective or correct the record as they see it. We contacted over 100 data brokers in North America. However, not one of them responded to our requests for interviews.

References

Abad, G. L., & Orón, L. C. (2016). How Social Networks and Data Brokers Trade with Private Data. *Redes. com: Revista de Estudios para el Desarrollo Social de la Comunicación*, 14, 84-103.

Anthes, G. (2014). Data Brokers are Watching You. *Communications of the ACM*, 58(1), 28-30.

Arai, Maggie. (2023). Five Things to Know about Bill C-27. Schwartz Reisman Institute. University of Toronto. April 17.

Arantes, J. (2023). Educational data brokers: Using the Walkthrough Method to Identify Data Brokering by Edtech Platforms. *Learning, Media and Technology*, 1-14, 1-14.

Baccaro, F. (2021) Economics of Privacy: The Role of Data Brokers. Thesis

Bircкan, G., Dutra, M. L., de Macedo, D. D., & Godoy Viera, A. F. (2020). Personal Data Protection and Its Reflexes on the Data Broker Industry. In *Data and Information in Online Environments: First EAI International Conference*, DIONE 2020, Florianópolis, Brazil, March 19-20, 2020, Proceedings 1 (pp. 103-117). Springer International Publishing.

Brill, J. (2013). Demanding Transparency from Data Brokers. *The Washington Post*, 15.

Brooks, N. (2001). Data Brokers: Background and Industry Overview. *Wall Street Journal*, 5(5), 552a.

Crain, M. (2018). The Limits of Transparency: Data Brokers and Commodification. *New Media & Society*, 20(1), 88-104.

Davidson, J. R., Austin, R., Troshchynsky, A., & Di Felice, V. (2022). Bill C-27, Proposed Amendments to Canada's Federal Privacy Legislation Affecting Private Sector Organizations. *Intellectual Property Journal*, 35(1), 71-97.

Department of Justice Canada. (2023, November 27). Bill C-27: An Act to Enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to Make Consequential and Related Amendments to Other Acts.

Dudley, C. (2015). Strange Intersections Between Data Brokers and the CFAA: A Financially Supported Attack on Privacy. SSRN.

Elvy, S.-A. (2017). Paying for Privacy and the Personal Data Economy. *Columbia Law Review*, 117(6), 1369-1459.

Fairclough, B. (2016). Privacy Piracy: The Shortcomings of the United States' Data Privacy Regime and How to Fix It. *Journal of Corporation Law*, 42(2), 461-480.

Federal Trade Commission. (2012). *FTC to Study Data Broker Industry's Collection and Use of Consumer Dat.*

Frederik, J. Z. B., Judith, M., Sanne, K., Ronan, Ó. F., Kristina, I., Tom, D., Balazs, B., & Claes, de V. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*, 14(1), 82-96.

Gervais, A., Filios, A., Lenders, V., & Capkun, S. (2017). Quantifying Web Adblocker Privacy. In *Computer Security–ESORICS 2017: 22nd European Symposium on Research in Computer Security*, Oslo, Norway, September 11-15, 2017, Proceedings, Part II 22 (pp. 21-42). New York: Springer International Publishing.

Glez-Peña, D., Lourenço, A., López-Fernández, H., Reboiro-Jato, M., & Fdez-Riverola, F. (2014). Web Scraping Technologies in an API World. *Briefings in Bioinformatics*, 15(5), 788-797.

Gu, Y., Madio, L., & Reggiani, C. (2021). Data Brokers Co-opetition. *Oxford Economic Papers*, 74(3), 820-839.

Kim, J. (2006). On the Data Trail: How Detailed Information About You Gets into the Hands of Organizations with

Whom You Have No Relationship: A Report on the Canadian Data Brokerage Industry, CIPPIC, 2006(82). *Information & Technology Law*, 10(1), 12-13.

Kim, J. (2023). Data Brokers and the Sale of Americans' Mental Health Data. Duke University Report

Kraus, D., (2020). Transparency as a First Step to Regulating Data Brokers.

Kuempel, A. (2016). The Invisible Middlemen: A Critique and Call for Reform of the Data Broker Industry. *Nw. J. Int'l L. & Bus.*, 36, 207.

Luscombe, A., Dick, K. & Walby, K. (2022). Algorithmic Thinking in the Public Interest: Navigating Technical, Legal, and Ethical Hurdles to Web Scraping in the Social Sciences. *Quality and Quantity*, 56, 1023-1044.

Martin, B. A. (2020). The Unregulated Underground Market for Your Data: Providing Adequate Protections for Consumer Privacy in the Modern Era. *Iowa Law Review*, 105(2), 865-900.

McCain, J. (2009). Applying the Privacy Act of 1974 to Data Brokers Contracting with the Government. *Public Contract Law Journal*, 38(4), 935-953.

Merzdovnik, G., Huber, M., Buhov, D., Nikiforakis, N., Neuner, S., Schmiedecker, M., & Weippl, E. (2017). Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 319-333). IEEE.

Mishra, S. (2021). The Dark Industry of Data Brokers: Need for Regulation? *International Journal of Law and Information Technology*, 29(4), 395-410.

Neally, D. (2019). Data Brokers and Privacy: An Analysis of the Industry and How It's Regulated. *Adelphia Law Journal*, 22, 30-46.

Nie, Y., & Han, X. (2019). Research On Consumers' Protection in Advantageous Operation of Big Data Brokers. *Cluster Computing*, 22, 8387-8400.

Oh, H., Park, S., Choi, J. K., & Noh, S. (2021). Deposit Decision Model for Data Brokers in Distributed Personal Data Markets Using Blockchain. *IEEE Access*, 9.

Palk, L., & Muralidhar, K. (2017). A Free Ride: Data Brokers' Rent-Seeking Behavior and the Future of Data Inequality. *Vand. J. Ent. & Tech. L.*, 20, 779.

Pardau, S. L. (2018). The California Consumer Privacy Act: Towards European-Style Privacy Regime in the United States. *Journal of Technology Law & Policy*, 23(1), 68-114.

Reviglio, U. (2022). The Untamed and Discreet Role of Data Brokers in Surveillance Capitalism: A Transnational and Interdisciplinary Overview. *Internet Policy Review*, 11(3), 1-27.

Rieke, A., Yu, H., Robinson, D., & Van Hoboken, J. (2016). Data Brokers in an Open Society.

Roderick, L. (2014). Discipline and Power in the Digital Age: The Case of the US Consumer Data Broker *Industry. Critical Sociology, 40(5), 729-746.*

Roose, K. (2020, July 26). Don't Ban TikTok. Make an Example of It. The New York Times.

Rostow, T. (2017). What Happens when an Acquaintance Buys Your Data? A New Privacy Harm in the Age of Data Brokers. SSRN.

Ruppert, E., Isin, E., & Bigo, D. (2017). Data Politics. Big Data & Society, 4(2).

Shenkman, C., Franklin, S. B., Nojeim, G., & Thakur, D. (2022). Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers.

Sherman, J. (2021). Data Brokers and Sensitive Data on US Individuals. Duke University Sanford Cyber Policy Program, 9.

Solove, D. J., & Hoofnagle, C. J. (2005). A Model Regime of Privacy Protection. *University of Illinois Law Review*, 2006(2), 357-404.

Spivak, Russell. (2020). Too Big Fish in the Digital Pond? The California Consumer Privacy Act and the Dormant Commerce Clause. *University of Cincinnati Law Review*, 88(2), 475-514.

Tsesis, A. (2014). The Right to be Forgotten and Erasure: Privacy, Data Brokers, and the Indefinite Retention of Data. *Wake Forest Law Review, 48,* 105-151.

Vashey, J. (2020). Data Broker Practices and Privacy Ethics: How to Take Back Control of Personally-Identifiable Information (Order No. 27961815).

Venkatadri, G., Sapiezynski, P., Redmiles, E. M., Mislove, A., Goga, O., Mazurek, M., & Gummadi, K. P. (2019). Auditing Offline Data Brokers via Facebook's Advertising Platform. In The World Wide Web Conference.

Wayne, L. D. (2012). The Data-Broker Threat: Proposing Federal Legislation to Protect Post-Expungement Privacy. *Journal of Criminal Law and Criminology*, 102(1), 253–282.

West, S. M. (2019). Data capitalism: Redefining the Logics of Surveillance and Privacy. *Business & Society*, 58(1), 20-41.

Zook, M., & Spangler, I. (2023). A Crisis of Data? Transparency Practices and Infrastructures of Value in Data Broker Platforms. *Annals of the American Association of Geographers*, 113(1), 110-128.