The Aftermath of The Snowden Leaks and Its Impact on Big Data: A Look at How Stakeholder Perspectives Influence a Technology's Application

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Felipe Barraza

May, 2021

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature Felipe Barraza Date 5/10/21
Felipe Barraza

Approved _____ Date _____
Sean T. Elliott, Department of Engineering and Society

The Aftermath of The Snowden Leaks and Its Impact on Big Data: A Look at How Stakeholder Perspectives Influence a Technology's Application

Edward Snowden sacrificed his entire life to do something he believed he was morally obligated to; in his mind, the consequences of his actions were outweighed by the benefits that would come from them. He proceeded to leak the most significant set of government information to the public that has ever been leaked (Chadwick & Collister, 2014). Whether his actions were productive or destructive depends on the individual. Citizens concerned about their privacy would deem Snowden a hero. Government officials whose work was sullied might consider Snowden a criminal, or worse, a terrorist. Telecommunications companies caught in the middle of this debacle might not even care about Snowden, and only concern themselves with maintaining their public image. These different perspectives and groups are fundamental in the development of new technologies and the rules that apply to them. This paper will utilize the Social Construction of Technology theory to dissect the case of Edward Snowden and apply those findings to the emerging field of data science, which is beginning to wrestle with ethical questions much like those present in the Snowden vs. NSA case.

Before delving into specific details about Edward Snowden's dilemma, it is necessary to understand the relationships between society and technology. Many frameworks exist to analyze this relationship, but for the purpose of this paper, the Social Construction of Technology (SCOT) theory will be applied. Scholars use this framework to identify the importance of structural concepts to an understanding of the social shaping of technology (Klein & Kleinman, 2002). That is to say, scientists believe that technology is shaped by human actions, and the application of said technology can only be understood by considering the context in which is it used. To better articulate this idea, social scientists use four core concepts that comprise the

SCOT framework: interpretative flexibility, relevant social groups, design flexibility, and lastly, problems and conflicts.

The first concept is interpretative flexibility, the idea that technological artifacts have unique meanings and applications for a variety of groups (Pinch & Bijker, 1984). An example of this is an automobile. Some individuals' livelihoods revolve around driving one all the time, and thus view automobiles as tools to make ends meet. On the other hand, others may see them as prized possessions that require the utmost care, and should only be used for special occasions. These different views of a given technology's application can be simplified into groups.

This group-by-group investment is the second concept of SCOT; relevant social groups are the embodiment of the unique meanings and applications of a technology (Klein & Kleinman, 2002). As such, technologies are constantly influenced by a variety of groups with different needs and expectations. In the case of the automobiles, there are groups like truckers and postal workers that utilize such devices in their careers, and thus interpret the technology as a tool. On the other hand, there exist automobile collectors that are content by merely owning a vehicle they deem valuable. To this group, an automobile is a luxury or a trophy. Just by comparing the two groups presented, a clear difference in their desires and needs can be seen. Balancing these needs is necessary for a technology's success, but oftentimes, appealing to all groups is impossible due to conflicting ideologies between them. Therefore, a technological idea must be malleable before it is concretely introduced into society. This malleability allows the technology to undergo improvements and adapt to the stakeholders' needs before it is finalized.

This idea of multiple designs is central to the third concept of SCOT. There will be conflicts between the different interpretations of a given technology, so continuously adjusting the design until all conflicts subside is the only way for the technology to be fully welcomed

(Klein & Kleinman, 2002). This concept is formally known as closure and stabilization (Klein & Kleinman, 2002). The automobile perfectly encapsulates this concept, since there are a multitude of vehicle classes catered to unique interpretations. Those who need a sturdy machine capable of carrying heavy objects can employ the might of a pickup truck. Those who desire a sleek design and exhilarating ride will enjoy super cars. Different interpretations of a technology allow stakeholder groups to benefit without having to compromise their needs and desires for that of another group. Beyond groups of individuals, technologies can also be influenced by abstract concepts present in society.

This wider context is the premise of the fourth and final concept of SCOT. Wider context includes all the other sociocultural and political milieu in which development takes place (Klein & Kleinman, 2002). This can include the relationships between sets of groups, the backgrounds of individuals groups, the rules that dictate how groups respond, and the social norms of how groups voice their opinions. An example of this would be the introduction of electric automobiles. There are groups like oil companies, that are opposed to electric automobiles; and there are groups like environmentalists that believe this new technology is the only acceptable interpretation. Between these groups there exists a power dynamic that affects the influence each has over other scientists, and by extension, the likelihood of a technology's implementation. Understanding how social groups and technologies interact can help scientists predict the reaction to a newly introduced technology. Therefore, it is the responsibility of its creators to be familiar with the field to which they wish to contribute to. Neglecting these responsibilities could lead to failure to uphold ethical responsibilities and could negatively impact society.

It is imperative then to apply SCOT to the study of data science, an ever-growing field that is the combination of computer science and statistics. This emergent field of work brings

with it a new set of ethical considerations, especially in relation to the subject matter, Big Data. Big Data encompasses any and all structured and unstructured information collected, stored, linked, and analyzed either online or offline (Chen & Quan-Haase, 2020). This large amount of information carries a proportionately large amount of value with it, and as such, ethical considerations arise when determining the best use of this data. However, there is yet no concrete set of rules for the proper use of Big Data. The development of a code of conduct is an on-going process. Addressing data ethics is an integral part of the work, but considerations must be made since Big Data and its meaning are socially constructed and influenced by social, political, and technological forces (Chen & Quan-Haase, 2020). The issue of working with Big Data is that it has introduced more complexities rather than solving issues; instead of providing objectivity, accuracy, veracity, and inclusivity, it has introduced biases, subjectivities, and forms of oppression (Chen & Quan-Haase, 2020). These issues need to be addressed through the scope of the SCOT framework because Big Data's nature is concerned with what is being processed, and who it's being processed for (Chen & Quan-Haase, 2020).

Much like any other technology, Big Data is relevant to a large set of social groups. Big Data projects could be developed for individuals, organizations, clients, or international governments (O'Leary, 2016). These various groups could be generalized into three different categories of stakeholders: collectors, utilizers, and generators (Zwitter, 2014). Big Data collectors determine which data is collected and for how long it is stored. This group also affects the utility of data (Zwitter, 2014). Big Data utilizers redefine the purpose for which data is used, and thus interact with big Data collectors (Zwitter, 2014). Big Data generators are separate from the other two groups since they are the source of the Big Data; these generators can vary from natural phenomena, to programs, to sentient actors such as humans (Zwitter, 2014). This group

generates data by performing their respective tasks, but in the case of humans, this group may not be explicitly aware of their production. Because of this, ethical challenges arise around the use of the Big Data generated by the other groups of stakeholders.

This conflict of ethical considerations between stakeholders can be seen in the case of Edward Snowden and his former employers, the National Security Agency (NSA). In short, Edward Snowden was alarmed at the collection of Big Data that the NSA had amassed, which included personal and demographical information up to the private conversations of citizens. He felt perturbed that one group had access to so much information from another group, especially when this data was being collected without the knowledge of the public. Given that the NSA is a government agency, whose purpose is to serve and protect its citizens, Snowden felt it was important to alert the other stakeholders of the Big Data being collected so that they may voice their opinions of such a technology being used that way.

This case is a prime example of the dynamic between Big Data utilizers and the rest of the stakeholders. The general public, the generator of the data, has its privacy breached by the NSA, the utilizers of the data, when they intervene in everyday mechanisms like phone calls and mail services. The providers for those services, like telecommunications companies or the postal service, are thrust into controversy because of an interloper that was exploiting their business. In the case of Snowden v. NSA, SCOT becomes especially useful in that it is capable of dissecting the situation into groups and interpretations, and seeing how this event influences the development of new technologies that involve Big Data. It is important to consider, then, the main stakeholders in this case, as they embody the general stakeholder groups of Big Data technologies.

Edward Joseph Snowden had humble beginnings, being born in North Carolina and joining the army as a young adult. He was discharged from service after suffering physical injuries and started working for the NSA as a security guard at one of their facilities. He eventually worked his way up into a position at the Central Intelligence Agency (CIA), and later transferred to a private contractor of the NSA (Verble, 2014). It was at this position that he managed to acquire access to classified government documents. Between his time at the CIA and the nature of the data contained in those documents, Snowden became disillusioned with the government he was serving and how their actions impact the world (Verble, 2014). So much so that he sacrificed his six-figure salary, his comfortable life in Hawaii and risked the ire of the United States of America, arguably one of the most influential countries in the world- all because he wanted to do what he believed was his moral responsibility (Verble, 2014).

Immediately after Snowden released 1.7 million documents of secret data to a slew of news organizations, the US government responded by revoking his passport and aggressively deterring a number of foreign countries from granting his asylum (Scheuerman, 2014). This essentially made Snowden a global target, something he had foreseen. "[He] not only predicted that he would be accused of illegality, espionage, and even treason, but also that the US government would marshal its imposing power resources to discredit and severely punish him" (Scheuerman, 2014). Despite all these difficulties, Snowden acted on what he believed was ethically correct. His own values and responsibilities dictated his course of action, and through a consequentialist perspective, he determined his sacrifice to be for the good of all mankind.

In contrast to Snowden's selflessness exists the NSA, an organization whose history is less than ethically sound. Starting as a code and cipher decryption unit in World War I, the NSA was later reformed into an intelligence centerpiece of the US Military, guiding troops in several

international conflicts (Verble, 2014). Despite honorable beginnings, the NSA later found itself in hot water, earning both public and government scrutiny (Verble, 2014). The first instance of public humiliation occurred in 1975, when the NSA's "Watch List" was brought to light. This list, originally intended to be for the protection of the American public, was found to be comprised of U.S. citizens who were thought of poorly by intelligence agencies. At first, the individuals on this list were communist sympathizers and threats to the president, but the list later turned into a roster of anti-war activists during the Johnson era, and then expanded to include characters deemed dubious by other government agencies (*National Security Agency Tracking of U.S. Citizens – "Questionable Practices" from 1960s & 1970s*, 2017). The negative perception of such a device furthered when it was discovered that telecommunications companies were sharing their customers' information with the NSA. This exact information is what facilitated the construction of the NSA Watch List, since it contained information like a customer's name, location, and outgoing call destination.

A similar situation occurred after the terrorist attacks on September 11th, 2001. The United States government gave permission to the NSA to monitor its citizens without the need of a permit, something prohibited by the Foreign Intelligence Surveillance Act of 1978, the direct result of the previous controversy. This meant that the NSA could again freely monitor citizens, even when they were not communicating with foreign nationals. Technology had advanced greatly since the times of the Watch List, and now the NSA had access to much more information thanks to computers and internet services, like email. Although this revelation did not result in a new civil protection act, the public was left violated by its own government – a feeling that is undoubtedly difficult to forgive and forget.

These two previous controversies are based on the NSA's seemingly limitless access to the public's personal information; therefore, the situation followed by the Snowden leaks should come as no surprise. The NSA has a history of overextending its authority into the privacy of individuals, both American and foreign. However, it is the agency's purpose to protect its citizens, and as such, it has evolved into an ethically questionable organization that will cross any boundary to achieve its goals. To this end, there exist some morally redeemable qualities in its actions, since the organization is acting out only to serve its purpose and secure safety for the population.

Despite diametrically opposed viewpoints of these stakeholders, there are additional groups that influenced the Snowden vs. NSA case. This controversy was first presented by journalists from various news sources. This group of stakeholders held an unprecedented amount of power in the situation since they could control the dominant narratives of the story while exploiting the networks it generated (Chadwick & Collister, 2014). These journalists could act as both the collectors and the utilizers of Big Data (Chadwick & Collister, 2014). While not explicitly clear, this control of the media presents ethical challenges. Absolute control of information is what Snowden wanted the NSA to avoid, but by providing so much knowledge to journalists and then purposefully removing himself from the picture (Verble, 2014), he was permitting the media to exercise absolute control over the topic. Likewise, it is difficult to ascertain to what end these journalists truly cared about advocating for personal privacy. It could have been quite likely they were merely interested in reaping the financial benefits of a momentous journalistic piece.

Another group that was involved in this case is the telecommunication agencies that provided the data to the NSA. These companies are primarily collectors of data and likely have

no ill intentions but they can still be compelled into acting unethically. This was probably the case, since the data the NSA was using information from phone calls and internet records- data that is normally collected by these companies. In terms of ethics, this group of stakeholders was acting out of legal duty, abiding to the laws of the country in which they reside. A government intelligence agency was requesting to use their data, so it could be said that the repercussions of denying them access to the data would cause more problems than it was worth. Refusal to do so could be seen as withholding information, and possibly even interference in judicial proceedings. Despite the honorable implications turning over information has; this could also be seen as an act of self-interest. These companies serve citizens by granting them access to a variety of services, and as such, should consider themselves providers for the public. Therefore, suddenly releasing personal information of clients could be interpreted as a violation of the client-to-provider contract they are meant to uphold.

It can be seen then, from the description of the relevant stakeholders in Snowden v. NSA, that balancing interests is a near-impossible task when creating a new technology. Each group has precise values and objectives, and would prefer to see technologies adhere to its idealized use of Big Data. Individuals that agree with Edward Snowden would champion for privacy, and condone the use of algorithms that rely on Big Data to function. Security agencies like the NSA need Big Data to function, since their security systems revolve around knowing everything they can about an individual. To them, this collection of information is the key to successfully accomplishing their purpose. In terms of journalists and service providers, their opinions on Big Data are clouded. Privacy is greatly favored, and at the individual level these groups likely agree with Edward Snowden. But when it comes to business, these two stakeholder groups might air on the side of (ab)using Big Data, since it can drive more traffic to their services, and thus,

greatly boost revenue. Journalists can benefit from search algorithms that recommend articles to people based on their information, and telecommunications agencies can sell customer information to Big Data compilers, like advertisement agencies.

To what end these stakeholders acted out in ethical compliance we will never know. All that can be concretely analyzed is the actions committed by each of these groups and the repercussions that they had. These actions and consequences should be examined through an ethical scope to determine the validity of their execution. It is here that the SCOT theory becomes most useful; it exposes the values and actions of each stakeholder group and evaluates them in the context of whatever sociocultural landscape they exist in. Every group was acting in its own self-interest, yet these actions impact society as a whole and as such need to be held accountable.

Scientists should inform themselves of this controversy when attempting to create new technologies for data science. At its core, any project dealing with Big Data will have the same stakeholder groups as the Snowden v. NSA case. There will exist corporations seeking to benefit from stolen information, businesses stuck in the middle, and individuals concerned about their privacy. Failure to mitigate conflicts and compromise on the best interpretation will invoke the last stakeholder group, the media. This group has the strangest relation to technology, due to its freedom to construct whatever narrative it desires around the situation, and thus influencing the success of an emergent technology.

In terms of the use of Big Data, scientists must also concern themselves of the sociopolitical environment in which they reside. These environments are in constant fluctuation; stakeholders could agree in the present, but at any moment their values and opinions might drastically diverge and cause difficulties for scientists. An example of this is how a small

company will initially agree with the public on privacy regulations, but as it grows into a global entity, its concern over privacy could diminish in favor of profits and control of the market. These relationships between groups are also influenced by the demographical region and history. Scientists in authoritarian nations could experience pressure from the governments to construct one-sided technologies that facilitate totalistic control; whereas, capitalistic societies likely care little for the nuances of the technology so long as profits are maximized. Regardless of the sociopolitical landscape, stakeholder groups are constantly interacting. As such, scientists must seek the best possible way of pacifying, or at least not disturbing, the relationships between these groups when introducing a new technology.

The presentation of the Edward Snowden case serves to articulate the ethical complexities that exist in fields containing Big Data, and how scientists must understand the dynamics between the stakeholder groups when designing a technology. There will always exist stakeholders with conflicting desires and perspectives, and it is the responsibility of data scientists to maximize benefit for all relevant social groups while maintaining ethical integrity. Through the presentation of the Snowden v. NSA case, and its subsequent analysis through the use of SCOT, data scientists are equipped with the foundations for the ethical considerations of any technology they develop. New challenges and frameworks are yet to develop, due to the field's infancy; but it is paramount that researchers, social scientists, and data scientists remain aware of all the influences present and keep every group accountable for its requests and actions. There is no need for espionage, treason, and extradition in the development of new technologies. This process should serve all without the need for extreme sacrifices and punishments.

References

Chadwick, A., & Collister, S. (2014). Boundary-Drawing Power and the Renewal of Professional News Organizations: The Case of The Guardian and the Edward Snowden National Security Agency Leak. *International Journal of Communication (19328036)*, *8*, 2420–2441.

Chen, W., & Quan-Haase, A. (2020). Big Data Ethics and Politics: Toward New Understandings. *Social Science Computer Review*, *38*(1), 3–9. https://doi.org/10.1177/0894439318810734

Klein, H. K., & Kleinman, D. L. (2002). The Social Construction of Technology: Structural Considerations. *Science, Technology, & Human Values*, *27*(1), 28–52.

National Security Agency Tracking of U.S. Citizens – "Questionable Practices" from 1960s & 1970s. (2017, September 20). National Security Archive. https://nsarchive.gwu.edu/briefing-book/cybervault-intelligence-nuclear-vault/2017-09-25/national-security-agency-tracking-us

O'Leary, D. E. (2016). Ethics for Big Data and Analytics. *IEEE Intelligent Systems*, *31*(4), 81–84. https://doi.org/10.1109/MIS.2016.70

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441.

Scheuerman, W. E. (2014). Whistleblowing as civil disobedience: The case of Edward Snowden. *Philosophy & Social Criticism*, *40*(7), 609–628. https://doi.org/10.1177/0191453714537263

Verble, J. (2014). The NSA and Edward Snowden: Surveillance in the 21st century. *ACM SIGCAS Computers and Society*, *44*(3), 14–20. https://doi.org/10.1145/2684097.2684101

Zwitter, A. (2014). Big Data ethics. *Big Data & Society*, *1*(2), 2053951714559253. https://doi.org/10.1177/2053951714559253