MDPI

Article

Data Collection in the US: Policy Recommendations for Protecting Individuals' Privacy

Audrey Bertin[1]

[1] Smith College - Department of Statistical and Data Sciences Northampton, MA, USA;
* Correspondence:

Version May 4, 2021 submitted to Water

Abstract: The US is currently facing challenges with data privacy. In the modern age, the personal information of millions of Americans is collected and tracked daily. Data has essentially become a currency, with companies putting ethics aside to try and collect as much data—and therefore as much profit—as possible, threatening individuals' rights to privacy and placing them at risk of harm. Unlike many other developed countries, the US has taken little action at the national level to address this issue. We consider why that is and propose recommendations for how the US can proceed to address the issue through policy.

The United States is currently facing a crisis of personal data privacy. In recent decades, a new era in the digital world—commonly termed the *Internet of Things*, or IoT—has emerged. The IoT refers to the technologies, devices, and people who enable the sharing of data worldwide, and is used to characterize the modern internet age as one whose focus is now on big data (Atzori *et al.* [1], Elvy [2]).

Improvements in computing power and internet speed, alongside the development of new technologies capable of storing and utilizing massive quantities of data, have ushered in a new economic age: the data economy. Data is now a hot commodity, with the power to be incredibly valuable to those with the technology to utilize them. According to the big data strategist at Oracle, a major software company, "data is in fact a new kind of capital on par with financial capital for creating new products and services" (MIT Tech Review and Oracle [3]). Data provides this value through several means: by enabling businesses to profile and target people, leading to higher success rates in attracting customers; by providing information that can be used to help optimize systems; by helping manage and control things; by allowing companies to model probabilities more accurately; and by allowing certain software to operate in a way that would not be possible otherwise (Sadowski [4]).

Much of this data comes directly from the general public: the people who use the goods and services produced by companies—like Amazon, Google, and Facebook—that participate in the data economy. These corporations collect user data constantly and on a massive scale. Amazon tracks users' purchases and voice commands—even going so far as to track the lines highlighted in books bought by Kindle readers (Paul [5]). Google tracks every search users make, every YouTube video they watch, their full calendar schedule, their Gmail messages, everywhere they go, how long they stay there, and what route they take—even if users do not have Google Maps open (Google Team [6]).

This large-scale data collection poses significant ethical implications when considering the potential effects on consumers. For one, there is the concern of data breeches. Since user data is collected and shared over the internet, it is at risk of being released—or taken for nefarious purposes—through cyber-hacking events. In 2019, hundreds of millions of facebook users' phone numbers, locations, and emails were stolen (Bowman [7]). Facebook is not alone; many other companies have seen serious data breaches in recent years: Yahoo, Experian, Twitter, and Microsoft, to name a few (McCandless and Evans [8]). This danger is heightened further in situations involving private health or financial data, leaving consumers at risk of negative impacts from the release of sensitive health information,

as well as possible identity theft and financial harm. Additionally, the misuse or unwanted release of information on polarizing issues such as religion, sexual orientation, or gender identity could potentially put vulnerable consumers in harms' way—making them targets for attack.

Even if data is not released in a breech, its use and collection can still harm consumers in other ways. Large scale data collection is often used to power machine learning algorithms, which use data to make predictions about people. These algorithms—though seemingly objective at first glance—are often negatively biased toward minoritized individuals, described in the book *Data Feminism* as those who are "actively devalued and oppressed by a dominant group," often including women, people of color, and the poor (D'Ignazio and Klein [9]). For example, facial recognition algorithms built in large part off of Facebook photos misidentify black women at a significantly higher rate than other groups (Buolamwini and Gebru [10]). The more data that is collected, the more algorithms that can be created—placing minoritized individuals at risk for potential harm due to algorithmic bias.

Data can also be used for psychological manipulation. Sites like Facebook and Twitter have a wealth of data on their users—enough to predict with high accuracy how they will react when exposed to certain stimuli. These sites can use that knowledge to spread targeted messages and actively change the beliefs held the public. This is what happened in the 2016 election, when Facebook's advertising system targeted those individuals it calculated to be likely susceptible to conservative messaging with advertisements that reflected the ideals of Donald Trump. Though not known for certain, it is widely believed that these advertisements may have convinced enough voters to support President Trump that he eventually won the election (Madrigal [11]).

The practice of mass data collection is far from ethically sound. In spirit—though not officially declared by law on the federal level—it violates one of the founding principles of the United States: individuals' right to privacy. The right to privacy is a crucial foundation of this country. It was first officially alluded to in the Fourth Amendment to the constitution—though it is implied to some degree in the First, Third, and Fifth as well—and has been reinforced at the highest levels of the judicial system for centuries.

Mass data collection can also put individuals' First and Fourteenth Amendment rights at risk. Freedom of speech and of expression are threatened by the possibility of data breeches—users who fear that their data may be taken and used by unauthorized individuals may refrain from sharing controversial opinions or personal information on religion, sexual orientation, or other sensitive matters online, for fear that it may be used against them when they would have otherwise done so if not fearful of an information breach. The equal protection clause of the Fourteenth Amendment is likewise threatened. Although not necessary intentional, the bias present in the mass public/consumer data algorithms has the effect of treating equal groups differently, which—if used in certain circumstances—can violate the principle that all people deserve equal protection.

Now, more than ever, Americans have lost faith in the ability of companies to protect their private data, and their trust only lessens year by year (Olmstead and Smith [12]).

Despite all of the dangers and ethical concerns inherent in mass data collection, companies still continue to practice it because of the financial benefits. Lawmakers in the United States have recognized the problem and attempted to solve it through legislation, but their efforts have fallen short. The US has failed to produce any comprehensive legislation on data privacy at the federal level and at the state level there exists only a patchwork of laws, plagued by inconsistencies, conflicting information, and sub-optimal enforcement procedures (O'Connor [13]).

Healthcare data is by far the most protected variety, though even it lacks a clear and comprehensive piece of legislation. The Health Insurance Portability and Accountability Act (HIPAA) only applies to certain "covered entities," and does not protect people in all situations (Department of Health and Human Services [14]). Student health records are covered under a different law, the Family Educational Rights and Privacy Act, which occasionally combines or conflicts with the Children's Online Protection Privacy Act (COPPA), meant only to protect the data of children under 13 years of age (Dept. of Education [15], O'Connor [13]).

This failure to protect personal data is unique in the industrialized world. 128 countries out of 194 worldwide have put in place national data protection laws, many of which are quite strict and offer citizens significant control over the use of their data (Conference on Trade and Development [16]).

Citizens of the European Union (EU) are protected under the Global Data Protection Regulation (GDPR), a comprehensive piece of data privacy legislation which gives individuals in the EU the power to control what data is collected on them and what is done with it in addition to restricting the transfer of personal data from EU members to other countries. The GDPR is relatively expansive, providing consumers with a wide variety of privacy protections, including requirements that individuals be quickly notified in the event of a data breach and that—at a minimum—identifiable data must be pseudonymised. Additionally, EU citizens are given rights to access and control their data. Under GDPR, individuals have the right to access information about what is being collected of them, the right to rectify mistakes, the right to have their data erased or to restrict/object to processing, as well as rights related to automated decision making and profiling. Companies (data "controllers") are required to meet high privacy standards, and there are systems in place to ensure compliance. These include Data Protection Officers, appointed to help observe and promote compliance, alongside a tiered fine system for GDPR violations (European Commission [17]).

The EU is not the only region in the world that has strong data protection laws. South Korea, for example, has the Personal Information Protection Act (PIPA). This law is very similar to the GDPR, with requirements for at least pseudonymising the data. PIPA also contains rules for measures that must be taken when handling personal data to ensure privacy. Express consent is required for the collection of personal data, with a specific focus on sensitive data. Information such as passport or drivers license number and information about ideology, religion, health, sexual orientation, or other sensitive subjects must be collected separately from one another *and* separately from any other consent. Citizens also possess similar rights to under the GDPR, including access, correction, suspension of use, and removal of personal data (Bae Park [18]).

Some countries have even gone so far as to declare data privacy a fundamental right. In Chile, the official constitution (Article 19, Number 4) establishes the individuals right to (i) respect and protection of private life, (ii) honor of the person and his/her family, and (iii) **protection of his/her personal data**. Anyone whose rights are threatened or disturbed has the power to file a Constitutional Protective Action in response (Molina [19]).

Despite the apparent constitutional focus on privacy in the country, it is clear that the United States trails far behind other developed countries in terms of data privacy protections. Why is this?

Perhaps the most important factor in preventing nationwide data privacy regulation is the role of the technology lobby in politics. Many of the world's most powerful technology companies are headquartered in the United States: Facebook, Google, Microsoft, Twitter, and Oracle, to name a few. Bringing in hundreds of billions of dollars in revenue, these companies play a pivotal role in the US economy, and as a result they have a lot of political power. Their large profit margins enable these companies to lobby the US government, preventing the passing of laws and policies that are not in their best interest financially. Alphabet Inc, the parent company of Google, spent 21.74 million dollars trying to affect US policy in 2018 (D'Souza [20]), and Facebook spent 19.68 million in 2020. (Chung [21]). Technology companies have now surpassed Big Oil and Big Tobacco—the previous most powerful lobbying groups—as the biggest spenders in US politics, and their spending is only increasing over time (Chung [21]). A privacy law is counter to these companies' best interests. Additional requirements for data security, limitations on what can be collected, and financial punishments for non-compliance risk damaging these companies' profits, which rely partially on large-scale data collection. Technology companies do not want to see a privacy law passed, and have the funding and power available to prevent any drastic changes from happening.

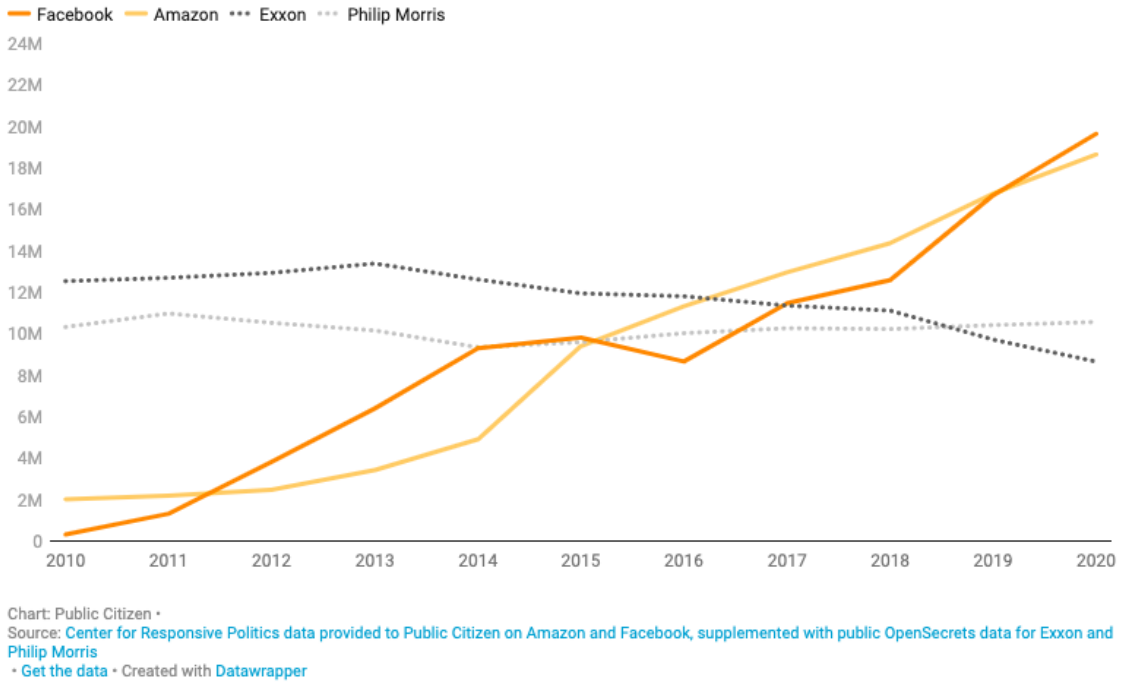

Figure 1. Big Tech replaces Big Oil and Tobacco as big lobbying spenders: Total lobbying spending in Washington vs Facebook and Amazon's lobbying spending (2010-2020). Source: https://www.citizen.org/article/big-tech-lobbying-update/

There is also the issue that the United States does not currently possess any agency that has the ability to fully manage and ensure compliance with a federal data privacy law. In the EU, all of the member states have independent privacy-focused authorities to enforce the GDPR. The US has nothing like that. The closest thing is the Federal Trade Commission (FTC), which has historically managed other aspects of consumer protection and privacy. However, it is relatively weak in its powers: it is constrained in its authority, with jurisdiction primarily in interstate commerce. The FTC can only regulate violations that meet standards for "unfairness and deception," but these are poorly defined and unclear. Furthermore, the FTC lacks a strong record of privacy enforcement—even in previous cases where it has had the jurisdiction to intervene—and does not have leadership with a strong level of technological expertise in the field of data privacy (Chao *et al.* [22]).

Due to the power of the technology lobby and the lack of enforcement ability, it is difficult to imagine a large-scale, universal data privacy law like the GPDR being passed at the national level any time in the near future. If one were pushed in congress right now, it would almost certainly fail, due to both challenges with its enforcement and lobbying pressure from big tech. Instead, lawmakers wishing to institute privacy policy will have to take a slower, step by step approach, taking careful consideration at each step to prevent dramatic counteraction from technology companies.

The first step necessary will likely be either the creation of a new data protection agency or the bolstering of FTC powers to include more oversight capability for privacy regulation. A data protection law without an agency capable of overseeing compliance will be useless, so this step should be completed first. As this action does not directly and immediately financially impact technology companies, it is also one of the more feasible policy actions available at the moment.

In the mean-time, states should continue to create their own data privacy regulations, though with additional thought to ensure that their proposed laws are consistent with those that have already been passed. Although a patchwork of state laws is not optimal, the more that individual states begin to push for privacy legislation, the more pressure is put on the federal government to enact a policy nationwide, as it becomes clear that the public desires change.

In the long run, however, a state-by-state patchwork will not be sufficient and we will need a comprehensive federal law to ensure that guidelines are consistent and all citizens are protected. The passing of this legislation should be done with great care. Lawmakers should focus primarily on methods for prevention of privacy failures, rather than on mechanisms for punishment after the fact. This will ensure that the financial harms to technology companies—in the form of large non-compliance fines—will be minimized, making the law more palatable to the technology lobby. Due to the significant political polarization in United States, it is also important to ensure that any data privacy regulations that are suggested are written in a way that is favorable to both the Democratic and Republican parties—and, if possible, put forward by a bipartisan coalition—so that the law is not put on hold for political reasons. There is evidence that members of both parties have been supportive of privacy regulation in the past (Committee on Commerce, Science, & Transportation [23]), so this goal is not un-achievable.

The road to comprehensive data privacy legislation in the United States is not an easy one, but it is important to uphold a standard of data ethics and ensure that our citizens do not face dangerous outcomes from the large-scale data collection happening in this country. Although the US is still far behind much of the industrialized world in terms of protecting its consumers' data privacy and many obstacles lie in the way of changing this, with the right focus and a carefully planned course of action, this has the potential to change for the better in the coming years.

References

1. Atzori, L.; Iera, A.; Morabito, G. The internet of things: A survey. *Computer networks* **2010**, *54*, 2787–2805. doi:https://link.springer.com/article/10.1007/s10796-014-9492-7.
2. Elvy, S.A. Commodifying Consumer Data in the Era of the Internet of Things. *B.C.L. Rev.* **2018**, *59*, 423.
3. MIT Tech Review and Oracle. The Rise of Data Capital, 2016.
4. Sadowski, J. When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* **2019**, *6*, 2053951718820549. doi:https://journals.sagepub.com/doi/full/10.1177/2053951718820549.
5. Paul, K. 'They know us better than we know ourselves': how Amazon tracked my last two years of reading, 2020.
6. Google Team. Safety Center — Google Privacy Documentation, 2021.
7. Bowman, E. After Data Breach Exposes 530 Million, Facebook Says It Will Not Notify Users, 2021.
8. McCandless, D.; Evans, T. World's Biggest Data Breaches & Hacks, 2021.
9. D'Ignazio, C.; Klein, L.F. *Data Feminism*; Mit Press, 2020.
10. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency. PMLR, 2018, pp. 77–91.
11. Madrigal, A.C. What Facebook Did to American Democracy — And why it was so hard to see it coming, 2017.
12. Olmstead, K.; Smith, A. Americans and cybersecurity. *Pew Research Center* **2017**, *26*, 311–327.
13. O'Connor, N. Reforming the U.S. Approach to Data Protection and Privacy, 2018.
14. Department of Health and Human Services. Covered Entities and Business Associates, 2017.
15. Dept. of Education. Privacy, 2021.
16. Conference on Trade and Development. Data Protection and Privacy Legislation Worldwide, 2020.
17. European Commission. GDPR — Official Legal Text, 2018.
18. Bae Park, K. South Korea - Data Protection Overview, 2020.
19. Molina, O. Personal data protection is a constitutional right in Chile, 2018.
20. D'Souza, D. Tech Lobby: Internet Giants Spend Record Amounts, Electronics Firms Trim Budgets, 2019.
21. Chung, J. Big Tech, Big Cash: Washington's New Power Players, 2021.
22. Chao, B.; Null, E.; Park, C. Enforcing a New Privacy Law — Who Should Hold Companies Accountable?, 2019.
23. Committee on Commerce, Science, & Transportation. Cantwell, Cassidy, and Klobuchar Introduce Bipartisan Legislation to Protect Consumer Privacy, Promote Public Health for COVID-19 Exposure Notification Apps, 2020.