Breaking Down Brokers: Anonymizing Your Personal Data

Authors: Elise Callzidilla, Jared Ponticelli, Connor West

Table of Contents

About Us

Name: Jared Ponticelli
Hometown: Buffalo Grove, IL
Major: Business
Interests: Chess, Management

"Privacy is of great importance to me. I gave my email to be notified when my suit arrived, and now I see upwards of 30 spam emails a day, and most are revolting in nature"

Name: Connor West
Hometown: Denver, CO
Major: Politics
Interests: Aviation, Finance

"This Public Policy challenge is a perfect blend of my interests in practical public policy and academic research, allowing me to grow as a student and team member."

Hometown: Arlington Heights, IL
Major: Neuroscience and Business
Occupation: Assistant Trainer and Barn Manager at OBF

"The Public Policy Analysis Challenge has allowed me to broaden my interests and deviate from my typical science-filled schedule. I have also found some of the research regarding Privacy and the Internet extremely useful to implement practices within the workplace and protect client information when marketing" - Elyse

Executive Summary:

Overview - The Quick Pitch

We need to prevent consumer data from being used in ways for which the consumer did not consent.

The Problem

· Once data is collected, collectors can sell or share it with frustrating, inappropriate, and dangerous places.
· Data is collected on all internet users, and everyone's data is at risk of being used by undesired 3rd parties.
· There is little regulation and oversight for the sale of data and 3rd parties.

Root Causes

· Businesses sell or share data on their consumers with data brokers to squeeze value out of each transaction.
· Data brokers segment and target consumers, using their personal information, by further sharing their data.

Future implications

· Users of technology will continue to get bombarded with unwanted attention from 3rd party efforts.
· Distrust of data collectors and class action lawsuits because of misuse will persist.

The Solution

• An encryption method used in the Census can be applied to online data to protect consumer's privacy.
· Applies to areas consumers do not give consent for, and doesn't interfere with efficiency or commerce.
· Expands on existing regulations enforced by the FTC.
· Effectiveness is reassessed and evaluated by the FTC every 5 years.

Cost and Benefit

- The cost is associated with the increased expense and added step to de-identify data and regulate this.
- The benefit is a comprehensive approach to assure no data misuse, and a reparation system should it occur.

Highlights

- The solution is scalable and customizable to allow a comprehensive approach that consumers can control.
- A user's consent is required to sell or share their data, so those without consent can't harm a user.
- Implied consent can be maintained, so there are no obstacles that impede efficiency to users or businesses.

Keys to Success

- A bipartisan effort to recognize privacy as an issue and act accordingly at the Federal level.
- Accessibility and acceptance for implementation by businesses.
- An increase to FTC funding of the Privacy and Identity Protection general operating budget

Practicality

- Privacy is a current subject for policy, so now is the time to attack the problem.
- There are already conversations and proposals to increase the funding and jurisdiction of the FTC.
- Impractical cases of the solution, such as key exemptions for application, have been included.

Expected Results

- All businesses, from small local shops to big tech will support easy integration.
- All consumers, from those without concern for privacy to those cautious will feel safer.
- The FTC will be able to attack the user data misuse and uncover the data being used without consent.

The Problem

Brokers

Businesses that collect, process, and license information to other corporations are known as data brokers (also known as information brokers). While the information that they sell may be of great benefit, such as research or useful promotional content, they also frequently aggregate information on specific individuals from several sources to produce lists of email addresses, phone numbers, or physical addresses that may be distributed to marketers (Guinness, 2022). According to a Forbes article written by the internationally renowned strategic business technology advisor Bernard Marr, some of the companies that know even more about consumers than Google or Facebook, are flying under the radar. A few of the largest brokers include Axicom, Nielsen, Experian, Equifax, and CoreLogic. The FTC has published reports claiming that some of these companies are among the top data-broking specialists in the entire industry Marr states that Axicom has “pioneered the business model of collecting data on people, segmenting them and selling it to other businesses to use in their marketing” (Marr 2017). So, what sets these other names apart? Well, Corelogic for example is focused on the mortgage and real-estate industries whereas Equifax and Experian are specialized in credit references. Nielsen can be considered the leader in market research and television ratings, similar to the aforementioned Axicom (Marr 2017).

These companies continue to collect personal information and sell the personal info or financial data to third parties without many of their users knowing.

In a democratic nation like the United States, shouldn't you have the right to know what is happening with your data, let alone control the hands it falls into? Glenn Greenwald,in his Ted talk titled *Why Privacy Matters*, states,

"The United States and its partners, unbeknownst to the entire world, has converted the Internet, once heralded as an unprecedented tool of liberation and democratization, into an unprecedented zone of mass, indiscriminate surveillance (Greenwald 2014)."

This shows that even though we live in a democracy and have the right to privacy, something bigger is going on outside of our control regarding the data market. It seems as though; we Americans are allowing the market to dictate the majority of what happens to our data and others exploit the market by forcing us as consumers to give up our data, and in this case privacy, for the usage of their services. Most businesses, corporations, and organizations all use, sell, and share user generated data significantly beyond the consent of the generator, even in the case of implied consent. But how do they get ahold of our data in the first place if we aren't intending or consenting to such practices? What role do they themselves play in this "data market"?

The Data Market

The issue of personal data sharing is one of incredible scale, but often hard to pinpoint. Out of the 257 billion dollar data brokerage industry (MMR 2021), it can be difficult to figure out which transactions of customer data reveal “too much” about a customer, but we know it is a significant and integrated issue. Hundreds of prominent companies like Google, Facebook, Equifax and many more have been known to “data mine” customer information and share it with third parties for varying purposes. Looking at Google alone, it was estimated that as of 2023 Google has 274 million active users, and is the leading search engine for web browsing (Statista 2023). When considering how an active Google user may give location information to Google Maps, email information to Gmail, search engine history to the Google browser and much more; it becomes very clear the power Google holds over millions of customers within its data collection. Factoring in the hundreds of companies that do the same as Google, the world of data begins to look very scary. However, the total elimination of data brokerage is not preferable either. Many customers appreciate the convenience of targeted advertisements as well as other benefits that stem from companies having in-depth knowledge of their customer base. The problem is not that data collection exists, it's that specific types of data, like addresses, names, social security numbers, etc. can be shared to third parties without penalty, regulation, or user knowledge.

Consent

Businesses have monopolized the sale of consumer data, and are clearly taking advantage of consumers. It is important to really understand the role of these companies before fully coming to any conclusions regarding the nature of their possession of consumer data. According to *The Cost of Reading Privacy Policies*, it would take roughly **244 hours annually** to read privacy policies which comes out to around 40 minutes a day per person reading (McDonald & Cranor 2008). There is no practical world where the average person would have time to read every single user agreement they come across. Additionally, even if all these people were to read every user agreement they come by, they would not necessarily be notified of where their data goes, if they even agree with the terms at all. In an article from The New York Times' Wirecutter it was found that "In most states, companies can use, share, or sell any data they collect about you without notifying you that they're doing so" (Klosowski 2021). Clearly, data brokers are the root cause as their practices are coercive in nature and users are essentially being forced to give up their privacy to use the necessary online products that make up their everyday lives. On top of the oppressive agreements put forth by businesses, they are also engaging in the sales and "gifting" of consumer data to third parties. Danielle McNamara of Loyola University Chicago states that "The data broker market generates approximately 200 billion dollars a year and is used by countless businesses" (McNamara 2021). As you can see, data brokerage is a very large and profitable industry, yet it can often be hidden behind a veil of software. In Meta's

Help Center for their popular social media app Facebook, it states that they do not sell user data. However, one quick skim of this article proves that the company is still working with advertisers by “gifting” said data to partners regarding user demographics and other personal information that could identify its users (Facebook). This corroborates the idea of corporations “gifting” data instead of outright selling it. On top of the roles that businesses play in the data broker market, there is also the stark exploitation of consumers through violations of agreements and data brokerage monopolies. In an article from TIME magazine, Joanna Plucinska writes, “Cable giant Comcast announced Monday that 200,000 of its customers will have to reset their login information after a suspected security breach, although the company denies it was hacked” (Plucinska 2015). Evidently, the cable company was exposed for violating its user agreement and ultimately its customers' privacy rights. Another company, Kochava, participated in the sale of its geo-location data, which was uncovered by the Federal Trade Commision (FTC) that filed a lawsuit against the company for violating consumer privacy rights (Staff & Gaynor 2022). Not only are there corporations exploiting consumers' data privacy, but there are now data brokerages which have consolidated, leading to the formation of monopolies over data, and these monopolies misuse the data. In a Forbes article, Bernard Marr writes of international corporations which have turned data farming into their core business strategy. He claims that the pace and scale of this market is exponentially rising which proves that these companies are exploiting users into surrendering their rights. With the company's role in mind, it is clear that

not only is privacy being violated, but corporations play a big part in the process of **causal relationships** which consumers are being taken advantage of (Marr 2017). Clearly, legislators have an opportunity to attack the issue of data privacy without destroying the industry of data brokerage, but while changing the much-needed status quo that continues to violate the privacy rights of internet users.

Possible Solutions

With a problem of the nature previously described, there are three significant factors that must be considered when solving a problem of this type. First is the probability of breach, with the lowest probability being the ideal solution (Worth). Second is the time until discovery, as shorter time can greatly reduce remediations and other costs that must be paid (Worth). Third is effective remediation for all parties impacted, with maximum benefits for all parties as the objective (Worth). In this section we will list four possible solutions to the issue of personal data sharing. The following proposals will heavily consider the three previously listed metrics to conduct an adequate cost-benefit analysis.

Abolish all Data Collection

A possible solution to the issue of data privacy could be explored is to ban the collection of any and all consumer data. This would in fact ensure that companies don’t leak or share personal information about customers as they wouldn't have access to the information in the first place. However, this proposal is full of absurdity and is a solution that goes far beyond the scope of the problem. First, the business models of tens of thousands of companies would be made obsolete. Anything from a maps app not being able to obtain your location data, to social media apps not being able to keep a digital record of the name and emails of users. Many companies need to collect and store consumer data to maintain the

functionality of a service and benefit customers. The issue is not the collection of data itself, it's the sharing of personal data beyond the company that a user intends.

In our cost analysis, we concluded that the value of consumer data provides too many benefits that would cease to exist should we disallow collection. The data can be applied to so many quality-of-life features for users of technology, and it would do substantial harm from the consumer perspective if disallowed (Brookman). Additionally, businesses use data to optimize efficiency and provide benefits to consumers, which directly influences their success (Zhou). Because data is a key determinant for the experience of both consumers and businesses, we believe the cost of this proposal is too high to be considered as a valid option for our solution.

Abolish Data Transfers

Another possible solution to address this is to make the sale and transfer of all consumer data illegal. On its face, this is a quick and easy solution, however, we have many objections. First off, there is no denying that the sale of digital data can benefit consumers. This data improves personalized advertising, and company research and development when creating new products or services. It is common for users to find products and services they are interested in through the backdoor selling of their data (Staff, 2017). Many organizations from political campaigns to academic researchers rely on the ability to buy information about consumers to

achieve their goals. The sale of data is not inherently dangerous, the sale of personal data is. A second objection we have is that this solution has absolutely no viability in congress. Not only would data brokers and large tech companies such as Facebook, Google, etc. fight tooth and nail against this through lobbying, as they profit significantly from it, but politicians themselves may be very hesitant to back a bill that would preclude them from collecting digital campaign information. This solution makes an impact beyond the intended problem and is incredibly unlikely to be implemented by Congress.

While this proposal still allows some benefits of consumer data to both users and businesses, we believe that the value lost by restricting data is significant enough to defer away from this approach. The spreading of data allows for many additional benefits, such as research, useful outreach, and improved marketing, which is the most influential factor in providing value to data (McGrann). These uses of data either need special consideration for omission, or must be included in the solution (Worth). In an effort to retain the value of data, in both the positive outcomes to consumers and as a profitable asset to businesses, we believe the cost of devaluing the data to be too high for this to be considered as a valid option for our solution.

Data Brokerage Registry

Another possible solution would be to mandate the data brokers be tracked on a national registry. Data brokers are essentially information resellers, so such a registry would fall under the jurisdiction of the FTC and “would identify companies that collect information about consumers who are not their own customers” (Abbott, 2019). One downfall with this solution is that even if companies are selling data that is anonymous, studies have proven that it takes just 15 characteristics (including age, gender, or marital status) to re-identify someone 99.98 percent of the time (Guinness, 2022). However, a national registry would still be helpful in this case to provide consumers with the proper information on who to contact to better understand where their information is being sold, what it is being used for, and how to opt out of such practices. Obviously, there is value to be found in selling such data, as marketing helps consumers and businesses alike. The key to successfully sharing such information however is a slippery slope as some data brokers don’t have the best reputations. One striking example of this is the time when Equifax sold a list of late mortgage payers' data to brokers who bought and resold the data, which violated the Federal Trade Commission's Fair Credit Reporting act (*Data brokers profile*). Clearly, a national registry could help to not only record and display such suspicious behaviors but also prevent future violations. Ultimately, a national registry for data brokers would be a solid starting point when it comes to protecting consumers’ data. Although, like the aforementioned

solutions there are some flaws with this national registry idea which may decrease the likelihood of such a law to be passed on the federal level.

While some believe that “shining light on an issue”, meaning to draw awareness to it, is the best first step in solving a problem, we believe that the benefits of a registry do not provide enough for this to be considered a solution (McGrann). The identification of data brokers may seem useful, but it is quickly diluted by the many big tech companies that have brokerage components. Furthermore, a registry does not restrict broker ability to sell or share your consumer data with undesirable parties. Just because a broker is registered, that does not address the quality of reputation that a broker has. We do not believe that the industry standards that may emerge from this proposal are enough of a benefit for this to be considered as a valid option for our solution, and we seek to further reduce the probability of undesirable parties gaining access to your data.

Our Recommended Solution

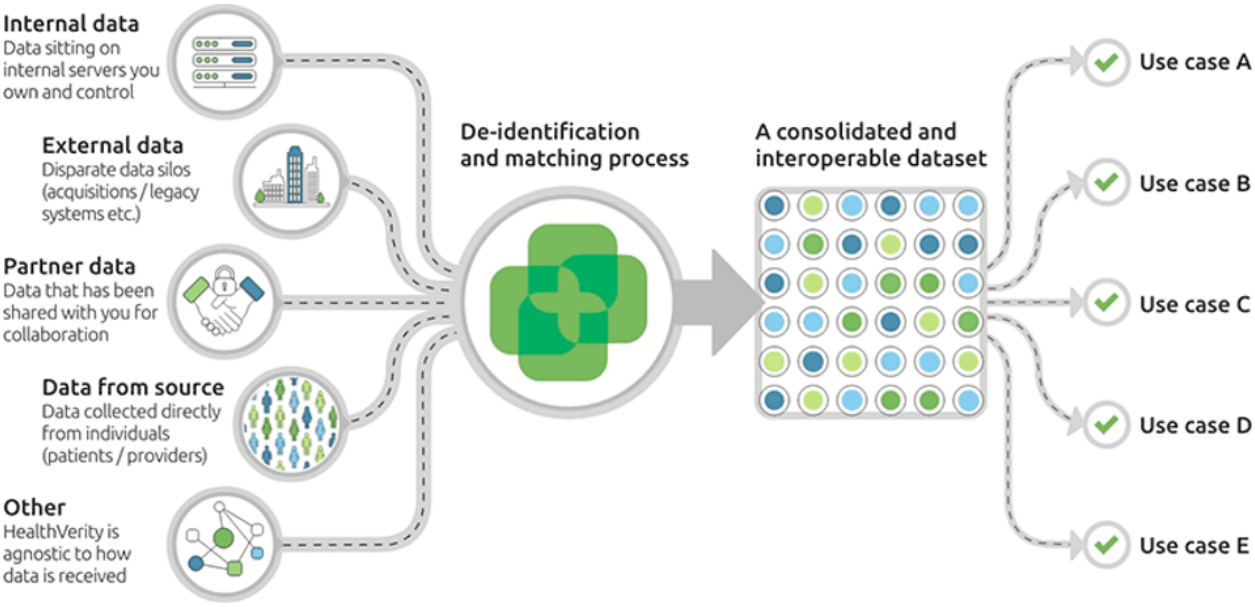
Our proposed solution is to mandate that data be de-identified before it is sold or shared between organizations. De-identification is the process of removing personal information from datasets, including addresses, email addresses, and names. A regulation like this would make it illegal for data brokers to buy data that has the consumers' personal information, without the user's explicit consent, and would essentially anonymize the process of data brokerage. The obvious advantage to this solution is that private companies should still be able to continue collecting data from their respective customers for marketing and research purposes, but the customer would not have to worry about valuable information leaving the organization that they agreed to give information to. This solution puts the responsibility on the buyers and sellers to protect consumer and user privacy, instead of the generators of data themselves. In fact, this solution has already been implemented in the California Consumer Privacy Act which mandates that companies "reasonably" remove data that would identify specific customers before sale (CCPA 2023). Similarly, the General Data Protection Regulation, an overarching data privacy law, outlaws the sale of data that contains "direct identifiers" such as name, postal code, phone number, and address, as well as "indirect identifiers" such as geolocation, employment history, and organizational affiliations (Wolford, 2022). Both laws have served as a valuable tool to protect consumers from companies that want to overextend their reach.

This solution would require federal implementation. If only applied to a state level, it would get rather confusing when determining who has jurisdiction over enforcing this policy. It would be difficult to determine if it is applicable to companies within the state, data generated by residents of the state, or all people in that state at the time, and then it would remain unclear if a consumer were to travel or disallow their exact location. For a comprehensive and unified approach, we decided that our proposal had to use a federal approach to be effective (Walker). Due to the necessity for widespread cooperation for this proposal, we decided to refute the usage of state governments for our solution, and we choose to apply our idea to the national level for greatest compliance and therefore success. Anything less than preventing undesirable access to consumer data across the nation would not exceed the threshold to qualify as a solution that best suffices all three main components of the problem. The anonymization of consumer data through encryption when selling or sharing the data is the solution that retains all the benefits that spreading consumer data provides, but then disallows undesirable parties from using this data to maliciously target consumers. The question is how?

Differential Data Encryption

Now that the solution we recommend has been presented in its entirety it is natural to have concerns over the viability of mandating companies to “de-identify” their collected data before sale. Luckily, there is already reliable precedent for approaching such a process; in fact, the US government uses it every year when collecting census data. When the United States government asks the American population to fill out the national Census Survey they are legally obligated to make results anonymous as the information included in the census (race, age, income level, marital status, etc.) is highly personal information (US Bureau of Labor Statistics 2022). This data is anonymized through the process of “differential privacy”, a technical term that refers to a data set that cannot be picked apart to find data about specific individuals. This is achieved through the addition of false but insignificant data inputs that are added as “**noise**” to make it impossible to identify specific individuals. The process does introduce some small errors in data collection but not significant enough to alter conclusive results (Lo Wang 2021). The differential privacy method is one that every company should have to adhere to before sharing or selling customer data, as it effectively removes identifiable personal information, and makes it nearly impossible to trace a specific individual to their data. With the current state of technology, differential privacy tools are easily accessible and available, and should not create any significant burden on data sellers who wish to commoditize consumer data. The process is well

documented and used in social science fields to protect personal information. There is no reason why private companies shouldn't follow suit.

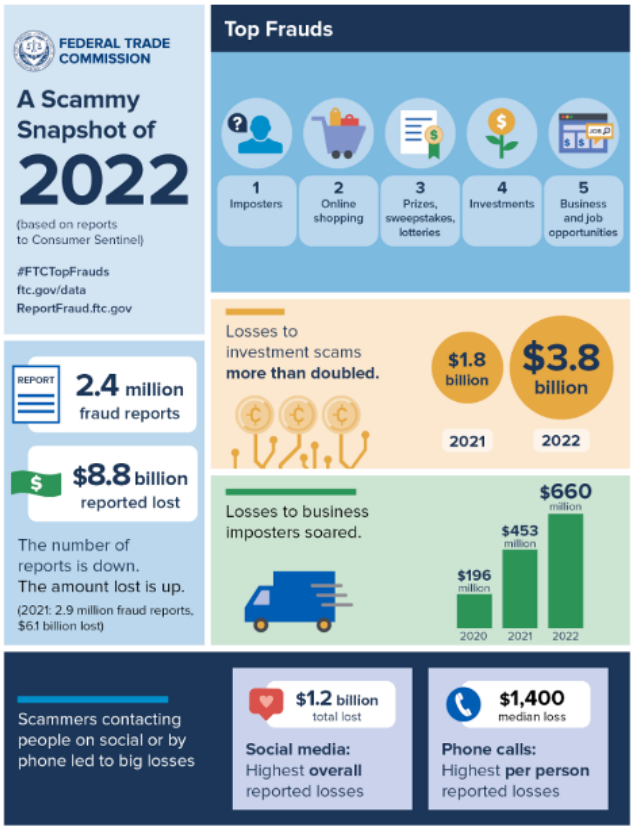


Enforcement

Considering data brokerage falls under information commerce, enforcement of any legislation related to de-anonymizing data would fall under the jurisdiction of the Federal Trade Commission. Since an increase in regulation of the data brokerage industry would expand the power and responsibilities of the FTC, we recommend that a budgetary increase be included in any legislative proposal containing our solution to account for increased bureaucratic labor for enforcing fines. Fines

related to data could also potentially be absorbed back into the FTC to account for increased costs.

Its also important to note that technology and the internet is ever changing, and doing so with incredible speed, therefore regulations must follow suit. If implemented, we recommend the FTC review their de-identification standards periodically every 5 years to account for changing market dynamics. Who knows the types of data collection we may see in the future? This answer is unknown but one thing is clear: consumer privacy must be protected even in a fast changing landscape.



Benefits

Our solution excels at satisfying the three major components that a solution of the problem should address. First, the probability of breach is not only reduced, it is nearly eliminated. With increased difficulty tracing a data point back to a specific consumer, and the disallowance of unauthorized selling or sharing of data without differential data encryption, consumer data is unlikely to get passed along to undesirable parties. This still gives the consumer an option to opt-in for their consumer data to be shared, research can be gathered from the anonymized form, companies can still collect, use, and sell or share data, and the probability of a breach is reduced across all circumstances. For the time until the discovery component, the aspect of our proposal that calls for increased regulation through data audits will fulfill this piece. Even if some big tech companies or data brokers are not audited, the possibility of an audit, pressure from industry standards, and the widespread compliance from purchasers of data will discourage misuse of consumer data, and make misuse increasingly noticeable. Discovery will be expedited by the ability of the FTC, and even consumers to an extent, to audit how their data is being used, and where or by whom their data may have been misused. Lastly, our solution establishes effective, direct, and comprehensive remediation. The establishment of a victim's fund through the collection of fees following misuse of consumer data by companies, remediation will be provided to the victims of the

misuse. This is effective as companies will be discouraged to avoid hefty fees, which have a compounding effectiveness with the reduced time to discovery, and these fees will generate a substantial fund to be paid out to the victims that had their data misused. This remediation method is direct in that it directly affects the consumer with compensation. Also, this approach is comprehensive as anyone that can generate data can also qualify for this remediation, so no victim is left uncompensated. It is because of the reduced probability of misuse or a breach, in part due to an increased likelihood of discovery with a decreased time until discovery, and effective remediation that will both dissuade misuse and properly compensate victims, that we believe our solution is best fit to address the problem instead of any other proposal.

In an article from The Rand Journal of Economics, Bounie and his colleagues wrote about various strategies used by data brokers when selling information to multiple firms (Bounie et al., 2021). They hypothesized that brokers would price-discriminate between customers. Using a model which simulates the process of data being distributed by brokers, the team was able to formulate different results based on various strategies employed between trials. In the results, they showed that brokers sell information on consumers to companies that pay the most, but will sell unidentified data to companies who pay the least. This proves that data brokers can use de-identified data, but may need to employ alternate strategies to reap full benefits.

An article written by Arnaud De Bruyn and Thomas Otterwhich touches on Bayesian profiling (De Bruyn & Otter, 2022). The colleagues suggested that the new method of presenting data provided more reliable predictions. Using their proposed method of Bayesian profiling, they were able to conclude that the current simple count method was inferior to that of the Bayesian. Overall, this proves that de-identified data has higher face validity and therefore may be more valuable than originally thought.

In an article on data encryption written by Basapur and his colleagues (Basapur et al., 2019) , they wanted to propose a method that would, “minimize the number of attributes to be obscured and enhance data usage while preserving information confidentiality” (Basapur et al., 2019). The team used two data sets and uploaded them into Apache Spark on a cloud environment to test and found that there is a correlation between confidentiality levels and the number of attributes to be obscured. This data encryption technique clearly preserves the privacy of data but allows for it still to be used for other purposes by third-parties, which is why such practices would be beneficial in the making of our policy.

Expert Consultation

Industry experts have also expressed direct approval, and offered additional insights when consulted about our solution.

In a conversation with Justin Brookman, we focused on the consumers. Not only would aggregate data remain effective and valuable should that be the limit of what could be collected, but our solution allows more value to be collected from the consumer data and still protects consumers from the leading problem: unwanted advertisements (Brookman). This problem is worsened by the fact that the sale of consumer data, which often results in unwanted spam, is unable to serve as the saving grace for struggling industries, such as news, so the irresponsible sale of consumer data to undesirable third parties is unnecessary for them as it is not a significant determinant of their business or their original intention (Brookman). Even in the case of the transfer of data between two companies that are owned by the same parent company, we must protect the consumer's data from being transferred across functionally different companies without their consent, and we should include the data broker components of big tech within our definition of brokers and our solution (Brookman). While Brookman believes that coercion to surrender one's rights to their data is present when consumers use the internet, and consent is still not enough to properly protect consumers as their privacy policy would forego the consumer's protection, we believe that there is pressure to assure the protection of consumers regardless of a company's privacy policy (Brookman).

In a conversation with Kevin McGrann, we focused on the likelihood of passage for our proposal. McGrann claims that a bipartisan effort is highly probable given the nature of the issue and the relevancy of our recommendation (McGrann). Data broker lobbyists are powerful, but the industry is under attack, and trust in them has been reduced, so a bipartisan effort can likely overcome that obstacle. McGrann recommends that our proposal utilizes the FTC, by either expanding current regulations and powers that keep brokers in check or adding new powers within the FTC, and assures that the relevant committees would be the Energy and Commerce Committee and the Senate Commerce Committee (McGrann). Following this conversation, and with further research assisted by the Applied Data Center of Lake Forest College, we have decided to propose an increase to the Privacy and Identity Protection general operating budget within the FTC, as that budget is used to protect personal information on the internet and the FTC sees value in its increase (FTC).

In a conversation with Enid Zhou, we focused on exceptions to the solution and determinants of its effectiveness. While Zhou seconds confidence in the likelihood and ability of a bipartisan effort, she expressed two specific concerns that must be addressed by our solution. First, politicians require access to the information within the voter files for campaign purposes, and any proposal that impedes their access might prevent their approval of our proposal (Zhou). As to avoid an extremely influential barrier to the passage of our proposal, we wish to include an exemption that grants implied consent to the use of the voter files and the digital data

collected for them. Second, the ability of some brokers to unencrypt the data would be a concern for the longevity of the consumer data. To account for this, we must incorporate a clause that data must remain in its anonymized state if that was the method by which it has been sold. Therefore, a broker is unable to unpack the data that they have been sold with differential data encryption as they have not received consent from the consumer. Data comes in many different types, and in vast amounts, so protection to disallow misuse of data can never be 100%, but privacy is a bipartisan issue, so an improvement with this proposal is probable (Zhou).

In a conversation with Steve Worth, we discussed the primary motivators of organizations and the prevalence of industry standards. Worth claims that there are three primary motivators for an organization to act. First, they are motivated to act in accordance with what feels right (Worth). We appeal to this motivator as privacy is a universal right that must be protected for users of technology. Second, they are motivated to act in accordance with policy (Worth). Pending the bipartisan support and implementation of our proposal, a policy will exist to pressure organizations to follow through with the proposal. Third, they are motivated to act in accordance with, or exceed, industry standards (Worth). In any given circumstance, privacy protection can be seen as an asset worth considering for organizations and consumers. We believe that there is added benefit to a superior privacy policy, and this may serve as a differentiator between competitive companies, so there will be an industry initiative to excel at privacy protection beyond what a policy may require (Worth). Ultimately, policy is used to enforce

compliance with what is believed to be the correct course of action, and the industry standards will reinforce the intent by establishing an asset worth representing beyond basic compliance of a policy.

Success Factors

Assuming our proposed solution has been implemented we must look for success factors; or indications that our policy has made a significant positive impact. To do this we must restate the primary issues associated with the sharing of personalized data. To the average person, they will be impacted by this issue through spam emails. For example, a user signs up for an account on a reputable website, but then that reputable site shares their personal email address with data buyers who will then use their contact to send marketing emails to the user. In September of 2021, it is reported that around 106 billion spam emails are sent per day worldwide (Dixon 2022). Our proposed solution would require that personal email addresses be anonymized before the sale. In this way, we would expect that users would receive significantly fewer spam or unsolicited emails from companies if this legislation is implemented. Ideally, this would lead to a declining success rate for online scams and more satisfied users that are not being barraged by overwhelming targeted messaging. The risk factors of our solution are primarily found in the externalities and the implementation of our proposal.

Externalities

External components will be impacted by the implementation of our legislation. While these externalities would include restrictions to the voter files and possibility for aftermarket deanonymization, the additional clauses of our legislation prevent the effects on these externalities. Instead, the significant externalities that would be affected include the need for a middleman to de-identify collected data, where the funding for that may come from, and a decrease in some broker options, and a loss of some jobs. While a decrease in some broker options is acceptable, as we are reducing the options to include only trustworthy data brokers that are compliant with the intention of our legislation, and the corresponding loss of jobs at the data brokers are compensated for with the creation of jobs in data surveillance, and that those lost jobs should not have existed in the first place as they threaten consumers, the other three externalities are significantly affected and worthy of mentioning as such. The first is the relationship between brokers and collectors of data. The burden of anonymizing data is a factor worth exploring further. The brokers are unable to purchase data not collected with differential data collection practices. However, not all data collectors can provide data in that way, so how should this work within our legislation? We must account for data being made available to purchase in that format, where does that responsibility fall, and what is the most effective way to apply the differential method that the Census uses? The second is the potential price gouging of data collected with differential data collection practices. With a legal requirement to purchase data only presented in an

anonymized state, companies can force the hand of brokers to pay a marked-up price as there is no longer healthy data competition. This externality is combated with an additional clause that the conceptual value of data cannot be raised based on the collection method. Lastly, we must determine if the differential data collection method is applicable to all types of data, or if we need a workaround for specific data types.

In order to retain most of the key benefits of the selling and sharing of data, and data brokers, we wish to grant exceptions to circumstances in which the industry of the collector is already regulated by policy, such as The Health Insurance Portability and Accountability Act, or in the cases outlined in virginia law (Lamont).

“Nothing in this chapter shall be construed to restrict a controller's or processor's ability to: Prevent, detect, protect against, or respond to security incidents, identity theft, fraud, harassment, malicious or deceptive activities, or any illegal activity; preserve the integrity or security of systems; or investigate, report, or prosecute those responsible for any such action” (Virginia Law)

Exceptions for these uses allow for a positive adjustment to our proposal that helps to account for the positive benefits of these practices and organizations, so we can omit any adverse impact in these cases.

Secondary Effects

Our proposal introduces a mandatory procedure for managing the sale or sharing of data, the secondary effects that may exist involve businesses implementing minimal compliance or simply paying the fine for violation of compliance. Should a business not become motivated by industry standards, and elect to follow the bare minimum procedure for policy, we must assure that the minimum is sufficient to serve as a solution to the problem. Additionally, we must consider both the threat and the intent of a fine to assure that this is not a commonly chosen route as our intent is to protect consumer data instead of punitive actions against brokers. With both of these secondary effects being ways to avoid prioritizing protection of data, we have chosen to adapt our solution to account for these possible effects.

As a result of the effects, our proposal is to maintain that minimal compliance is still enough to properly protect consumers, and intending to pay the punitive fine is not a desirable enough course of action that it is often considered. Our proposed policy is comprehensive enough to assure that even a business that gains consent to sell and share consumer data in any possible way is unable to provoke misuse of that consumer's data. This is possible as consent to sell and share that data, even with implied consent, ends with the business that collected it, so a buyer of that data is unable to resell or share it without following our proper encryption procedure. Those that wish to violate this method are to be charged a hefty fine. To best enforce our solution, fines must be calculated based on the egregiousness of

the offense, the significance of the data that was misused, and this enforcement should be the authority of the FTC (Lamont). Our punitive action works like this to assure that it is neither a one time fine for being caught nor a flat fine for misuse, but instead one that accounts for damage to the victim, which is also the current standard within government . Each aspect of misuse will drastically increase the punitive action against the organization, and this will assist in the proper reparation to the victims of misused data, or to do that by assisting in the upkeep and improvement of the FTC. Money collected from fines may also be assigned as agency funding for the FTC to further support their operation (Lamont). To avoid the secondary effects of refusing earnest implementation of our ideas, we have enhanced our solution to assure that compliance is the most sensible option, and there are mechanisms in place to follow up on that concept.

Unintended Consequences

The first unintended consequence that may follow the implementation of our proposal would be an adverse impact to small businesses. We anticipate one major consequence for any business outside of big tech; the inability to encrypt their own data for sharing or sale. The inability to encrypt data would become a barrier that disallows the spread of data, which prevents all of the benefits that spreading data responsibly can result in, including profits from its sale. The intention of our

proposal is not to adversely affect small businesses, so we are relieved that we can count on some additional externalities to circumvent the unintended consequence.

Given this, small businesses will be able to thrive from our proposal, given the help of some probable additional externalities. First, the utilization of data encryption middlemen will be able to assist in allowing those otherwise unable collectors of data to sell or share their data (Worth). As a lucrative opportunity, we believe middlemen will be accessible to help package data with differential data protection, and this will expand the capabilities of all collectors. Additionally, this will also combine the efforts of data collectors under one organization for data brokers and researchers to access as a primary touchpoint. Second, businesses that have successfully integrated into a market, and have earned the trust or following of a consumer base, will be able to maintain the consent of their consumers to sell or share their data, and these businesses need not respond to our proposal, so the benefits of our proposal will affect their consumer data at the secondary broker stage. Lastly, small businesses that do not collect consumer data can only experience positive change. These additional externalities of our proposal allow for the safer spreading of data, greater opportunities to share the data, especially for collectors that were previously not offering their consumer data, and these factors provide a substantial opportunity for the group that we were initially concerned about.

Implementation

Challenges

The need and effectiveness of this solution are unquestionable, but there will be two major challenges for implementation. First is the compliance of big tech companies, which are responsible for significant portions of data and have previously determined most data policies. Second, we will need our bill to bypass broker lobbyists integrated within the federal government. While we believe that both barriers are insufficient to stop our proposal, they serve as the two primary determinants of success with implementing our solution, which will easily persist past both significant barriers to entry, and will gain increased strength because of the challenges that affect it.

Big Tech Companies

For big tech companies and compliance, we believe that they will be fully supportive of adopting this solution and the three key aspects that affect them. First, the

companies won't need to change much. Consumers will continue to agree to their private and individual privacy policies, and the companies can maintain current practices of data usage. This data will still be protected from second hand sharing or sale, because consumers only gave consent to the policy of the big tech company, and the recipient of the company's spreading must protect the consumer data with encryption. Second, big tech can boast a more secure platform. The third parties are corrupting big tech with spam and other nuisances that they can live without, and third parties are less capable of targeting their users after our solution. Lastly, big tech can avoid class action lawsuits by adhering to a federal policy for data management. These companies, thought of as platforms, have been continuously prosecuted for data and privacy related hearings. We believe that big tech companies will be supportive of our solution as it does not demand much change from them, it can increase the safety of their users by addressing the three components of the problem, and it can protect them from further legal issues regarding privacy.

Broker Lobbyist

For broker lobbyists, we discount their ability to reject our proposed solution as they are perceived as untrustworthy and data privacy is perceived as a priority

action item. While we would normally be concerned about the inability to gain the support of a simple majority, we take comfort in knowing that there is a distrust of broker representatives, data privacy is a bipartisan issue, and it has been a current topic within state governments, congress, and the supreme court. Recent hearings and class action lawsuit settlements further reinforce the need for additional privacy regulation. With an increase to FTC funding existing as an ongoing discussion, we like to propose that the privacy concerns be added as a further push in the direction to motivate bipartisan support that furthers FTC oversight and the protection of privacy.

Conclusion

Privacy is an inherent right; it must be protected in all instances. Currently, privacy for users of technology is not protected as implied consent allows the misuse of user data, and this is perpetually causing further harm. As an unregulated area, we need to implement both funding and a procedure to protect user data from misuse by data brokers and those looking to sell or share data. By expanding an already existing function of the FTC, to allow for data audits that assure compliance of our proposal, we are able to assure that consent is the key that locks up your data, and nobody else can access it without your permission.

Bibliography

Abbott, J. (2019, September 13). *Time to build a national data broker registry*. The New York Times. Retrieved February 26, 2023, from https://www.nytimes.com/2019/09/13/opinion/data-broker-registry-privacy.html

Basapur, S. B., Sylaja, B. S., & Venkatesh. (2019). Privacy preservation using (L, D) inference model based on Dependency Identification Information Gain. *International Journal of Engineering and Advanced Technology*, *8*(6S3), 1170–1173. https://doi.org/10.35940/ijeat.f1196.0986s319

Bianchi, T. (2023). *Topic: Google*. Statista. Retrieved April 6, 2023, from https://www.statista.com/topics/1001/google/#topicOverview

Bounie, D., Dubus, A., & Waelbroeck, P. (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*(2), 283–313. https://doi.org/10.1111/1756-2171.12369

Brookman, J., Ponticelli, J., & West, C. (2023, February 9). Interview with Director of Privacy and Technology Policy at Consumer Reports. personal.

California Consumer Privacy Act (CCPA). State of California - Department of Justice - Office of the Attorney General. (2023, February 15). Retrieved February 26, 2023, from https://www.oag.ca.gov/privacy/ccpa

Cate, F. (n.d.). *Privacy and consent*. Fred Cate: Privacy and Consent | TED Talk. Retrieved February 26, 2023, from https://www.ted.com/talks/fred_cate_privacy_and_consent

Data broker market: Global industry forecast (2022-2029) by data category, data type, pricing model, end use sector, and region. MAXIMIZE MARKET RESEARCH. (2023, February 28). Retrieved April 6, 2023, from https://www.maximizemarketresearch.com/market-report/global-data-broker-market/55670/

Data brokers profile. Data Brokers Profile | Privacy International. (n.d.). Retrieved February 26, 2023, from https://privacyinternational.org/taxonomy/term/543

De Bruyn, A., & Otter, T. (2022). Bayesian consumer profiling: How to estimate consumer characteristics from aggregate data. *Journal of Marketing Research*, *59*(4), 755–774. https://doi.org/10.1177/00222437211059088

Dixon, S. (2022) Average Daily Spam Volume Worldwide from October 2020 to 2021 (in billions). Statista. https://www.statista.com/statistics/1270424/daily-spam-volume-global

Guinness, H. (2022, May 25). *How data brokers threaten your privacy*. Popular Science. Retrieved February 26, 2023, from https://www.popsci.com/technology/data-brokers-explained/?amp

Lamont, K., & Ponticelli, J. (2023, April 7). Interview with Director at Future of Privacy Forum. personal.

Lo Wang, H. (2021). For The US Census, keeping your data anonymous and useful is a tricky balance. NPR. https://www.npr.org/2021/05/19/993247101/for-the-u-s-census-keeping-your-data-anonymous-and-useful-is-a-tricky-balance

Marr, B. (2017, September 7). *Where can you buy Big Data? here are the biggest consumer data brokers*. Forbes. Retrieved March 26, 2023, from https://www.forbes.com/sites/bernardmarr/2017/09/07/where-can-you-buy-big-data-here-are-the-biggest-consumer-data-brokers/?sh=4bdc98c86c27

McGrann, K., Calzadilla, E., Ponticelli, J., & West, C. (2023, February 21). Interview with Senior Vice President at Forbes Tate Partners and former Boehner Political Chief. personal.

Pattison-Gordon, J. (2022, March 20). *Reform to federal internet legislation must learn from past mistakes*. Governing. Retrieved February 26, 2023, from

https://www.governing.com/security/reform-to-federal-internet-legislation-must-learn-from-past-mistakes

Ponticelli, J., & Walker, C. (2023). Interview with Professor of Public Policy. personal.

Ponticelli, J., & Zhou, E. (2023, March 7). Interview with Senior Counsel at EPIC. personal.

Staff, the P. N. O., & Nguyen, S. T. (2022, March 11). *Federal Trade Commission act*. Federal Trade Commission. Retrieved February 26, 2023, from https://www.ftc.gov/legal-library/browse/statutes/federal-trade-commission-act

US Bureau of Labor Statistics (2022). Consumer Expenditures and Income: Collections and Data Sources. US Department of Labor. https://www.bls.gov/opub/hom/cex/data.htm#:~:text=Survey%20notification%20and%20collection%20method,-A%20selected%20sample&text=The%20Census%20Bureau%20conducts%20both,the%20Interview%20and%20Diary%20Surveys.

Worth, S., & Ponticelli, J. (2023, March 28). Interview with Tech Entrepreneur. personal.