

**They Only Want You For Your Data: Data brokers are turning everything about us into a  
commodity market**

Wesley Jones

Iowa State University

POLS 534: Legal and Ethical Issues in Cyber Security

Dr. Matt Behrens

11 May 2023

### **Abstract**

Consumers *feel* the omnipresence of data sharing — one Amazon search turns to a deluge of ads on Facebook. A baby registry spawns fliers for minivans. Moving to a new home garners alarming calls about the perils of not having proper mortgage insurance. Is my phone listening to me? During the process of navigating online you are alerted that sites are tracking you (“*Please accept these cookies to continue!*”). These are raw facts collected about you, but analyzed information can be so much more, it has become a traded good. It exists as a nearly completely unregulated commodity. Categories with names like “Group O: Singles and Starters” contain subsets such as “O55 - Family Troopers: Families and single parent households living near military bases” shroud the concept that entire swaths of potential customers have been distilled to a homogeneous set of likely similar (and accurate) data points. The commodity marketplace already exists and it has an entire framework of regulations to draw on. Rather than inventing a complicated set of policies – one which would allow for exploitable loopholes! – the focus can be redirected at the process of exchanging data for money (or other data).

*Keywords:* data broker, big data, information, ethics, regulation, transparency

**They Only Want You For Your Data: Data brokers are turning everything about us into a commodity market**

**Contents**

<b>Introduction</b>	<b>5</b>
<b>Literature Review</b>	<b>6</b>
Terminology . . . . .	6
History and Function . . . . .	7
An early industry . . . . .	7
How does this work? . . . . .	9
Using Data Brokers . . . . .	10
Pre-internet and economy development . . . . .	10
Intersection of data brokering and “Big Data” . . . . .	11
Consumer advantages . . . . .	11
Ethical Considerations . . . . .	12
Transparency and avoiding inequities . . . . .	12
Modern regulation (and lack thereof) . . . . .	13
The 2014 Federal Trade Commission Report . . . . .	14
<b>Procedures of Data Exchanging</b>	<b>16</b>
Basic background on variance issues . . . . .	17
Exploring the consumer profiles . . . . .	18
Experian Mosaic Consumer Profiles . . . . .	18
A McKinsey report . . . . .	21
<b>Advantages of Treating Data Exchanges with Commodity Regulations</b>	<b>22</b>
What does commodity policy look like in the real world? . . . . .	23

DATA BROKERS AND INFORMATION AS A COMMODITY 4

**Issues When Studying Data Brokers and Exchanges 24**

**Conclusion 24**

Continuing the conversation . . . . . 25

**References 27**

## Introduction

In Chicago, grain traders yell out, bidding on futures and wrestling to participate in a marketplace of the United State's abundant agricultural products. The trades are tallied, evaluated, and published. Farmers sell goods to contracts owned by sprawling international traders and shippers. Processors work their margins to capitalize against the commodity market. Data brokers skim information from every aspect of our lives, processing these into trade-able goods — selling compiled information to a myriad of corporate buyers, who can interpret, analyze, and market whichever conclusions they determine. A common focus in this subject area is on the violated consumer — striped of autonomy by behemoth data collectors. This focus has merits. Consumers deserve transparency and the dignity of being able to browse online without becoming a target of a marketing campaign. Consumers *feel* the omnipresence of data sharing — one Amazon search turns to a deluge of ads on Facebook. A baby registry spawns fliers for minivans. Moving to a new home garners alarming calls about the perils of not having proper mortgage insurance. Is my phone listening to me? During the process of navigating online you are alerted that sites are tracking you (“*Please accept these cookies to continue!*”). These are raw facts collected about you, but analyzed information can be so much more, it has become a traded good. It exists as a nearly completely unregulated commodity.

All of this is cause for concern, but we are decades into data brokering. How does a well established industry change? I searched for literature about data brokers that helped to provide a context for the origins of the data selling industry. I provide some context for how data broker services are built for selling information and what the ethical complications of their techniques are. Finally, I focus on the economy of this brokering industry. This involves highlighting some regulations that are in place, as well as areas that are relatively exposed. To establish this field is worth pursuing I will walk the reader through some “procedures”. This involves providing some examples of the consumer profiles offered by a data broker with an explanation of the complications this relies on. Also, I will present a case of consumer data extracted to an industry informational report. Again, this report draws on collected data and can provide great context to

areas where problems with data assumptions become exacerbated over time. With these procedures laid out I hope to show that regulation goes beyond ensuring user privacy rights, instead society relies on regulation as a means to check the economic *value* traded data intends to offer.

The naturally secretive business of data brokers prevents effective large statistical analysis. Instead conclusions must be extracted from the marketing material. It is not a good solution, admittedly, but consider this process a starting point.

### **Literature Review**

Reviewing recommendations for long-term responses to the ever-growing data brokers, I settled on industry-wide regulation, in a way. I could not find a reliable approach to handling the ethical concerns that the data brokering industry raises by relying on personal data governance. The simplest solution is one that might seem complicated, but when reviewing the structure of data brokering and coming to see the philosophy of information exchanges, I see that focusing on economic regulation policy is likely more viable than any transparency or user-based solution. Regulating the implementation of acquired data from brokers also offers a methodology to assess the social implications of compiled data sets. What future policies does a suggestion like this provide? The answer is unclear, but the facts are there, we're likely too far in to turn back the clock on data brokers.

### **Terminology**

These are common terms I will use throughout this discussion. There is not a large formality in this research area, but a key source from the Federal Trade Commission — so these definitions are extracted from their discussion on data brokers (Ramirez et al., 2014).

- **Data Broker:** While there's no single definition, there is a good consensus that it is a company that collects, aggregates, analyzes, and then sells data about consumers.
- **Big Data:** Big data is an information technology industry term that denotes the process of collecting, storing, and analyzing data. This can be a subset of a data broker business or a

generalized term to refer to the combination functions of collecting, storing, and analyzing data.

- **Data points, consumer profiles, and buckets:** These terms combine together to make a cohesive process. A data point or variable is a single piece of either direct data or an extrapolated single piece of data. These pieces are combined to form a consumer profile. That is either a specific individual who possesses all of the data points used in the profile or an extrapolated concept again. These consumer profiles are placed into buckets. Typically a business would buy a bucket or range of buckets to either sell directly to or use the metrics to create products and services to sell to lucrative buckets.<sup>1</sup>
- **Data mining:** The process of *collecting* source data for the brokers. Mined data is sourced literally from everything, everywhere, all the time. There is little point in being more specific about that. I typically refer to data as collected rather than mined — the focus of this paper is on the value of information after it has been aggregated.

## History and Function

### *An early industry*

Broughton, Blackburn, and Vickers focus on some of the historical associations of information brokering. (1991) They describe the onset of brokerage services from first part of the 20th century to the latter (this article is written in 1991). A key set of concepts emerges in their discussion — concepts that even early in the information era were recognizable.

- **Commodification of information.** They recognize that a commodity marketplace is developing surrounding information access. This marketplace is tuned to *organization* and *accuracy*.

---

<sup>1</sup> I discovered during the readings that the actual aspects of *how* the profiles and buckets are made seem to be the proprietary sauce of the brokers. One concept is key — they focus on anonymizing the data, something that while it might seem moral, I expect they are using this to shield the fact that they are extrapolating (read: making up) many data points from a smaller source set of data.

- **Business continuity relies on information.** They discuss revelations following curtailing of information departments during economic shifts as businesses sought to save their bottom line only to realize the impact that reliable information access provides. These changes ultimately lead to the centralization of information brokers. They also argue that the United States' role in public funding shifts contributed to the privatization focus of the information brokering. This component is paired with public library funding — in essence as public services were restricted businesses increasingly internalized or paid for brokerage services.
- **Ethical qualities provide business assurances.** This justification seems to provide a logical connection. Businesses trust information from brokers with high levels of efficacy and, in turn, ethics. The authors acknowledge lack of vested interest businesses have in maintaining clear sourcing for their data. They indicate that the value of information supersedes the trust in collection — essentially if it's correct, what business is going to turn it down?

The authors are not satisfied with the amount of data available, they, from our modern perspective, naively look forward to a growing marketplace of information (Broughton et al., 1991). A fallacy might exist here — one discussed throughout this paper. The assumption that a businesses ethics *persist* after a sale has been completed. The data brokers might have implemented a highly effective code of ethics in their data processing, but this assurance cannot be transferred with the data — it is up to the purchaser to ensure that their handling of the data (and resulting conclusions drawn from the data) are ethical.

Acquiring information from a public resource is typically more involved than online. Any pedestrian experiences at your local government offices can help you understand the challenges. Government records sourced to data brokers will typically require knowing enough information to gather the record. For instance, the United States Postal Service sells address change lists to data brokers. Brokers will release this data to anyone who can providing the original address for a



postal “customer” (Crain, 2018; Tanner, 2013).<sup>2</sup>

The original player in the consumer profiling game was the credit scoring system (Reviglio, 2022). The United States economy increasingly relies on an implicitly trusted credit scoring, but data brokers who manage the scoring have aggregated these financial points with additional factors that denote social indicators. The evolution of data brokers and businesses together has resulted in implicit trust in the scores with no accounting for the accuracy of the pseudo-financial rating (Mishra, 2022).

### ***How does this work?***

Internet users are likely more aware of the existence of data harvesting than ever before. Smartphone users today benefit from the alerting of microphone, camera, and location usage. Websites require acceptance policies for device access and prompt for cookies. This illuminates the entire not so public aspect of data mining that the data brokers rely on (Reviglio, 2022).

A key component to what makes the data brokers tick is the combination of the data from disparate sources — today that might mean several online sites or even offline public records (Ramirez et al., 2014). This combination then must be analyzed and functional for businesses to report on. This process involves creating a profile of the consumer based on specific insights collected. This profile is measured against the “ideal profile” of the most desirable consumer of any good or service. Ideally the data is anonymized and provided in this profiled sense as a generic object. Profiling to create targeted market is an endgame of the trade (Yeh, 2018; Mishra, 2022; Kuempel, 2016). The variables used to measure the applicability of a consumer profile can be indirect. Since these brokers have such a vast array of data points they typically offer to extrapolate consumer profiles (Acxiom, 2018). A person looking at a data set about themselves

---

<sup>2</sup> The USPS has made a habit of referring to US mail users/receivers as “customers”, which isn’t always a clear term since you’re required to interact with the federal mail system. The USPS is effectively side-stepping its own duty — as required by federal code! — to the it’s customers’ data. Read more from James McCain’s, “Applying the Privacy Act of 1974 to Data Brokers Contracting with the Government” (McCain, 2009). This subject area was compelling, but ultimately out of scope of this paper.

would not necessarily know if the data was directly collected from online interactions or if the content is inferred from other data points (Reviglio, 2022; Ramirez et al., 2014).

Because of the nature of how these brokers collect their data, consumers are oblivious to the process (Kuempel, 2016). Internet users would likely agree with a sentiment that “*someone* is tracking me” and “*someone* is collecting my data.” Aside from the credit reporting agencies who have been in the news over the years for mishandling collected data the nine companies in the Federal Trade Commission report are opaque and unknown (Ramirez et al., 2014). The commodity market for data is not determined by the individual user — a bonus for the collectors, since users are relatively disconnected from what information about them ultimately ends up compiling a set of factors that will then be used to market to them in the future, the people who the data is collected from are disengaged from the process (Crain, 2018). As Reviglio (2022) puts it; “Users are considered commodities. Their activity on the web is the unconscious work for the benefit of internet companies” (p. 10)

## **Using Data Brokers**

### ***Pre-internet and economy development***

Broughton, Blackburn, and Vickers (1991) provided an analysis of the state of information brokering in 1991, offering much historical context to the burgeoning industry. Since their discussion predates the internet, they provide a “pure” list of usages for data services (— this usage is probably only innocent in the eyes of a jaded consumer from today looking back in time). These services are nearly entirely focused on internal data management (existing company records surrounding clients and customers). It seems that the “data harvesting” methodology that users are well adjusted to today have not yet been employed.

Models for the development of an information commodity market have existed for decades prior to the internet. Advertiser funded industries developed naturally around newspapers, magazines, and television. These provided insight into the consumer as the commodity (Crain, 2018). A distinction should be made — it is not *us* as the consumer, being

traded on the commodity exchanges, but instead our *thoughts*.<sup>3</sup> Optimization in capitalism has trended these brokers to be come ever larger (Reviglio, 2022). By embracing commodity-style economy the data broker industry has provided our data directly to nearly every company in the world. (Crain, 2018)

### ***Intersection of data brokering and “Big Data”***

“Big Data” does not have a technical definition, but it typically involves a combination of collecting, storing, analyzing, and distributing (selling) consumer data (Ramirez et al., 2014). The term has been common in the information technology world as both a viable descriptor of technology and a buzzword. Due to the nature of data brokers they have become participants (or even controllers) of the big data market space. Typically these companies are massive and sell to other large companies — thus, “big data” (Crain, 2018). Yeh asserts that a key component to the success of data brokers is their data management and analytic methods (2018). The innovation provided by big data to the data brokers was the significant increase in efficiencies of collection and storage techniques and shortly followed by improved data analytics technology. By improving this infrastructure these companies are able to build specialized data sets — often quite specific. Different brokers have engaged in areas of specialty (Yeh, 2018).

### ***Consumer advantages***

Personalized advertising might help both small businesses and consumers navigate an exceedingly crowded marketplace. The accessibility of consumer profiles and the subsequent accuracy of the contained data points can allow for a small budget to reach wider audiences (Mishra, 2022). The FTC (2014) notes this relationship, but as Kuempel (2016) notes their assessment agrees with the arguments provided by traditional data brokers. The general sentiment when making proponent arguments for big data aggregation focuses on consumer spending habits. The ability to target products to specific consumers for purchase seems to be the start and

---

<sup>3</sup> If you're not familiar with your attention as a platform available for “colonization”, I invite you to watch Bo Burnham speak about this concept from Douglas Rushkoff in a panel;  
<https://www.youtube.com/watch?v=UmUm7oBqCVw?t=2484>

end consumer advantages. At the time of this report the FTC is discussing a new phenomenon where *online* behaviors had begun to follow targeted consumers *offline* (Ramirez et al., 2014).

## **Ethical Considerations**

### ***Transparency and avoiding inequities***

A dire world coalesces when reading into the collections and analytics of the large data brokering services. Yeh combats this with two oft returned to concepts (2018).

1. Transparency is an easy way for consumers to weigh in on the “correctness” of the data that is being stored on them and their habits. These services can offer accounts and agreements that itemize the selling of data to third parties.
2. Utilize audits and regulations to ensure inequities and rights are not stripped during the processing of the data. In a system where certain biases persist into the data collection and analyzing process it is inevitable that an unequal picture of the consumer is constructed. When data brokers sell their biased findings the institutions of inequities become still more entrenched. Relying on government intervention to help account and audit the data would prevent this disastrous cycle.

The issue here is common in the information technology industry. The same people who build a consumer profile are typically not trained on the social understanding required to create said categories. Since the method of analyzing and bucketing consumer profiles is part of the technology that drives data brokers’ profits they are unlikely to expose any consequential details about how the data points are categorized. Before the FTC report came out, a large data broker provided an online portal that allowed users to review the data collected on them — the portal even allowed for editing and discretionary opt-outs. This portal service is no longer active <sup>4</sup> (Crain, 2018).

<sup>4</sup> The service was called “About the Data” (note the passive voice — *the*, not *yours* or *ours*). It was offered by Acxiom starting in 2013, but has since disappeared. In its place, Acxiom offers a (lame) explanation about changes in company structure. Read more here, <https://www.acxiom.com/about-us/privacy/atd/>

Mishra suggests that transparency will not go far enough. The sheer amount of data collected on individuals surpasses any reasonable level of personal management. This is compounded with the large number of brokers that are in the market today. Mishra calls out some of the data points as sources of crime as well, where profiles for easy targets are extracted and sold to perpetrators of fraud (2022). Crain describes this marketplace as incompatible with transparency (2018).

The methodology used to aggregate data points to a common profile, “bucketing”, implicitly involves discrimination. Introducing biases into this discrimination process can further inequities. The brokers do not attempt to mask any of this process, instead they embrace the societally imposed disadvantages of lower income urban minorities, for example, providing targeted bucketing for credit weak individuals (Kuempel, 2016). Drawing on zip codes, education level, political affiliation popular data brokers utilize “analyzing” to further flush out buckets of users (Crain, 2018). This is speculation and complicates not only the sourcing of the data, but its trustworthiness.

### ***Modern regulation (and lack thereof)***

Kuempel compiles literature surrounding legislation of data privacy specific to the United States (2016). Currently, regulation has been left to the economy and self-regulation. This method discounts individual user privacy protections in lieu of (potential) business advantages. The government is more dynamic though — there are pockets of accountability scattered around the federal bureaus. Various agencies have developed audit controls that account for more proactive approaches to protecting user data than the business-first approach of Congress. We see this in history with the protections the federal government affords to your data under *its own purview* (McCain, 2009). These protections focus on narrow aspects such as driver’s license records (specifically address disclosures) and in 1988 a relevant protection for video tape rental details (Kuempel, 2016).

There are minor improvements in recent legislative history, but they are still relegated to specific areas of coverage (albeit — less specific than one’s video rentals) (Crain, 2018). The

Consumer Financial Protection Bureau comes to mind. The Bureau's stated mission includes a provision for creating transparency. By instituting various protections on the credit system this government agency essentially offers slight data protections over a consumers credit information (or at least transparency into the system for the consumer)(CFPB, 2023). The problem with relying on these groups to provide any tooling for consumer protection is that they do not apply to the *actual* data broker, "...data brokers' marketing activities largely fall outside the scope of existing laws like the Fair Credit Reporting Act, which, among other provisions, requires consumer reporting agencies to provide individuals with access to credit reports" (Crain, 2018, p. 5).

Kuempel analyzes the current model of regulation within the data brokerage community: self-regulation (2016). Businesses in this space implement privacy policies which the FTC will enforce (to an extent), however, for companies *without* these policies, no enforcement is possible. This leaves self-regulation via any official channels unreliable at best — often it is demonstrably negligent (Kuempel, 2016; Yeh, 2018). The model of self-regulation has become standard across the internet at this point (Crain, 2018). This is complicated because data brokers can follow a well defined code of ethics — one even privacy advocates would approve of, but since they are ultimately selling crafted data, the interpreting party has little to do with the standards of the even the best brokerage. As Reviglio puts it; "...it is essential that individuals are also protected against outputs of data processing, namely the potential harms that could result based on inferences" (2022, p. 16).

### ***The 2014 Federal Trade Commission Report***

In 2014, the Federal Trade Commission released a report dissecting nine data brokers. This report is integral to the current understanding of this industry. Every article cited here utilized the report (as long as the article was written after the publication of the FTC report). This contribution is immense and invaluable. The commission built upon existing reporting efforts from the commission in previous years in an effort to lay out the industry (Ramirez et al., 2014). Some key points in a review of the report:

- Purchasing between brokers has made transparency more complicated. Data is not wholly owned by a single broker (Kuempel, 2016; Crain, 2018). Disturbingly, a service that has been hacked has also leaked data to the dark web. By the nature of the data hacked from a data broker the theft would likely link the data points to individual users (Reviglio, 2022).
- Commonly protected data (e.g. medical) is entangled with website registration data — presumably this occurs when users register for access to their medical data from a healthcare provider’s portal.(Kuempel, 2016)
- Consumer opt-out for data sharing is nullified by technical hurdles or ambiguous language. In some cases opting out does not guarantee the data will not continue to be shared and collected. (Kuempel, 2016)
- The report did not address the data miners like Google and Facebook. They can act as the initial sources of data, but are not (at least by the FTC) considered brokers. (Crain, 2018)

In a series of *Washington Post* editorials a member of the FTC laid the groundwork from some of the arguments made in the full report. Julie Brill sought to inform readers about the vast data collected by brokers and help to strike an understanding about where consumer data exchanges boundaries need to exist (Crain, 2018). The recommendations of the FTC are common thoughts — or at least have become such since it’s release, but the methodology of addressing concerns raised in the report is more haphazard than one might expect. A sample of their vague approach is here:

Some industry members have expressed concern that such a centralized portal would be unwieldy, given the sheer number of data brokers in the marketplace and the fact that consumers may be overwhelmed by the breadth of information. To address this concern, in creating such a portal, Congress could consider limiting the portal to a number of the largest data brokers (fifty, for example, or other number deemed appropriate). Ramirez et al., 2014, p. 54

The research into the brokers was groundbreaking, but nearly nine years later and these unfulfilled policy recommendations have likely become increasingly infeasible.

The FTC does recognize there is an onus on purchasers to not perpetuate discrimination with the application of analyzed consumer profile buckets.

Finally, the Commission recommends that data brokers take reasonable precautions to ensure that downstream users of their data do not use it for eligibility determinations or for unlawful discriminatory purposes. For example, while the data segment of “Smoker in Household” could be used to market a new air filter, a downstream entity also could use the segment to suggest that a person is a poor credit or insurance risk, or an unsuitable candidate for employment or admission to a university. Ramirez et al., 2014, p.55–56

This last recommendation is their most timeless, but it is likely that the FTC should have been more cognizant of the predatory nature of the industry when this was written. They continue to point out that while the race, gender, color, religion, and etc. categories are technically prohibited by law – issues can arise as interpretations are made to apply these categories to the aforementioned groups of marginalized people (Ramirez et al., 2014). It is good for them to tie these potentials of perpetuating inequities in – however their recommendations simply do not address the root problems facing these groups.

### **Procedures of Data Exchanging**

Forging an understanding of what consumer profiles look like is an important first step to determining the viability for regulation of this marketplace. With this understanding we can construct an expected sequence of data and describe issues arising from convergence. People typically are apathetic towards the collection of their data, but when it comes to describing erroneous conclusions being drawn ideally the source subject would engage in a more empathetic response. Businesses and advertisers consistently build models that are informed by potentially affected data sets and as a well constructed replication algorithm can show – repeated variances



will result in widespread issues. Individuals should not feel compelled to be overtly concerned with *who* is stashing their information or *what* is being done with it. If the literature has indicated anything reliably – it is that the individual focus is wrought with issues. A better focus needs to be on the exchange of this data and what happens as it is replicated throughout society.

### **Basic background on variance issues**

A basic<sup>5</sup> statistical approach to the data sourcing problem can start with pointing out the issues with sample size. For every overtly factual data point assigned to an individual there is a chance that the data source has an inconsistency at that point. Imagine two options for an address:

- **Source A:** 1600 Pennsylvania Avenue
- **Source B:** 1600 Pennsylvania Avenue

Without relying on external validation tools (remember: often data points are highly individualistic in these sets!), there is no clear way to rely on this data. There are some conclusions to be drawn.

- The addressee moved between source collection time frames.
- Source A is wrong.
- Source B is wrong.
- Both sources are wrong.

A data set could be constructed to contain both facts, but that could ultimately inhibit analysis. Without a reliable address the geolocation of an individual cannot be relied on for correct bucketing. A model constructed to correct for this would be at the mercy of the

---

<sup>5</sup> The term "basic" is doing *a lot* of legwork here. This paper will **not** delve into the mathematics of variance within these types of data sets. This section aims to provide enough due diligence to assure the reader that indulging the following discussions on data are worthwhile. I relied on MIT's computer science course lecture, "Statistical Thinking" some years ago for enough background, it is still available and worthwhile today (MIT OpenCourseWare, 2012)

intelligence used to build it, artificial or human. In this case an address book containing all streets in the United States would be a worthwhile reference – unless that data set contains the incorrect fact already <sup>6</sup>.

Data analytics is a massive field of study. Dropping the allure of cynicism, assume that the majority of these issues are resolved. A data set with 100 points and a 99% reliability rate has **one** wrong fact. Every time this fact is exchanged to another broker with an overlapping inconsistent fact they have to decide: are we wrong or are they wrong? Every time a bucket is built the facts can get skewed. The severity of this minor, but can cause issues with perpetuating existing inequities – something covered later in this paper.

### **Exploring the consumer profiles**

#### *Experian Mosaic Consumer Profiles*

Experian offers a break down of their customer segmentation categories (Experian, 2019). We can walk through these categories, looking into what factors they might rely on and decipher how the complications discussed in the literature review – namely those of data reliability – during the devising of consumer buckets via exchanged data.

Categories with names like “Group O: Singles and Starters” contain subsets such as “O55 - Family Troopers: Families and single parent households living near military bases” shroud the concept that entire swaths of potential customers have been distilled to a homogeneous set of likely similar (and accurate) data points. Among this flowery language is a good indication for how Experian sees consumer correlation. Looking at the process to assign a segment to an individual can show the inaccuracies. Let us lay out an adult male (informed by yours truly):<sup>7</sup>

- 30-35 years old
- Rural city/town dweller

---

<sup>6</sup> *Pensylvania* was an acceptable spelling of the state during the naming of that street (1790), the spelling can be seen this way on the Liberty Bell (National Park Service, n.d.).

<sup>7</sup> It might feel like I am oversharing, but remember, this is just between you, me, and every data broker.

- House owner of >1 year
- Married
- No children

With just this information there is a number of categories to fit into like “F22 - Fast Track Couples: Active, young, upper established suburban couples and families living upwardly-mobile lifestyles” (p.87-8) or “N47 - Countrified Pragmatics: Countrified Pragmatics Modest income couples and singles living rural, casual lives” (p. 135-6). These are fitted well based on the prescribed values like what is listed for F22: "Credit-aware, Comfortable spender, Active lifestyles, Tech-savvy, Music lovers, Football fans" (Experian, 2019, p. 88). Within this bucket there is indication that the data and conclusions drawn do not align. A core tenant is "Tech-savvy", but the technology adoption category is *journeymen* – which is shy of the highest rating of *wizard*. Reviewing the in-depth marketing suggestions for F22 continues to display the problems this method can have in bucketing individuals (or households as is often referenced). I invite the reader to ponder groups of Millennials who share this makeup:

Fast Track Couples consist of young couples and families already on the road to upward mobility. Although ninety percent are under the age of 35, most are married, work at good jobs and own their homes. Two-thirds have kids. Most have at least some college education, which has helped them obtain decent-paying jobs in sales, white-collar professions and technology. Experian, 2019, p. 253

The prescription continues to describe the most common bucket member. Should this be the select cohort advertised to, some deviance might occur. If younger generations fail to buy homes at steady rates, this consumer bucket will become outdated as advertisers see less returns in home-focused efforts to push products. Similar issues can arise for every characteristic as they are subject to change. Ideally this is an expected result of this data broker service, after all it encourages purchasing new models all the time.

What if a consumer meets many, but not all of the qualities of a bucket, but current

consumer trends have a greater push to advertise that *missing* category? For example, in the F22 listing, two-thirds of consumers in that profile were homeowners. The remaining one-third are likely going to be receiving quite a bit of overlap for home-related purchases advertised to them. Consider the direct-to-consumer marketplace: items like camera-equipped doorbells and internet connected security systems and thermostats. These products are common, but not something a renter would typically engage with. A question arises then, is the advertisement going to influence a renter to purchase and mount a security camera? The technology would certainly work, after all it is wireless and non-intrusive. The bucket seems to fit, unless of course, now the consumer application has become overstated and advertisements are *informing* consumer behavior that is *informing* consumer profiling.

Under the model of regulation we have now for data brokers (self-regulation), data exchanges can be freely participated in to *improve* this categorization. In instances where an individual's facts have been determined unreliable, existing data and correlating data can be used to control for the predictive analysis required to fit the consumer into a bucket. Households in lower cost of living states can be *worth* as much as those in urban adjacent areas where cost of living decreases the value of a paycheck. Individuals who do not spend as much on technology are bucketed with groups they otherwise do not fit within – after all is there a community RadioShack to participate in the consumption of technology? Does a 80-inch flat screen fit within the individual's car? Barriers to access of resources when creating consumer profiles might be a bigger underrepresented problem.

Some of these natural variances are irrelevant. The systems data brokers construct to manage their data sets can be assumed to work – after all, consumers can decipher whether the advertisement is relevant to them quite easily. Others are more insidious. Something the FTC warns about is a dystopia where the presumed reality appears markedly different to differing social classes. Different brands, qualities of goods, and entirely different products are marketed to individuals across social classes (Ramirez et al., 2014). If this world of data perpetuates continued micro-data targeting, inequities become ingrained since exposure to external influences become

rarer.

### ***A McKinsey report***

Access to resources like grocery stores, something out of the control of individual residents, can enter a problematic cycle of institutionalized information. Take a McKinsey report detailing consumer expectations for grocery stores for example (Aull et al., 2022). While it is hard to categorize reports like these as “Big Data” that we are used to seeing from data brokers, they are certainly comprised of individual survey data compiling buckets of consumers to fit each statistical conclusion McKinsey is looking to exhibit. They are comparing consumer survey responses (remember in a corporate studies, data is not published, but they indicate they had 4000+ respondents) from 2021 vs. 2020, or other similar ranges. One specific statistic stuck out to me: “Buy groceries in large stores, where I can buy everything in one place” an figure that McKinsey lists as a +22 of “net intent”, or what surveyed participants indicated they felt strongly about (Aull et al., 2022, p. 10). Assuming this figure is not simply made up, it is the second strongest attribute in the exhibit. A grocery store under the advisement of McKinsey would certainly look to be a one-stop shop for consumers if they trusted this assessment. What is not accounted for in this data is the social situation that consumers find themselves in:

- Are they having to *commute* outside their normal working/home route to get to a grocery store. Reacting to the advice of McKinsey by creating large suburban presences might mean that the desire to have a one-stop shop goes from a nice-to-have to must-have for shoppers. In a scenario where small stores have closed, ones that were within reasonable distances of a residence, forcing all shoppers to travel to their grocery store, it makes sense that the consumer weight of having a single point of purchase would go up.
- There is no indication that the survey questions would successfully contextualize positive responses. A consumer that rates the ability to easily find all their groceries in one location might also indicate that that is not a feature they are willing to drive for. There is also no indication for the time frame of which the respondent is purchasing groceries for. A consumer could be expecting wholesale type shopping for a one-stop grocery store visit

every two weeks. As with so much of brokered data, the sourcing is not transparent. Should these survey results be combined, again there is no clarity provided to contextualize the responses provided by consumers.

There are other simple complications to data presented like this. This paper is not intending to be a “take-down” of McKinsey’s consulting efforts, but rather show the complications this data profiling can introduce. If this data presented was aggregated or distilled out of other data sets (exchanged), it could perpetuate a grocery store access issue for urban residents. Look to a fundamental sociology question: *how does inequality become institutionalized?* This is it. Without a way for to assess data from a data brokers’ compiled consumer profiles, there is no way to ensure that aggregated data is not poisoning corporate interest. In a philosophical sense: what came first, the data indicating the consumer interest (chicken) or the data informing the consumer interest (egg)?

### **Advantages of Treating Data Exchanges with Commodity Regulations**

When dealing with regulation for data brokers, there are two approaches to regulation. A user right to privacy focus is commonly suggested. This is shown frequently in popular media with exposés displaying the egregious data collection happening across the sprawling internet-connected lives lead today. The major drawback to this approach is to enforce proposed regulation requires individual action. The policy would be a boon to data brokers as requiring individual action against a world of unknowns is a good way to ensure that no action is taken. Further complicating this is the issue that if today *anonymized* data is being exchanged from broker to broker to buyer, then after user privacy rights are established *non-anonymized* data would have to exist or identity tracking would have to exist in order to retain user rights down the chain.

The realistic approach of data broker regulation is to address the system that these brokers rely on: raw data and a marketplace. Who, where, and what are all circumstantial to the already existing data. Crain provides an excellent summary of the individual versus commodity approach:

The consumer empowerment policy model sidesteps commodification and neglects

its basic insight: people are the products, not the consumers, of the data broker industry and commercial surveillance at large. Consumer empowerment looks to transparency to correct an imbalanced interaction between more or less equal parties. Commodification suggests that unfairness between parties is not a glitch in the system – it is the system. (Crain, 2018, p 13)

The commodity marketplace already exists and it has an entire framework of regulations to draw on. Rather than inventing a complicated set of policies – one which would allow for exploitable loopholes! – the focus can be redirected at the process of exchanging data for money (or other data).

### **What does commodity policy look like in the real world?**

Dare I suggest tangible policy changes? Here are some conclusions for actionable policies I have drawn while reviewing this subject.

- Establish a value system based on (a) data quality rating, (b) number of unique data points for profile (c) amount of individual data sources that were compiled to make a profile – this could be *individuals* or *sources* (eg. a website selling user data). A valued good can be taxed, either by tariffs at exchange time or by assessed currency value as a owned product.
- Establish a trading system to track data owners. This would require the first assessment, but the advantage of a trading system is that investment into trading would help to create a realistic evaluation of the data.
- Force data producers and sellers (eg. websites selling user data) to contract their sale *before* selling data. Essentially anyone warehousing un-marketed user data cannot simply choose to sell it one day. They must negotiate a contract to sell data in the future based on a market value assessment of the data and what the commodity is valued at in order to sell it. This is as much a barrier as it is a regulation system to ensure that frivolous market growth is checked.

These might seem overbearing to create, but remember our economy already has a

mercantile exchange for trading commodities and futures. Farmers, processors, and commodity traders already participate in this system, buying and selling trillions of U.S. dollars in everything from corn to water and electricity.

### **Issues When Studying Data Brokers and Exchanges**

Until a reliable process of data collection, combination, profiling, and exchange can be fully compiled, researchers are left to speculate what the process is. A cynical reading of the Experian and McKinsey marketing material is helpful, but the accuracy of conclusions drawn from it are likely only worth a grain of salt. Speculating in this area provides motivation to encourage pressure on developing marketplace regulations.

When it comes to proposing entering the data broker into a commodity market, there is a slew of unknowns. The argument is obviously tenuous. The only consolation I can offer to concerns like this is that the economics arguments can be laid out, but until a marketplace is established *no one* can definitely predict the potential unintended consequences. I am not naive enough to suggest there will not be a host of them, this paper does not seek to assert anything along those lines. I offer this instead: currently the process of buying and selling data is not only opaque, but unregulated. Transparency offers no regulation.

### **Conclusion**

Reading through current literature on this subject one cannot but help be struck with a sense of helplessness. The FTC report recommendations do not instill confidence in the advice provided to legislators. Have we come too far down this path of data accessibility to turn around? Proponents of the EU's General Data Protection Regulation (GDPR) policies will tell you no<sup>8</sup>. What I have seen in the selected readings does provide some hope that in the US our policy will at least have firm footing. There is general consensus on some important issues — issues that in the past would have been wholly ignored.

A common consensus exists that the consumer profiling targeting is more damaging to

---

<sup>8</sup> I don't delve into GDPR here. More here: <https://gdpr.eu>



consumers who are non-white and low income. They stand to lose more in the bucketing systems. A glossy view for those who could have foreseen the information commodity market might have been to assume that increased data would create more *accurate* profiling for the traditionally misrepresented among us. This is a fallacy in the marketplace though. The markets do not tune their modeling to accurately represent the *sources* of information — rather the data is compiled to reinforce existing business biases, thus leading to the sale of the data set. Establishing that these inequities can be unconsciously perpetuated because of inherent biases both in the organization and interpretations of profiles is vital to agreeing on the future regulations.

The original process of data collection is affronting to many. This is a rather sad side effect of our current information system, but we should not get distracted by it. Reform can be more effective when placed in a way to curtail the trading markets. The goal of regulation should be to reduce the commodity value of consumer information. Accurate data must be worth more than vast quantities of data. Treating the data brokering industry as a commodity market will ensure that regulation focuses on the manageable aspects of the process of collecting, analyzing, and selling user data.

Intrusions into consumer privacy are a compelling and fraught issue, but that should be dealt with because consumers deserve transparency and ownership of their information. The FTC recommendations make this need for transparency central to their report (Ramirez et al., 2014), but even had Congress followed through with every recommendation provided by the FTC consumer data could still be mishandled and abused by the *purchaser*. As long as the data is for sale the response should be economic. If the goal is to really curtail the data brokering (it should be!) — the economic approaches focus on the real issues.

### **Continuing the conversation**

I alluded to some concepts and ideas or gleaned them from reading. These are areas where I see the conversation on data brokering and its subsequent economy can be expounded on. Some of these topics are beyond the scope of my research, but others might be lacking studies.

- **Statistical analyzing of consumer profile buckets.** A breakdown of the metrics contained

within some of the common profiles. These profiles often consist of various juxtaposed facts (30 y/o male, owns a Toyota, lives in a town of less than 35,000 people in the Midwest, and streams TV shows five or more times per week (Acxiom, 2018).) A savvy effort could compare non-data brokered information to the consumer buckets to assess accuracy. This was started here, but with hard data the mathematical assessment could back the assumption that motives among data brokers lies in selling products more than having correct data points.

- **Assessment of data based studies in the social sciences that rely on data from large brokers.** Is the data set relied on by some academic research further “poisoned”? As Reviglio points out, often it is easier to purchase data from the brokers, especially considering the need for a population match, than it is to procure from government resources (Reviglio, 2022).
- **Delve into the motivations of data-based marketing online.** When data brokers sell their consumer profiles, they are assisting in a targeted sales tactic — what are the motivations of large companies engaging in such targeted methods? This can range from quarterly earnings to youth engagement. I mentioned in a footnote above (3) a concept that compared this level of engagement as “colonization of the mind”.
- **Improving data brokering by restricting the market.** Does the quality improve if the quantity decreases? What is the market for more accurate subsets of data? Can applying taxation and regulation to the commodity marketplace of data brokering affect the reliability of consumer profiles?

## References

- Acxiom. (2018, October). *Acxiom Digital Audience Variables (UK)*. United Kingdom.
- Aull, B., Coggins, B., Kohli, S., & Marohn, E. (2022, May). *The state of grocery in North America* | McKinsey (Tech. Rep.).
- Broughton, D., Blackburn, L., & Vickers, L. (1991, June). INFORMATION BROKERS AND INFORMATION CONSULTANTS. *Library Management*, 12(6), 4–16. doi: 10.1108/EUM00000000000838
- CFPB. (2023, February). *Consumer Financial Protection Bureau: About Us*. <https://www.consumerfinance.gov/about-us/>.
- Crain, M. (2018, January). The limits of transparency: Data brokers and commodification. *New Media & Society*, 20(1), 88–104. doi: 10.1177/1461444816657096
- Experian. (2019). *Mosaic USA E-Handbook*.
- Kuempel, A. (2016). The Invisible Middlemen: A Critique and Call for Reform of the Data Broker Industry. *International Law*.
- McCain, J. (2009). APPLYING THE PRIVACY ACT OF 1974 TO DATA BROKERS CONTRACTING WITH THE GOVERNMENT. *Public Contract Law Journal*, 38(4), 935–953.
- Mishra, S. (2022, March). The dark industry of data brokers: Need for regulation? *International Journal of Law and Information Technology*, 29(4), 395–410. doi: 10.1093/ijlit/eaab012
- MIT OpenCourseWare. (2012, March). *Lec 15 | MIT 6.00SC Introduction to Computer Science and Programming, Spring 2011*.
- National Park Service. (n.d.). *The Liberty Bell: From Obscurity to Icon* (Tech. Rep.).
- Ramirez, E., Brill, J., Ohlhausen, M. K., Wright, J. D., & McSweeney, T. (2014, May). *Data Brokers: A Call For Transparency and Accountability: A Report of the Federal Trade Commission (May 2014)* (Tech. Rep.). Federal Trade Commission.
- Reviglio, U. (2022, August). The untamed and discreet role of data brokers in surveillance capitalism: A transnational and interdisciplinary overview. *Internet Policy Review*, 11(3).

doi: 10.14763/2022.3.1670

Tanner, A. (2013, July). How The Post Office Sells Your Address Update To Anyone Who Pays (And The Little-Known Loophole To Opt Out). *Forbes*.

Yeh, C.-L. (2018, May). Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers. *Telecommunications Policy*, 42(4), 282–292. doi: 10.1016/j.telpol.2017.12.001