Who Transferred My Data?

Ethical Concerns in Data Transfer

Team C37
Archit Gupta
Jean-Christophe Santens
Weijing (Ellen) Lu
Zifan Liu
Zhengyi (Andrew) Zhang

11/25/19

Abstract

The past decade has witnessed an increased awareness regarding unwarranted flow of data within various organizations. But still these organizations are able to get away legally because of lax regulations, without making amends on ethical front. In this whitepaper, we raise one such grave ethical issue regarding organizations transferring data that includes personal information without taking consent from concerned individuals. It is a fundamental right to know how information is being put to use as companies are notorious for the way they use data for monetization purposes, while disguising their true intent by claiming to work for the 'public good'. It is a clear-cut case of deception and we try to debunk this by making use of two case studies and then provide recommendations – first, legally defining the ownership of data to clearly express the corresponding accountability and rights and then, building upon that idea by introducing the concept of treating data as a commodity. By the end of this whitepaper, we hope to make our user aware about the gravity of situation and provide viable solutions to put this grave matter to rest.

TABLE OF CONTENTS

I. INTRODUCTION

What is common among the so called 'FAANG' group (Facebook, Apple, Amazon, Netflix, Google) apart from the fact that all of them offer high performance technology stocks? Simple, each one of them is facing multi-million-dollar lawsuits, related to unjustified access and usage of information. It is true that data is a free-flowing entity, but organizations have taken this concept quite literally as they don't feel the need to intimate the data subject, individuals whose information is captured, before making use of data and the overall intent behind it. Companies are right in legal sense (because of lax regulations, a point that we'll discuss in the later sections) but fail miserably when ethicality of their actions is taken into consideration.

The objective of this white paper is to make the reader aware about the unfathomable ways through which their PII (Personal Identifiable Information) is travelling across shores of different organizations and being used, all without consent. Starting off in the Problem Statement section, we'll provide a background regarding the unethical ways through which organizations indulge in information sharing without notifying and/or taking consent from the individuals involved using real-life case studies. Moving on to the solution section, we will try to propose a clear distinction in data ownership by dividing ownership into 3 types and then concentrate on the scenario where data is wholly owned by an individual and provide the idea of treating data as commodity. Finally, we will tie back the problem statement with the proposed solution, thereby showcasing applicability of our approach and conclude with a brief wrap-up.

II. BACKGROUND & PROBLEM STATEMENT

Data transfer is inevitable with emerging data economy. However, because of vagueness in laws and regulations, many instances of data transfer are legally in place, but ethically questionable. One recent example of it is the Google's project 'Nightingale', which will be discussed in detail in subsequent section. For greater understanding, following is the way we define certain terms -

Data: "information in digital form that can be transmitted or processed".[1]

Data transfer: "the collection, replication, and transmission of large datasets from one organization or business unit to another". [2]

The core issue organizations are facing lies in the very definition of who owns the data and to what extent can they exchange/make use of it. Even the laws followed

across continents are not clear and consistent, which attracts a lot of confusion. As an example - General Data Protection Regulation (GDPR) in EU is more comprehensive in nature as it includes the likes of right to be informed in comparison to the outdated data regulations followed in the US. As a result, companies in the US are able to manipulate the system by tweaking the core definition of regulations and using it to suit their purpose.

This results in a very serious issue as organizations are prone to release and make use of data, majorly for monetary purposes. As law abiding citizens and equal shareholders in this ecosystem, it is in our best interest to be kept in the 'loop' which includes knowing in what form our information is being transferred (if at all), who is the intended recipient and what purpose will it serve. Knowing all this will provide answers to concerning problems at individualistic level namely-

- Is my information being put to use in a legalized manner?
- Is it bringing any value to an existing process or being used for greater human good?

Having knowledge about these fronts will establish faith and confidence among individuals. It will also be in the best interest for corporations to answer these questions as trust is a two-way street; if they act reasonably, so will the data generators i.e. the concerned individuals. This added faith can result in greater profitability for the companies as well because customers will choose to buy products and services offered by their 'trusted' sources. On the other hand, ignoring the importance of protecting data subject's sensitive information and transferring it without consent or notification will eventually lose people's trust, subsequently leading to significant monetary loss as is the case witnessed by global giants such as Facebook. Therefore, complying with the ethicality of free flow of data can potentially help information seekers (organizations) avoid causing potential harassment to data providers (individuals).

In order to fortify our position, we have included case studies involving business giants which have breached wall of trust of their user base by indulging in ethically wrong activities -

1. Data Transfer of Personal Health Records: Google's Foray in Healthcare

Google has been an implicit leader when it boils down to innovation and standards for more than a decade now. Other organizations tend to follow its lead, even when the deed is not so ethical in nature. Of late, Google has been questioned by various governments and in turn, people regarding its unfair tactics to acquire and make use of highly sensitized data. Latest in the series of acquisition is highly personal health records from a company called Ascension under pretext of a project codenamed 'Nightingale'[3].

In order to create a strong hold in the highly lucrative healthcare industry, all the major tech companies are treading on an expansionist pathway, acquiring and making deals with different health care companies. In the wake of it, Google made a deal with Ascension, a popular hospital chain, to transfer all its patient records on Google Cloud platform. To give some background, health records are labelled as PII and thus, come under HIPAA regulation which directs for de-identification of data fields before sharing. Under the same law, record keeping organization (Ascension in this case) cannot share this data related to patients with its business partner (Google), unless the data is used for helping record keeper to carry out its health care function. Google and Ascension maintain that they fulfill the later condition but for some reason entirely known to them, didn't de-identify the data which encompasses lab results, doctor diagnosis and hospitalization records. The most frightening thing above all is that everything happened without the consent of stakeholders (patients in this case), that too when the projected intent behind this act is 'public good'. After this incident, people have started questioning the true intention behind it as both organizations treaded on an unethical path.

This case highlights how organizations take benefit of the free-flowing nature of data as well as lax regulations, which lets them use data without even notifying the concerned party. It is a direct violation of human rights as the data owner has a right to know where his/ her data is being put to use, how any people/entities have access to it and what is the intended use of it. By solving this issue, organizations can engage in using data whichever way they deem fit. Along with that, they can save ample amount of time spent in fighting lawsuits accusing them of unlawful and unethical usage of data.

2. Data Transfer During Merger & Acquisition(M&A): Walmart's acquisition of Jet.com

An alternative scenario of indirectly collecting data from the individual or entity does exist. Through the acquisition of a company, the acquirer can, under current legislation, get its hands on the datasets that the acquired firm owns without any restrictions. Consequently, it happens whether the data subjects (individuals whose data is captured) like it or not. The question we should ask ourselves is whether from an ethical point of view, this should be allowed or not.

The acquisition of Jet.com by Walmart for $3.3 billion in 2016 is a well-documented example of this case[4]. The main reason for the deal is that to this day, retail behemoth Walmart is still looking for ways how to compete with its online counterpart Amazon. Walmart is struggling to reach the urban millennial consumer, the exact type of consumers Jet.com and Amazon (in a partial manner) are targeting. The acquisition make sense from a strategic perspective – Walmart would get information about their target customers and use its online retail knowhow, while Jet would benefit from Walmart's supply chain infrastructure. To a certain extent, the purchase paid off. While Jet.com isn't doing well, the product offering of Walmart.com surged and e-commerce sales grew by 40% in 2018.

However, whether Walmart's investment in Jet.com a success story is not the point we're trying to make. It is the ease with which corporations can obtain personal identifiable information (PII) through acquiring the proprietor of such data that should raise questions regarding ethicality. Even if a system is implemented wherein companies must obtain consent from the individuals of whom they are trying to obtain the PII of, we are not accounting for the case when the same company is acquired by another company and in the process, the PII gets transferred too. We as users, may be comfortable in giving information to the former company but may act skeptical when the same information is transferred to the acquiring organization. It is a case of unethical behavior as the acquiring company have an intent which may or may not be acceptable to the user whose information is being transferred.

The solution we propose will solve the ethical front of the problem and will help organizations take decisions without any worry or fear of doing wrong to the society.

IV. RECOMMENDED SOLUTION

The controversy between legal and ethical aspect of data transfer is largely due to the ambiguous definition of data ownership - both data subjects (individuals) and data

collectors (organizations) treat data as their own property. Data subjects believe that PII is their private property and they should have the right to determine the extent of data usage regardless of which entity collects and processes it. It is evident in both case studies as even though the organizations took shoddy legal route, they were wrong on the ethical front by not intimating the concerned user. Therefore, in order to settle this conflict, there is an urgent need for a clear and concise data ownership definition.

The basis of ownership depends on the type of data one is taking into consideration. Broadly, data could be categorized into three types: PII, factual data and fiction data (Figure 1)[5]. We suggest using a continuum-based approach depending upon the level of personal information present in the dataset, which directly corresponds to the risk of identification. If the data contains highly sensitized information which can essentially be used to trace back an individual, then the ownership should shift towards data subject.

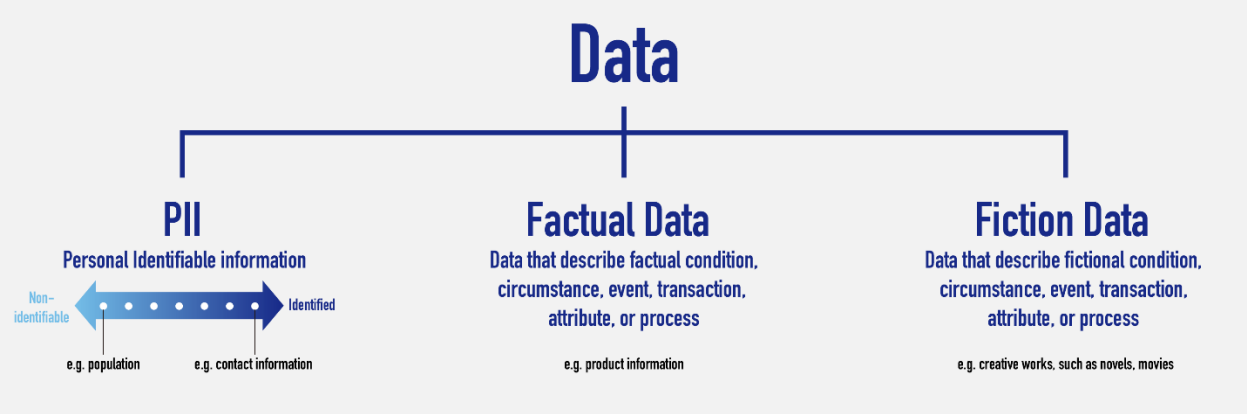

Figure 1. Categories of Data

For wide applicability of this paper, rather than using standard definition of PII in EU or in the US, PII is represented by a continuum, suggested by Schwartz and Solove [6]. On one end of the continuum, the risk of identification is very low, even though the data is collected from individuals. The example of such PII could be population of a certain state collected during census. On the other end of the continuum, the data is highly identifiable or identified, such as the data that includes name, date of birth or contact information.

On the other hand, if the data includes combination of both identifiable PII and factual data, such as the data present in Walmart case, then the ownership should be shared by both data subject (individual) and data collector (Jet.com in the above case in particular). Finally, if the data is completely generated using fictional information, the data generating entity would be the data owner.

Depending on the type of data explained above, a spectrum showcasing ownership can be created:

1) owned mainly by data subject

2) owned by both data subject and data collector

3) owned mainly by data collector.

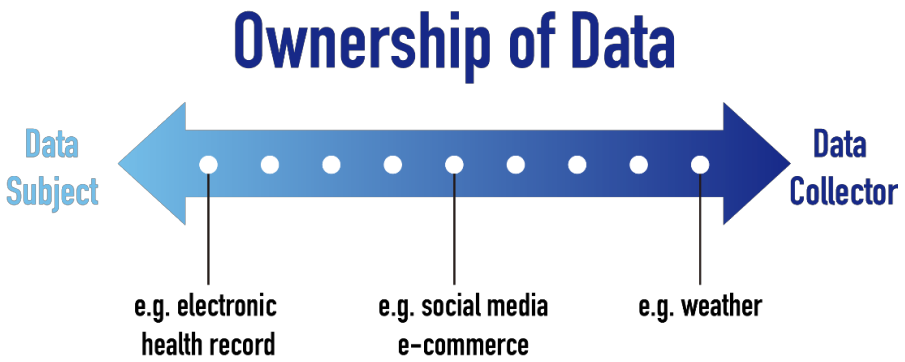

Figure 2. Ownership of Data

Data as a Commodity

Under our suggested definition, if the data is characterized as identified/highly identifiable PII, its sole ownership lies in the hands of the concerned individual. As a result, the individual 'owns' the data and so, the data collector cannot use the data without consent. We propose the idea of treating data as a commodity which essentially means that once the data subject legally owns the data, it should have the ability to 'sell' the data, including the recipient they want to sell to. According to Professor Tonetti from Stanford university, giving ownership of data to the concerned individual is also giving back the control of privacy to his hands[7]. That way, it would also depend upon the person whether he wants to sell the data or simply donate it for a humane cause.

But the concern lies in the fact that how would such an exchange take place, for which we introduce a concept called 'Data Exchange'. It would essentially work in a similar manner as a stock exchange works – Just as value of a stock is determined by its historic run and on the current events, the price of PII can be industry specific. Taking today's scenario into consideration, a PII representing healthcare might able to fetch more money than a PII representing music preferences. Taking further this concept, companies will act as buyers as they vie to gain the precious pieces of information from sellers (individuals in this case). Whichever company bids the highest for PII would finally get hold of that data. Also, the price of an information may depend on its inherent nature – whether the information is for one-time use only or reusable. If

it's the former case, value of PII may be traded at a higher rate as compared to the later one.

Treating data in such a manner can bring respite to all major problems related to data usage as then the user will relinquish his/her right over the piece of information, which subsequently won't obligate companies to share their plans with the general public. It's a win-win situation as data owners will get an acceptable value for their piece of information and companies won't be held wrong from both legal and ethical perspective.

Applicability in Real World

In the first case study involving Google & Ascension, the information in consideration should be wholly owned by an individual as it's highly sensitized in nature, thus coming under the umbrella term of 'PII'. As a result, it should be left to the owner's discretion whether he wants to sell his information or not and if he wants to, then whether to Google or not. Because Google is bound to monetize from that information (however much it claims to do public welfare), it can provide a reasonably priced compensation and therefore, save itself from the obligation of informing people about its intent. This way, both the information provider and information seeker can be satisfied and data flow can take place without any constraint.

In the second case study involving Walmart & Jet.com, as the information can be termed as 'Factual', the ownership is jointly shared by both the individual whose information is collected and the Data Collector (Jet.com in this case). In a hypothetical scenario, Walmart should have had offered a reasonable compensation to the Jet.com users whose data was part of the transaction, when it acquired the website. Had the user accepted the offer, then it should have been retained by Walmart, otherwise should be removed on ethical grounds.

V. CONCLUSION

It has been long since companies have taken advantage of ambiguities (and resulting loopholes) present in the prescribed data usage laws. The objective of this paper has been to illuminate the reader about such ethically wrong practices involving data usage without consent. Through this literature, we have tried to provide a new direction in which regulations may take shape by providing a clear definition of data ownership and recommending solution involving 'Selling of Data' to remove the core issue of consulting before usage. It would essentially eliminate the concern behind

usage of data for profiting purposes as the data owner would willfully relinquish the control over his/her data, which would give the right to data acquiring company to make use of data as it please. There's no denying of fact that going down this path would present its own set of difficulties and complicacies, but it is certainly an achievable solution which can bring respite to this ever-growing problem of unethical usage of personal data.

Organizations hold the power to do unimaginable things with all the data around them, but that doesn't mean they should ignore the public well-being and misuse the power they have been bestowed upon. As part of the natural ecosystem, they have a moral obligation to protect the interests of their stakeholders. In conclusion, a quote that perfectly captures the sentiment - 'With great power comes great responsibility'.

REFERENCE

[1] Data. (n.d.). In Merriam-Webster's online dictionary. Retrieved from https://www.merriam-webster.com/dictionary/data

[2] "What is Data Transfer: Definition | Informatica"

[3] "Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans"

[4] "Walmart's $3.3 billion acquisition of Jet.com is still the foundation on which all of its e-commerce dreams are built"

[5] Ritter, J., & Mayer, A. (2018). Regulating Data as Property: A New Construct for Moving Forward. *Duke Law & Technology Review, 16*, 220-277.

[6] Schwartz, P. M., & Solove, D. J. (2014). Reconciling Personal Information in the United States and European Union. *California Law Review, 102*, 877.

[7] Should Consumers Be Able to Sell Their Own Personal Data?