Back on the Data Trail:

The Evolution of Canada's Data Broker Industry

Samuelson-Glushko Canadian Internet Policy and Public Interest Clinic

Draft Report
March 31, 2018

Executive Summary

This report follows up on CIPPIC's 2006 study of the Canadian data brokerage industry. This groundbreaking study pulled back the curtain of this poorly understood industry. Data brokers are intimately familiar with consumers yet practically unknown to the average Canadian. The past decade has seen radical changes to the industry, its technology, and society itself. Simply, the world has changed. The emergence of big data analytics, behavioural advertising and social networking has transformed the shape of this industry. In this report, we take a look back at these changes and the impact they have had on the shape, practices, and products of the data broker industry in Canada.

What Are "Data Brokers"?

We will again use the definition for "data brokers" employed in our 2006 report, "On the Data Trail", as a company "whose primary business involves the trading and analysis of personal information". We observe at the outset that this is a contested term. The industry is complex. Key actors include:

- large online platforms such as Google and Facebook;
- marketers in the personal data and analytics industry;
- marketers in the risk analysis industry; and
- Other businesses spread throughout the economy, ranging from retailers to internet of things developers.

Sources of Personal Information

We categorize data brokers' sources of consumer data into four categories:

1. "Public" sources where data brokers take the position that consumer consent is not needed at all. These include public directories such as telephone directories and business and trade directories;
2. "Vendor sources" that sell personal information outright. These include:
    a. Subscription services, such as newspaper & magazine publishers, and book, music and movie clubs
    b. Retailers, including mail order retailers, retailers dealing with requests for product information, or managing warranty card and product registrations, and service providers such as telcommunications and financial institutions,
    c. Marketers involved in surveys, contests, and similar offers
    d. loyalty card service providers; and
    e. Non-Profit and Charitable Organizations that sell their lists;
3. "Utilizing sources" that use personal information for financial gain, such as social networks, although these services typically do not share their raw consumer data, but instead provide advertisers access to consumer profiles built on this data; and
4. "Anonymized" sources that sell consumer information that is not identified with individual persons.

The Data Brokers and their Services

The data broker industry has undergone drastic changes over the past decade due to technological advances related to the rise of the smartphone and the maturation of data analytics technology. New players have emerged to claim their stake in the data broker industry.

1. Marketing Cloud & Data Management Platforms - Marketing Cloud providers offer internet-based, “Software as a Service” (SaaS) “one-stop shops” for all marketing needs.

(a) Device Matching - These services identify individual customers across devices, allowing marketers to streamline ads and content to an individual user regardless of whether the consumer is using a phone, tablet, or computer.

(b) Data Management Platforms - These services offer a marketing database and interface for all types of consumer data, regardless of source, making that data actionable, offering insights into markets, and to serve targeted advertising

2. Social Media & Search - Social media platforms offer information to brokers not available in the past. People search tools and data append services “scrape” social media sites to add data to their products.

(a) Self-Service Onboarding - Social media services offer “online marketing” data services, in the form of “onboarding”. Onboarding allows marketers to use data from offline sources to find and target customers with online advertising. Social networks offer marketers “self-service” tools to locate their customers on these platforms and serve them with advertising.

(b) Consolidation and Strategic Partnerships - Social media companies have partnered with data brokers to enhance their services, allowing allowing marketers to use offline data to target social media users with advertisements.

3. Retargeting / Behavioural Advertising
Ad “retargeting” serves ads across the web to users who have visited or performed a certain action on a specific website or app.

4. Scoring & Risk Mitigation Products - Many data broker services score and categorize current and prospective customers for marketing and risk-mitigation products. Marketing applications score consumers as, for example, “high-value” prospective customers or “under-performing” customers, prompting business to target these consumers with appropriate sales strategies. Risk-mitigation services, in contrast, are used to identify risky clients, such as those likely to default on loan payments or those more likely to commit fraud.

(a) Industry-Specific Scoring - Some data brokers offer industry-specific scoring solutions, catering to the needs of specific industries, such as automobile sales.

(b) Scoring to Determine Level of Service Provided - Data brokers may use scoring to determine the level of service customers receive when making an inbound call to a company, allowing companies to prioritize the inbound calls of high value customers over those less valuable to the company.

5. People Search Products - People search products allow individuals to locate other individuals both physically and online and to discern between individuals with similar names. These products offer information about consumers obtained from publicly available sources, including social media. In Canada, people search products are reserved in their offerings than in the US.

Social Networks and Data Brokers

Social media sites are “social graphs” – networks of social connections that can be mapped. Social media is a source and purveyor of data in the data broker ecosystem. They are also themselves unique players in our data ecosystem, and in fact are acquirers of commercial data. Social networks’ massive user base provides for a rich source of data.

1. Social media as a source of personal information - Social networking platforms are among the richest and most accessible sources of personal information on consumers. This information is used by the social network to develop its own products and to provide marketers with the ability to directly target advertising at specific segments of the user base. Terms of use generally purport to provide consent to all of these activities. Commercial data brokers also mine user activity on social networking sites for publicly available information that users post to social networks themselves.

2. Social media sites and the commercialization of data -
Social media platforms not only collect user data, they also profit from it. These sites use data to group their users into lists based on characteristics such as age, gender, lifestyle, and education. Advertisers pay social networks to target their advertising at those groups of users who are most likely to have an interest in their product.

3. Social media data products - Social networks analyze user data to provide targeted advertising and measure its performance. Social networks are searching for new ways to commodify their users’ data through ad services. Analytics tools allow advertisers to better understand their customers and the people who are interested in their product. The platforms compete to provide advertisers with the most detailed analysis of the effectiveness of their ads – what type of people engage with the ad, for how long, and how often does the ad lead to an online purchase.

4. Social media data users - Social media platforms rely on commercial data brokers to supplement their own user data with “offline data” about users’ lives away from their sites. Search engines and social media sites rely on advertising as their primary source of revenue. By acquiring additional offline data, social media platforms are able to develop a

more precise profile of their users and the ads that they are most responsive to. This allows advertisers to target their ads to specific audiences that are most likely to be interested in a certain product.

Conclusion: Data Brokers and the Law

PIPEDA, regulates data brokers' collection, use, and disclosure of personal information. The practices we have identified in this report raise significant compliance issues under this law.

Consent - Organisations must obtain the consent of and individual for the collection, use and disclosure of personal information. Data brokers often obtain such information through retail partners or other third parties. It is difficult to obtain proper consent through third parties, particularly where the intended use is a secondary purpose.

Security - Personal data has become the object of data thieves' attention. However, we are seeing that data security can be difficult. While PIPEDA obliges organisations to implement security measures appropriate to the sensitivity of the information, it is an open question as to whether this obligation is sufficient to incentivize businesses to take consumer data security seriously. Imposing legal liability for inadequate security practices is worthy of consideration.

Exclusions from Legal Protection - PIPEDA only applies to certain kinds of personal information, and to commercial practices. Information deemed "publicly available" by regulation is excluded, as are non-commercial dealings, including those by political parties. While provincial privacy legislation can close these gaps, few such laws do. Similarly, while anonymous data does not constitute "personal information" under PIPEDA, it can act as a proxy for personal information when combined with geographically-based information such as postal codes (with added risk of inaccuracy).

Inappropriate Purposes - PIPEDA requires that organisations may only collect, use or disclose personal information for purposes that "a reasonable person would consider appropriate in the circumstances". Data brokers use of data analytics risks the use of personal information for profiling and discriminating among consumers on grounds that violate human rights law, or for unfair or unethical purposes.

Where industry practices exceed what a reasonable consumer would be comfortable with, these concerns raise genuine issues. Greater transparency and accountability is required, and privacy regulators must wield meaningful tools to ensure that organizations collect, use and share consumer data in accurate, fair, and appropriate manners.

Introduction

Over a decade ago, CIPPIC undertook a detailed study of the Canadian data brokerage industry in our 2006 report, “On the Data Trail: How detailed information about you gets into the hands of organizations with whom you have no relationship”.[1] The study was the first of its kind in Canada to pull back the curtain of this massive but little understood industry. Data brokers are peculiar businesses: they are intimately familiar with consumers yet practically unknown to the average Canadian.

Over a decade later, and despite radical changes in the structure and players in Canada’s data broker industry, this report continues to be the leading analysis of Canada’s data broker industry. Indeed, the Research Group of the Office of the Privacy Commissioner of Canada’s 2015 discussion paper on the industry, “Data Brokers: A Look at the Canadian and American Landscape”,[2] relied heavily on our report from the previous decade to describe the industry.

In the time that has passed since publication of our first report, the world has changed. The emergence of big data analytics, behavioural advertising and social networking has transformed the shape of this industry. Old players have been purchased and merged. New sources for our personal information – our phones, our social networking accounts – have complicated the picture. In this report, we take a look back at these changes and the impact they have had on the shape, practices, and products of the data broker industry in Canada.

In Part I, we undertake the definitional work and define our subject: what, exactly, is a “data broker”? What businesses that deal with individuals’ data are not data brokers? Where do social networks fit in this space? In Part II, we revisit sources of personal information. We answer the question, where do data brokers get their information about us? In Part III, we identify the data brokers, or at least some of them, and their service offerings. We identify categories of companies in the data broker business, and provide profiles of individual businesses to offer a clearer picture of these categories. And we turn to the big questions: how do data brokers make money? What do they sell, and who are the buyers? In Part !V, we look at social networking businesses and offer an understanding of where they fit in the data broker world. We conclude with a look at the Canadian regulatory environment, with a particular view to the challenges PIPEDA, Canada’s commercial private sector privacy protection legislation, encounters in attempting to regulate data brokers.

Throughout the report, we offer asides and examples to better illustrate the challenges the data broker industry poses to consumers, legislators and regulators. For example, two major news stories book-ended our studies: in September of 2017, news broke of a major security breach at the credit bureau Equifax, one of the largest data brokers and holder of reams of sensitive financial information about individuals. In March of 2018, news broke of Cambridge Analytics’ work on Facebook’s platform, and the potential for the use of the data it gained through Facebook to influence the American election of 2016. We look at these

---

[1] https://cippic.ca/sites/default/files/May1-06/DatabrokerReport.pdf.

[2] (September 2014), https://www.priv.gc.ca/media/1778/db_201409_e.pdf.

stories in these pages, and identify the challenges posed by, and lessons learned from, these dealing with individuals' data.

Part I What Are “Data Brokers”?

In our 2006 report, “On the Data Trail”, we defined a data broker as a company “whose primary business involves the trading and analysis of personal information”.[3] We will again use that definition to define the scope of our report. For our purposes, we will examine the industry that both trades in and analyzes personal information.

However, we observe at the outset that this is a contested term. How do companies that merely provide data, or serve as a platform for the collection of data, fit into the data broker ecosystem? How do service providers who, for example, offer data management platforms, fit? Are social networks “data brokers” - while they certainly analyze our personal information, do they trade in that data?

Other more recent analyses of data brokers have taken a broader approach to their subject matter. The US Federal Trade Commission, for example, in its 2014 report on data brokers, defined them as:

> Companies whose primary business is collecting personal information about consumers from a variety of sources and aggregating, analyzing, and sharing that information, or information derived from it, for purposes such as marketing products, verifying an individual’s identity, or detecting fraud.[4]

No actors in the data-driven ecosystem self-identify as data brokers. Some industry terms include: “Data Provider” (Facebook Partners; Google Ad Words Partners), “Database Marketer” (LUMA); “Identity Resolution & People-Based Marketing” (Axciom); Marketing analytics (Environics Analytics);”Data-as-a-Service (DaaS) provider” (Oracle Data Cloud); and “Global risk information provider” (Transunion).

Key Players in Personal Data Ecosystem

Definitions aside, there is no denying that the personal data ecosystem is large and complex. Below, we describe categories of actors in this ecosystem.

Large online platforms

One of the most important developments in the personal data ecosystem is the rise of pervasive online platforms. The Center for Global Enterprise describes these entities as follows:

> Platforms have unique characteristics, with a central feature being the presence of network effects. Network effects are prevalent in platforms, and they mean that more

[3] CIPPIC, “On the Data Trail” at 4.
[4] FTC, “Data Brokers: A Call for Transparency and Accountability” (May 2014) at , https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf,

> users beget more users, a dynamic which in turn triggers a self-reinforcing cycle of growth. Further, most of today's platforms are digital: they capture, transmit and monetize data, including personal data, over the Internet. They may not be purely digital; in that they may have physical elements included in the product offering, but most successful platforms today take advantage of the power of pervasive Internet connectivity in the hand of billions of users and have at their heart a software engine. [5]

There are relatively few major players in this category, and these encompass the likes of Google and Facebook.

Personal data and analytics industry

(1) Marketing

The online marketing industry fuels both the demand for data, and the technology to optimize data. Online, marketers employ sophisticated advertising technologies to service both ad Sell-Side and Demand-Side Platforms. Data Management Platforms are "one-stop shops" for marketers to organize and manage collection of first party data and access third party data. These sophisticated platforms are offered by major vendors such as Acxiom and Oracle and offer functionality that includes ad retargeting, analytics, and ad supply services. Features of modern Data Management Platforms include:[6]

- Cross Device Identity Management (CDIM) - This feature seeks to tie consumers, across all of their devices, back to the platform's consumer identifying code. The feature typically works on both certainties ("this is the consumer") and probabilities.
- CRM Onboarding - Data Management platforms aim to match website-side data collection with offline data-sets.
- Programmatic Segmentation - DMPs employ machine learning to automatically identify potential market segments for marketers.
- Attribution - DMPS are able to identify how well marketing is working, by tying placement of ads to sales of products.
- Tag Management - tag management layers website-side outcomes with a layer of audience intelligence, enabling marketing tools such as split testing, A/B testing, and multi-variate testing.

---

[5] Peter C. Evans and Annabelle Gawer, The Rise of the Platform Enterprise: A Global Survey (January, 2016) at 5; https://www.thecge.net/app/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf.

[6] Tim Norris, "What are the main differences between these three products: BlueKai, Krux and Lotame?" (3 July, 2015), https://www.quora.com/What-are-the-main-differences-between-these-three-products-BlueKai-Krux-and-Lotame,

(2) Risk Data

Financial risk analysis businesses have long been among the biggest players in the data broker industry. These include credit bureaus such as Transunion and Equifax, but also Identity verification and fraud detection service and insurance companies.

(3) Other Business in the Personal Data Ecosystem

In addition to these major categories of data brokers, there is an enormous range of businesses involved in data brokerage. Retailers sell consumer purchase data to market research companies and consumer data brokers. In fact, one of the world's largest retailers, Walmart, has stopped "participating in the data selling" à because they see themselves as essentially a large platform competing with the likes of Amazon and have adopted a walled-garden approach to their customers' data.[7] Media organizations and digital publishers similarly trade in their users' data. For example, Spotify sells streaming data - "unique listening preferences and behaviors of Spotify's 100 million users in 60 countries" - to the advertising firm WPP.[8] Telecom companies and Internet Service Providers have long sought to glean greater insights about the activities of their users, even in Canada.[9] the emergence of the internet of things promises to offer data collection opportunities to an extremely wide range of device manufacturers, all of which may feed into the data ecosystem and the hands of data brokers.

---

[7] Wolfie Christl, "Corporate Surveillance in Everyday Life: How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions", Report by Cracked Labs (June 2017) at 14, http://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf.

[8] "WPP's Data Alliance and Spotify announce global data partnership", (15 November, 2016), https://www.wpp.com/wpp/press/2016/nov/15/wpp-data-alliance-and-spotify-announce-global-data-partnership/.

[9] See, e.g., "Bell faces $750-million lawsuit over tracking of customer's cellphone Internet usage" Globe and Mail (17 Aopril, 2015), https://www.theglobeandmail.com/report-on-business/industry-news/the-law-page/bell-faces-750-million-lawsuit-over-tracking-of-customers-cellphone-internet-usage/article24001810/.

Part II Sources of Personal Information

Perhaps the most troubling aspect of the data broker industry to consumers is the mystery of where the industry obtains its data. If a business is selling information about us, ought we not to be engaged in the initial collection of that data?

In out 2006 Report, we identified numerous sources of data, including:[10]

- Newspaper & Magazine Publishers
- Book, Music and Movie Clubs
- Mail Order Retailers
- Service Providers
- Surveys
- Warranty Cards and Product Registrations
- Contests, Offers and Loyalty Cards
- Seminars and Conferences
- Requests for Information
- Websites
- Non-Profit and Charitable Organizations

While these sources remain relevant, much has changed since 2006. Social media platforms now offer fine-grain detail about individual preferences and personalities. The explosion of data flows (e.g. more devices (Internet of Things), cashless economy) and improved techniques for re-identifying anonymized data have challenged consumer anonymity.

Below, we categorize all of these new and old sources of consumer data into four categories:

1. "Public" sources where data brokers take the position that consumer consent is not needed at all;
2. "Vendor sources" that sell personal information outright;
3. "Utilizing sources" that use personal information for financial gain; and
4. "Anonymized" sources that sell consumer information that is not identified with individual persons.

Data brokers get information about you from…

(1) "Public" information

Canada's federal commercial-sector privacy legislation generally requires consent to the collection, use and disclosure of personal information. However, the law effectively excludes from these protections information deemed "publicly available" by regulation. These

[10] CIPPIC, "On the Data Trail" at 29-36.

regulations exclude from protection personal information published directories such as telephone directories.[11] Indeed, telephone directories are a rich source of data about consumers.

The Regulations Specifying Publicly Available Information

PIPEDA's regulations carve out four categories of publicly available personal information from protection under the law. It is worth reproducing the relevant provisions in full:

1 The following information and classes of information are specified for the purposes of paragraphs 7(1)(d), (2)(c.1) and (3)(h.1) of the Personal Information Protection and Electronic Documents Act:

- (a) personal information consisting of the name, address and telephone number of a subscriber that appears in a telephone directory that is available to the public, where the subscriber can refuse to have the personal information appear in the directory;
- (b) personal information including the name, title, address and telephone number of an individual that appears in a professional or business directory, listing or notice, that is available to the public, where the collection, use and disclosure of the personal information relate directly to the purpose for which the information appears in the directory, listing or notice;
- (c) personal information that appears in a registry collected under a statutory authority and to which a right of public access is authorized by law, where the collection, use and disclosure of the personal information relate directly to the purpose for which the information appears in the registry;
- (d) personal information that appears in a record or document of a judicial or quasi-judicial body, that is available to the public, where the collection, use and disclosure of the personal information relate directly to the purpose for which the information appears in the record or document; and
- (e) personal information that appears in a publication, including a magazine, book or newspaper, in printed or electronic form, that is available to the public, where the individual has provided the information.

Note that not all publicly information is excluded from protection, but only those classes identified. In PIPEDA Case Summary #2005-297, the Assistant Privacy Commissioner found that an organization that collected the business e-mail address of an individual of his employer's website and used it to contact him for marketing purposes without his consent contravened PIPEDA.[12]

(2)(a) Sources that sell your personal information to others

In our 2006 Report, we identified a number of classes of organizations that sell consumer data to third parties. These include

- Subscription services, such as newspaper & magazine publishers, and book, music and movie clubs

[11] Regulations Specifying Publicly Available Information, SOR/2001-7, https://www.canlii.org/en/ca/laws/regu/sor-2001-7/latest/sor-2001-7.html,

[12] PIPEDA Case Summary #2005-297(2005), https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2005/pipeda-2005-297/.

- Retailers, including mail order retailers, retailers dealing with requests for product information, or managing warranty card and product registrations, and service providers such as telcommunications and financial institutions,
- Marketers involved in surveys, contests, and similar offers
- loyalty card service providers; and
- Non-Profit and Charitable Organizations

All of these activities involve the generation of lists of individuals along with contact details, which may be supplemented by information about the consumer generated and recorded by the list manager.

(3) Sources that use your personal information for financial gain (but don't sell it)

An interesting category of commercial actors worth considering are those that use consumer information for financial gain but don't actually sell the data. These organizations use the data internally, but emphasize that they do not sell it externally. Social networks are an excellent example of these kinds of organizations. They use their data to profile their users, then sell access to those profiles to advertisers. These companies say they don't sell your data to others, but they don't need to: personal data powers their ad networks and ad inventories. This practice "gets around" the consent needed to share data by vertically integrating their platform and ad network; they monetize the data without needing to share it externally.

The model is in may ways similar to that of a 20th century newspaper or television network: they sold access to their audiences, and so took great pains to understand who their audience was. They are attention merchants.[13]

We address the interesting position of social networks within the data broker economy in much greater detail in Part IV, below.

(4) Sources that "anonymize" your data before selling it to others

PIPEDA requires consent to the collection, use and disclosure of personal information. The difficulty of obtaining the consent of a consumer to inclusion of their personal information in the databases of a data broker leads many organizations to rely on anonymous data to populate their databases. This is effectively, a two-step manoeuvre to avoid falling afoul of the law. First, consumer data is collected and associated with a territory, such as geographic areas associated with postal codes or census dissemination areas. All data points associated with that geographic area is then aggregated across the area. All individuals within that geographic area are then associated with that aggregated data.

The shortcomings of anonymization techniques and the potential for re-identification of individuals is well documented.[14] Simply, it is surprisingly difficult to achieve perfect

[13] See Tim Wu, The Attention Merchants: The Epic Scramble to Get Inside Our Heads, Knopf (2016).
[14] See Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2010) 57 UCLA L.R. 1701 https://www.uclalawreview.org/pdf/57-6-3.pdf and Latanya Sweeney, "Simple Demographics Often Identify People Uniquely", Carnegie Mellon University, Data Privacy Working Paper 3. (2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf.

anonymization. This means that even where data brokers use anonymized data, consumers face the risk of identification. Indeed, the Privacy Commissioner of Canada has concluded that "Personal information that has been de-identified does not qualify as anonymous information if it is still possible to link the de-identified data back to an identifiable individual."[15]

The Public Interest Advocacy Centre's 2011 Report, "Consumers Anonymous",[16] identified many problems with anonymized data, including:

- The cumulative disadvantage to consumers resulting from rational discrimination;
- Privacy loss due to online tracking and targeting;
- A loss of consumer and social autonomy;
- The potential for misuse and abuse or anonymized or reidentified data; and
- Harms to democratic and liberal values resulting from the expansion of laws, policies and procedures mandating consumers to provide personal information.

* * *

Having identified the various sources by which data brokers collect their data, it is time to turn to the brokers themselves and the services they offer to understand how they use our data.

---

[15] OPC Case Summary #2009-018: "Psychologist's anonymized peer review notes are the personal information of the patient", https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2009/pipeda-2009-018/.

[16] PIAC, "Consumers Anonymous? The Privacy Risks of De-Identified and Aggregated Consumer Data" (6 October 2011), http://www.piac.ca/wp-content/uploads/2014/11/piac_consumers_anonymous_paper_final_6oct2011.pdf.

Part III The Data Brokers and their Services

Since CIPPIC's 2006 report, the data broker industry has undergone drastic changes. Most have resulted from technological advances made over the past decade. The rise of the smartphone has exponentially increased the amount of information available to data brokers. Individuals now disclose vast amounts of personal information publicly on social media platforms. Simultaneously, increases in computing power have made it possible to process all of this new and existing data and glean insights that once would have been impossible for humans or older technologies to uncover. New players have emerged to claim their stake in the data broker industry. These range from tech behemoths such as Google, Adobe, and Oracle to smaller, more specialized startup data broker companies, many of which are quickly acquired by the industry's big players once establishing their data broker credentials. Below is an overview of some of the products currently available. Almost all products discussed below are available in both Canada and the U.S.

Marketing Cloud & Data Management Platforms

Marketing Cloud providers offer an array of internet-based data broker services. Marketing Cloud services are, for the most part, a "Software as a Service" (SaaS) product; they are computing services offered to marketers over the internet. These services span the three major groups of data broker marketing products identified in the FTC's 2014 report, Data Brokers: A Call for Transparency and Accountability.[17] They include those traditionally associated with direct marketing, such as "data append" services and "marketing lists," as well as "online marketing," and "marketing analytics" products. Often, products are bundled as part of a suite such as the Adobe Marketing Cloud.[18] The idea is to offer marketers a one-stop marketing solution. These products simultaneously act as a client database and analytics platform, tying in client data from past purchases, website visits, social media, and second and third-party sources into one profile; ad server; email marketing solution; web content management system; customer relationship management system; and broker for the exchange of client information with second and third parties. Major providers of marketing cloud solutions include Oracle, Salesforce, and Adobe. The marketing cloud and data management platforms discussed below are offered in both Canada and the US.

The past ten years have seen a blurring of the lines between activities traditionally associated with offline, direct marketing products and online products. What is now on offer are 360 degree, holistic marketing solutions that allow marketers to reach their clientele on virtually any platform they wish, at any time. Take the above-mentioned Adobe Marketing Cloud for example; marketers can append their marketing lists with second and third party data or gain access to third party marketing lists using Adobe's "Audience Marketplace;"[19]

---

[17] The FTC, in its 2014 report, drew three broad categories of marketing products offered by data brokers: direct marketing products, online marketing products, and marketing analytics products. See FTC report 23-31.

[18] http://www.adobe.com/ca/marketing-cloud.html

[19] https://helpx.adobe.com/audience-manager/how-to/audience-marketplace.html; a similar data marketplace is offered by Salesforce through its DMP (Data Management Platform).

they can then use this information to deliver targeted advertising on a number of platforms and across devices and locate their brick-and-mortar customers online.

The Audience Marketplace offers both traditional third party data, such as demographic and geographic data compiled by data brokers, and more granular "second party" data, direct from other marketers. Second party data is often touted as allowing marketers to gain deeper insight into their customers' purchase and media habits. The availability of second party data on a wide scale is a relatively recent phenomenon. Second party data was historically unavailable to those outside of a given organization, but now services such as Adobe's "Audience Marketplace" and Salesforce's Data Management Platform facilitate the sharing of such information between unrelated companies. For instance through one of these marketplaces, a hotel chain may purchase an airline's customer or website visitor dataset, giving it access to detailed information not usually available to those outside of the airline. This dataset may reveal detailed information about a customer's flight preferences, if they are likely to purchase upgrades, or what ads spurred the customer to purchase.

Adobe Marketing Cloud users can also access third party marketing lists through the Audience Marketplace. Such lists are often provided by traditional data brokers such as Acxiom, with whom Adobe has entered into a "Premier Partnership." Audience Marketplace allows users to access Acxiom's data to "enrich and extend their existing 1st party data audience segments with [Acxiom's] demographic, lifestyle and purchase intent data."[20] Such third party data can also be used as the basis for serving ads to prospective customers based on demographic, geographic, and psychographic characteristics. Information gained from Adobe's Audience Marketplace can similarly be used for registration targeting. This involves augmenting the information held about a website's registered users with third or second party data to serve the users with relevant ads or content.

Device Matching

Oracle, Adobe, and Salesforce all provide products that identify individual customers across various devices. The notion behind these products is that by identifying related devices, marketers can serve streamlined ads and content to an individual user. Since the advent of smartphones and tablets, marketers have struggled to identify their customers online. The same customer may access a marketer's website or applications from several devices, all without ever having logged in. Traditionally, this one customer would appear to the marketer as several different website visitors. With device matching, marketers can identify a group of devices associated to one individual.Adobe's product, known as the "Marketing Cloud Device Co-op," identifies related devices by using anonymized login data.[21] Co-op members integrate Adobe's code into their website. When a visitor to that website logs in from a device, Adobe adds that device to a "device cluster" associated to the login. Other marketers who are members of the co-op, and whose websites the user may not have logged into, can then identify related devices through the device cluster and serve relevant ads and content to those devices as a group.

---

[20]http://www.businesswire.com/news/home/20160927005480/en/Acxiom-Announces-Premier-Partnership-Adobe-Audience-Data

[21] Information about the the coop and its members is found here: https://cross-device-privacy.adobe.com/

Data Management Platforms

Companies such as Adobe, Salesforce, Oracle, and Nielsen have all positioned themselves as leaders in the Data Management Platform (DMP) market. DMPs serve as a marketing database and interface for all of the types of data mentioned above (first, second, and third party). The aim of DMPs is to make the vast quantities of data now collected by marketers actionable, to gain insights into existing and prospective markets, and to serve targeted advertising. Several platforms use artificial intelligence to analyze and segment customer data, serve ads to prospective customers, and optimize ad campaigns.[22] Adobe's Audience Manager helps marketers build "unique audience profiles so you can identify [the marketer's] most valuable segments and use them across any digital channel."[23] Each of these major players also seem to be expanding and improving their product offering through proprietary development and strategic acquisitions.[24]

Social Media & Search

One of the greatest shifts that has occured in the past ten years in the data broker industry is the rise of social media platforms. These platforms offer an abundance of information to brokers not available in the past. People search tools "scrape" social media sites to add data to their search results. Data append services do the same, allowing marketers to add detailed demographic and psychographic information to customer profiles. Furthermore, social media companies themselves have stepped into roles traditionally played by data brokers. The social media services below are all offered in both Canada and the US.

Self-Service Onboarding

For several years now, social media services such as Facebook, have offered "online marketing"-types of data broker services, specifically in the realm of "onboarding." Onboarding allows marketers to use data from offline sources to find and target customers with online advertising. In the past, data brokers would take a list of customers provided by a marketer, and run that list against lists of registered users purchased from certain websites to find matches. The broker would then serve ads to the matches. Social media platforms such as Facebook and Twitter, with their own massive lists of registered users, now offer marketers "self-service" tools to locate their customers on these platforms and serve them with advertising on social media services and beyond.[25]

---

[22] Both Nielsen and Salesforce's platforms employ artificial intelligence.

[23] http://www.adobe.com/data-analytics-cloud/audience-manager.html

[24] Since 2014, Oracle acquired both Bluekai and DataLogix, Salesforce purchased Krux, and Nielsen purchased eXelate to bolster their offerings.

[25] It may be of interest to some that social media onboarding services have not only been used by marketers, but also allegedly by Russian state actors to influence the 2016 US election (https://www.washingtonpost.com/business/economy/russians-took-a-page-from-corporate-america-by-using-facebook-tool-to-id-and-influence-voters/2017/10/02/681e40d8-a7c5-11e7-850e-2bdd1236be5d_story.html).

Facebook's onboarding tool is known as "Custom Audiences."[26] This tool allows marketers to use offline and other data to serve ads to custom segments across Facebook's advertising network. As of late 2017, marketers can create a custom audience from three different sources: customer contact lists (phone numbers or email), website visitors, and app users. To make a custom audience from a contact list, advertisers upload a list of customer email addresses or phone numbers to Facebook's advertising platform. Facebook then runs the list against its user database and matches the emails and phone numbers to existing Facebook users. The users are then grouped together as a custom audience that the advertiser can then target for advertising purposes. Facebook states that the unmatched emails and phone numbers are deleted.[27] Advertisers can also create custom audiences of visitors to their websites. This is achieved by uploading the "Facebook Pixel," a small snippet of code, to the advertiser's website. The pixel allows Facebook to identify Facebook users who have visited the website. These visitors can then be targeted with advertising across Facebook's ad network. Advertisers can create Custom Audiences made up of all website visitors or base an audience on certain actions performed or not performed by visitors on the website. For instance, an advertiser may create a Custom Audience of visitors who added an item to their online shopping cart, but did not complete the transaction, or they may create an audience of visitors who clicked on a specific link. Lastly, companies with their own mobile apps can create Custom Audiences from their app users. Similar to website Custom Audiences, advertisers can create audiences from their app users at large or target users who have taken specific actions on their apps (e.g. viewing a certain product or achieving a certain level in a game). To create these audiences, app makers must integrate certain Facebook code into their apps. The code matches app users with Facebook users to enable ad targeting across Facebook's ad network.

While Custom Audiences is a powerful tool on its own, Facebook also offers a service known as "Lookalike Audiences."[28] Lookalike Audiences allow marketers to expand their advertising reach through segments of Facebook users that resemble a specified group. Marketers can create a Lookalike Audience of Facebook users with similar characteristics to a Custom Audience, a website's visitors, or the fans of a company's Facebook Page.[29] Marketers can select the geographic location and desired size of a Lookalike Audience. The smaller the audience, the more closely it will reflect the characteristics of the marketer's base audience; the larger the audience, the more general it will be in its resemblance.

Google, Twitter, and Snapchat all offer similar tools to Facebook's Custom and Lookalike Audiences.[30] Marketers can upload lists of client emails to each of these services. The lists are then cross referenced with Google, Twitter, or Snapchat's user database, identifying any matches. Marketers can then serve ads to the matched users on each of these companies' platforms. In Google's case, its "Customer Match" audiences can be targeted across

---

[26] https://www.facebook.com/business/a/custom-audiences
[27] https://www.facebook.com/business/help/112061095610075
[28] https://www.facebook.com/business/a/lookalike-audiences
[29] https://www.facebook.com/business/a/lookalike-audiences
[30] https://support.google.com/adwords/answer/6379332?hl=en; https://business.twitter.com/en/targeting/tailored-audiences.html; https://forbusiness.snapchat.com/audiences.

YouTube, Gmail, Google Search Network, and Google Shopping.[31] As of late 2017, however, it was not possible to use Customer Match to serve ads through Google's massive Google Display Network at large.

Consolidation and Strategic Partnerships

Social media companies have partnered with established and start-up data brokers to enhance their services. For instance, Twitter partners with Acxiom and Oracle's DataLogix, allowing marketers to use offline data to target Twitter users with advertisements.[32] Through what are known as "Partner Audiences," marketers can now "target Twitter Ads to users who have shown powerful signals of intent off of Twitter."[33] More than 1,000 partner audiences are available to marketers directly through the Twitter Ads user interface.[34]

Retargeting / Behavioural Advertising

Over the past 10-15 years, several companies specializing in ad "retargeting" have emerged. Two leaders in the field are AdRoll and Criteo. These companies specialize in serving ads across the web to users who have visited or performed a certain action on a specific website or app. Typically this is achieved by embedding code in one's website that places a cookie file in a visitor's web browser upon visiting the site. When the visitor leaves the site without completing a certain action, such as purchasing an item in their shopping cart, they are served with ads as they browse the web.

Whereas in the past, most retargeting was achieved without using personally identifiable information, that has changed with the introduction of onboarding tools. For example, Criteo offers "Audience Match," an onboarding tool that allows marketers to serve ads to their existing customers across the web using the marketer's Customer Relationship Management (CRM) data. This tools works similarly to those of the social media companies mentioned above, with marketers uploading customer lists to Criteo's user interface. Customers are then matched to "deterministic IDs" within Criteo's "Shopper Graph" database. Criteo boasts a match rate of over 60 percent.[35]

Shopper Graph is one of Criteo's foundational technologies, serving as the backbone to several products. Shopper Graph pools identity, interest, and measurement data. As Criteo puts it:

> "Identity data connects shoppers and their devices, apps, and online/offline environments. Interest data links shopper interest to [a marketer's] product catalog. Measurement data provides an understanding of sales and conversions across [Criteo's] network of retailers."

---

[31] https://support.google.com/adwords/answer/6379332?hl=en
[32] https://blog.twitter.com/marketing/en_us/a/2015/introducing-partner-audiences.html
[33] https://blog.twitter.com/marketing/en_us/a/2015/introducing-partner-audiences.html
[34] https://blog.twitter.com/marketing/en_us/a/2015/introducing-partner-audiences.html
[35] https://www.criteo.com/news/press-releases/2017/10/criteo-empowers-retailers-and-brands-successful-future/

Criteo uses these three types of data to serve ads to individuals at the optimal place and time to elicit a purchase or desired action. Shopper Graph also powers Criteo's Customer Acquisition tool. Customer Acquisition uses data on "aggregated and anonymized historic shopping and browsing events" to target promising new customers with ads on the web.

AdRoll, in addition to retargeting, offers a similar customer acquisition tool that it calls "AdRoll Prospecting." According to AdRoll, Prospecting,

> "finds audiences using the IntentMap™, the largest proprietary data co-op that advertisers can access by contributing their site data. Nearly five thousand advertisers of all sizes have opted into the IntentMap™, pooling more than 1.2 billion digital profiles from across the web and mobile."[36]

Marketers who use Prospecting embed an AdRoll pixel into their site code. AdRoll analyzes visitor data from the marketer's website and then identifies internet users who act similarly to the marketer's existing customers. The marketer can then target these prospected future customers with ads across the web. AdRoll does not charge its clients a fee for the data it uses to power Prospecting. Instead, those who wish to have access to Prospecting must opt-in to AdRoll's data co-op and exchange their own site data with AdRoll in return for access to this service.

Both AdRoll and Criteo's full array of services are available in Canada and the USA.

Scoring & Risk Mitigation Products

Many data broker services score and categorize current and prospective customers. Scoring is used in both marketing and risk-mitigation products. In the marketing context, scores may be assigned to "high-value" prospective customers, or "under-performing" customers, prompting marketers to "activate" these individuals or segments to unlock their full profit potential. In the risk-mitigation realm, scoring is used to identify risky clients, such as those likely to default on loan payments or those more likely to commit fraud. The process for how one major player, eBureau, assigns scores to customers was described in the New York Times as the following: (1) the client submits a data set containing the names of sales leads it has already bought, along with the names of leads who were converted into customers; (2) eBureau appends "several thousand" details such as "age, income, occupation, property value, length of residence and retail history" from its database to the sales leads; (3) from this raw data, eBureau "extrapolates up to 50,000 additional variables per person;" (4) eBureau then processes the data for common factors among the existing customer base; (5) scores are then assigned to prospective customers "based on their resemblance to previous customers."[37] The same process may be used to assess the riskiness of potential customers or their propensity to commit fraud.

---

[36] https://www.adroll.com/learn-more/prospecting

[37] http://www.nytimes.com/2012/08/19/business/electronic-scores-rank-consumers-by-potential-value.html

In the US, scoring products have been criticized for operating just out of the reach of existing credit reporting laws, and for creating a two-tiered system that puts more vulnerable consumers at a disadvantage.[38] At first glance, it seems possible that Canadian privacy and credit reporting laws have kept at least some players in this industry from entering the Canadian market. The prevalence of scoring services appears to be less widespread in Canada than in the US, however, that may change. Large, multi-national corporations with Canadian operations such as TransUnion have begun to acquire scoring services, and may roll out these services in whole or in part within Canada.

Industry-Specific Scoring

Data brokers such as Corelogic have found profits in offering industry-specific scoring solutions. Corelogic, a large US-based broker, provides data and analytic services based primarily on property information as well as consumer and financial information. Some of Corelogic's services include Automotive Credit Reporting, Background Data, and Direct Marketing Solutions. Automotive Credit Reporting combines traditional credit reporting, ID verification and customer acquisition tools -- this product aims to help auto dealers reduce risk and improve business performance.

eBureau, which was acquired by TransUnion in October, 2017, provides predictive scoring and analytics services for marketers, the financial services industry, and online retailers, among others. eBureau's flagship product is its "eScore" system–mentioned above–which may be used for marketing, lead management, fraud prevention, credit assessment, and collections decisions. eScores "leverage eBureau's big data assets and automated technology to develop highly predictive statistical scores"[39] in each of these areas.

Scoring to Determine Level of Service Provided

At least one data broker studied offers a product that appears to use scoring to determine the level of service customers receive when making an inbound call to a company. eBureau's "Inbound Call Routing" solution, based on "eScores," "helps call center operations managers improve inbound call routing decisions and results with automated scoring and actionable data insights."[40] According to eBureau:

> "eScores identify the inbound callers most likely to convert, those with the highest lifetime value and those most receptive to cross sell-offers, improving your call routing decisions and your agents' ability to successfully present the best offer to the caller."[41]

The implication here seems to be that eBureau's technology allows companies to prioritize the inbound calls of high value customers over those less valuable to the company. As the New York Times notes, "[scores such as eBureau's] can determine whether a customer is

---

[38] http://www.nytimes.com/2012/08/19/business/electronic-scores-rank-consumers-by-potential-value.html

[39] http://www.ebureau.com/b2c/escore

[40] http://www.ebureau.com/b2c/marketing-lead-management

[41] http://www.ebureau.com/b2c/marketing-lead-management

routed promptly to an attentive service agent or relegated to an overflow call center."[42] eBureau also allows companies to append demographic and contact details to call records, with the aim of streamlining approval and call engagement processes.

People Search Products

People search products are exactly what their name would imply. They allow individuals to locate other individuals both physically and online and to discern between individuals with similar names. These products "offer information about consumers obtained from government and other publicly available sources."[43] Just as it has with other types of data broker products, social media has had a significant impact on this industry. People search services scour public social media profiles in order to add depth to their search results, linking information from other sources to an individual's social media profiles. Searches may be conducted using a person's name, phone number, username, or social security number (US). In the US, people search products provide a wide array of information on individuals such as aliases, birthdays, news stories, telephone numbers, gender, interests, address history, education information, death records, relatives, employment history, marriage and divorce records, email addresses, criminal records, civil records (bankruptcies, liens, judgements), property ownership and sales history.[44]

In Canada, people search products are more reserved in their offerings. While they search a wide array of public records in the US, in Canada, American people search services are limited to information available from social media and public websites – they do not provide Canadian property and arrest records. This is likely due the difficulty of accessing arrest records in most provinces, which can only be obtained through court documents.[45] That being said, companies such as Sterling Talent Solutions offer specialized background check services which it offers to employers, landlords and organizations, but these services are not as readily available to consumers as their US counterparts. Services such as Canada 411 offer telephone and address information, and invite users to log into their services using Facebook for more accurate results, but do not build an aggregate profile from public records.

---

[42] http://www.nytimes.com/2012/08/19/business/electronic-scores-rank-consumers-by-potential-value.html

[43] FTC 2014 Report at 34.

[44] FTC 2014 Report at 34.

[45] http://www.cbc.ca/news/canada/hard-to-check-criminal-records-of-others-1.1145038

Part IV Social Networks and Data Brokers

In this Part, we look at social media as a source of data; social media as a purveyor of data; social media data products; and social media as an acquirer of commercial data. Social media sites are "social graphs" – networks of social connections that can be mapped. Application Programming Interface (API) developers essentially tap into the social graphs in designing their apps, and social media sites obtain information in return based on users' use of the apps. Social networks' massive user base provides for a rich source of data.

Social media as a source of personal information

Social networking platforms like Facebook, YouTube, Instagram, and Twitter are among the richest and most accessible sources of personal information on consumers. Users disclose personal details about themselves as soon as they begin creating profiles. To create a Facebook profile, for example, users must provide their name, email address, date of birth, and gender, and must consent to the company's Terms of Use and Data Policy.

Facebook's Data Policy acknowledges that the company retains an array of personal information voluntarily submitted by its users, including their location, interests, preferences, social connections, commercial transactions, and the kinds of devices that they use to access Facebook services. The company's policy also acknowledges that it receives data from its users' web browsing activities on third-party websites and apps that have integrated Facebook's Social Plugins, such as the Like and Share buttons.[46] This information is used by the company to develop its own products and to provide marketers with the ability to directly target advertising at specific segments of the social network's nearly 2 billion users. Facebook's Terms of Use further reinforce the social network's ownership of its users' personal information. By consenting to these terms, users give the company permission to use their personal information in connection with "commercial, sponsored, or related content" provided or promoted by Facebook.[47]

Facebook is not an outlier when it comes to harvesting user information for commercial purposes. Google's Privacy Policy permits the company to collect similar user data, including profile attributes, interests, post and login locations, and the types of devices used to access its services.[48] Acquired by Google in 2006, YouTube also collects information from users' Google accounts and their interactions with websites that incorporate Google services and YouTube content. The company's Google+ social network is governed by the same privacy policy. Google, like Facebook, obtains consumer data from sites and pages that have integrated its Google+ and YouTube sharing features, as well as its advertising and analytics tools. Through these integrated features, Google obtains the web address of the site being visited and the IP address of the visitor.[49] These records can be added to the profiles of existing users.

---

[46] Facebook, "Data Policy" (29 September 2016), online: Facebook <h ps://www.facebook.com/about/privacy>.

[47] Facebook, "Terms of Service" (30 January 2015), online: Facebook <h ps://www.facebook.com/legal/terms>.

[48] Google, "Privacy Policy" (17 April 2017), online: Google <h ps://www.google.com/intl/en/policies/privacy/>.

[49] Google, "How Google uses data when you use our partners' sites or apps" (17 April 2017), online: Google <h ps://www.google.com/policies/privacy/partners/>.

Instagram, acquired by Facebook in 2012, similarly collects all data submitted by users, as well as device information, user content metadata, and tracking cookie data.[50] Twitter likewise collects data on every aspect of its users' use of its services, from profile information to ad clicks.[51] As with Google and Facebook, Twitter also obtains data on users when they visit third-party sites that include Twitter features, such as its Feed widget or Share button.[52]

Professional networking site LinkedIn, acquired by Microsoft in 2016, similarly acquires data from its 500 million users' direct and indirect interactions with the site, including voluntarily submitted profile information; users' synced contact and calendar information; the content and metadata of direct messages; and tracked data obtained through cookies, web beacons, and device identifiers.[53]

Social media upstart Snapchat billed itself early on as a privacy-friendly alternative to the established social media giants, but its current Privacy Policy is just as expansive as the other major platforms. The company lays claim to the same kinds of user information as its competitors, including log information comprised of usage details such as device and browser type, access times, page views, IP address, and pages visited prior to using the service.[54] Despite Snapchat's reputation for privacy, the company also collects the information its users send through its service, such as Snaps and Chats.[55]

In this context, users voluntarily serve as a direct source of personal information for social media companies, which themselves supply advertisers with aggregated user information for marketing purposes. Commercial data brokers also mine user activity on social networking sites for publicly available information that users post to social networks themselves – online profiles and posts that have knowingly or unknowingly been made public based on a user's privacy settings. As Twitter's Privacy Policy states, "What you share on Twitter may be viewed all around the world instantly. You are what you Tweet."[56]

Social media sites and the commercialization of data

Social media platforms not only collect user data, they also profit from it. These sites use data to group their users into lists based on characteristics such as age, gender, lifestyle, and education. Advertisers pay social networks to target their advertising at those groups of users who are most likely to have an interest in their product. Advertising is by far the largest source of revenue for sites like Facebook and YouTube. In 2016, Facebook reported $26.9 billion in advertising revenue, while Alphabet—YouTube and Google's parent company—reported $63.8 billion in ad revenue.[57]

---

[50] Instagram, "Privacy Policy" (19 January 2013), online: Instagram <h ps://help.instagram.com/155833707900388/?helpref=hc_fnav>.

[51] Twi er, "Privacy Policy" (18 June 2017), online: Twi er <h ps://twi er.com/privacy?lang=en>.

[52] *Ibid* at "Using Our Services".

[53] LinkedIn, "Privacy Policy: Informa on We Collect" (7 June 2017), online: LinkedIn <h ps://www.linkedin.com/legal/privacy-policy#collect>.

[54] Snapchat, "Privacy Policy: Informa on We Collect" (5 June 2017), online: Snap <h ps://www.snap.com/en-US/privacy/privacy-policy/>.

[55] *Ibid*.

[56] *Supra* note 6.

[57] Facebook Investor Rela ons, "Facebook Reports Fourth Quarter and Full Year 2016 Results" (1 February 2017), online: Facebook

Social networks auction advertising space on their pages and apps. Advertisers set the maximum price that they're willing to pay for their ad to appear on a particular page. Google's auctioning system ranks bids based on the highest bid combined with a 'Quality Score' that takes into account the ad's past performance and its relevance to the audience that will view it.[58] Facebook's ad auctions take into consideration the value of the advertiser's bid, the likelihood of users interacting with the ad, and the ad's relevance to the intended audience.[59] Twitter and other social networks similarly allocate advertising space based on the advertiser's bid, the relevance of the ad to the target audience, and the likelihood that users will engage with the ad.[60]

Facebook states that it does not sell its users' personal information, but the company's privacy policy states that it uses data to show users relevant ads and measure their effectiveness. Google also states that it does not sell personal information such as names, email addresses, or payment information, but the company's AdSense service delivers personalized advertising to users based on their interests and demographics.[61] Like Facebook, Google sorts its users into different advertising audience categories based on the websites that they visit, their tracked internet activity, past ad clicks, and their account activity on other sites and devices.[62]

Twitter appears to be the most open about how it commercializes its users' data, including users' biographical information, the metadata associated with their tweets, and the accounts that they follow. The social network's Privacy Policy states that it "broadly and instantly disseminates" users' published information to a variety of organizations, including search engines, software developers, and market research firms.[63] "When you use Twitter to follow, Tweet, search, view, or interact with Tweets or Twitter accounts, we may use these actions to customize Twitter Ads for you," the site's advertising FAQ states.[64]

The company also logs data on anyone who does not have a Twitter account but who visits the social network anonymously or visits websites that have integrated Twitter plugins like its Share button. This data includes IP address, browser type, referring web page, search

---

<h ps://investor.. .com/investor-news/press-release-details/2017/facebook-Reports-Fourth-Quarter-and-Full-Year-2016-Results/default.aspx>; Alphabet Investor Rela ons, "Alphabet Announces Fourth Quarter and Fiscal Year 2016 Results" (26 January 2017) online: Alphabet <h ps://abc.xyz/investor/news/earnings/2016/Q4_alphabet_earnings/>; "Adver sing revenue of Google websites from 2001 to 2016 (in billion U.S. dollars)", online: Sta sta <h ps://www.sta sta.com/sta s cs/266242/adver sing-revenue-of-google-sites/>.

[58] Google, "AdSense Help: How does the ad auc on work?", online: Google <h ps://support.google.com/adsense/answer/160525?hl=en#2>.

[59] Facebook Business, "About the delivery system: Ad auc ons", online: Facebook <h ps://www.facebook.com/business/help/430291176997542>.

[60] Twi er, "Bidding and auc ons FAQs: What is a quality adjusted bid?", online: Twi er <h ps://business.twi er.com/en/help/troubleshoo ng/bidding-and-auc ons-faqs.html>.

[61] Google, "How Ads Work: We do no sell your personal informa on to anyone", online: Google <h ps://privacy.google.com/how-ads-work.html>.

[62] Google, "AdSense Help: About Google Ads", online: Google <h ps://support.google.com/adsense/troubleshooter/1631343>.

[63] *Supra* note 6 at "Tweets, Following, Lists, Profile, and Other Public Informa on".

[64] Twi er Business, "How Twi er Ads Work", online: Twi er <h ps://business.twi er.com/en/help/troubleshoo ng/how-twi er-ads-work.html>.

terms, mobile carrier, and cookie information.[65] This information is used to make inferences about consumer preferences and deliver targeted advertising based on those inferred preferences. In 2012, the company began licensing archived tweets and user information for commercial data analysis.[66]

As with other social media sites, LinkedIn directs advertising to site members and visitors based on tracked data, user-submitted profile information, site usage, and internet browsing data.[67]

Social media data products

Social networks analyze user data to provide targeted advertising and measure its performance. Users are grouped into different categories based on information gleaned from their online profiles and their interactions with a given social media platform. Google's AdWords service, for example, allows advertisers to deliver targeted video advertising to YouTube users based on their age, gender, location, and interests. AdWords offers similar targeted advertising services for its search results and Gmail inbox. Advertisers are able to gauge the effectiveness of their ads through Google's analytics tools.[68]

Facebook's Adverts Manager service allows advertisers to target certain segments of Facebook and Instagram users based on demographics, such as education, relationships, and political affiliation; interests, such as fitness, hobbies, and shopping; and behaviours, such as purchase patterns and life events.[69] Each of these categories is made up of selectable sub-categories. For example, the 'Fitness and Wellness' interest category allows advertisers to target groups of users who share an interest in specific activities, such as bodybuilding, meditation, dieting, or Zumba.[70]

Robert Sherman, Facebook's manager of Privacy and Public Policy, explained the business model of the company—and the social media industry in general—when he appeared before the House of Commons Ethics Committee in 2012, stating that Facebook offers its users a free service paid for by ad revenue. "In general, when you post information on Facebook… that's information we might use to decide which ads to show you. Advertisers will come to us

---

[65] *Supra* note 6 at "Using Our Services".

[66] Mitch Lipka, "What you should know about Twi er's data sales" *Reuters* (1 March 2012), online: <h p://www.reuters.com/ar cle/us-twi er-data-idUSTRE8201IU20120301>; Julie e Garside, "Twi er puts trillions of tweets up for sale to data miners", *The Guardian* (18 March 2015), online: <h ps://www.theguardian.com/technology/2015/mar/18/twi er-puts-trillions-tweets-for-sale-data-miners>.

[67] *Supra* note 8 at "Adver sing".

[68] Google, "AdWords: A preview of video adver sing", online: Google <h ps://adwords.google.com/intl/en_ca/home/how-it-works/video-ads/?subid=ca-ww-et-awx_aw>.

[69] Facebook Help Centre, "Does Facebook sell my informa on?", online: Facebook <h ps://www.facebook.com/help/152637448140583?helpref=uf_permalink>; Facebook Adverts Manager, online: Facebook <h ps://www.facebook.com/ads/manager/crea on/crea on/?act=10153830167535392&pid=p1>.

[70] *Ibid* Adverts Manager.

and will say 'I'd like to show this ad to people who are interested in a particular topic.' We'll show the advertising to the users," he told the committee.[71]

Facebook's advertising services also allow companies to target specific customers who they already have an established relationship with offline, away from the social networking site—a process known as "onboarding". The site's Custom Audience Ads feature allows companies to identify customers based on a combination of email address, Facebook account, phone number, name, date of birth, gender, location, and device ID.[72] Advertisers can also target Facebook users who share similar characteristics with their existing client base.

Google offers a similar feature called Customer Match through its AdWords service. Advertisers upload their customers' email addresses and create an ad campaign to target those customers. When the target customers sign in to Google and access Google Search, YouTube, or Gmail, the ad is delivered.[73]

Twitter's Tailored Audiences feature similarly allows advertisers to upload lists of email addresses, phone numbers, or Twitter usernames, and deliver ads targeting those specific users.[74] Advertisers can also identify visitors to their websites and users of their apps and target them with Twitter advertising.[75] LinkedIn offers Contact Targeting, which allows advertisers to upload contacts and deliver customized advertising through the site.[76]

'Closing the loop'

Social networks are searching for new ways to commodify their users' data through ad services. Each site's ad service offers analytics tools that allow advertisers to better understand their customers and the people who are interested in their product. The platforms compete to provide advertisers with the most detailed analysis of the effectiveness of their ads – what type of people engage with the ad, for how long, and how often does the ad lead to an online purchase.

Facebook's Audience Insights offers aggregated data on an audience's demographics, interests, site usage, and purchase behaviour.[77] Google offers its own Analytics tool that provides advertisers with demographic analysis of their audiences based on age, gender,

---

[71] House of Commons, *Evidence of the Standing Committee on Access to Information, Privacy and Ethics*, 41st Parl, 1st Session, No 57 (27 November 2012) at 1550 (Robert Sherman), online: <h ps://www.ourcommons.ca/DocumentViewer/en/41-1/ETHI/mee ng-57/evidence>.

[72] Facebook for Developers, "Marke ng API: Custom Audience", online: Facebook <h ps://developers.facebook.com/docs/marke ng-api/reference/custom-audience>.

[73] Google AdWords "AdWords Help: About Customer Match", online: Google <h ps://support.google.com/adwords/answer/6379332?hl=en>.

[74] Twi er Business, "Tailored Audiences", online: Twi er <h ps://business.twi er.com/en/help/campaign-setup/campaign-targe ng/tailored-audiences.html >.

[75] *Ibid*.

[76] LinkedIn Marke ng Solu ons, "Ad Targe ng: Contact Targe ng" (18 April 2015), *LinkedIn Pulse* (blog), online: <h ps://business.linkedin.com/marke ng-solu ons/ad-targe ng/contact-targe ng>.

[77] Facebook Business, "Learn More About the People that Ma er to Your Business with Facebook Audience Insights" (8 May 2014), online: Facebook <h ps://www.facebook.com/business/news/audience-insights>.

and interests.[78] Efforts to track ad performance among target audiences has become increasingly sophisticated. In 2013, Facebook acquired Atlas Advertiser Suite, an online advertising network that allows marketers to measure the effectiveness of their ads across multiple online platforms by linking users' browsing data with their online and offline purchase history.[79] This essentially allows advertisers to determine the number of times a purchaser has viewed an ad and engaged with an ad before purchasing the product being advertised. When Facebook announced its acquisition of Atlas, the company stated that it would help its client advertisers "close the loop" by comparing the performance of its Facebook ads with ads appearing elsewhere online.[80]

Google's DoubleClick service similarly allows advertisers to deliver targeted online advertising and measure its effectiveness beyond Google's own websites and services through browser cookies.[81] Twitter's MoPub, which specializes in mobile app advertising, similarly allows advertisers to target advertising and measure its effectiveness beyond Twitter's social network.[82]

Social media data users

Social media platforms are increasingly relying on commercial data brokers to supplement their own user data with "offline data" about users' lives away from their sites. Facebook, for example, obtains data on its users' interests, relationships, personal attributes and location, as well as information gathered through websites and apps that incorporate Facebook services. The company is able to supplement this information with additional data that is collected, sorted and sold by "third-party partners" to create even more detailed profiles of its users.[83] Twitter and Snapchat's respective privacy policies acknowledge that the companies obtain data from third party providers, while Google's privacy policy is less clear about the company's relationship with data brokers.[84]

Given that social networking sites already have a wealth of information on their users, it may seem odd that the companies are acquiring additional information about their users' online

---

[78] Google, "Analy cs Help: About Adver sing Features", online: Google <h ps://support.google.com/analy cs/answer/3450482?hl=en>.

[79] Atlas Solu ons, "About Atlas", online: Atlas <h ps://atlassolu ons.com/about-atlas/>; Alex Kaveh Senemar, "Facebook Partners With Shadowy 'Data Brokers' to Farm Your Informa on" (18 April 2015) *LinkedIn Pulse*, online: <h ps://www.linkedin.com/pulse/facebook-partners-shadowy-data-brokers-farm-your-alexander-k-senemar>.

[80] Josh Cons ne, "Facebook Confirms It Will Acquire Atlas Adver sers Suite From Microso to Close The Ad Spend Loop" *TechCrunch* (28 February 2013), online: <h ps://techcrunch.com/2013/02/28/facebook-acquires-atlas/>.

[81] Joanna Geary, "DoubleClick (Google): What is it and what does it do?" *The Guardian* (23 April 2012), online: <h ps://www.theguardian.com/technology/2012/apr/23/doubleclick-tracking-trackers-cookies-web-monitoring>.

[82] MoPub, "Overview", online: MoPub <h ps://www.mopub.com/publishers/overview/>.

[83] Facebook, "What kinds of informa on do we collect?", online: Facebook Privacy Policy <h ps://www.facebook.com/policy.php>.

[84] Twi er, "Informa on Sharing and Disclosure", online: Twi er Privacy Policy <h ps://twi er.com/privacy?lang=en>; Snap Inc, "Informa on we collect from third par es", online: Snap Privacy Policy <h ps://www.snap.com/en-GB/privacy/privacy-policy/>; Google, "Informa on we collect", online: Google Privacy Policy <h ps://www.google.com/policies/privacy/#infocollect>.

activities. As online marketing consultancy Gartner explains, the business world is beginning to recognize the value of acquiring data from as many sources as possible. Profiting from data requires "having the chops to improve marketing and sales, improve business performance, reduce risks, develop new products and services, and/or license, barter or trade the data with others."[85]

Search engines and social media sites rely on advertising as their primary source of revenue. Facebook reported $27.6 billion in revenue in 2016, $26.9 billion (97 percent) of which was from advertising payments.[86] Alphabet, Google's parent company, reported $89.7 billion in revenues in 2016, including $63.8 billion (71 percent) in ad revenue.[87] Twitter's total advertising revenue in 2016 was $2.25 billion.[88]

By acquiring additional offline data, social media platforms are able to develop a more precise profile of their users and the ads that they are most responsive to. This allows advertisers on sites like Facebook to target their ads to specific audiences that are most likely to be interested in a certain product. The more that a company like Facebook knows about a particular user, the more choice it can offer advertisers to target their ads to specific audiences. As the company's privacy policy states, user information allows the company to improve advertising and measure the effectiveness of ads that appear on its site.[89]

The data broker industry appears to be far more competitive than social media. There are many more firms, with larger companies obtain data from smaller firms specializing in niche forms of data. Oracle is one of the largest purveyors of personal consumer information, posting $37 billion in revenues in 2016, although it is unclear how much of its revenues were based on the sale and leasing of consumer information.[90] Epsilon parent company Alliance Data posted $7.1 billion in revenue in 2016, attributing 30 percent of that amount to its data brokering subsidiary.[91] Experian, which is entirely focused on data collection and brokering, posted $4.6 billion in revenue in 2016.[92]

[85] h p://blogs.gartner.com/doug-laney/microso -links-into-a-treasure-trove-of-informa on/
[86] Facebook Investor Rela ons, "Facebook Reports Fourth Quarter and Full Year 2016 Results" (1 February 2017), online: Facebook <h ps://investor. .com/investor-news/press-release-details/2017/facebook-Reports-Fourth-Quarter-and-Full-Year-2016-Results/default.aspx>.
[87] Alphabet Investor Rela ons, "Alphabet Announces Fourth Quarter and Fiscal Year 2016 Results", online: Alphabet <h ps://abc.xyz/investor/news/earnings/2016/Q4_alphabet_earnings/> ; "Adver sing revenue of Google websites from 2001 to 2016 (in billion U.S. dollars)", online: Sta sta <h ps://www.sta sta.com/sta s cs/266242/adver sing-revenue-of-google-sites/>.
[88] Twi er Inc., "Annual Report 2017", online: Twi er < h p://files.shareholder.com/downloads/AMDA-2F526X/4458419093x0x935049/05E6E71E-D609-4A17-A8BD-B621324A950D/TWTR_2016_Annual_Report.pdf>.
[89] *Supra* note 1.
[90] Oracle Corpora on, "Fiscal 2016 Year to Date Financial Results", online: Oracle <h p://s1.q4cdn.com/289076952/files/doc_financials/4Q16/Q416_Form8K_Exhibit99-1_Earnings_Release_Tables.pdf>.
[91] Alliance Data, "2016 Annual Report", online: Alliance <h p://phx.corporate-ir.net/External.File?item=UGFyZW50SUQ9Mzc0ODc5fENoaWxkSUQ9LTF8VHlwZT0z&t=1&cb=636284031380795829>.
[92] Experian, "Annual Report 2016", online: Experian <h ps://www.experianplc.com/media/2733/experian-ar2016.pdf>.

Issues

- What's the commercial relationship between social media and search engine sites and commercial data brokers?
- Why are social media sites acquiring additional data beyond what they obtain directly from users?
- What are the privacy and consumer protection implications of such practices?
- What is the current legal framework regulating the collection and sale of data by data brokers?

Who

- Social media (Facebook, LinkedIn, Twitter, Instagram, YouTube) and search engines (Google) acquire data from brokers that specialize in collecting data and assigning that data to individual consumers. Data brokers also sort consumers in to particular lists based on specific attributes, which can be bought, leased, or traded.
- Data brokers also rely on sites like Facebook and Google to acquire publicly available consumer data, and then attach "offline" information to specific users – credit card purchases, loyalty card transactions,
- A 2014 FTC Report examined the data practices of nine data brokers: Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf, Recorded Future.
- Facebook provides links to opt-out pages for its main data brokers (Acxiom, Datalogix (Oracle Data Cloud), Epsilon, Experian, Quantium) https://m.facebook.com/help/494750870625830?helpref=uf_permalink.
- ProPublica notes that opting out of Facebook's data providers' gathering is often complicated and the company occasionally changes the firms that it obtains data from. (https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them)

What

- Sites like Facebook and Google collect data on users' personal online habits – search terms; 'likes', 'favourites', and 'follows'; geographic indicators (GPS and profile information) – to match specific online advertisements with targeted audiences.
- Data brokers collect consumer data from numerous sources, largely without consumers knowledge: bankruptcy information, voting registration, consumer purchase data, web browsing activities, warranty registration, consumer purchase data, web browsing activity, warranty registrations.

When

- In 2012, Financial Times profiled the partnership between Datalogix and Facebook to measure the effectiveness on ads that appear on social network. (Facebook's data partner joins the dots but privacy questions linger, FT 24 Sept 2012)
- In 2013, AdvertisingAge reported that Facebook was testing targeted advertising based on consumer data provided by firms such as Epsilon, Acxiom, and Datalogix to match loyalty program data with individual Facebook profiles.

(http://adage.com/article/digital/facebook-partner-acxiom-epsilon-match-store-purchases-user-profiles/239967/)
- In 2016, ProPublica published a report describing “detailed dossiers” about users’ offline lives that Facebook obtains from commercial data brokers to supplement data already gathered from Facebook usage. Jeffrey Chester, executive director of the Center for Digital Democracy, is quoted: “Facebook is bundling a dozen different data companies to target an individual customer, and an individual should have access to that bundle as well.” Article states that Facebook began working with data brokers in 2012 starting with Datalogix deal. (https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them)

Where

Acxiom’s corporate HQ is Conway, Arkansas, with seven additional offices in US and twelve international offices, none of which are found in Canada (UK, Poland, China, Germany, Australia)
https://www.acxiom.com/about-us/locations/
Acxiom confirms that it collects “business and consumer telephone directory listing information, including name, address, telephone number and other information… of published Canadian telephone directories” and enhances this info with census data… “From this data, Acxiom creates business and consumer telephone directories that we license to companies and non-profit organizations for their international use as an automated and inexpensive form of directory assistance or for direct mail and telemarketing purposes.”
privacy@acxiom.com, call 1-877-774-2094
https://www.acxiom.com/about-us/privacy/canadian-data-products-privacy-policy/

Oracle Canada (which acquired Datalogix in 2014) has offices in Mississauga, Calgary, Dartmouth, Edmonton, Markham, Montreal, Ottawa, Toronto, Quebec City, Richmond BC, Vancouver, and Waterloo. According to the company’s privacy policy, the company collects information including, but not limited to: name and physical address, email address, phone numbers, demographic attributes, past transactional behavior, corporate and employment information, cookie data, IP address, behavioural data and web search info related to a consumer’s use of the company’s websites and apps. The company states that it uses personal information for a range of purposes, including tailoring marketing to consumer interests.
https://www.oracle.com/legal/privacy/privacy-policy.html
Oracle Canada
www.oracle.com
(800) 363-3059
100 Milverton Dr.
Mississauga, ON
L5R 4H1

Epsilon has offices throughout the US and one office in Toronto
The company collects information voluntarily provided through registration with its websites, publicly available information, and tracking cookies.
https://www.epsilon.com/en_US/privacy-policy.html

Toronto
111 Gordon Baker Rd
3rd floor
Toronto, ON M2H 3R1

North America
privacy@epsilon.com
Epsilon
2550 Crescent Drive
Lafayette, CO 80026
Attn: Privacy Manager

Experian
- Owns and manages extensive databases that provide geographic, demographic, financial and lifestyle information on millions of consumers around the world… assists marketing campaigns by using advanced analytic tools to predict who is most likely to respond to ads.
http://www.experian.ca/corporate/experian-marketing-services.html
- North American HQ in Costa Mesa, CA.
- Experian Canada Offices – 2 Bloor St. East, Suite 3501, Toronto, ON Canada M4W 1A8
https://www.experian.com/privacy/

Why

Types of Products (FTC Report, pgs 23-29)
- Marketing – direct marketing (data append, marketing lists)
- Risk Mitigation – registration targeting, collaborative targeting, onboarding (combining online and offline data for segmentation, matching, targeting consumers online), marketing analytics, identity verification, fraud detection
- People Search
Pg 39 – Data brokers have a wide range of clients (insurance companies, banks, entertainment companies, governments). Data brokers sell to other data brokers for the purposes of: direct marketing, online marketing, marketing analytics, identity verification, fraud detection. Data brokers supply data to advertising firms for the purposes of direct marketing, online marketing, marketing analytics, identity verification, fraud detection, people search. Data brokers supply data to tech companies for the purposes of: direct marketing, online marketing, marketing analytics, people search.

How

Sources (FTC Report, pgs 11-16)
- Government Sources: publicly accessible licensing databases, property records, voter registration information, legal records.
- Publicly Available Information: phone listings, press releases and news reports, publicly available online activity (comments, public social media activity, blogs).
- Commercial Data Sources: purchasing data (item, price, brand, date of purchase, payment type) collected by retailers and credit card companies; subscription data collected by magazine publishers; voluntarily provided consumer survey information

Sources (G&M, How Big Data profits from personal information, 24 Feb 2014 http://www.theglobeandmail.com/opinion/follow-your-data-from-your-phone-to-the-marketplace/article17056305/)
- Internet (tracking cookies), Social Networks (Facebook, Google, Twitter), Mobile Phones (GPS & apps), Loyalty Cards (AirMiles)

Conclusion: Data Brokers and the Law

Canada's private sector privacy protection legislation, the federal Personal Information Protection and Electronic Documents Act ("PIPEDA"), regulates data brokers' collection, use, and disclosure of personal information.[93] PIPEDA defines "personal information" broadly as any "information about an identifiable individual", subject to certain exceptions discussed below. PIPEDA requires organizations to obtain the informed consent of individuals for the collection, use and disclosure of their personal information. Those purposes must be ones that a reasonable person would consider appropriate in the circumstances. And organizations must adopt safeguards appropriate to the sensitivity of the information held.

Compliance

As we have seen in undertaking our studies for this report, data broker practices pose compliance challenges to each of these elements of PIPEDA.

Consent

Organisations must obtain the consent of and individual for the collection, use and disclosure of personal information. Data brokers often obtain such information through retail partners or other third parties. It is difficult to obtain proper consent through third parties, particularly where the intended use is a secondary purpose. CIPPIC has long advocated for organizations to adopt the best practice of providing notice of the use or disclosure to consumers at the time of data collection along with an explicit, opt-in consent (or as a second best, a clear and simple opt-out).

Security

Personal data has become the object of data thieves' attention. However, we are seeing that data security can be difficult. While PIPEDA obliges organisations to implement security measures appropriate to the sensitivity of the information, it is an open question as to whether this obligation is sufficient to incentivize businesses to take consumer data security seriously. Imposing legal liability for inadequate security practices is worthy of consideration.

Below, we take a detailed look at the experience of Equifax following its historic 2017 data breach.

The Equifax Data Br each

Equifax, one of only two credit bureaus operating in Canada (and one of three in the United States), has been in the business of collecting, analysing, and selling consumer information for nearly 120 years. In its infancy, the Atlanta-based company helped lenders gauge the

[93] Some Provinces enjoy similar privacy legislation.

trustworthiness of borrowers by conducting overt, targeted consumer surveillance. Correspondents tracked individuals and recorded their indiscretions and liabilities — financial or otherwise — in reports later sold to businesses and banks. These reports, which could not be consulted or revised by the public until the early 1970s, combined financial information and moral assessments of individuals. Plans to digitize these these records in the late 1960s were met with significant public outcry, with Professor Alan Westin warning that easy access to computerized credit records posed a risk to Americans' privacy, civil liberties, and basic humanity.

Amidst advances in computer technology and a shift from qualitative to quantitative assessments of creditworthiness, credit bureaus have dramatically increased, digitized, and diversified their data holdings since the mid-1980s. To date, Equifax alone now manages 1200 times more data than the United States Library of Congress, and stores information about millions of consumers worldwide. With that in mind, its hard to overstate the impact of the massive data breach Equifax suffered in 2017, which impacted over 143 million adults in the United States and at least 19,000 individuals in Canada.

Dubbed the "largest leak of personal information in history," the Equifax breach has attracted significant regulatory, political, and legal attention since it was first announced on September 7, 2017. Attackers breached Equifax's servers through a vulnerability in the company's online disputes portal in May 2017. After reportedly discovering the breach on July 29, 2017, the company waited an additional two months before disclosing the incident to the public. In written testimony before the U.S. House Committee on Energy and Commerce Subcommittee on Digital Commerce and Consumer Protection, former Equifax CEO Richard F. Smith explained that the breach was caused by "both human error and technology failures." Equifax IT staff reportedly failed to patch the affected Apache Struts software after a vulnerability was publicly announced in March 2017, and network scans failed to detect the mass exfiltration of consumer data during the attack.

After infiltrating Equifax's systems, attackers stole a wide range of personal information including social security numbers, birthdates, addresses, and driver's license numbers. The attackers also stole 200,000 individuals' credit card numbers. Given the volume of data stolen, and the fact that hacked data has yet to surface on online black markets, some have speculated that the breach may have been part of a nation-state level attack.

Equifax was widely criticized for responding clumsily in the days following the breach. Consumers struggled to get information about whether or not they were affected, and many waited for hours to speak with Equifax representatives on understaffed phone lines. The website that Equifax set up to provide information about the breach, www.equifaxsecurity2017.com, required users to provide detailed personal information in online forms, and attracted the suspicion of some users. The site was soon spoofed by a web developer unaffiliated with Equifax, and the company's Twitter account repeatedly directed people to the fake website. New York state attorney general Eric Schneiderman also slammed Equifax for including an "unacceptable and unenforceable" arbitration agreement in the terms and conditions of credit protection tools the company released after the breach. Equifax has since made its credit lock tools free, waived credit protection fees, and removed the offending arbitration agreement and class action waiver.

The Equifax breach has since prompted over a hundred class action lawsuits in the United States and Canada. In Ontario, a class action for breach of privacy is ongoing. Affected individuals have also sought recourse through lawsuits in small claims court, with some plaintiffs in the United States opting to use an online chatbot to help draft their legal documents.

Since the breach, international regulators have responded by launching investigations and issuing notices for affected individuals. The U.S. Federal Trade Commission has issued general guidance, providing next steps for affected individuals, warning against Equifax-related phone scams, and explaining credit freezes, alerts, and locks to American consumers. The Canadian Office of the Privacy Commissioner similarly opened an investigation in September 2017, and is working with Equifax to issue notices and provide information to affected Canadian consumers. Public discussion about cybersecurity best practices and potential legislative and policy reforms in response to the Equifax breach remains ongoing in the United States, Canada, and abroad.

Exclusions from Legal Protection

"Publicly available" - PIPEDA effectively excludes from certain legal protections information deemed "publicly available" by regulation.[94] In the case of personal information published in a telephone directory, organizations may collect, use and share such information without consent.[95] The collection, use and disclosure of this information can have consequences for us. For example, we may well choose to make our telephone and address available through a telephone directory to members of the public so that friends and colleagues are able to contact us; it does not follow that we are fine with that same data being used by a data broker to profile us for marketing purposes. The manner in which the law treats personal information in this particular context violates our usual view of privacy as a normative and contextual right. Protection of "personal information" under PIPEDA ordinarily requires consent, not secrecy. It is past time to revisit this exclusion.

"Non-commercial actors" (e.g., political parties) - PIPEDA does not apply to non-commercial activity. For that reason, much of the voter data gathering and use practices engaged in by non profit organizations such as political parties, for example, lies outside of the jurisdiction of the federal Privacy Commissioner's oversight powers. As we have seen with the Cambridge Analytics/Facebook news story, political engagement with personal information through social networking businesses and private data analytic firms has proved deeply troubling. We look at this situation in some detail below.

[94] See Regulations Specifying Publicly Available Information, SOR/2001-7 (13 December, 2000), https://www.canlii.org/en/ca/laws/regu/sor-2001-7/latest/sor-2001-7.html.
[95] PIPEDA, paragraphs 7(1)(d) , 7(2)(c.1) and 7(3)(h.1). See, generally, Office of the Privacy Commissioner, "Publicly Available Information" (March 2014) https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda-compliance-help/pipeda-interpretation-bulletins/interpretations_06_pai/; see also Englander v. Telus Communications Inc., 2004 FCA 387, http://canlii.ca/t/1j6r7.

BC's Personal Information Protection Act applies to both commercial and non-commercial activities, including to non-profit organizations. The former BC Information and Privacy Commissioner took the position that she had jurisdiction over the privacy practices of political parties and investigated the BC NDP for its information collection practices undertaken in vetting potential candidates.[96]

Cambridge Analytica, Facebook, and Democracy

A Canadian by the name of Christopher Wylie has blown the lid off of an immense data collection operation by a company named Cambridge Analytica. The revelations, which indicate their involvement in supporting both the Brexit campaign and the Trump campaign, have shone new light on the prevalent role of data brokers and the use of big data in support of political aims.

Who is Cambridge Analytica?

Cambridge Analytica is a data analytics firm under the parent company Strategic Communications Laboratories (SCL). They specialize in "psychological operations", the art of changing people's minds through disinformation and propaganda.

Who are the major players?

Christopher Wylie was a data scientist who worked for and helped set up Cambridge Analytica and is the whistleblower who came forward to The Guardian about Cambridge Analytica's activities.

The CEO of Cambridge Analytica was Alexander Nix, who also was the CEO of SCL Elections. BBC's Channel 4 news revealed undercover footage of Nix outlining services that the company provides, including bribery and blackmail of political figures. Cambridge Analytica has since suspended Nix pending an internal investigation.

Cambridge Analytica obtained their data from Dr. Aleksandr Kogan, a Russian-American data scientist who lectures at Cambridge on psychology and social media psychometrics. Cambridge Analytica contracted Aleksandr Kogan's company, Global Science Research, to harvest Facebook user data. Kogan harvested the data of 50 million users and sold it to Cambridge Analytica for commercial profit, though this was in violation with his license with Facebook as the licence was to collect user data for academic purposes only. Kogan is also an associate professor at St. Petersburg State University and has received grants from the Russian government for research into Facebook users' emotional states.

The company has ties to the Republican party. Steve Bannon, the former executive chairman of Breitbart and Donald Trump's former chief strategist, met Nix and Wylie in 2013 and immediately saw the potential use of Cambridge Analytica. Bannon went on to introduce them to Robert and Rebekah Mercer. The father-daughter duo, famous for donating millions to right-wing causes and the Trump campaign, became the primary investors behind Cambridge Analytica.

What data was Cambridge Analytica collecting?

[96] BC Information and Privacy Commissioner, P11-01-MS, "Summary of the Office of the Information and Privacy Commissioner's Investigation of the BC NDP's use of social media and passwords to evaluate candidates", https://www.oipc.bc.ca/mediation-summaries/1399.

Cambridge Analytica contracted GSR to harvest facebook user data. To do this, Kogan set up an app named thisismydigitallife, which harvested data from its users, and most importantly, also harvested the data of their friends without their knowledge or consent. Though only around 320,000 people voluntarily took the test, GSR was able to harvest the data of over 50 million people by expanding data collection to the friends of the test-takers.

What were they using it for?

Using this data, Cambridge Analytica created sophisticated psychological profiles on each user so they could selectively target users with tailored messages and political advertisements. While it remains unclear what impact their operations have had events like the 2016 US Presidential election and the Brexit vote, they were certainly embroiled in them and both British and American authorities have launched investigations following these revelations.

What has Facebook done about this?

The Guardian published a report on Cambridge Analytica's use of Facebook data to support the campaign of Ted Cruz in December of 2015. Facebook did not take action until August of 2016, at which point they asked Cambridge Analytica and Wylie (who had left Cambridge Analytica by this time) to delete the data obtained by GSR. Wylie describes the process of certifying that this had been done as simply "tick a box and sign it and send it back, and that was it."

Facebook CEO Mark Zuckerberg has recently acknowledged that Facebook made mistakes with the handling of Cambridge Analytica. He announced a three-step plan: (1) to investigate all apps that have had access to "large amounts of data" prior to 2014, when they limited app access to data; (2) to further restrict developer access to data; and (3) release a tool which will allow Facebook users to easily see what apps have their data.

Are Canadians involved?

It is unclear whether the private information of Canadians has been affected by Cambridge Analytica's operations, and the Privacy Commissioner has launched an investigation to determine the extent to which Canadians have been affected. However, a shocking number of Canadian actors are already wrapped up in this story.

The Canadian-born whistleblower Christopher Wylie got started in politics in the Liberal Party at the age of 17. The next year, he went to work for Obama's campaign, where he honed his skills as a data scientist. Liberal sources have been quoted by the CBC as saying that Wylie pitched a similar version of what he did at Cambridge Analytica in 2009, and that they rejected him. The Liberal Party has since issued a statement stating that they contracted Wylie to work for them as late as 2016, for a $100,000 pilot project for their caucus research bureau. Melissa Cotton, managing director for the bureau, said that only preliminary work was done, and that, "after seeing what was offered, Liberal Caucus Research Bureau decided not to move forward."

A Canadian advertising firm, AggregateIQ, has been linked by the Guardian to the Brexit Leave campaign, Wylie, the Mercers, Steve Bannon, and Cambridge Analytica. It has recently come to light through another whistleblower that AggregateIQ may have knowingly helped the Leave campaign hide spending that would have pushed them over their campaign spending cap.

Key takeaways

There is an immense amount of data on each of us available through social media to organizations seeking to collect and use such data for their own purposes. Despite Mark Zuckerberg's statement, this is unlikely to change. Most of us know Facebook and its advertisers primarily makes its profits by analyzing the data we generate when we use its services to provide us with targeted advertising, but are content to hand over data to seemingly innocuous applications and to Facebook itself on a regular basis nonetheless. So long as it's to provide us with services and marketable goods, it's entirely with Facebook's mandate as a private corporation.

The changes proposed by Zuckerberg aren't sufficient to protect Facebook users from actors who want to weaponize their data and turn it against them for political means, and it certainly won't stop advertisers from knowing a tremendous amount about us. Facebook's attempts to rectify the situation indicate something deeper, that this industry is still mostly self-regulating and it is time for that to change.

"Anonymous Data" - Anonymous data does not constitute "personal information" under PIPEDA. However, when combined with geographically-based information such as postal codes,it can act as a proxy for personal information, with the problematic added risk of inaccuracy. Use of such data to make decisions about us may affect us. It may be time to look at the privacy implications of such use of "anonymized data" as a proxy for personal information.

Inappropriate Purposes

PIPEDA requires that "An organization may collect, use or disclose personal information only for purposes that a reasonable person would consider are appropriate in the circumstances."[97] With the emergence of big data analytics, the risk of the use of personal information for profiling and discriminating among consumers. Where organizations engage in such practices to profile or draw inferences about consumers on grounds that violate human rights law, they will violate the prohibition against inappropropriate purposes. Similarly, unfair or unethical dealings will likely violate this prohibition.[98]

Concluding Thoughts

The data brokerage industry occupies in a region of the economy that is opaque to consumers, its objects of commerce. It is difficult for consumers to appreciate the mechanisms by which data brokers collect, use and trade in consumers' personal information, and so the usual mechanisms by which markets discipline businesses are not in place. The industry is complex, with multiple kinds of actors collecting, processing, and aggregating data to create and use consumer profiles. Reporting by CIPPIC and others on the activities of the industry are insufficient to overcome this difficulty.

[97] PIPEDA, s. 5(3).
[98] See Office fo the Privacy Commissioner, "Draft guidance: Inappropriate data practices – interpretation and application of subsection 5(3)", (28 September, 2017), https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/consultation-on-consent-under-pipeda/gd_53_201709/.

Is concern over the practices of the data brokerage industry justified? We would argue that where industry practices exceed what a reasonable consumer would be comfortable with, they are. The ever-growing volume of data - and its increasingly personal and biographic nature - amplifies these concerns. Consumer data has value, and the recent front-page news about the targeting of large brokers demonstrates, this value has become the object of desire of data thieves. For these reasons, we are of the view that is is time for privacy regulators to look on this industry with a sharper eye. Greater transparency and accountability is required, and regulators must wield meaningful tools to ensure that organizations collect, use and share consumer data in accurate, fair, and appropriate manners.