2022

Data on Demand

Isaac Zarate

It is 2022, and for many, the concept of data tracking is nothing new. Perhaps you have heard that same story from friends and family about how companies like Facebook and Google use data to track your behavior. Maybe you have had that funny coincidence where you post about your new workout routine and suddenly find yourself inundated with ads for new exercise equipment. Whatever the case may be, this is not that fleeting remark about how tech giants are looming over our shoulders. Neither is this a technical dissection of what data is. Instead, this is a dispassionate and simplified analysis of the *other* side of data that lies behind the concerns of many being raised today. Read on knowing that this analysis is truly one half of the topic of data, not out of bias but because this side of the topic has yet to be critically conveyed in a manner that is digestible enough for consumers of all backgrounds. The goal of this essay is to bridge that gap, providing an explanation to the rationale of such concerns, and informing you of the role your data plays in today's digital climate. This essay will not assume value judgements or conclude what is taking place is right or wrong; the question of ethics and legality is up to you. If, however, in the end, you are motivated to make a change, some potential options will be listed at the end of this essay. Lastly, if you have nothing to hide or do not use mainstream social media, don't click away, you may find more relevance in this essay than you might have expected.

To examine the role your data plays in today's climate we must first establish a rudimentary understanding of what such data is. This essay will primarily be discussing data in terms of information technology at the surface level. What won't be discussed is raw data (such as the binary 1's and 0's television depicts hackers are so interested in), as the technical side of this topic is less crucial to the purpose of this essay. In general, data is the digital footprint you leave behind every time you Tweet, text, call, search, and so on. More specifically, in the context of collection and tracking, data is divided into two main categories: data and metadata. Using an email as an example, data may consist of the actual content of the email, such as what you typed. Alternatively, metadata may consist of when the email was sent, who sent and received it, what device it was sent from and where, but not the actual content. There is much contention about whether metadata is less intrusive to privacy than data is, as it is often described as data about data, but as will soon be discovered, there is little which can't be uncovered within metadata. With this basic understanding established, unless otherwise stated, the word data will be used in its colloquial manor as an umbrella term.

So, what is the concern with data? To begin, an identification of some of the ways your data is potentially unsecure is in order. The truth is that data is harvested on a scale which makes it virtually inescapable. Starting from the ground up, many phone carriers that provide cell service like Verizon, T-Mobile, and AT&T, collect the data users are emitting (McAuliffe). While it is becoming more common for carriers to allow consumers to opt out of such practices, the implicit consent of consumers' usage often gives these companies the green light. Additionally, because the wireless emission of data is dependent upon service providers, any attempt to limit data collection within a device is futile, as all data must first pass through their filter. Cell providers aren't alone in this either. Internet service providers (ISPs) also participate, meaning that Wi-Fi based devices (computers, tablets, etc.) are also being screened (Howard). Picture these services as the first checkpoint your data passes through.

Moving on to what happens within your device, many apps and services may also be collecting and storing your data, composing a second checkpoint. In 2018, one Oxford University study examined nearly one-million free Android apps, revealing that a majority of the applications contained utilities from Alphabet (Google's parent company), Facebook, Twitter, Microsoft, Verizon, and Amazon, which enabled the transmission of user data to these companies ("Mobile Apps"). Android users aren't alone, another study found nearly two-thirds of the apps they examined on the Apple store collected at least some form of data (Klosowski). Moreover, you might find some of the apps participating to be unsuspecting. The same study also found that weather apps were some of the most commonly data driven software available, demonstrating the fact that everyone is interested in consumer data (Klosowski).

Not only may your data be collected, but it may also be tracked via your device ID. A device ID is a unique and specific strand of numbers that identifies each device within the digital world. Think about your device ID like the VIN number for your car. Using this ID, companies may track mobile activity across websites, applications, and services (Klosowski). This is one of the reasons you might get ads for new footwear on Instagram after searching for shoe stores near you on Safari or Google Chrome. To top it all off, one study has shown that some preinstalled apps on Android phones are collecting and transmitting user data as well, regardless of whether the app is in use (Sekhose). These apps come straight out of the box with the phone, often times embedded in the device's read-only memory (ROM), meaning they cannot be deleted (Sekhose).

Now, a third and distinct checkpoint of data in this examination is web browsing. What makes browsing the internet so unique is that one application (Bing, Google Chrome, Safari, etc.) can open the door to a virtually infinite realm of software. This means that while you may only be intending to expose yourself to one company's database, you are likely sharing your data with every site you visit. Web browsing may be its own topic, but its role in this analysis is vital, nonetheless. One survey found that 79% of websites use trackers to collect user data (Crawford). Like mobile apps, not only may these sites be collecting your data, but they are likely tracking your activity as well.

One form of internet tracking many of us are familiar with is cookies. If you have ever wondered what these actually are, this essay is for you. Cookies are small segments of data that websites store on a user's device, essentially creating a link between their device and the site (Crawford). The duration these cookies are active can span the length of time the site is open, and potentially even years after the original visit, depending on the application (Crawford). This is just one of the ways companies may be able to track your activity from site to site. Another way websites may follow your digital trail is via fingerprinting. Rather than utilizing the surface information of a user's online activity, fingerprinting creates a "profile" of the device's properties, such as what operating system the user is on, what browser version they are running, what language their device is set in and so on (Crawford). This method may allow your device to not solely be tracked but also identified within the sea of other users. Lastly, web beacons are yet another potential tracking function, acting as small and hidden tags which follow how you interact with your content (Crawford). A common use for web beacons is in email marketing, providing companies with information such as when and if a user opened an email, as well as how many times they have accessed it (Crawford).

Service providers, applications, websites… these are only the tip of the iceberg, and not even its entirety. Each of these fronts are a rabbit hole in their own right and deserve equal attention independently. However, that is not the goal of this essay. The goal of this essay is to understand the other side of your data and the role it plays in today's climate, and the picture all of this should be painting is that your data is not *yours*. The checkpoints that have been examined compose a dragnet that may be used to fish your data in massive quantities. Not only may your data be gathered in bulk, but it may also be analyzed and sold as well. However, before moving forward, it is necessary to stop and reevaluate what your data is.

At this point in the process, some of you may be thinking, why does my data matter anyways? The truth is, not every bite of information you leave behind does matter. Furthermore, much of the data you share via your devices is actually very beneficial. The free and democratized system of digital interaction that has been created is fundamentally dependent upon the collection of users' data. Think about Google, when have you ever paid to use their search engine? Nonetheless, Google is a billion-dollar company. The exchange is your data and attention, both of which are lucrative assets. A popular adage used by economist goes "there is no such thing as a free lunch." Consider the possibility that if you aren't paying for the product, you *are* the product. However, the concern today has to do with the other, untold side of your data that is coming to light. So let us uncover what this "other" is.

The missing piece to the puzzle is personal data. The data that is being collected today may be revealing more about you than ever before, composing a category of digital information aptly named personal data. Often, under the pretext that it is less intrusive to privacy, metadata is becoming the primary source of information companies look to for personal data. Unraveling this falsification is the first step in recognizing the significance of data today. Returning to the earlier definition, metadata is described as data about data. However, it is becoming clear that such a definition is drastically oversimplified and ignores the sensitive nature of the information which can be extracted from within it. One of the reasons metadata is considered more secure is because it allegedly lacks personally identifiable information (PII), making it an anonymous source of data. However, numerous studies have shown that tracing metadata back to an individual can be a relatively trivial task. In one study, researchers were able to retrace and identify individuals' phone numbers from metadata with up to an 80% accuracy (Mayer et al.).

Once your metadata is traced back to you, much more sensitive information can be inferred about you. Using the dataset acquired by retracing telephone metadata as a launching point, researchers in the same study were able to predict, with a fairly strong accuracy, the relationship status, religious affiliation, health status, and home location of various users (Mayer et al.). The extent to which researchers were able to extrapolate this information was even more surprising. In one of the cases, researchers were able to predict and later confirm an individual had purchased a firearm (Mayer et al.). In another case, researchers were able to infer that one woman had an abortion, examining her call record which showed a call to her sister one morning, several calls to a Planned Parenthood clinic two days later, additional, brief calls to the clinic two weeks later, and one follow up about a month after (Mayer et al.). Again, all of this was from the more conservatively labeled metadata.

Alternatively, an article by the *New York Times* examined mobile devices and the location data that is emitted. The article raises questions about how "anonymous" data can be when it is a product of users' every move. Reportedly, one company revealed to the *Times* that it tracked individuals' movements accurate to within just yards and updated it more than 14,000 times a day (Valentino-DeVries et al.). The article later follows the location of a device which presumably belongs to a child, spending time at a playground just outside of a middle school in New York around 8a.m., before entering the school and remaining there until 3pm (Valentino-DeVries et al.). This device was one of more than 40 at that particular school (Valentino-DeVries et al.). As technology continues to develop, the ability to screen data and make inferences such as the ones above is becoming increasingly efficient. We have only scratched the surface of what your data may tell, and all it takes is the subscription to a commercial database to begin.

So, if your data is so sensitive, why might companies be collecting it? What might they be doing with such information? As alarming as some might find the scope of data harvesting to be, the truth is that on paper, it isn't as creepy as it may seem. The primary motive for many of these companies is simply the monetization of user data. Your data is today's digital gold mine. One study estimated that the U.S. data market in 2021 was valued at approximately $50.1 billion, with expectations for rapid growth ("Global Big Data"). Consequently, the monetization of data means that not only may the companies you interact with be harvesting your data, but many others might be following it as well. To elaborate, your information may be explicitly and consensually collected by the company you are directly interacting with (first-party), but then subsequently sold to a plethora of marketing firms, data brokers, and other businesses (third-party), composing an entire other segment of your data's lifecycle (Klais).

To put things into perspective, a recent study conducted on Apple's app store found that roughly 80% of the domains an app contacted on average were to third-parties (Klais). This statistic raises concern about what occurs in the background of users' devices, as users are unlikely to know who their data is being sold to, or what data is being sold as well (Klais). Generally, the data that is being sold today is analyzed for patterns to better understand consumer behavior. Understanding consumer behavior allows businesses to tailor to specific audiences that may be more likely to generate revenue.

Picture it this way; you are the new business owner of a clothing company. In a world of almost 8 billion people, surely there are some who are willing to buy your product. If it was possible for you to know who was in the market for clothing, shared the style you are curating, and lived within a profitable distance to sell to, you could increase your profits exponentially. This is the approach many businesses today are taking, categorizing users based on a curated profile of their accumulated data. The information personal data reveals allow businesses to reduce the friction of matching buyers with sellers, creating a more efficient marketplace.

Now, while the logic to this methodology may be sound, the implications of creating an entire market based upon the personal information and behavior of millions of people can be problematic. Let us examine some of the concerns that are being raised about your data today. The sale of consumer data may be the motive for the data industry, but what it has enabled is a highly specific, detailed analysis of our everyday lives, made widely available with little regulation. In the U.S., there is no federal institution, law, or amendment *specifically* dedicated to regulating digital privacy. Instead, a patchwork of laws and institutions have been tasked with providing ancillary support for protecting your online privacy ("Internet Privacy"). Some of these overseers include the Federal Trade Commission (FTC), Electronic Communications

Privacy Act (ECPA), and the Fourth Amendment, however, none of the current systems in place explicitly regulate the collection, tracking, and sale of your personal data as a whole ("Internet Privacy").

Beyond the regulation of data itself, consumers may not be properly informed as to how their data is being handled. A study conducted by Pew Research Center found that 63% of American adults say they do not understand the laws and regulations that protect their data privacy (Auxier et al.). Furthermore, 81% of those adults reported they are asked to agree to the privacy policies (terms and conditions pop-ups) of businesses on a monthly basis, while only 9% read the terms and conditions every time (Auxier et al.). This lack of understanding has raised concerns over whether or not there is a power imbalance in terms of consumer data, with some arguing full consent is given when users click "accept" on the terms and conditions pop-ups, while others believe the legal terminology is too esoteric for everyday consumers to decipher.

With the increasing ability to understand user data, companies are learning how to leverage it to their advantage as well. One example comes from techno-sociologist Zeynep Tufekci, who revealed that advertisers, using a combination of user data and artificial intelligence algorithms, have the potential to target bipolar individuals who are about to enter the manic phase—a condition often associated with compulsive behavior, making the group a potentially lucrative audience for businesses to exploit (Tufekci). This is an extreme example of the potential of targeted ads, but it demonstrates the power of today's technology to utilize your personal data in ways it has never been used before.

On another end of the spectrum, federal, state, and local governance in America are utilizing user data. In a controversial act, police forces across the U.S. are beginning to turn to third-party companies to provide them with data-based algorithms which can allegedly predict where crime will happen, as well as who may commit it ("Case Study"). In California, the Fresno Police Department is using data from social media, property records, arrest reports and more to calculate "threat scores" of suspects ("Case Study"). Predictive policing, as it is termed, raises concerns about false targeting and the constitutional dilemma of assuming all are innocent until proven guilty. At the federal level, the National Security Agency (NSA), has been collecting data from major telecommunication companies including Facebook, Yahoo, AT&T, Verizon, and more, under a program titled PRISM (Glory). PRISM is a top-secret national security program that filters data to track potential terrorists and other criminals (Glory). Leaked in 2013, the program has since garnered immense backlash due to its seemingly unfettered access to the information of not just potential terrorists but innocent civilians as well (Glory).

The list can go on, but the point is clear. These examples are a cherry-picked selection which is intended to demonstrate the spectrum of implications mass data collection can have. At this point, you, the reader, may be anywhere on the continuum of privacy policy from full-fledged data display to Faraday cage in a basement. Wherever you may be, it is beneficial to know your options when it comes to preventing and reducing the collection of your data. Beginning in the order of which the "checkpoints" were presented, here are few ways to reduce and perhaps even prevent the harvest of your data. When it comes to cell carriers, you may not be able to fully stop the collection of your data, but you can often stop the sale of it to third-parties. Simply call your carrier and ask how you can opt out of the sale of your data or visit their website and search for a tab titled somewhere along the lines of "manage privacy settings" or "how your data is used," where you can modify your account preferences (McAuliffe).

As for internet service providers, similar steps can be taken, but an additional method of preventing data tracking is the use of virtual private networks (VPNs). VPNs are becoming one of the more common methods of preventing online data harvesting, acting as an intermediary between your device and the internet. Using a VPN, your wireless connection and data are first sent to a VPN provider's servers where they are encrypted (hidden), and then sent out to your ISP and subsequently the web (Hoffman). Some benefits of a VPN are the prevention of ISP data collection, tracking, and sale; the ability to bypass geographic limitations like governmental restrictions on content; and the concealment of your IP address (digital identifier for computers, much like your mobile device ID) from both ISPs and websites (Hoffman). Some popular VPNs include ExpressVPN, Surfshark, and NordVPN.

Preventing data collection by apps and services can be hit and miss, as your acceptance of those terms and agreements pop-ups is an explicit consent and often requirement of many companies. However, like many others, even if you can't prevent it entirely, you can often reduce the amount of data that is collected. A good place to start is by beginning with your device itself. Based on an article by the *New York Times*, iPhone users should go to "Settings > Privacy > Tracking, and then [disable] Allow Apps to Request to Track" (Klosowski). This feature automatically denies any *new* application's request to track you across various apps and services on your device. An additional option is to open specific apps within the home page of the Settings application, allowing you to fine tune app privileges on a case-by-case basis. For Android users, go to Settings > Privacy > Permission manager, where you will find a list of privileges that each of your applications have access to, and can alter them as you like. Alternatively, if you'd like to tailor the access of a specific app, you can go to Settings > Apps & notifications, and click on each app to view and alter its privileges.

Beyond the direct prevention of data harvesting, a good practice would also be to begin reading those terms and agreements pop-ups when you get them. That way you can make an informed decision about which apps and services you would like to use. If you are an Apple user, recent updates have introduced a “privacy nutrition label” on the app store, providing you with a detailed list of what trackers an app is potentially using, organized into three categories: data used to track you, data linked to you, and data not linked to you (Klosowski). This new feature may also be a useful tool in determining which apps you want to download. Lastly, sifting through and deleting unused and preinstalled apps may reduce the amount of trackers active on your device.

Looking to web browsing, there is one simple but major step you can take towards preventing data collection and tracking. The answer is simply changing which web browser you use. Today, there are a plethora of free and safe browsers available to download that prevent the operation of cookies, fingerprinting, and web beacons, among other features. Some of these browsers include Brave, Mozilla Firefox, and DuckDuckGo. Speaking from experience, DuckDuckGo conveniently provides a grade for each site you visit based on the evaluated privacy of the site, as well as lists each third-party company that tried to track you. Coupled with a VPN, the decision to use a privacy-oriented browser can drastically reduce your digital footprint on the internet, if not erase it entirely. Now, understandably, not everyone is willing or ready to abandon the convenience and reliability of major web browsers like Google Chrome and Safari. A viable alternative is to download an extension for your existing browser, giving a similar result as the aforementioned software. Some of these free extensions include Adblock Plus, DuckDuckGo, uBlock Origin, and Ghostery (O'driscoll). On top of downloading an extension, consider entering your web browser and changing your preferences to reduce the tracking and sale of your data as well.

Utilizing all of these methods in conjunction will provide you with the best possible chance of minimizing the collection and tracking of your data. However, managing your digital footprint doesn’t have to be an all or nothing dichotomy. Moreover, it would be incorrect to assume that enabling all of these measures does not come with any drawbacks. The reality is that the more of these steps you take, the more the initial functionality and convenience of your digital experience may diminish. For example, disabling internet cookies entirely may reduce the functionality of certain websites and sometimes outright break them. Or consider when you search for restaurants near you on your computer. The immediate results that pop up require the automatic sharing of your IP address—a utility that is potentially lost with the use of VPNs and privacy-oriented browsers. These are real factors that will need to be weighed if you choose to adopt these privacy measures.

When, in the future, you hear one of those remarks about how tech giants are looming over our shoulders you can now, at the very least, understand the rationale behind the comment. There is a billion-dollar data industry collecting intimate details about each of our lives on a massive scale, and there is little regulation currently in place to ensure it is being managed properly. Regardless of where you stand on the privacy debate over data, hopefully you can walk away from this essay with a greater understanding of the *other*, potentially alarming side of your data. Perhaps consider reevaluating your current digital footprint and whether adopting more rigorous privacy standards is the right decision for you. As big data is an emerging industry, stay informed about your privacy rights. Read those arduous terms and agreements pop-ups when you get them and take a moment to vet the services you use. Lastly, remember that data is not bad, it is how we use it that defines its meaning. With that definition in mind, what does today's data mean to you?

Works Cited

Auxier, Brooke, et al. "Americans' attitudes and experiences with privacy policies and laws." *Pew Research*, 15 November 2019, https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/.

“Case Study: Your Tweet Can and Will Be Used Against You.” *Privacy International*, 30 August 2017, https://privacyinternational.org/case-studies/745/case-study-your-tweet-can-and-will-be-used-against-you.

Crawford, Eliza. “Website Tracking: Why and How Do Websites Track You?” *Cookie Pro*, 16 November 2020, https://www.cookiepro.com/blog/website-tracking/#:~:text=The%20beacon%20is%20usually%20a,or%20information%20about%20the%20browser.

“Global Big Data Market to Reach $234.6 Billion by 2026.” *PR Newswire*, Global Industry Analysts, Inc., 29 June 2021, https://www.prnewswire.com/news-releases/global-big-data-market-to-reach-234-6-billion-by-2026--301322252.html.

Glory, Nwachukwu. “PRISM Program: Here is all you need to know about it.” Privacy Savvy, 14 December 2020, https://privacysavvy.com/security/spying/prism-program/.

Hoffman, Chris. “What Is a VPN, and Why Would I Need One?” *How To Geek*, 11 August 2021, https://www.howtogeek.com/133680/htg-explains-what-is-a-vpn/.

Howard, Ari. “Is your ISP tracking your personal data?” *Allconnect*, 16 December 2021, https://www.allconnect.com/blog/is-your-isp-tracking-your-personal-data.

“Internet privacy laws revealed - how your personal information is protected online.” *Thomson Reuters*, https://legal.thomsonreuters.com/en/insights/articles/how-your-personal-information-is-protected-online#:~:text=There%20is%20no%20single%20law,unfair%20or%20deceptive%20commercial%20practices. Accessed 28 March 2022.

Klais, Brian. “New Research Across 200 iOS Apps Hints that Surveillance Marketing is Still Going Strong.” *URL Genius*, 20 January 2022, https://app.urlgeni.us/blog/new-research-across-200-ios-apps-hints-surveillance-marketing-may-still-be-going-strong.

Klosowski, Thorin. “We Checked 250 iPhone Apps—This Is How They’re Tracking You.” *New York Times*, 6 May 2021, https://www.nytimes.com/wirecutter/blog/how-iphone-apps-track-you/.

Mayer, Jonathan, et al. “Evaluating the privacy properties of telephone metadata.” *PNAS*, 16 May 2016, https://www.pnas.org/doi/full/10.1073/pnas.1508081113.

Works Cited Continued

McAuliffe, Zachary. “Your Phone Carrier Is Selling Your Personal Data. Here's How to Tell It to Stop.” *CNET*, 17 December 2021, https://www.cnet.com/tech/services-and-software/your-phone-carrier-is-selling-your-personal-data-heres-how-to-tell-it-to-stop/.

O’driscoll, Aimee. “How to stop browser tracking: 5 free anti-tracking browser extensions.” *Comparitech*, 20 January 2021, https://www.comparitech.com/blog/vpn-privacy/free-anti-tracking-browser-extensions/.

Sekhose, Marcia. “Research reveals Android phones constantly send data from pre-installed apps even if they’ve never been used before.” *Business Insider*, 13 October 2021, https://www.businessinsider.in/tech/news/android-phones-constantly-send-data-from-pre-installed-apps-even-if-theyve-never-been-used-before/articleshow/86988368.cms.

Tufekci, Zeynep. “We're building a dystopia just to make people click on ads.” *Ted*, September 2017, https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads?referrer=playlist-who_s_watching_us&autoplay=true.

Valentino-DeVries, Jennifer, et al. “Your Apps Know Where You Were Last Night, and They’re Not Keeping It Secret.” *New York Times*, 10 December 2018, https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html?mtrref=undefined&gwh=743EFE5AE8B17C9A7F906DE92F6E0CC5&gwt=pay&assetType=PAYWALL.

“Your mobile apps are tracking you.” *Internet Health Report*, April 2019, https://internethealthreport.org/2019/your-mobile-apps-are-tracking-you/.