Journal of the Midwest Association for Information Systems

Special Issue
Information Security, Privacy, and Ethics
Editor: David P. Biros

Table of Contents

Volume 2020, Issue 2, July 2020 **www.JMWAIS.org**

Contact

Daniel J. Power, Ph.D.
Professor of Management and Information Systems
Department of Management
University of Northern Iowa
Cedar Falls, IA 50613
(319)273-2987
daniel.power@uni.edu

Rassule Hadidi, Dean
College of Management
Metropolitan State University
Minneapolis, MN 55403-1897
(612)659-7295
rassule.hadidi@metrostate.edu

Journal of the Midwest Association for Information Systems (JMWAIS) at http://jmwais.org is a double-blind, peer-reviewed, quality focused, and open-access online journal published by the Midwest United States Association for Information Systems at http://www.mwais.org/. The infrastructure for online publication of this journal is currently provided by the Metropolitan State University, St. Paul, Minnesota.

Journal of the Midwest Association for Information Systems

Table of Contents

Editorial Board

Journal of the Midwest Association for Information Systems

Volume 2020 | Issue 2 Article 1

Editor's Comments

Date: 07-31-2020

The Challenges of New Information Technology on Security, Privacy and Ethics

David P. Biros
Oklahoma State University, andyluse@okstate.edu

Abstract

The rapid rate of growth and change in Information technology continues to be a challenge for those in the information sector. New technologies such as the Internet of Things (IoT) and wearables, big data analytics, and artificial intelligence (AI) are developing so rapidly that information security and privacy professionals are struggling to keep up. Government and industry call for more cybersecurity professionals and the news media make it clear that the number of cybersecurity breaches and incidents continues to rise. This short article exams some of the challenges with the new technologies and how they are vulnerable to exploitation. In order to keep pace, information security education, ethics, governance and privacy controls must adapt. Unfortunately, as history shows us, they are slow to evolve, much slower than the technologies we hope to secure. The 2020s will usher in vast advancements in technology. More attention needs to be given to anticipating the vulnerabilities associated with that technology and the strategies for mitigating them.

Keywords: Internet of Things (IoT), big data analytics, Artificial Intelligence (AI), security ethics, privacy, risk

DOI: 10.17705/3jmwa.000057

1. Introduction

I have been working in information security for nearly 30 years. In the fall of 1991, I was a young Captain in the US Air Force and just beginning a Master's degree program in Information Resource Management at the Air Force Institute of Technology (AFIT). I had just purchased a new PC with a whopping 1 MB RAM and a 40 MB hard drive and was setting it up to connect to AFIT with my 2400 baud modem, when a friend called to tell me about a computer virus going around. He said it was called the Michelangelo virus and it affected master boot record of your computer if it was infected. The virus would activate on Michelangelo's birthday (March 6). It was then that I decided to learn more about information security or as we called it in the military, information assurance. Over the years, I was part of a military information security inspection (audit) team, served a DoD task force for complying with the Federal Information Security Management Act (FISMA, 2020), drafted Air Force information assurance policy, served as the Chief Information Security Officer (CISO) for the Air Force CIO, taught dozens of information security and risk assessment coursed, and researched and published articles on security vulnerabilities.

One thing is certain. The domains of information security and privacy are evolving and doing so rapidly. Technologies are changing and more data is being collected than ever before. We used to count viruses in the thousands. Now we count them in the millions. We used to connect to the Internet with a home computer using dial-up connections for a limited period of time. Now everything connects to the Internet all the time (i.e. Internet of Things). Data storage used to be expensive, relatively speaking, and policy makers and developers limited how much data was collected and how many lines of code were written. Now, there is more data collected than ever before and that data is used in ways never imagined (big data analytics, training artificial intelligence).

Yet with all the advances in technology, our information security risk continues to grow. Everything connected to the Internet is a potential target and, in turn, a security vulnerability. Extremely large data sets hold massive amounts of sensitive and valuable information that are vulnerable to loss. Artificial Intelligence (AI) is poised to change the world again but has its own vulnerabilities (Biros, et al., 2019). Not only that, there are also implications for privacy, ethics, and education both at home and in the workplace. Fortunately, the authors of the papers in this special issue (introduced later) have done a great job at investigating the changes to these all-important facets of information security. First, a closer look at the aforementioned technologies is in order.

2. Rapid Growth of IoT and Wearables

Probably the greatest source of new information security vulnerabilities and loss of privacy is with IoT and wearables. Oberländer et al (2018), define IoT as "connectivity of physical objects equipped with sensors and actuators to the internet via data communication technologies." We have smart speakers, doorbells with cameras connected to our wifi; smart watches to tell us how much to walk and how well we slept; autonomous vehicles, and AI systems that monitor our driving, vet applicants for jobs, and prevent us from jaywalking (Sharma and Biros, 2019; Biros, et al, 2019). Further, as two of our authors in this issue note, businesses also use IoT for performance monitoring. In many cases, it is not the technology itself, but the user's inability to properly configure either the device or their home wifi system. While some earlier, high-profile attacks were directed toward Internet cameras and baby monitors (Wang 2018), in 2019 information security breaches due to IoT devices as starting point for deeper networking hacks became a major concern (Buntz, 2019). In short, the addition of IoT to our world has led to millions of new targets to be exploited.

A major subset of the IoT is wearables or internet connected devices that can be incorporated into clothing or worn as an accessory. Apple watch and Fitbit devices are common examples. They have the ability to monitor heart rate, calories burned, steps or mile walked or ran, and sleep patterns and quality. Most often the devices are paired with apps that collect and analyze the data before reporting results back to their owners. It is estimated that by next year 378.8 million devices will be in use in the US alone. (Statista 2019). As noted, the data is collected and analyzed for pre-defined purposes, but it can also have unintended consequences. This is not to say wearable are bad. On the contrary, there are a number of reports of wearables alerting users of previously undiagnosed and potentially fatal heart conditions (Reisinger 2018), however there have been serious cases of wearable data used for purposes that the manufacturer never intended. In a recent report, military troops using wearables and a fitness-oriented social networking app that track users running routes compromised the locations of some secret US military bases (Miller 2018). According to Hsu (2018), a popular fitness app company, Strava, used heat maps to graphically represent the locations of runners and others

exercising. While it was expected that densely populated areas would show high heat, some areas thought to be uninhabited depicted patterns of heat. In turn, these areas were correctly surmised to be secret military bases in remote parts of the world. The massive amounts of data generated by IoT and wearables can be analyzed for both positive and negative purposes.

3. Big Data Analytics

IoT and wearables make up only part of the data collection picture. All toll, in 2018, about 2.5 quintillion bytes of data were created every day (Marr 2018). This included data produced by Internet searches, social media, communications, digital photos, services, and IoT and wearables. All of this data is analyzed to produce meaningful information impacting many areas of society (Gupta, et al., 2018). It has been used in health care, firm performance, society activities and more. Big data analytics can help diagnose medical conditions, analyze comorbidities in health care, and help to optimize traffic flow.

With all the potential capability some very serious concerns remain, one of them being privacy implications (Jenson, 2013). Almost anyone who has ever used Amazon knows it targets individual customers based on their purchasing history. Searches on Google result in advertisements that reflect the subject of a recent search. All of the data collected can be sold or transferred without the consumer's knowledge or consent (or awareness of consent). The world of big data analytics continues to eat way at the privacy rights of individuals. While there are some laws such as the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act of 2018, there continues to be significant privacy challenges and implications. In some cases, such as AI, even developers and analysts do not fully understand how the data will be used and interpreted.

4. Artificial Intelligence

Another technology worth noting is AI. It is a rapidly going field with many applications. Organizations are using AI to help select potential employees, build autonomously driving cars, monitor prescription drug doses and prevent credit card fraud (Biros, et al., 2019). AI tools are trained by very large data sets to recognized patterns and trends and act on them. It has the potential to reduce production costs, cut time to accomplish tasks (e.g. sifting through thousands of resumes) and help enforce laws and statutes. The technology requires big data to adequately train systems that use decision trees, neural networks and other advancements to make decisions on behalf of human decision makers.

Like the other technologies mentioned AI has its challenges as well. As Biros and colleagues (2019) demonstrated, AI projects suffer from problem fit, input data issues, and application problems. For example, Amazon created a recruiting tool to help it recruit and hire more women. However, the data used to train the tool came from the previous ten years of resumes from male-dominated tech companies. As a result, the tool showed a bias against women. In Ningbo, China an AI application designed to catch jaywalkers erroneously cited a well-known business woman for jaywalking because her picture was on the side of bus in the street (problem fit). Also, a company is Israel developed a face recognition program they claimed could identify pedophiles and terrorists by analyzing their physical traits. Multiple scientists have questioned the validity of the tool and the ethical implications enormous. If such technology gets into the wrong hands or if it should be hacked and manipulated, the results could be disastrous.

5. Summary and Recommendations

The three major technologies mentioned above have the potential to help us live longer and healthier lives, understand more about health care issues and production optimization, and reduce the number of tedious tasks we must perform. However, we cannot overlook the potential they have for security, privacy and ethical implications that come with them. They bring with them vulnerabilities of too much data sharing, aggregating massive amounts of data in one location, and lulling us into relying on them for all of decision-making tasks. It is important that we investigate the privacy concerns and ethical consideration of such technology. Fortunately, the papers included in this special issue do just that.

6. Overview of the contents of this issue

This issue contains four articles on information security, privacy and ethics.

Luse, A and Burkman J. investigate the use of RFID wearables in the context of a corporate environment using privacy boundary research. Their findings show that while being monitored negatively impacts employee satisfaction, greater

transparency in implementation may alleviate some of the negative aspects of implanting such technologies in the workplace.

Young J., Smith T., and Zheng, S. extend Mason's information ethics framework of privacy, accuracy, property and accessibility (PAPA) to capture some of today's new technology considerations focused around big data. Their extension includes the concepts of behavioral surveillance, governance and privacy.

Maunula, G. demonstrates that the analysis of sharing economy processing activities uncovers potential privacy, security and data protection concerns related to a platform's disclosure of personal data to end-users. This has considerable implications for compliance with the GDPR and she posit that is correction requires a multi-disciplinary approach.

Weiser M. and Bowman A. round out the articles with a content analysis of the leading information security textbooks as compared to the needs of employers with respect to the skillset of the university graduates. The results of study found that coverage of terms associated with security knowledge areas demanded by the marketplace requires attention if schools are going to turn out well qualified information security knowledgeable graduates.

7. References

Biros, D., Sharma, M., and Biros, J., (2019) Vulnerability and risk mitigation in AI and machine learning" *Cutter Business Technology Journal*, 32 (8)

Buntz, B. (2019) A year in review: 12 IoT security considerations. *IoT World Today*, Retrieved from. https://www.iotworldtoday.com/2019/08/15/a-year-in-review-12-iot-security-considerations/

Gupta, A., Deokar, A., Iyer, L. et al. (2018) Big data and analytics for societal impact: Recent research and trends. *Information Systems Frontiers*, 20**,** 185–194.

Hsu, J. (2018) The Strava heat map and the end of secrets. *Wired*. Retrieved from https://www.wired.com/story/strava-heat-map-military-bases-fitness-trackers-privacy/

Jensen M. (2013) Challenges of privacy protection in big data analytics, *2013 IEEE International Congress on Big Data*, pp. 235-238

Marr, B. (2018) How much data do we create every day? The amazing stats everyone should know. *Forbes.* Retrieved from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#7fa5414360ba

Oberländer, A. M., Röglinger, M., Rosemann, M., and Kees, A. J. E. J. o. I. S. (2018). Conceptualizing business-to thing interactions–a aociomaterial perspective on the Internet of Things," (27:4), pp. 486-502.

Reisinger, D. (2018). Apple Watch Credited with Saving a Man's Life." Fortune. Retrieved from http://fortune.com/2018/05/03/applewatch-saves-life/

Sharma M. and Biros, D. (2019) Building trust in wearables for health behavior *Journal of the Midwest Association of Information Systems*, (2019:2)

Statista (2019). Wearable device unit sales worldwide by region from 2015 to 2021 (in Millions). Retrieved from https://www.statista.com/statistics/490231/wearable-devices-worldwide-by-region/

Taylor, G. L (2018). Pentagon reviewing troops' use of fitness trackers in light of security concerns, *The Wall Street Journal*.

Wang, A. (2018) I'm in your baby's room: A hacker took over a baby monitor and broadcast threats, parent say" *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2018/12/20/nest-cam-baby-monitor-hacked-kidnap-threat-came-device-parents-say/

Author Biography

Dr. David Biros is an Associate Professor of Management Science and Information Systems and Fleming Chair of Information Technology Management at Oklahoma State University. A retired Lieutenant Colonel of the United States Air Force, Dr. Biros' last assignment was as Chief, Information Assurance Officer for the AF-CIO. His research interests included deception detection, insider threat, information system trust and ethics in information technology. He has been published MIS Quarterly, the Journal of Management Information Systems, Decision Support Systems, Group Decision and Negotiation, MISQ Executive, the Journal of Digital Forensics Security and Law and other journals and conference proceedings.

This page intentionally left blank

Journal of the Midwest Association for Information Systems

Volume 2020 | Issue 2 Article 2

Date: 7-31-2020

Wearables in the Workplace: Examination Using a Privacy Boundary Model

Andy Luse
Oklahoma State University, andyluse@okstate.edu

Jim Burkman
Oklahoma State University, jim.burkman@okstate.edu

Abstract

Wearable technologies have become a popular consumer product for health, entertainment, etc., but the use of such wearables in the workplace is still somewhat new. Wearables also offer the potential to provide benefits for both employer and employee in the workplace but the implementation of such technologies creates privacy implications that may affect worker attitudes. Wearable types can take many forms but this study focuses on RFID wearables due to their low cost, proven durability and reusability (Zhu & Hou, 2020). This research investigates the use of RFID wearables in the context of a corporate environment. Utilizing privacy boundary research, findings show that while being monitored negatively impacts employee satisfaction, this satisfaction further varies based on the voluntary nature of the implementation and the gender of the employee. Findings suggest that greater transparency in implementation may alleviate some of the negative aspects of implanting such technologies in the workplace.

Keywords: privacy, wearable, gender, monitoring, voluntary

DOI: 10.17705/3jmwa.000058

1. Introduction

This research investigates the implementation of RFID wearable technologies within the workplace in relation to privacy. Radio Frequency Identification (RFID) has been utilized by corporations to provide business advantages for several years. These devices can be used for access control, inventory management, supply chain management, etc. (Wu, Nystrom, Lin, & Yu, 2006) and as a method to track assets and aggregate data to minimize human intervention and reduce costs (Asif, 2005). While RFID has been used to assist in certain business processes for many years, RFID can also be used to improve the experiences of employees in the workplace or help manage and monitor employee activities utilizing RFID wearables (Brady, 2018). This would generate benefits (both tangible and intangible) for the business (Wu et al., 2006) by implementing RFID to alter employee behavior by enabling enforcement of compliance with certain business processes (Kim & Garrison, 2010; Staats, Dai, Hofmann, & Milkman, 2017).

While wearables may provide benefits to both the business and employees, the impact on and attitudes of the employees regarding the implementation may differ due to privacy issues. These employee attitudes may negatively impact the satisfaction of employees with the implementation and the workplace. Previous research suggests that privacy is a social process whereby individuals interact with others implementing a regulation process depending on the individual's identified social group, those outside this group, and the control of the user over situational aspects (Altman, 1975; Palen & Dourish, 2003). Computer performance monitoring (CPM) systems used to measure, record, store, and compile data on the activities of employees (Schleifer & Shell, 1992) may improve employee performance but this increased monitoring runs the risk decreasing employee satisfaction due to a loss of privacy (Chalykoff & Kochan, 1989).

This research helps to contribute to the overarching question of what are the boundary conditions in relation to privacy in the area of wearable tracking in the workplace. We extend Palen's (2003) research on privacy boundaries given its applicability to the area of privacy and boundary conditions. We extend the theory by further explicating the identity boundary through division into both the *self* and the *other* boundary conditions. We then utilize an experiment to investigate the three boundary conditions of disclosure, self, and other on the satisfaction of employees with the technology. The research attempts to better understand the three interacting forces of control, privacy, and self within the context of wearable technology. Our findings show that the satisfaction of users with the wearable technology is affected by a combination of all three variables, providing usable findings in the workplace for such implementations.

2. Background

Privacy research has spanned a variety of contexts over a wide range of topics for several decades. With regard to individual privacy, Altman's privacy regulation theory is highly regarded as a seminal piece in social psychology (1975, 1977). Privacy regulation theory aims to understand individual privacy intentions within a larger social context rather than through a dichotomous lens. The theory posits that privacy is not just about the individual avoiding social situations (Palen & Dourish, 2003), but is a dynamic process of controlling access to one's self or socially identified group (Altman, 1975). Overall, Altman hypothesizes that an individual's desired level of privacy changes over time in response to the specific environment.

While privacy is not a new area for research, the impact of new information technologies on privacy is still fairly novel and fluid due to the rapid pace of technological change. Altman's research was analyzed within the context of information technologies by Palen and Dourish (Palen & Dourish, 2003). In their work, they extended Altman's idea of the changing nature of privacy per environmental circumstance to define three specific boundaries that effect privacy with the environmental context: the disclosure boundary, the identity boundary, and the temporal boundary. The disclosure boundary deals with the selective disclosure of personal information by individuals to others and is most closely associated with the traditional view of privacy. The identity boundary deals with the boundary set between self and other, specifically within the individual's social groups and affiliations. Lastly, the temporal boundary deals with differences in privacy over time where past actions impact current actions with regard to privacy. These boundaries can differ substantially from the historical understanding of privacy due to the impact of information and communication technologies in the environment (Petronio, 2002).

For this research we investigate and adapt two of the boundaries in Palen's model: the disclosure and identity boundaries. The third, temporal, boundary is likely relevant and applicable to understanding the impact of new information technologies on privacy. The challenges of capturing longitudinal data at this early exploratory stage have been foregone in this study in order to use the expedient methodology of written scenarios and survey questions. The disclosure boundary is utilized in line with Palen's original definition. Conversely, the identity boundary is expanded. Palen described the identity boundary as "self vs. other" and is therefore the only boundary condition within the model

consisting of two different subsections. Many research streams have looked at the delineation between self and others. Ecologists study the interaction of separate organisms within an environment (Malmstrom, 2012). Personality psychology investigates differences between separate individuals (Friedman & Schustack, 2016). Social psychology, while looking at individuals within social environments, still delineate an individual from others (Myers, 2012). For this reason, we separated the identity boundary to include two separate boundary conditions: self and other. Figure 1 displays the research model as used in this research.

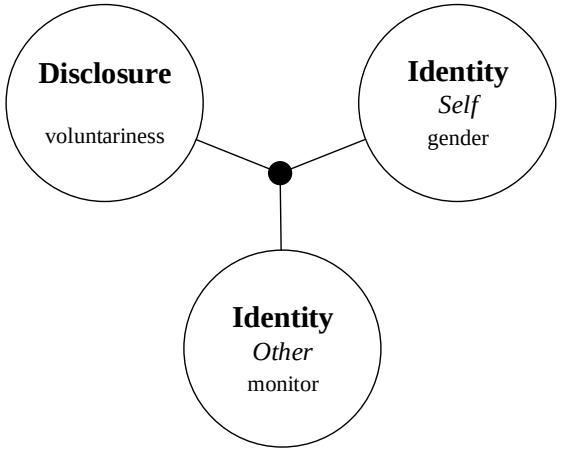


Figure 1. Research model with operationalizations.

The three boundaries in the proposed theoretical model are operationalized in this research in the context of electronic performance monitoring using RFID wearable devices. Specifically, the *other* involves the organizational use of these devices for monitoring. *Self* is defined as a social construction with respect to the individual and their perceived group. For this research we employ the socially constructed gender grouping of the individual (Myers, 2012) utilizing two self-identified facets, namely male and female. *Disclosure* pertains to the voluntariness of the wearable use. The interaction of these three items is explained below.

First, the *other* identity boundary refers to the environmental context within which privacy is assessed by the individual. If the environment is specifically monitoring individuals, this would make the boundary between self and other more apparent and would increase tension between the other and those in the environment (Palen & Dourish, 2003). As noted by McNall and Roch (2007), Electronic Performance Monitoring (EPM) raises privacy concerns that can differ by the type of EPM. Direct types of EPM surveillance such as an RFID wearable have a larger impact on employee feelings of privacy invasion than indirect methods such as computer logging; however, any use of EPM leads to some significant degree of employee perceptions of privacy invasion. They suggest that this could be due to participants experiencing a loss of perceived control that then adds to the feelings of privacy invasion (Eddy, Stone, & Stone-Romero, 1999). Yost et al. (2019) also point out that privacy entails autonomy, or control over one's personal information and utilize psychological reactance theory (Brehm, 1966) as a lens to explain the negative behaviors that arise from loss of that control. They find that perceptions of privacy invasion are positively related to state reactance, an aversive motivational state. This set of research points to a negative impact when the environment (other) imposes privacy reducing measures. Given that employee perceptions of autonomy and control have repeatedly shown to affect job satisfaction (Chung-Yan, 2010; Ganster & Fusilier, 1989; Hackman & Oldham, 1975, 1980; Spector, 1986) we expect that:

H1: Individuals have higher job satisfaction when not being monitored as compared to monitored regardless of the mandated nature of the use or gender.

Conversely, when individuals are not being monitored, the tension of the other in the environment is not as high and the boundary between self and other lessens (Palen & Dourish, 2003). Given this decrease in tension from the other, disclosure would have a less salient impact on an individual's privacy decisions. Furthermore, it has been established that there are no significant differences in job satisfaction between men and women in the workplace when all other environmental conditions are held constant (Mason, 1995). Therefore, given the lack of monitoring from the other, we hypothesize:

H2: When not being monitored, there is no difference in job satisfaction between men and women regardless of the mandated nature of the use.

As noted, the impact of *other* on both *disclosure* and *self* is hypothesized to have an overarching effect such that both *disclosure* and *self* are impacted. The disclosure boundary within Palen's model implies a degree of choice with regard to the individual's decision to disclose information, but a problem occurs when the level of volition is lessened or complexly stricken. Palen notes that "problems emerge when participation in the networked world is not deliberate, or… not within one's total control" (2003, p. 4). Mandated use of a technology engenders a loss of personal control with regard to technology use. Furthermore, Venkatesh et al. (2000) conducted a longitudinal study of individual technology adoption specifically looking at gender differences in decision making. Their study found the impact of perceived behavioral control on intention to adopt and use a new technology to be significantly moderated by gender. Specifically, women's perceptions of their level of control is far more relevant than those of men in the context of adopting a new technology. This loss of control is exacerbated when they know they are being monitored which, as discussed earlier, leads to some sense of privacy invasion that when combined with the feeling of control loss, may drive negative affect even further. This leads to a confounding effect of both the loss of privacy when being monitored and a loss of control when use is mandatory that is different between men and women. Therefore:

H3: When being monitored, the level of job satisfaction for men and women differs based on the mandated nature of the use.

H3a: When being monitored, there is no difference in job satisfaction for men whether the use is mandatory of voluntary.

H3b: When being monitored, there is a difference in job satisfaction for women depending on the mandated nature of the use such that women have higher job satisfaction for voluntary use as compared to mandated use.

Given the intersection of the three boundaries in the model and the above arguments, the overarching hypothesis for this research is a three-way interaction on individual job satisfaction with RFID wearable tracking that will depend on 1) the monitored nature of the environment (other), 2) the voluntary nature of the implementation (disclosure), and 3) the self-identified gender of the subject (self). The delineated hypotheses discuss the simple main effects and simple-simple main effects to paint the specific interaction effects, yet the above hypotheses are all dependent on an overarching significant three-way interaction.

3. Data Collection

Subjects were solicited from two sections of a core undergraduate business course at a large Midwestern university. The course was a required course for all majors in the college. Students were offered extra credit by the course instructor for participating in the research. The course instructor was the same for all sections of the course and did not discuss the specifics of the study with the students. RFID technology was selected as the focal research artifact due to its familiarity with the subjects and clear disassociation from biometric monitoring devices like a smart watch, yet also novel as a personal tracking device, thereby presenting little preexisting bias. Also, the student sample was chosen as a conservative sample for the research. Given surveys showing less privacy concerns among Gen Z'ers as compared to other generations,[12] significant results in this study would provide more definitive proof that there would be privacy issues among other generations that have a higher degree of privacy concerns.

The experiment was a 2 (voluntary vs. mandatory) x 2 (not monitored vs. monitored) factorial design, with an added non-experimental independent variable of self-identified gender, making the entire study a 2x2x2 factorial study. Subjects for this experiment were randomly assigned to both levels of the experimentally implemented independent variables. The subjects were sent an email with a link to the survey that consisted of one of the four randomly-assigned scenarios and identical questions for each subject regardless of scenario. The scenarios consisted of a short paragraph

[1] https://www.f5.com/labs/articles/threat-intelligence/are-gen-z-ers-more-security-savvy-online-than-millennials-#:~:text=While%20the%20majority%20of%20respondents,86%25%20for%20Gen%20X).

[2] https://www.fintechbusiness.com/industry/1265-gen-z-value-personalisation-over-data-privacy

describing the individual working for a company that has decided to implement RFID wristband technology for employee use. From here each subject received one of four scenarios from the categories in Table 1.

Table 1. 2x2 experimental manipulations.

Voluntary Not Monitored	Voluntary Monitored
Mandatory Not Monitored	Mandatory Monitored

The mandatory-monitored scenario would require the employee to adopt the technology and its purpose would be to monitor the employee. A mandatory-not monitored scenario would also require the employee to adopt the technology, but the purpose of the technology would be to improve access control through the unlocking of doors, devices, etc. A voluntary-monitored scenario would give the employee the option to adopt, but the technology would be used to monitor the employee. Finally, a voluntary-not monitored scenario would give the employee the option to adopt, but the purpose of the technology would be to improve access control through the unlocking of doors, devices, etc. Following the scenario, subjects answered questions pertaining to their perceived job satisfaction adapted using previous research (Bowling & Hammond, 2008; Weiss, Dawis, & England, 1967).

4. Results

369 subjects participated in the study. The dependent variable of job satisfaction had a Cronbach alpha value of 0.85, showing good internal consistency. The three items that made up the dependent variable were then averaged to create a single dependent measure. The two experimental independent variables of voluntariness and monitoring and the third independent variable of self-identified gender were coded using 0 and 1 as per previous research (Luse, Townsend, & Mennecke, 2018). The sample also contained 49 percent females, providing a good proportion of men and women across the four experimental conditions.

To test the hypotheses, an ANOVA-based general linear model was run. The results showed a three-way interaction of mandated use, monitored use, and gender on job satisfaction ($F = 5.50$, $p = 0.02$). To examine the presence of an overall effect of monitoring regardless of the voluntary nature of the RFID use and the gender of the individual, simple-simple main effects were examined. Results found that job satisfaction was higher when individuals are not monitored for men when use is voluntary ($F = 35.78$, $p < 0.001$) or mandatory ($F = 38.75$, $p < 0.001$) and for women when use is voluntary ($F=21.91$, $p < 0.001$) or mandatory ($F = 95.15$, $p < 0.001$), thereby supporting H1. Furthermore, simple-simple main effects found that when not being monitored, there are no differences in job satisfaction for men versus women whether RFID use is voluntary ($F=0.26$, $p=0.614$) or mandatory ($F=2.88$, $p=0.090$), supporting H2. Conversely, when being monitored, the level of job satisfaction is dependent on both the voluntary nature of the RFID use and the gender of the individual ($F=5.95$, $p = 0.015$), supporting H3. Results (see Figure 2) show that while the job satisfaction of voluntary versus mandatory use is not significantly different for men when being monitored ($F=0.029$, $p=0.864$ – supporting H3a) (i.e. the solid lines are basically on top of one another) there is a significant difference in job satisfaction of voluntary versus mandatory use for women when being monitored ($F=13.05$, $p<0.001$) (the dashed lines) with women showing higher job satisfaction for voluntary use when being monitored, supporting H3b. Figure 2 graphically depicts the effects.

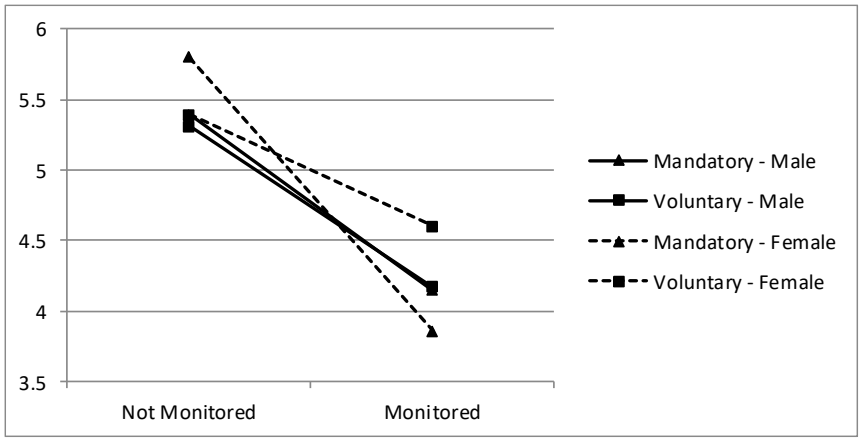


Figure 2. Graphical results.

5. Discussion

This research investigates the use of wearables within the organization. Results show that monitoring by others in the environment has a differential impact on *disclosure, self,* and *other* boundaries. Specifically, when not being monitored, both the impact of the voluntariness of use of wearables within the workplace and the gender of the subject have no combined or separate impact on individual satisfaction with the wearables. Conversely, when being monitored, individual satisfaction is differentially impacted by voluntariness of use and gender of the individual such that women are significantly less satisfied with mandatory use whereas men are not.

Theoretically, this research provides several important contributions. First, we adapt Palen's model of privacy boundaries (Palen & Dourish, 2003), by differentiating between the two opposing forces within the identity boundary yet providing an interacting link between all three. This separation in the model provides a delineation of the self from others that has been researched in various areas (Friedman & Schustack, 2016; Malmstrom, 2012; Myers, 2012). Furthermore, we provide a study operationalizing each of the three boundaries in the model and demonstrating that all three interact.

This research also provides valuable findings for corporations. While wearables might provide benefits for the company, employee satisfaction may vary. While using these technologies to monitor employees will lower employee satisfaction, the voluntary nature of the implementation can have differing impacts. Men may not be significantly affected by the voluntary nature but caution should be used for women when the monitoring is mandatory as this may adversely affect their satisfaction, thus leading to a negative impact on work morale. Given the tension in this boundary between the individual and the corporate other (Palen & Dourish, 2003), methods should be used to better convey the interaction of the corporation with the system and its overall use to allow employees to better understand the boundary and hopefully alleviate some qualms they may have with the system.

One avenue for further exploration is in the longitudinal aspect of this research. While we used and modified two boundaries within the boundary model of Palen, this research did not investigate the original model's temporal boundary (Palen & Dourish, 2003). Altman's original privacy regulation theory postulates that privacy levels change with the environment over time (Altman, 1975). Our model adds to the theoretical literature by more fully articulating the duality of the identity boundary, but this interaction may be impacted longitudinally. Future research should investigate the impact of temporal disparities in tandem with the developed model in this research.

6. Conclusion

This research investigates the impact of the use of wearables to monitor employees. Palen's privacy boundaries are adapted to better understand the interplay of *disclosure*, *self*-identity, and *other* identity on individual privacy attitudes. Specifically, we look at the impact of voluntariness, self-identified gender, and monitoring on employee satisfaction. Results show that while monitoring always has a negative effect on satisfaction, the voluntariness of the implementation differentially impacts men versus women with women having significantly less satisfaction when the monitoring is mandatory. The research provides an updated theoretical model to build on other privacy research. Practically, the results

provide organizations with guidelines for aiding in more successful implementation of monitoring systems.

7. References

Altman, I. (1975). *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. Monterey, CA: Brooks/Cole.

Altman, I. (1977). Privacy regulation: Culturally universal or culturally specific? *Journal of social issues, 33*(3), 66-84.

Asif, Z. (2005). Integrating the supply chain with RFID: A technical and business analysis. *Communications of the Association for Information Systems, 15*(1), 24.

Bowling, N. A., & Hammond, G. D. (2008). A meta-analytic examination of the construct validity of the Michigan Organizational Assessment Questionnaire Job Satisfaction Subscale. *Journal of Vocational Behavior, 73*(1), 63-77.

Brady, T. M. (2018). Wrist band haptic feedback system. In: Google Patents.

Brehm, J. W. (1966). *A theory of psychological reactance*: Academic Press.

Chalykoff, J., & Kochan, T. A. (1989). Computer-aided monitoring: Its influence on employee job satisfaction and turnover. *Personnel Psychology, 42*(4), 807-834.

Chung-Yan, G. A. (2010). The nonlinear effects of job complexity and autonomy on job satisfaction, turnover, and psychological well-being. *Journal of occupational health psychology, 15*(3), 237.

Eddy, E. R., Stone, D. L., & Stone-Romero, E. E. (1999). The effects of information management policies on reactions to human resource information systems: An integration of privacy and procedural justice perspectives. *Personnel Psychology, 52*(2), 335-358.

Friedman, H. S., & Schustack, M. W. (2016). *Personality: Classic theories and modern research*: Pearson.

Ganster, D. C., & Fusilier, M. R. (1989). Control in the workplace. In C. L. Cooper & T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 4, pp. 235-280). Chichester, England: Wiley.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied psychology, 60*(2), 159.

Hackman, J. R., & Oldham, G. R. (1980). *Work Redesign*. Reading, MA: Addison Wesley.

Kim, S., & Garrison, G. (2010). Understanding users' behaviors regarding supply chain technology: Determinants impacting the adoption and implementation of RFID technology in South Korea. *International Journal of Information Management, 30*(5), 388-398.

Luse, A., Townsend, A. M., & Mennecke, B. E. (2018). The blocking effect of preconceived bias. *Decision Support Systems, 108*, 25-33.

Malmstrom, C. (2012). Ecologists study the interactions of organisms and their environment. *Nat Edu. Knowledge, 3*, 88.

Mason, E. S. (1995). Gender Differences in Job Satisfaction. *The Journal of Social Psychology, 135*(2), 143-151. doi:10.1080/00224545.1995.9711417

McNall, L. A., & Roch, S. G. (2007). Effects of Electronic Monitoring Types on Perceptions of Procedural Justice, Interpersonal Justice, and Privacy 1. *Journal of Applied Social Psychology, 37*(3), 658-682.

Myers, D. G. (2012). *Social Psychology* (6th ed.). New York: McGraw-Hill.

Palen, L., & Dourish, P. (2003). *Unpacking" privacy" for a networked world.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Petronio, S. (2002). *Boundaries of privacy: Dialectics of disclosure*: Suny Press.

Schleifer, L., & Shell, R. (1992). A review and reappraisal of electronic performance monitoring, performance standards and stress allowances. *Applied Ergonomics, 23*(1), 49-53.

Spector, P. E. (1986). Perceived control by employees: A meta-analysis of studies concerning autonomy and participation at work. *Human relations, 39*(11), 1005-1016.

Staats, B. R., Dai, H., Hofmann, D., & Milkman, K. L. (2017). Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science, 63*(5), 1563-1585.

Venkatesh, V., Morris, M. G., & Ackerman, P. L. (2000). A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational behavior and human decision processes, 83*(1), 33-60.

Weiss, D. J., Dawis, R. V., & England, G. W. (1967). Manual for the Minnesota satisfaction questionnaire. *Minnesota studies in vocational rehabilitation*.

Wu, N.-C., Nystrom, M., Lin, T.-R., & Yu, H.-C. (2006). Challenges to global RFID adoption. *Technovation, 26*(12), 1317-1323.

Yost, A. B., Behrend, T. S., Howardson, G., Darrow, J. B., & Jensen, J. M. (2019). Reactance to electronic surveillance: a test of antecedents and outcomes. *Journal of Business and Psychology, 34*(1), 71-86.

Zhu, H., & Hou, M. (2020). Research on the Application of RFID in Equipment Management in Universities. In *Recent Trends in Intelligent Computing, Communication and Devices* (pp. 591-595): Springer.

8. Appendix - Scenarios

Mandatory-Monitored
The company has informed you that the technology will have the ability to track what you are doing at the company. The company will be able to know your location at all times while you are in the building. The company has also informed you that this technology is required to be used by all employees.

Mandatory-Not Monitored
The company has informed you that the technology will be used in order to improve security throughout the building. You will be able to scan into the building as well as unlock doors and devices that you are normally allowed to access. The company has also informed you that this technology is required to be used by all employees.

Voluntary-Monitored
The company has informed you that the technology will have the ability to track what you are doing at the company. The company will be able to know your location at all times while you are in the building. The company has also informed you that this technology is optional. You are under no obligation to adopt the technology.

Voluntary-Not Monitored
The company has informed you that the technology will be used in order to improve security throughout the building. You will be able to scan into the building as well as unlock doors and devices that you are normally allowed to access. The company has also informed you that this technology is optional. You are under no obligation to adopt the technology.

Author Biographies

Andy Luse received a B.A. degree in Computer Science from Simpson College, M.S. degrees in Information Assurance, Computer Engineering, Business Administration, and Psychology, and Ph.D. degrees in Human Computer Interaction, Computer Engineering, and Information Systems from Iowa State University. He is currently an Associate Professor in Management Science and Information Systems at Oklahoma State University. Andy's research has focused on computer security and research methods. He has been published in the *Journal of Management Information Systems*, *IEEE Transactions on Visualization and Computer Graphics*, *ACM Transactions on Computing Education*, *IEEE Transactions on Education*, *Decision Sciences Journal of Innovative Education*, *Computers and Human Behavior*, and many other outlets.

Jim Burkman received a B.A. degree in Business Administration from Western Colorado University and a Master's of Business and Ph.D. degree in Management Information Systems from Indiana University. He is currently an Associate Professor of Professional Practice in Management Science and Information Systems at Oklahoma State University. Jim's research has focused on behavioral aspects of information system use and security. He has been published in the *European Journal of Information Systems*, *Journal of the AIS*, *Statistics Education Research Journal*, *Journal of the Midwest Association for Information Systems* and other outlets.

This page intentionally left blank

Journal of the Midwest Association for Information Systems

Volume 2020 | Issue 2 Article 3

Date: 07-31-2020

Call Me BIG PAPA: An Extension of Mason's Information Ethics Framework to Big Data

Jacob A. Young
Bradley University, jayoung@fsmail.bradley.edu

Tyler J. Smith
Bradley University, tjsmith3@fsmail.bradley.edu

Shawn H. Zheng
Bradley University, hzheng@fsmail.bradley.edu

Abstract

In 1986, Richard Mason proposed the PAPA framework to address four ethical issues society would likely face in the information age: privacy, accuracy, property, and accessibility. In this paper, we propose an extension to the PAPA framework by appending three additional issues relevant to information ethics in the big data era. First, we outline the four components of Mason's original PAPA. Second, we briefly review the major technological changes that have occurred since Mason proposed his framework. Third, we outline concepts relevant to the big data context. Fourth, we propose and discuss our extension by appending three ethical issues related to behavioral surveillance, interpretation, and governance to Mason's original PAPA framework, forming BIG PAPA. Lastly, we discuss how these issues impact practice and how they can inform future research.

Keywords: big data, ethics, privacy, security

DOI: 10.17705/3jmwa.000059

1. Introduction

Just as society entered the information age, Richard Mason proposed his PAPA framework, comprised of *privacy*, *accuracy*, *property*, and *accessibility* (Mason, 1986). Mason's seminal paper provided a much-needed warning that has remained relevant as the foundation of information ethics for decades (Peslak, 2006, p. 117). Unfortunately, it appears much of Mason's message was largely ignored as news of massive security breaches, targeted marketing, and invasive tracking technologies litter today's headlines.

Although Mason's original arguments are just as appropriate today, we offer an extension in the context of big data. First, we outline the four components of Mason's original framework. Second, we briefly review the major technological changes that have occurred since Mason proposed PAPA. Third, we outline concepts relevant to the big data context to establish a contextual foundation. Fourth, we propose and discuss our extension by appending three ethical issues related to *behavioral surveillance*, *interpretation*, and *governance* to Mason's original PAPA framework, forming BIG PAPA. Lastly, we discuss how these issues impact practice and how they can inform future research.

2. Mason's PAPA Framework

Before discussing our extension, we must first ground our paper by outlining the key elements of Mason's framework. Although Mason recognized that several ethical issues would become increasingly important in the information age, he focused his discussion on four primary areas of concern: privacy, accuracy, property, and accessibility. These issues were organized into the PAPA acronym. Mason posed several key questions as he introduced an issue, which we have summarized in this section.

2.1. Privacy

> *What information about one's self or one's associations must a person reveal to others, under what conditions and with what safeguards? What things can people keep to themselves and not be forced to reveal to others? (Mason, 1986, p. 5).*

Mason provided two examples in his discussion on privacy. First, he recounted an effort by the legislature of the state of Florida to assess whether building codes were resulting in underutilized buildings. One study tasked a researcher with observing and recording the usage of toilets, mirrors, and sinks in bathroom facilities at Tallahassee Community College. As one would expect, the school's students, faculty, and staff argued that the study invaded their privacy and violated their rights. Instead of apologizing, the state claimed that the data being collected was more valuable than the harm it caused. Amazingly, it was not until the American Civil Liberties Union got involved that the state finally suspended the study. Of course, by that time the state had already collected enough information.

Mason then illustrated what he coined the *threat of exposure by minute description*. He noted that users regularly provide information to certain parties, but without consenting to that information being shared with others or merged into a central database. In his second example, Mason recounted an unsanctioned investigation that occurred when curious programmers at the city of Chicago's computer center began cross-referencing several databases based upon employee name and I.D. At first, they identified employees with unpaid parking fines. Next, they discovered employees who owed various fees associated with the alcohol and drug abuse program. This was understandably met with outrage once news of their activity was leaked to the public. In response, the city then established new rules governing the computer center's operations to better protect employee privacy.

At that time, minute description was still largely based upon cross-sectional data captured at infrequent intervals. As technology evolved, however, collection frequency increased, and the level of detail was enhanced. Mason described these connections as *threads* that would ultimately result in the formation of all-knowing dossiers. These advancements have allowed for more threads to be woven within and among datasets, forming a permanent record for every individual. Such knowledge is primed for abuse. Not only is someone's privacy likely to be invaded, if embarrassing information falls into the wrong hands it could be leveraged for blackmail.

2.2. Accuracy

> *Who is responsible for the authenticity, fidelity and accuracy of information? Similarly, who is to be held accountable for errors in information and how is the injured party to be made whole? (Mason, 1986, p. 5).*

As large datasets become more interconnected, data integrity becomes paramount. Therefore, Mason stressed the importance of information accuracy by highlighting specific consequences caused by errors in computerized bank transactions and weather forecasts. The first example pertained to a bank's refusal to acknowledge receipt of a mortgage

payment simply because a new computer system did not show it as paid. To prove his claim, the customer, Louis Marches, presented his coupon book to show that it had clearly been stamped as "paid" by the bank teller. Instead of acknowledging the error, the bank's employee's put their faith entirely in the computer system. Compounding the matter even further, the bank refused to accept subsequent payments until the "unpaid" payment was satisfied. This continued until the bank eventually foreclosed on the property. Making matters even worse, Marches' wife, who had been in bed recovering from a heart attack, suffered a stroke upon learning of the foreclosure from a debt collector.

The second example tells the tragic story of a man lost at sea in 1980. A weather forecast produced by the National Weather Service had stated that a nearby storm would not impact the ship's course, yet they soon found themselves in 80 knot winds and seas cresting at 60 feet. The erroneous forecast failed to predict the turbulent conditions due to a faulty buoy. Since the forecasting model was without the buoy's data, it miscalculated the storm's trajectory by several miles. Although both incidents resulted in large settlements to the injured parties, Mason recognized that we run the risk of repeating these mistakes if information systems are not carefully developed and tested.

2.3. Property

> *Who owns information? What are the just and fair prices for its exchange? Who owns the channels, especially the airways, through which information is transmitted? How should access to this scarce resource be allocated? (Mason, 1986, p. 5).*

Mason's focused his discussion on property around developments in artificial intelligence and the scarce resource of communication bandwidth. First, he worried about extraction of novel human thought and ingenuity and its subsequent implantation into machines. For example, will those who discover new insights be justly compensated if their ideas are replicated into billions of machines? When information is easily copied and transferred, controlling its exchange becomes increasingly difficult.

To illustrate his second concern with communication channels, Mason draws a parallel to Garrett Hardin's essay, "The Tragedy of the Commons." Since herdsmen directly benefited from each additional animal added to a pasture, but only indirectly experienced the costs of grazing, the pasture would eventually be destroyed by overuse. If the infrastructure of our data networks is also treated as a commons, we too run the risk of abusing it. On the other hand, when a communication medium is owned by a single entity, they wield considerable control over what conversations can take place. Therefore, we must find ways to ensure that a balance is found between ownership and access.

2.4. Accessibility

> *What information does a person or an organization have a right or privilege to obtain, under what conditions and with what safeguards? (Mason, 1986, p. 5).*

As Mason points out, literacy is critical to one's participation in, as well as the advancement of, any society. Mason explains that literacy goes much further than the ability to read. First, intellectual skills, including the ability to reason and calculate, must be developed through education. Second, one must have access to the necessary technology. Lastly, information must be accessible, else it cannot be consumed. Mason argues that one's own knowledge and economic status determines whether these three requirements have been met, and to what degree. Although technology has advanced rapidly, not all have benefited from it at the same rate. This has resulted in a large segment of the population becoming increasingly information poor. Mason continues by describing the necessary steps to access information stored in modern databases. Those who cannot complete the steps are likely to become what Mason refers to as "information drop outs" who will likely need greater assistance in the future.

2.5. New Social Contract

In his concluding remarks, Mason called for a new social contract that would preserve everyone's right to maximize their human potential. He made it clear that the dawn of the information age represented a critical junction for society, with the fate of future generations at stake. Mason stressed the importance of developing information systems that would "enhance the dignity of mankind" (Mason, 1986, p. 11). Mason hoped that keeping PAPA in the forefront of our minds would lead us to make wiser decisions. He encouraged developers to ensure that future information systems:

- would not "unduly invade a person's privacy;"
- produce and maintain accurate records;
- protect the available bandwidth to avoid repeating the "Tragedy of the Commons;"
- protect intellectual property; and,
- are widely accessible to foster information literacy.

In Mason's words, failing to abide by these guidelines would risk "information bankruptcy or desolation" (Mason, 1986, p. 11). Unfortunately, despite his best efforts, it seems these lessons were either largely ignored or forgotten over the following decades.

3. Evolution of Technology

> *"The real danger is the gradual erosion of individual liberties through automation, integration, and interconnection of many small, separate record-keeping systems, each of which alone may seem innocuous, even benevolent, and wholly justifiable" (U.S. Privacy Protection Study Commission, 1977).*

Much has changed since Mason's PAPA framework was published in 1986, especially as it pertains to data production. As Smolan & Erwitt (2012) point out, human civilization had only produced a cumulative total of 5 billion gigabytes of data from the beginning of recorded history until 2003. In 2011, we produced 5 billion gigabytes of new data every two days. By 2013, the interval was reduced to 10 minutes, and in 2015, just 10 seconds. Such rapid growth in data production propelled humanity into the big data era. Although it might sometimes feel like it, we did not get here overnight. The big data phenomenon was made possible due to revolutionary technological advancement over several decades.

In the 1980s, personal computing had truly arrived, putting the power of digital creation in the hands of the average consumer. The 1990s saw the birth of the Internet as the World Wide Web was made available to the public in 1991. Search engines, such as Yahoo! Search and Google, helped users find specific information across the vast sea of newly available data.

Although cellular devices emerged in the 1990s, it was not until the 2000s that mobile devices became readily available. Throughout his career as a science fiction writer, David Gerrold naturally made several predictions about future technology. Perhaps none rings as true today as Gerrold's (1999, p. 61) contribution to "Smart Reseller" magazine:

> *I've got a cell phone, a pocket organizer, a beeper, a calculator, a digital camera, a pocket tape recorder, a music player, and somewhere around here, I used to have a color television. Sometime in the next few years, all of those devices are going to meld into one. It will be a box less than an inch thick and smaller than a deck of cards…I call this device a Personal Information Telecommunications Agent, or Pita for short. The acronym also can stand for Pain In The Ass, which it is equally likely to be, because having all that connectivity is going to destroy what's left of everyone's privacy.*

Clearly, Gerrold's vision could not have been more prophetic. At the turn of the millennium, roughly 37 percent of Americans had a mobile phone, reaching 90 percent by the end of the decade (Kornstein, 2015). Smartphones, such as Apple's iPhone, significantly elevated mobile device capabilities, and ultimately realized Gerrold's Pita fears.

In 2009, electronic health records were mandated in the United States through the Health Information Technology for Economic and Clinical Health Act (HITECH Act) to streamline medical care. The 2010s saw wide adoption of social media platforms, such as Facebook and Twitter, promising to "connect you with the people around you," and the growth of artificial intelligence and predictive analytics (Power, 2016). Despite utopian ideals, technology was so disruptive during the 2010s that it has been considered by some as "The Decade Tech Lost Its Way" (Rahimian & Kelly, 2019).

4. Big Data Concepts

To deal with the emerging challenges created by these technology advancements, Doug Laney (2001) proposed a three-dimensional data challenge framework with fast increasing data volume, velocity and variety, which has become ubiquitous in definitions of big data. More recently, Van Rijmenam (2014) suggested including veracity, variability, value, and visualization. In this section, we briefly define and discuss these terms to form our contextual foundation.

Data **volume** is increasing exponentially. Gantz and Reinsel (2012) estimated that the total size of the digital universe in 2010 stood at 1.2 zettabytes (ZB). It was estimated to increase to 4.4 ZB by 2013 and was expected to grow 40% per year into next decade, almost doubling the size every two years (Gantz & Reinsel, 2012). By their estimates, the total size of digital universe in 2020 will likely reach 40 ZB (Gantz & Reinsel, 2012).

Van Rijmenam (2014, p. 5) referred to **velocity** as "the speed at which data is created, stored, analyzed, and visualized." The rapid growth and abundance of emerging technologies such as sensors, connected devices, smart personal/home/office devices, 5G networks, mobile devices, autonomous automobiles, cloud services, e-health, and virtual reality, collectively accelerated the need for real-time data accumulation (Ariyaluran Habeeb et al., 2019). Velocity is generally captured through two traits: (1) frequency of generation; (2) frequency of handing, recording, and

publishing (Kitchin & McArdle, 2016). However, to holistically understand the velocity of big data, we must not only embrace technological advancement, but also simultaneously address potential privacy and security issues.

Variety refers to a wide range of data types, which include structured, unstructured, and semi-structured, as well as various data sources. Structured data usually resides in relational databases and is characterized by pre-defined and searchable fields. For instance, customer records such as name, phone numbers and Zip codes stored in a customer relationship management system or sensory device seeks to create a continuous, longitudinal record of usage. Data may be human or machine generated within the relational databases. Unstructured data on the other hand is not constrained by pre-defined data models or schema and resides in NoSQL databases. Common examples of unstructured data are text, audio, video files and images. Semi-structured data is a hybrid of structured and unstructured data. It usually has some defining characteristics but does not conform to a structure as rigid as what we see in relational databases (Kitchin & McArdle, 2016). For instance, pictured taken on a smart with mobile devices, are unstructured but are often geo tagged and time stamped.

Veracity describes the degree to which data is truthful and trusted. Veracity is increasingly being recognized as it is an essential component to extract value from big data. Veracity comprehensively describes data quality and can be translated to the reliability or consistency of the data (Gupta & Rani, 2019), or confidentiality, integrity, and availability of the data (Kepner et al., 2014). A wide range of factors can sacrifice data veracity depending on the type of data and the stage of data analysis, for example, inherent biases in data processing, untrustworthy data sources, and abnormalities. Common examples include user entry errors, duplication, and corruption, all of which pose threats to the potential value of big data.

Variability considers the inconsistencies of the data flow. Data loads become challenging to maintain at the same speed, especially with an increase in social media usage, which generally causes a peak in data loads when certain events occur (Katal, Wazid, & Goudar, 2013). Other variability issues relate to the multitude of data dimensions, types, and sources that eventually lead to the inconsistencies in the data.

Data **validity** in a statistical sense addresses the degree to which the tool measures what it claims to measure (Kelley, 1927). In the big data context, it refers to the correctness of data for its intended use (Gupta & Rani, 2019). For instance, sentiment analysis has recently become extremely popular in dealing with user-generated text data across social media platforms as it provides either positive, negative, or neutral sentiment predictions. While sentiment analysis can be a powerful tool in certain contexts, such as political campaigns, it has unknown validity for emotions of interest incapable to detect mood. For example, a customer with the verified purchase tag on Amazon writes a sarcastic review on a product. Although the review possesses veracity, it can easily be interpreted as a positive review. In this case, the sentiment analysis fails to capture the intended sarcasm and thus undermining the data validity.

Data's perceived **value** largely drives corporations to collect as much data as frequently as possible. While technology provided the spark, the insatiable appetite added the necessary fuel to spawn the big data inferno. Data's potential value is rooted in a desire to gain new insights, and subsequently, transform these insights into proper and timely actions that were not feasible before. When data is collected, sold, exchanged, data becomes a commodity that possesses value, the true value of big data, however, lies in the process of finding insights (Gupta & Rani, 2019). For example, the department store Macy's adjusts nearly 73 million items based on demand and supply in near-real-time for them to maximize profit. In a similar vein, retail giant Walmart utilizes semantic and text mining techniques to create better search results for its website that results in increased revenues (Desjardins, 2015).

Visualization refers to the use of visual elements such as charts, graphs, and maps to represent information and data (Tableau Software). The volume, veracity, and variety that characterizes big data make it particularly challenging to conduct visualization (Chen & Zhang, 2014). Firms not only use data visualization tools (e.g., Tableau, Chartist, Grafana, and D3.js) to transform large and complex data sets into simplistic, yet interactive pictures and dashboards to aid decision making, they also integrate user data with in-app visualizations. This is a common practice in social media platforms, such as the Snap Map from Snapchat, Find My Device service from Apple, as well as interactive filters from Snapchat and Instagram.

5. BIG PAPA

Twenty years after Mason's PAPA was published, Peslak (2006) reaffirmed the relative importance of the four original components. Privacy was perceived as the most important ethical issue, followed by accessibility, accuracy, and property. In this section, we suggest adding three issues (behavioral surveillance, interpretation, and governance) to the PAPA framework and propose an extension that we refer to as BIG PAPA.

Although Mason explored these interrelated ideas in his original work, we believe that our proposed extensions should be considered distinct issues in the context of big data. As shown in Figure 1, the behavior issue is primarily rooted in Mason's discussion on privacy, whereas interpretation is a subset of Mason's accuracy issue. Our standalone governance issue is formed from Mason's call for adequate safeguards throughout PAPA.

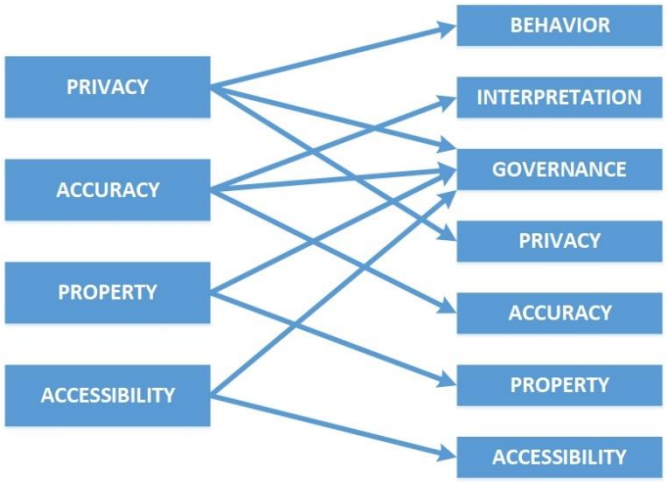


Figure 1.

While these issues are certainly not new, early concerns were focused on small silos of data. Therefore, any negative impact that could be caused by unethical behavior was naturally constrained. As technology and data consumption grew, these issues became more important. We illustrate this development by loosely overlapping each issue with data prior to the big data era (Figure 2), whereas each issue has become fully entrenched by big data in recent years (Figure 3).

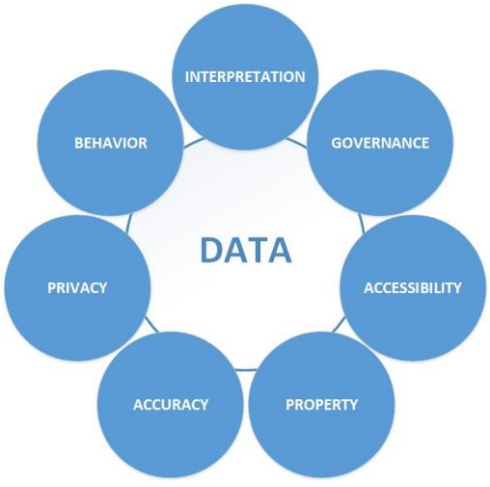


Figure 2. Pre-Big Data Era

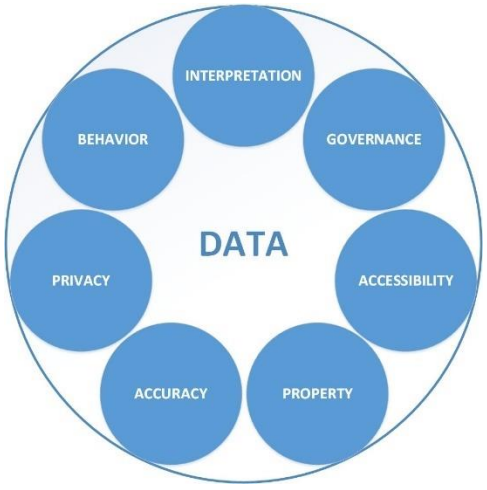


Figure 3. Post-Big Data Era

The seven issues that form BIG PAPA, as well as related questions, are provided in Table 1. We discuss each issue in the context of how big data has impacted society. We provide several examples to highlight the importance of our extension, as well as how Mason's original issues (privacy, accuracy, property, and accessibility) have evolved.

Issue		Related Questions
B	Behavioral Surveillance	• How can users be better informed before making an active choice to share data? • How do we preserve individual liberty when behavior is constantly monitored?
I	Interpretation	• How can we avoid developing flawed models built upon erroneous or incomplete data? • How can we reduce the likelihood of drawing erroneous conclusions? • How can we educate others to recognize poor analyses?
G	Governance	• What controls are in place to handle ethical challenges related to big data? • Who will watch the watchmen?
P	Privacy[1]	• What information about one's self or one's associations must a person reveal to others? • Under what conditions and with what safeguards? • What things can people keep to themselves and not be forced to reveal to others?
A	Accuracy[1]	• Who is responsible for the authenticity, fidelity, and accuracy of information? • Similarly, who is to be held accountable for errors in information? • How is the injured party to be made whole?
P	Property[1]	• Who owns information? • What are the just and fair prices for its exchange? • Who owns the channels, especially the airways, through which information is transmitted? • How should access to this scarce resource be allocated?
A	Accessibility[1]	• What information does a person or an organization have a right or a privilege to obtain? • Under what conditions and with what safeguards?
[1] Mason's original PAPA framework		

Table 1. BIG PAPA Framework

5.1. Behavioral Surveillance

> *How can users be better informed before making an active choice to share data? How do we preserve individual liberty when behavior is constantly monitored?*

When our every move, message, or even thought is being surveilled, it can have a chilling effect on our behavior (Lashmar, 2017; Richards, 2012; Stoycheff, 2016). Therefore, we focus this issue on any data that can reveal individual activity, such as location, communication, and digital interaction. Mason's discussion of privacy is certainly related to behavior, but technology at that time did not allow for the refined, targeted, and continuous data collection that is conducted today. Instead, most data were consciously provided by users themselves. Thus, we discuss behavior that is observed by virtue of active choices made by a user, as well as behavior data acquired through passive means of observation and collection. For example, a user might knowingly elect to share such data by making an active choice to opt in to obtain the benefits of a given product or service, or the behavior data could be passively collected without any clear user awareness or consent.

5.1.1. Active Choice

> *"Historically, a conversation that you might have in the hallway is private by default, public through effort…Conversely, when you engage online in equally public settings such as on someone's Facebook Wall, the conversation is public by default, private through effort. You actually have to think about making something private because, by default, it is going to be accessible to a much broader audience" (Boyd, 2010).*

When users make conscious choices to engage in certain behavior, they must assume some level of risk, even if their assessment of the risk does not reflect reality. For example, many subscribe to the idea that they are doing nothing wrong, and therefore have "nothing to hide" to justify their oversharing behavior (D. J. Solove, 2007). Yet, when questioned on the issue, most will quickly concede that they wear clothes, put curtains on their windows, and lock the door when they go to the bathroom. Until users come to the realization that they should care about protecting digital behavior just as much as their physical activity, it is difficult to demonstrate how their actions can jeopardize their privacy and security.

Sadly, there are countless instances of technology use resulting in serious consequences that users never fully anticipated yet would never have been possible without their active involvement in adopting the technology. Nude photos have been stolen from cloud backups and distributed online (Peterson, Yahr, & Warrick, 2014). Residences have been broken into after occupants reveal on social media that they will not be home for an extended period (Sanchez-Garrido, 2016). Smartwatch data uploaded to fitness websites has exposed sensitive locations, including military bases (Hern, 2018). Hackers have spied on and harassed vulnerable children when insecure baby monitors and cameras are connected to the Internet (Chiu, 2019).

Of course, we are not blaming victims for the illegal actions of others, but we do advocate for the development of basic security and privacy competencies before choosing to adopt certain technologies. Unfortunately, it is usually only after personally experiencing negative effects that they are motivated to modify their behavior to reduce future risk. Ignorance may be bliss, but that does not change the fact that individual behavior regularly undermines their own privacy, security, and liberty. Obviously, much of this issue can be attributed to poor consumer awareness, which we will discuss later.

5.1.2. Passive Observation and Collection

> *"No one disputes that the deployment of cheap, ubiquitous video cameras has made an environment of near total surveillance technologically feasible. Whether that's a good thing or a bad thing, however, depends on how much you trust the cameraman" (Grossman, 2001).*

After it was discovered that the National Football League had secretly agreed to allow police to use facial recognition technology to search for criminals among attendees of Super Bowl XXXV, the game became scornfully known as the "Snooper Bowl" (Brey, 2004; Grossman, 2001). Today, companies such as Clearview AI are scraping image data from publicly available resources, such as social media, to improve the accuracy of their facial recognition system (Hill, 2020). These efforts have caused many to question whether participating in seemingly innocent social media fads, such as the Facebook's "10 Year Challenge," is only helping to train facial recognition algorithms to better predict aging effects (O'Neill, 2019).

While it is understood that individuals should not expect privacy in public spaces, the invasive surveillance technology that has developed in the big data era has taken 'Big Brother' to the extreme (Power, 2016). The precision and pervasive nature of the countless devices designed to track identifiable individuals today has only amplified its impact on society (Joh, 2016). For example, computer vision can be used to track vehicles through license plate readers (Lum, Hibdon, Cave, Koper, & Merola, 2011; National Law Enforcement and Corrections Technology Center, 2010; Ozer, 2010; Zmud, Wagner, Moran, & George, 2016) and law enforcement agencies have deployed International Mobile Subscriber Identity-catchers, also known as stingray devices (Bates, 2016; Boyne, 2016; Norman, 2016; Pell & Soghoian, 2014). Stingray devices pose as cell phone towers to eavesdrop on cellular communication, uniquely identify citizens, and track their movements. Fortunately, as Power (2016) points out, U.S. courts have found several innovative investigatory techniques to be unconstitutional, such as the use of thermal imaging to peer inside homes to detect marijuana plants or the placement of GPS tracking devices onto vehicles without warrants.

However, the government does not fund the bulk of these surveillance systems. Much of the tracking technology is developed, financed, and implemented by technology companies under a "surveillance as a service" business model (West, 2019). Because the data is already being collected and stored by the private sector, access can easily be granted to a variety of interested parties, such as law enforcement and data brokers—companies who buy, sell, and exchange data (Otto, Anton, & Baumer, 2007). For example, Amazon has been criticized for sharing home surveillance videos captured by its Ring products with police departments without obtaining warrants (Cox, 2019). Instead, investigators simply need to acquire consent from owners of the recording device, effectively creating a surveillance network that bypasses court oversight. Considering that Amazon incentivizes police departments to promote Ring products, consumers are essentially encouraged to fund a budding surveillance state out of their own pocket. Not only do consumers have to worry about how their data will be collected and shared by companies without their knowledge, they should also be concerned about how vulnerabilities in these devices can expose such data to other parties ("Ring under fire over weakness in video device security," 2020).

Another example is Nextdoor, a social media platform that claims to be "the best way to stay informed about what's going on in your neighborhood" and states the following in their privacy policy (Nextdoor, 2019):

> *"If you decide to invite new members to join Nextdoor, you can choose to share their residential or email address with us, or share your contacts with us, so we can send an invitation and follow-up reminders to potential new members on your behalf."*

Rather than solicit the information directly, Nextdoor actively encourages new and existing members to divulge the personal information of others, without their neighbors' knowledge or consent. The neighborhood monitoring made possible through NextDoor can be referred to as a form of lateral surveillance, where citizens take an active role in monitoring the behavior of others, which was traditionally left to professionals (Andrejevic, 2004). While many recognize the inherent privacy risks associated with using such services (Masden, Grevet, Grinter, Gilbert, & Edwards, 2014), those who use Nextdoor are unlikely to recognize how their actions have the potential to cause real harm to others, such as those who do not wish to have sensitive information shared when relocating to escape abusive relationships and/or domestic violence. Further, such services have the potential to perpetuate discrimination, effectively building a "digitally gated community" (Kurwa, 2019).

In Mason's call for a new social contract, he sought information systems that would not "unduly invade a person's privacy to avoid the indignities that the students in Tallahassee suffered" (Mason, 1986, p. 11). Regardless of whether an individual actively consents to their behavior being monitored, the level of detail captured by modern devices is not consistent with this noble goal. Is this the role that we want technology to play in our society? If not, we must make an active stand to reclaim our privacy.

5.2. Interpretation

> *How can we avoid developing flawed models built upon erroneous or incomplete data? How can we reduce the likelihood of drawing erroneous conclusions? How can we educate others to recognize poor analyses?*

Although obtaining a census is far more possible in the big data era (Kitchin & McArdle, 2016), most analyses are still dependent upon samples. Models are built to infer relationships and predict future events. Therefore, as analytical capability advances, we must first guard against developing flawed models built upon erroneous or incomplete data, then avoid acting upon biased or incorrect conclusions. While Mason's inclusion of accuracy might appear to satisfy this issue, we argue that one can have accurate data, yet still form incorrect conclusions due to poor modeling and/or flawed interpretations. Therefore, we have teased out this issue from Mason's broader discussion on accuracy.

5.2.1. Garbage In, Garbage Out

In *Weapons of Math Destruction*, Cathy O'Neil (2016) explains how reliance on algorithms based upon flawed data can have disastrous consequences for society, such as reinforcing discriminatory practices, and at a massive scale. For example, facial recognition algorithms have been shown to have unacceptable accuracy rates and to further perpetuate inherent biases (Snow, 2018).

Further, because most algorithms are developed to gain a competitive advantage, and thus considered proprietary, we have limited visibility into how they were formed and are being employed. Efforts such as NIST's Facial Recognition Vendor Test (FRVT) have focused primarily on algorithm efficiency and effectiveness, not detecting flaws due to biased data (Introna & Wood, 2004). If algorithms are implemented before independent and comprehensive testing can be conducted, the impact of big data is more likely to result in harmful outcomes, even if the original motivations are well intended.

Given the constant creep towards predictive policing (Benbouzid, 2019; Joh, 2016), such as attempting to identify terrorists based upon whether they have purchased a life insurance policy (Van Rijmenam, 2014, p. 85), it is clear that this issue could not be more critical today. Therefore, it is essential for those dependent upon big data to ensure that analytic output meets information quality standards (Lee, Strong, Kahn, & Wang, 2002), especially before such results are applied to practice.

5.2.2. Lies, Damned Lies, and Statistics

The rush to employ analytics can also place tremendous power in the hands of untrained individuals. As Wheelan (2013, p. 95) plainly put it, "statistics cannot be any smarter than the people who use them. And in some cases, they can make smart people do dumb things." One such example is referred to as the prosecutor's fallacy (Fenton & Neil, 2011).

Assume that an expert witness from a scientific laboratory correctly testifies that a given piece of biometric evidence, such as a fingerprint or DNA, can be attributed to one out of one million Americans. When a prosecutor later argues that the chances of the defendant not being the guilty party are 0.0001%, they are incorrectly interpreting the statistic. If there are 300 million people in the United States, the evidence could be attributed to a pool of 300 potential suspects. Therefore, the probability that the evidence does not belong to the defendant is 99.67% (299/300). These types of damning errors are simply not acceptable, especially when life and liberty is at stake.

In the legal arena, the onus is on the states and the federal government to train their prosecutors on how to interpret and use statistical data. Even so, the Innocence Project and the Innocence Network, a group of independent organizations that work to exonerate wrongfully convicted people often by using DNA to prove innocence. Since the project began, 367 people have been exonerated, including 21 that have served time on death row (Innocence Project, 2020). Approximately 44% of those exonerated are reported to have been wrongly convicted due to a misapplication of forensic science.

5.2.3. Calling Bullshit

As data becomes more accessible and as society grows increasingly reliant on data-driven products, we must not only guard against conducting poor statistical analyses, but also better educate the average citizen to recognize flawed claims. To combat this emerging issue, Carl Bergstrom and Jevin West developed a course at the University of Washington entitled "Calling Bullshit: Data Reasoning in a Digital World" (https://callingbullshit.org). Bergstrom and West (2020) West define such *bullshit* as "language, statistical figures, data graphics, and other forms of presentation intended to persuade by impressing and overwhelming a reader or listener, with a blatant disregard for truth and logical coherence." Regardless of the method of delivery, we must ensure that both producers and consumers of data understand the inherent risks associated with statistical analyses.

5.3. Governance

What controls are in place to handle ethical challenges related to big data? Who will watch the watchmen?

Governance involves ensuring that necessary controls are in place to handle future ethical challenges (Smith, Milberg, & Burke, 1996). While Mason (1986) did call for safeguards to protect individuals from unethical behavior, these arguments were embedded within each issue. We believe that governance, or the lack thereof, in the big data era is an issue in and of itself. We discuss governance with respect to the roles industry and government should play in mitigating unethical and illegal behavior (Richards & King, 2014).

5.3.1. Industry

Governance currently operates under an idealistic model of shared responsibility among industry, government, and consumer. Industry groups and professional associations are expected to adhere to codes of conduct, government regulators provide enforcement and ensure adequate safeguards are in place, and consumers must make informed choices. Self-regulation is certainly preferable, but when self-regulation fails, the consequences of illegal and unethical behavior must be severe enough to deter others. Unfortunately, as with all ideals, they are rarely achieved as envisioned.

Although Equifax agreed to pay up to $700 million as part of a settlement in response to their 2017 data breach, only $425 million would go towards compensating approximately 147 million affected consumers (Federal Trade Commission, 2019). Equifax exposed names, dates of birth, Social Security numbers, physical addresses, and other personal information. Considering the severity of the breach, averaging under $3.00 per victim is offensive, and certainly does not make anyone whole, as is a major purpose of bringing a tort lawsuit.

Perhaps more egregious is the fact that Google was only fined $7 million after its Street View mapping project was caught secretly capturing passwords, e-mail addresses, medical and financial records, and other personal information as Google's vehicles traveled nearly every road in the country (Streitfeld, 2013). Considering the value this data holds, such a small fine does not even register as a slap on Google's wrist.

Clearly, self-regulation under the shared responsibility approach has not been effective in discouraging unethical practices, particularly when penalties are not proportional to offenses. This reality has led to calls for much needed reform of the data broker industry (Kuempel, 2016). However, most companies in the era of big data could be classified as data brokers, especially when one considers the volume, velocity, and variety of the data they collect. Therefore, we argue that far more comprehensive reform is needed for society to have any hope of meaningfully curtailing unethical data practices.

5.3.2. Governments

"So this is how liberty dies…with thunderous applause" – Queen Padmé Amidala in Star Wars: Episode III - Revenge of the Sith (Lucas, 2005).

For as long as industry produces data, governments at all levels will be waiting to tap in. In the U.S. alone, the federal government has attempted to achieve key escrow by building backdoors into phone hardware with the Clipper chip

(Abelson et al., 2015), used the terrorist attacks of September 11, 2001 to justify global mass surveillance (Kirk, 2014), and pressured technology companies to break their own devices following the terrorist shooting in San Bernardino, California (Hack, 2016; Newkirk, 2018).

Despite passionate pleas in the name of national security, these programs are regularly found to be wasteful, ineffective, and unconstitutional once they are subject to critical review and oversight from independent bodies. For example, the government was forced to admit that the mass surveillance programs revealed by Edward Snowden failed to thwart any terrorist threats that were not already detected through traditional methods (Medine, Brand, Cook, Dempsey, & Wald, 2014a, 2014b).

As one might expect, Dinev, Hart, & Mullen (2008) found that users' willingness to provide personal information is negatively influenced by privacy and government intrusion concerns yet is positively influenced by users' perceived need for government surveillance. These concerns can be seen in the effort that ultimately led San Francisco, ironically in the heart of Silicon Valley, to ban the use of facial recognition (Barber, 2019). However, these restrictions only prevent city agencies from purchasing surveillance technology, leaving the door open for corporations to share data with government officials. In 2009, Eric Schmidt, then CEO of Google, had the following to say with respect to industry cooperation with government demands:

> *"I think judgment matters. If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place. If you really need that kind of privacy, the reality is that search engines -- including Google -- do retain this information for some time and it's important, for example, that we are all subject in the United States to the Patriot Act and it is possible that all that information could be made available to the authorities." (Schneier, 2009).*

This mindset is consistent with the "surveillance as a service" concept (West, 2019), where private firms are complicit in doing the government's bidding. However, given society's reliance on technology, if our only hope to maintain privacy is by "not doing" something, we would be prevented from essential activities, such as making financial transactions, receiving medical treatment, and merely existing in public places. Clearly, such a standard is not acceptable in a free society, yet similar statements are commonly repeated by government leaders and those responsible for the world's largest technology companies.

As Scott McNealy, cofounder of Sun Microsystems, famously proclaimed in 1999, "You have zero privacy anyway…Get over it" (Sprenger, 1999). Yet, in 2015, he stated that "It scares me to death when the NSA or the IRS know things about my personal life and how I vote. Every American ought to be very afraid of big government" (Noyes, 2015). Likewise, Mark Zuckerberg, founder of Facebook, who has built his social media empire on "making people more open and connected" with little regard to limiting the sharing of data to third parties, paid more than $30 million for four houses surrounding his own residence in an effort to protect his own personal privacy (Nicks, 2016).

Given that industry struggles to self-regulate and some of the most invasive practices have been employed by governments, perhaps we should look elsewhere for a solution to the governance problem? One intriguing option is self-sovereign identity (Mühle, Grüner, Gayvoronskaya, & Meinel, 2018; Tobin & Reed, 2017). Self-sovereign identity allows a user to share the minimum level of detail for a given situation by presenting verifiable claims.

For example, consider a situation where an individual must prove that they are of a certain age. Instead of being forced to produce a driver's license or other official document that unnecessarily reveals sensitive information, he or she could simply present a verifiable claim that has been digitally signed by a government agency. The claim could be as simple as stating that the individual is over 18 years old. Not only does this prove eligibility, it does so without sharing any other sensitive information typically present on other authoritative documents, such as birthdate or home address.

While this is an extremely simple example, self-sovereign identity can be used for just about any conceivable use case. Ultimately, self-sovereign identity puts control back into the hands of the individual rather than continue to populate centralized repositories maintained by increasingly powerful data brokers and governments. Widespread adoption of self-sovereign identity would dramatically curb the current appetite for consumer data by starving data brokers of the minute description that makes a dataset so valuable.

5.4. Privacy

Mason's discussion on privacy highlighted the early beginnings of what would later become known as data brokers. These companies are motivated to obtain as much data as possible due to both perceived and real value. While data

brokers certainly present a significant threat to privacy, we approach the issue by focusing on consumer awareness, calls for the "right to be forgotten," as well as how privacy rights have been viewed in both common law and statutes.

5.4.1. Consumer Awareness

Several measures have been developed to assess consumer privacy concerns: Smith, Milberg, & Burke (1996) developed their global information privacy concern (GIPC) scale; Stewart & Segars (2002) developed the Concern for Information Privacy Instrument; and, Malhotra, Kim, & Agarwal (2004) developed the Internet Users' Information Privacy Concerns (IUIPC) instrument. Studies have also focused on particular use cases. For example, Bélanger, Hiller, & Smith (2002) investigated how four trust indices impacted consumer perceptions of privacy in e-commerce and Awad & Krishnan (2006) found that consumers who valued information transparency were less likely to provide information to aid in personalized marketing.

While this stream of research has yielded insights regarding privacy concerns, improving consumer awareness in practice has been a challenge, especially when the primary method of delivery has been through terms of service and privacy policies. Milne, Culnan, & Greene (2006) evaluated over three hundred privacy notices and concluded that in just two years, notice readability had declined, while length had increased. McDonald & Cranor (2008) later estimated that it would take approximately 244 hours per year for an individual to read the privacy policies for each website visited. Further, most terms of service agreements must be accepted in their entirety, which prevents users from negotiating fairer exchanges (Walker, 2012). Although there have been attempts to simplify terms of service, such as "Terms of Service; Didn't Read" (https://tosdr.org/), placing the privacy burden entirely on the consumer is not only unreasonable, but completely impractical for those who value their privacy, yet do not understand the risks associated with certain technology.

Boritz & No (2011) discovered that much of the privacy research had been conducted in the early 2000s, creating a void that was not accounting for rapidly developing technology. Several other studies have reviewed privacy literature in information systems (Bélanger & Crossler, 2011; Pavlou, 2011; Smith, Dinev, & Xu, 2011) in an effort to spur future privacy research. Lowry, Dinev, & Willison (2017) encouraged researchers to treat security and privacy as an IS artifact. However, despite greater attention to the privacy issue in academia, it is difficult to see any substantial progress being made in practice with respect to increasing consumer awareness.

5.4.2. Right to be Forgotten

One of the reasons it has been difficult for the average consumer to anticipate negative consequences of certain technologies is due to the impressive speed at which they have advanced. Without having enough time to observe and process how these developments might impact them, many participate, only to eventually find themselves regretting previous behavior, such as posting personal information on social media (Rosen, 2012).

Because of this, many have called for an online undo option, or Ctrl+Z (Jones, 2016), known as the "right to be forgotten," that would allow users to remove information from the seemingly permanent record made possible through the Internet. In January 2012, the European Commission announced that they would recognize this right, which was formally adopted in 2016. Viviane Reding, European Commissioner for Justice, Fundamental Rights, and Citizenship, explained the basic premise behind the right to be forgotten: "If an individual no longer wants his personal data to be processed or stored by a data controller, and if there is no legitimate reason for keeping it, the data should be removed from their system" (Reding, 2012, p. 5).

Despite the obvious benefits, valid criticisms against the right to be forgotten have been raised regarding how it might conflict with the rights of others, namely rights to free speech and freedom of expression (Walker, 2012). However, proponents recognize that requests to remove information under the right to be forgotten should be limited to content that the subjects of the data uploaded themselves (Ausloos, 2012; Walker, 2012). This approach respects individual privacy yet prevents the right to be forgotten from becoming a magic eraser that can be used to censor others.

Complicating the issue from a practical standpoint, however, is that content submitted by users can propagate to other online services (Ausloos, 2012), whose right to share would then be protected by their own freedom of expression. Further, content marked for deletion could still exist on data backups for a considerable amount of time. Therefore, the unfortunate reality is that the only reliable way for users to prevent others from maintaining a permanent record is to avoid sharing or allowing the data to be collected in the first place.

5.4.3. Right to Privacy in Common Law

In their seminal law review article, Samuel Warren & Louis Brandeis (1890) recognized that U.S. law was beginning to distinguish physical and psychological harms. The common law up to this point only recognized physical harms, such as battery. As Warren and Brandeis pointed out, the law began to recognize psychological harm that can result from the reasonable apprehension of physical harm. Warren and Brandeis applied this extension to scenarios involving publishing psychological damaging information in tabloids. The traditional common law notions of property, contracts, and copyright were inadequate to remedy psychological harms. Common law remedies still may not be adequate today.

Nickelodeon Consumer Privacy Litig., 827 F.3d 262, 293 (3rd Cir., 2016), involved a class action filed on behalf of 13-year-old plaintiff's alleging that Google and Viacom unlawfully collected information on the websites and videos the children viewed. The Plaintiffs claimed Google and Viacom committed the common law tort of intrusion upon seclusion. One of the four types of invasion of privacy torts, a defendant is found liable for the tort of intrusion upon seclusion when the plaintiff can show (1) an intentional intrusion (2) upon the seclusion of another that is (3) highly offensive to a reasonable person. Though the plaintiff's other claims failed, the plaintiffs in Nickelodeon were successfully able to prove that Google and Viacom had committed the tort of intrusion upon seclusion. What is highly offensive in the context of this tort to a reasonable person is guided by court decisions and lawyer arguments.

5.4.4. Right to Privacy in Statutes

Though some sectors in the United States are regulated in terms of privacy, some are not. The "sectoral approach" to privacy regulation creates a patchwork of legislation that overlaps in multiple areas of the economy and does not at all regulate in others (D. Solove, 2015). Some members of Congress have supported legislation to create a federal privacy law, though the efforts have yet to go anywhere. Giving a private cause of action and the preemption state privacy laws are some continual sticking points (Feiner, 2019).

The other approach to privacy regulation is the "omnibus approach" taken by the European Union. This approach is designed to regulate privacy no matter the sector and impacts American business, or any business that does business with citizens of the EU. The General Data Protection Regulation (GDPR) has been in effect for over a year and is a significant step in improving privacy regulation (Burgess, 2019). The GDPR attempts to alter how businesses handle their information and give consumers more control over their information.

While Congress has been unable to pass comprehensive bi-partisan legislation, some states have filled the gap. The California Consumer Privacy Act is the most recent state level legislation to directly address information privacy. This law is designed to allow consumers the ability to ask companies that collect data on consumers to see the information. Consumers can see personal data to include smartphone data, physical locations, and biometric data (Cowan & Singer, 2020).

The right to privacy is right difficult to define in the law, yet should be central to discussion on data and its use. Many laws dealing with privacy are centered on the premise that consumers will understand how their data is being used simply by being told. This fails to account for the true understanding of the content in the disclosure and does not adequately protect consumers. The right to be forgotten, much like the general right to privacy, is not easily implemented. The approach taken in the United States, based primarily on informed consent, is a confusing patchwork that leaves gaps in some areas and overlaps in others. The omnibus approach taken by the EU with the GDRP, is a significant step forward in creating an effective framework for the protection of consumer data.

Given the speed at which technology advances, it is unlikely for any law to fully protect privacy. The right to be forgotten is a desirable option, but the only guaranteed defense is to never share sensitive information in the first place. Therefore, it is imperative for consumers to understand how to protect their own interests.

5.5. Accuracy

Mason's accuracy issue primarily related to misinformation; errors and omissions that undermine data authenticity, fidelity, and accuracy. However, when it comes to the big data era, privacy advocates and industry should also be interested in how countermeasures can undermine all three characteristics. For example, obfuscation and disinformation can be employed by users to protect their privacy by shielding activity or information from others.

5.5.1. Misinformation

Mason focused his discussion on misinformation, such as inaccurate or incomplete data, and this issue could not be more relevant today. For example, a common remedy to data breaches, as will be discussed below, is to offer credit

monitoring services. However, merely being told about a potentially material error on one's credit report is not enough as it is often difficult to correct. Under the Fair Credit Reporting Act, firms must "adopt reasonable procedures for meeting the needs of commerce for consumer credit, personnel, insurance, and other information in a manner which is fair and equitable to the consumer, with regard to the confidentiality, accuracy, relevancy, and proper utilization of such information…" (15 U.S.C § 1681 (b)).

Those "reasonable procedures" however, often result in significant errors that can require consumers to expend great energy, often to no avail, to resolve grave errors on their credit reports. More than one in five consumers have a "potentially material error" on their credit report (Klein, 2017). The errors may not originate from the reporting agency since they merely collect what data they are given, as was the case when the Social Security Administration mistakenly lists 6,000 people a year as deceased. One consumer did not discover that her credit report indicated she was deceased until she went to apply for an auto loan (T. Anderson, 2017).

5.5.2. Obfuscation

Given the overwhelming number of ways that one's privacy and security can be violated, it is certainly understandable why many resign themselves to living in what can appear to be an inescapable panopticon. Yet, for those who seek out solutions, there are several tools available to increase user anonymity and obfuscate their activity (Bazzell, 2019).

For example, Howe (2015) surveyed several countermeasures that could be used to mitigate various contemporary threats to privacy in both the digital and physical worlds. TrackMeNot (http://trackmenot.io/) was designed to limit the effectiveness of collecting a user's search engine history by shrouding each query amongst additional searches, hiding the true search amongst useless noise. *I Like What I See* (https://github.com/sklise/i-like-what-i-see), was developed to automatically click every instance of the word "like" present on a given website, preventing others from knowing a Facebook user's true interests. *ScareMail* (https://bengrosser.com/projects/scaremail/), another Chrome extension, appends nonsensical text to emails that is expected to be flagged by surveillance monitoring systems, such as the National Security Agency's PRISM and XKeyscore programs.

The *Facial Weaponization Suite* (http://zachblas.info/works/facial-weaponization-suite/) attempts to render facial recognition useless by wearing masks constructed based upon aggregated facial data. The *Invisible* kit (http://biogenfutur.es/) is an open source project consisting of two sprays, one designed to remove as much DNA as possible, and the other to obfuscate any that remains. Each of these highlighted countermeasures not only limit the usable information that is shared with certain systems or people, but intentionally seek to undermine and sabotage the intended purpose of collecting such data.

5.5.3. Disinformation

In addition to obfuscation, users can also employ disinformation to increase privacy (Whang, 2012). Alexander & Smith (2011, p. 58) defined disinformation as "intentional deception in communications scenarios" and described two primary types: destructive and constructive. Destructive disinformation would involve removing key information through redaction, whereas constructive might involve adding false information to a document. A simple example of disinformation would include using a fictitious persona to create an online account or subscribing to magazines to populate databases with inaccurate information as to who resides at a physical address.

Although obfuscation and disinformation techniques can increase one's privacy, employing them could potentially cause algorithms to be less reliable, undermining the value and effectiveness of big data. Therefore, organizations who desire a high level of accuracy must be aware that a growing percentage of consumer data will no longer be trustworthy.

5.6. Property

Mason's conceptualization of the property issue focused primarily on questions related to data ownership, equitable exchange, and access to transmission channels. In the big data era, the reliance on cloud storage and Bring Your Own Device (BYOD) policies have further complicated the already difficult ownership issue. Secondary information use has resulted in data exchanges that are largely inequitable for the consumer. Debates over net neutrality also threaten to impede access to electronic information and resources.

5.6.1. Bring Your Own Device

Instead of issuing company-owned equipment, such as smartphones, laptops, and tablets to employees, organizations have largely adopted the BYOD approach (Shim, Mittleman, Welke, French, & Guo, 2013). While providing several benefits (e.g., accessibility, convenience, flexibility, employee satisfaction, productivity, innovation, and cost-savings),

this phenomenon also leads to several privacy and security implications, especially as it pertains to data breaches stemming from lost and stolen devices (Garba, Armarego, Murray, & Kenworthy, 2015).

However, as Mason pointed out, there are even cloudier situations that have no obvious answers. For example, who owns the data stored on an employee's personal phone when it is used for business purposes? If the company owns the data, how can they access it if they do not own the device? Further, should organizations be able to compel employees to install applications on employee owned devices?

With these questions in mind, it is easy to see how the BYOD approach also leads to numerous legal implications. Which laws and regulations apply to company BYOD policies largely depend on the sector the business is in and type of data collected. Companies need to sift through the data breach notification laws at the state and federal levels, state and federal laws and regulations on data security, international data protection laws, legal procedures that relate to eDiscovery, confidentiality obligations, contractual obligations, trade secret protection, and employment law related issues (Privacy Rights Clearninghouse, 2014).

5.6.2. Right to Share

Secondary information use refers to "the use of personal information for other purposes subsequent to the original transaction between an individual and an organization when the information was collected" (Culnan, 1993, p. 342). However, since terms and conditions are difficult and time consuming to read (McDonald & Cranor, 2008; Milne et al., 2006), users are unlikely to be fully informed as to what the organization can do with their data.

Further complicating this issue is the third-party doctrine, which essentially eliminates a user data rights, while granting organizations the "right to share" consumer data as they please (President's Council of Advisors on Science and Technology, 2014). The third-party doctrine is the legal proposition that people are not entitled to an expectation of privacy for information they voluntarily give to third parties. Therefore, the third-party doctrine currently applies to any data stored or processed through cloud services. Considering the digitization of every aspect of society, coupled with the shift back to centralized computing, users are left with little to no choice in preserving their privacy rights.

For example, in *Katz v. United States*, the Supreme Court proclaimed that "what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection" (Thompson, 2014). Numerous subsequent cases explained the relatively limited reach of the doctrine in the pre-digital age. The debate as to whether the third-party doctrine unnecessarily constricts the privacy rights of Americans has continued well into the digital age.

5.7. Accessibility

Mason's original concern with respect to accessibility was that poor and disenfranchised communities would fall further behind due to limited access to information. While this fear did materialize, some could argue that the ease in which users can access information about others carries grave threats to personal privacy and safety.

5.7.1. Access to Knowledge

The Internet has drastically increased the amount of information available, yet many areas still lack access to reliable broadband connectivity due to infrastructure and transmission costs. Several efforts, such as OneWeb and SpaceX's Starlink, are currently underway to reduce these costs and provide global access to broadband Internet by placing massive satellite constellations in much lower orbits to achieve lower latency than existing satellite Internet service providers (del Portillo, Cameron, & Crawley, 2018). As primary and secondary education continues to shift toward digital learning through laptops and tablets, the accessibility gap between the haves and have nots will only continue to grow. The accessibility issue has never been more evident than during the 2020 COVID-19 pandemic. With educational institutions and businesses suddenly forced to transition to online delivery and work-from-home conditions, those without a reliable Internet connection or access to necessary hardware and software suffered severe hardships.

5.7.2. Too Much Access?

> *"Just because something is publicly accessible does not mean that people want it to be publicized" (Boyd, 2010).*

The technology that spawned the big data phenomenon has also made it possible for personal and sensitive information to be easily collected and disclosed on the Internet. When done intentionally, this form of cyberbullying behavior is commonly referred to as "doxing" due to the release of documents or information without the victim's consent (Douglas, 2016; Li, 2018). Such information is commonly obtained through open-source intelligence (OSINT) gathering

techniques. Hackers regularly use OSINT to research individuals associated with their target organizations to increase the effectiveness of cyberattacks (Hayes & Cappa, 2018).

These same techniques have also been used in "swatting" attacks where someone calls the police to make a false report, such as claiming that a hostage situation is taking place at the target's residence, to cause a heavy police response (Jaffe, 2016; Li, 2018). The response typically involves deploying heavily armed SWAT (special weapons and tactics) teams, hence the name. Swatting attacks have been made against several celebrities, politicians, and other public figures, such as cybersecurity journalist Brian Krebs. The attack against Krebs was in retaliation for his story that identified a group of cybercriminals responsible for doxing public officials (Vaas, 2013, 2016). While these incidents were resolved without major harm, others have not been so fortunate. In 2015, the police chief for Sentinel, Oklahoma was shot by a homeowner when responding to a bomb threat called in by someone else (Slipke, 2015). In 2017, a dispute between Casey Viner and Shane Gaskill over an online video game resulted in an innocent man, Andrew Finch, being shot and killed by police as he exited his home in Wichita, Kansas (Jaffe, 2020).

Other examples can be found in the documentary series *Don't F**k with Cats: Hunting An Internet Killer* (Lewis, 2019). Concerned citizens took it upon themselves to track down those responsible for producing and uploading several viral videos showing a man torturing and killing kittens. Members of a Facebook group conducted extensive OSINT activities and identified several potential suspects. A social media firestorm was waged against one of the early suspects after his information was doxed among the Facebook group. It was later determined that the suspect was not involved, but not before he tragically committed suicide under the weight of false accusations. While their efforts were underway, a group member who had used a pseudonym on her Facebook profile was sent a link to an unsettling video that showed someone walking through the Las Vegas casino where she worked. Although the group eventually identified the individual responsible for the videos, their efforts were not successful in preventing Luka Magnotta from escalating his crimes. He continued to post additional videos until he was ultimately arrested and convicted of murdering Jun Lin, an international student at Concordia University, in 2012.

As these instances demonstrate, an affected individual or organization might be completely unaware that anyone is targeting them or that the sensitive information was available online until it has already resulted in harm. Thus, enhanced accessibility to data can also lead to disturbing and heartbreaking outcomes.

5.7.3. Unauthorized Access

While the personal computer ushered in distributed computing in the 1980s, emerging technologies are reverting to centralized computing due to the processing power needed to leverage big data. This trend increases the need for adequate connectivity for those who wish to adopt such systems. However, centralized databases also introduce several issues. For example, the push to convert patient data into Electronic Medical Records (EMRs) and increase access through health information exchanges can lead to security and privacy issues (Angst, 2010; Angst & Agarwal, 2009; Burns, Young, Ellis, Courtney, & Roberts, 2015). As we continue to desegregate data in the name of accessibility, we amplify breach magnitude and severity.

Although there is a clear need for anonymity (Bellaby, 2018), anonymization methods and policies, such as statistical disclosure control, are routinely proven inadequate in protecting consumer privacy in the big data era (Riederer, Kim, Chaintreau, Korula, & Lattanzi, 2016). In 1997, when Massachusetts' Group Insurance Commission decided to release records of hospital visits to researchers, Massachusetts governor William Weld gave reassurances that the data was anonymized and would not violate patient privacy (N. Anderson, 2009; Ohm, 2010). Sensing a challenge, Latanya Sweeney leveraged her research on U.S. Census and voter registration data (Sweeney, 2000), managed to identify Weld's health information, and his diagnoses and prescriptions to him.

Narayanan and Shmatikov (2008) demonstrated how users could be identified when using two separate data sources. Netflix, as part of their Netflix Prize data mining contest, had released over 100 million of what they believed to be sufficiently anonymized movie ratings from approximately 500,000 subscribers. When Narayanan and Shmatikov analyzed the ratings with similar data from Internet Movie Database (IMDb), they were able to identify anonymized Netflix ratings based upon public IMDb ratings. In discussing the implications of their study, Narayanan and Shmatikov (2008, p. 123) explained how seemingly worthless information could impact an individual's privacy:

> *First, his political orientation may be revealed by his strong opinions about "Power and Terror: Noam Chomsky in Our Times" and "Fahrenheit 9/11," and his religious views by his ratings on "Jesus of Nazareth" and "The Gospel of John." Even though one should not make inferences solely from someone's movie preferences, in many workplaces and social settings opinions about movies with predominantly gay themes such as "Bent" and "Queer as folk" (both present and rated in this person's Netflix record) would*

> *be considered sensitive. In any case, it should be for the individual and not for Netflix to decide whether to reveal them publicly.*

Successful reidentification also occurred when researchers at the Whitehead Institute were interested in assessing the risk of sharing anonymized DNA data (Gymrek, McGuire, Golan, Halperin, & Erlich, 2013; Van Rijmenam, 2014). The team found that DNA data sets that have been stripped of identifiers can still be attributed to surnames by examining specific genetic data and then cross-referencing freely available data sources on the Internet. What is particularly dangerous is that individuals can jeopardize relatives by sharing their own DNA, all without informing or obtaining consent. Given the rise of genetic testing services that are marketed to the average consumer, such as 23andMe and Ancestry.com, these findings have serious implications for the privacy and security of genetic information.

5.7.4. Data Breaches and Reporting

With more and more data stored, used, and transferred for business purposes, data breaches are increasingly common. Despite the staggering number and scope of these data breaches, significant challenges for consumers and corporations are created due to a lack of consistent data protection framework (Tschider, 2015). Much like other areas regarding data and privacy, federal laws related to cybersecurity are sector specific. These laws include provisions in Gramm-Leach-Bliley and the HIPPA Breach Notification Rule.

Unfortunately, a comprehensive federal data breach notification and protection law does not seem likely anytime soon. Numerous bills, such as the Data Accountability and Trust Act (Rush, 2019), have been introduced recently in Congress, but have failed to gain traction. This bill, H.R. 1282 would require the Federal Trade Commission (FTC) to require certain businesses and organizations to establish security practices for the treatment and protection of personal information and provide specified notice and offer credit-monitoring services in the event of a breach.

Adding to the complication of sector specific laws, all 50 states currently have their own form of data breach notification statutes. Some of the definitions and requirements of the laws that can vary greatly include: 1) what type personally identifiable information triggers a breach notification obligation to individuals, 2) what form of data that information is in, 3) when notice must be given to individuals, 4) what form of that notice is permitted, 5) what must information about the breach must be included in the notice, 6) what states require notification to state agencies, and 7) when notification to the credit reporting agencies is required (Millar & Marshall, 2019).

6. Discussion

6.1. Contributions to Literature

Although Mason's PAPA is just as relevant today as it was in 1986, we have highlighted several modern examples that we must consider as we usher in a new wave of technological advancement. Our extension also distinguished three issues that were deeply embedded within Mason's framework but deserve increased attention. First, we examined how behavioral surveillance has become a significant threat to individual privacy and liberty. Second, we illustrated how data errors and misinterpretations can lead to devastating outcomes, especially when applied in high stakes situations. Third, we also explored how both industry and government have failed to properly regulate unethical behavior. Lastly, we revisited the four original issues comprising Mason's PAPA by examining how technology has not only impacted our relationship with information, but more importantly one another.

6.2. Practical Implications

As we enter the 2020s, the mass proliferation of the Internet of Things (IoT) will likely be its most defining characteristic. Nearly every category of electronic device sold today has built-in connectivity, essentially eliminating any remaining firewall between the digital and physical worlds. We encourage professionals in all fields to assess the ethical implications of their work. Identifying potential issues during the infancy of any new endeavor allows for safeguards to be embedded during the development process. Employing an "ethical by design" approach to technology development would also better protect stakeholders by ensuring that adequate thought has been put into the ethical implications for each element of the system (Beard & Longstaff, 2018).

6.3. Legal Implications

With the confusing web of state and federal privacy laws, the onus remains primarily on individuals and corporations to reevaluate their behavior through the interpretation of statistical models and its application to governance. Corporations and industries have a direct impact on the behavior of individual consumers and should take steps to educate consumers beyond bare minimum informed consent requirements on how their data is collected and used. Individuals

must be able to understand the value their data has to corporations and how their everyday behavior contributes to the enhancement of that value. Additionally, corporations and even consumers still must advocate for strengthened privacy laws that protect consumers from forms of exploitation using their data.

7. Conclusion

We have provided several contemporary examples that continue to demonstrate the importance of Mason's seminal work. Each issue that Mason identified in his PAPA framework over 30 years ago is just as critical, if not more so, in the big data era. We also proposed our BIG PAPA extension to address emerging issues with respect to behavioral surveillance, interpretation of analytical models, and much needed governance. It is our hope that greater attention will be paid to these areas in both research and practice. We simply cannot afford to continue to ignore Mason's warnings any longer.

8. References

Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., Diffie, W., … Weitzner, D. J. (2015). Keys under doormats: Mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity*, *1*(1), 69–79. https://doi.org/10.1093/cybsec/tyv009

Alexander, J. M., & Smith, J. M. (2011). Disinformation: A Taxonomy. *IEEE Security & Privacy Magazine*, *9*(1), 58–63. https://doi.org/10.1109/MSP.2010.141

Anderson, N. (2009, September 8). "Anonymized" data really isn't—and here's why not. Retrieved from https://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/

Anderson, T. (2017, January 13). You may be dead: Every year, Social Security falsely lists 6,000 people as deceased. Retrieved from https://www.cnbc.com/2017/01/12/social-security-falsely-lists-6000-people-a-year-as-dead.html

Andrejevic, M. (2004). The work of watching one another: Lateral surveillance, risk, and governance. *Surveillance and Society*, *2*(4), 479–497. https://doi.org/10.24908/ss.v2i4.3359

Angst, C. M. (2010). Protect My Privacy or Support the Common-Good? Ethical Questions About Electronic Health Information Exchanges. *Journal of Business Ethics*, *90*(S2), 169–178. https://doi.org/10.1007/s10551-010-0385-5

Angst, C. M., & Agarwal, R. (2009). Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly*, *33*(2), 339–370.

Ariyaluran Habeeb, R. A., Nasaruddin, F., Gani, A., Amanullah, M. A., Abaker Targio Hashem, I., Ahmed, E., & Imran, M. (2019). Clustering-based real-time anomaly detection—A breakthrough in big data technologies. *Transactions on Emerging Telecommunications Technologies*, *e3647*(April), 1–27. https://doi.org/10.1002/ett.3647

Ausloos, J. (2012). The 'Right to be Forgotten' – Worth remembering? *Computer Law & Security Review*, *28*(2), 143–152. https://doi.org/10.1016/j.clsr.2012.01.006

Awad, N. F., & Krishnan, M. S. S. (2006). The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quarterly*, *30*(1), 13. https://doi.org/10.2307/25148715

Barber, G. (2019, March 14). San Francisco Bans Agency Use of Facial-Recognition Tech. Retrieved from https://www.wired.com/story/san-francisco-bans-use-facial-recognition-tech/

Bates, A. (2016). *Stingray: A New Frontier in Police Surveillance*. Washington, D.C. Retrieved from https://www.cato.org/publications/policy-analysis/stingray-new-frontier-police-surveillance?

Bazzell, M. (2019). *Extreme Privacy: What It Takes to Disappear in America*. (A. Martin, M. S. Williams, & J. Engstrom, Eds.).

Beard, M., & Longstaff, S. (2018). *Ethical by Design: Principles for Good Technology*. Sydney, Australia.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly: Management Information Systems*, *35*(4), 1017–1041. https://doi.org/10.2307/41409971

Bélanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *The Journal of Strategic Information Systems*, *11*(3–4), 245–270. https://doi.org/10.1016/S0963-8687(02)00018-5

Bellaby, R. W. (2018). Going dark: anonymising technology in cyberspace. *Ethics and Information Technology*, *20*(3), 189–204. https://doi.org/10.1007/s10676-018-9458-4

Benbouzid, B. (2019). To predict and to manage. Predictive policing in the United States. *Big Data & Society*, *6*(1), 1–13. https://doi.org/10.1177/2053951719861703

Bergstrom, C. T., & West, J. D. (2020). Calling Bullshit: Frequently Asked Questions. Retrieved from https://callingbullshit.org/FAQ.html

Boritz, J. E., & No, W. G. (2011). E-Commerce and Privacy: Exploring What We Know and Opportunities for Future Discovery. *Journal of Information Systems*, *25*(2), 11–45. https://doi.org/10.2308/isys-10090

Boyd, D. (2010, March 13). Making Sense of Privacy and Publicity. Retrieved from http://www.danah.org/papers/talks/2010/SXSW2010.html

Boyne, S. (2016). Stingray Technology, the Exclusionary Rule, and the Future of Privacy: A Cautionary Tale. *West Virginia Law Review*, *119*(3), 915–939. https://doi.org/10.2139/ssrn.2911844

Brey, P. (2004). Ethical aspects of facial recognition systems in public places. *Journal of Information, Communication and Ethics in Society*, *2*(2), 97–109. https://doi.org/10.1108/14779960480000246

Burgess, M. (2019, January 21). What is GDPR? The summary guide to GDPR compliance in the UK. Retrieved from https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018

Burns, A., Young, J. A., Ellis, T. S., Courtney, J. F., & Roberts, T. L. (2015). Exploring the Role of Contextual

Integrity in Electronic Medical Record (EMR) System Workaround Decisions: An Information Security and Privacy Perspective. *AIS Transactions on Human-Computer Interaction*, *7*(3), 142–165.

Chen, C. L. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Chiu, A. (2019, December 12). She installed a Ring camera in her children's room for 'peace of mind.' A hacker accessed it and harassed her 8-year-old daughter. Retrieved from https://www.washingtonpost.com/nation/2019/12/12/she-installed-ring-camera-her-childrens-room-peace-mind-hacker-accessed-it-harassed-her-year-old-daughter/

Cowan, J., & Singer, N. (2020, January 3). How California's New Privacy Law Affects You. Retrieved from https://www.nytimes.com/2020/01/03/us/ccpa-california-privacy-law.html

Cox, K. (2019, August 6). Police can get your Ring doorbell footage without a warrant, report says. Retrieved from https://arstechnica.com/tech-policy/2019/08/police-can-get-your-ring-doorbell-footage-without-a-warrant-report-says/

Culnan, M. J. (1993). "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly*, (September), 341–363.

del Portillo, I., Cameron, B. G., & Crawley, E. F. (2018). A Technical Comparison of Three Low Earth Orbit Satellite Constellation Systems to Provide Global Broadband. In *69th International Astronautical Congress* (pp. 1–15). Bremen, Germany.

Desjardins, J. (2015, July 29). Order From Chaos: How Big Data Will Change the World. Retrieved from https://www.visualcapitalist.com/order-from-chaos-how-big-data-will-change-the-world/

Dinev, T., Hart, P., & Mullen, M. R. (2008). Internet privacy concerns and beliefs about government surveillance – An empirical investigation. *The Journal of Strategic Information Systems*, *17*(3), 214–233. https://doi.org/10.1016/j.jsis.2007.09.002

Douglas, D. M. (2016). Doxing: a conceptual analysis. *Ethics and Information Technology*, *18*(3), 199–210. https://doi.org/10.1007/s10676-016-9406-0

Federal Trade Commission. (2019, July 22). Equifax to Pay $575 Million as Part of Settlement with FTC, CFPB, and States Related to 2017 Data Breach. Retrieved from https://www.ftc.gov/news-events/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related

Feiner, L. (2019, December 4). A federal privacy law is starting to crystallize, but Democrats and Republicans can't agree on how to do it. Retrieved from https://www.cnbc.com/2019/12/04/a-federal-privacy-law-is-starting-to-crystallize-senators-remain-divided-over-details.html

Fenton, N., & Neil, M. (2011). Avoiding probabilistic reasoning fallacies in legal practice using Bayesian networks. *Australian Journal of Legal Philosophy*, *36*(2011), 114–150.

Gantz, J., & Reinsel, D. (2012). *THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. Framingham, Massachusetts.

Garba, A. B., Armarego, J., Murray, D., & Kenworthy, W. (2015). Review of the Information Security and Privacy Challenges in Bring Your Own Device (BYOD) Environments. *Journal of Information Privacy and Security*, *11*(1), 38–54. https://doi.org/10.1080/15536548.2015.1010985

Grossman, L. (2001, February). Welcome to the Snooper Bowl. *TIME*, *157*(6), 72.

Gupta, D., & Rani, R. (2019). A study of big data evolution and research challenges. *Journal of Information Science*, *45*(3), 322–340. https://doi.org/10.1177/0165551518789880

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying Personal Genomes by Surname Inference. *Science*, *339*(6117), 321–324. https://doi.org/10.1126/science.1229566

Hack, M. (2016). The implications of Apple's battle with the FBI. *Network Security*, *2016*(7), 8–10. https://doi.org/10.1016/S1353-4858(16)30068-X

Hayes, D. R., & Cappa, F. (2018). Open-source intelligence for risk assessment. *Business Horizons*, *61*(5), 689–697. https://doi.org/10.1016/j.bushor.2018.02.001

Hern, A. (2018, January 28). Fitness tracking app Strava gives away location of secret US army bases. *The Guardian*. Retrieved from https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

Hill, K. (2020, January 18). The Secretive Company That Might End Privacy as We Know It. Retrieved from https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

Howe, D. C. (2015). Surveillance Countermeasures: Expressive Privacy via Obfuscation. *A Peer-Reviewed Journal About*, *4*(1), 88–98. https://doi.org/10.7146/aprja.v4i1.116108

Innocence Project. (2020). DNA Exonerations in the United States. Retrieved from https://www.innocenceproject.org/dna-exonerations-in-the-united-states/

Introna, L. D., & Wood, D. (2004). Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems. *Surveillance & Society*, *2*(2/3), 177–198.

Jaffe, E. M. (2016). *Swatting: The New Cyberbullying Frontier after Elonis v. United States*. *Drake Law Review* (Vol. 64).

Jaffe, E. M. (2020). From Terrorists to Trolls: Expanding Web Host Liability for Live-Streaming, Swatting, and Cyberbullying. *Boston University Journal of Science and Technology Law*, *26*(2), 51–66.

Joh, E. E. (2016). The new surveillance discretion: Automated suspicion, big data, and policing. *Harvard Law & Policy Review*, *10*, 15–42.

Jones, M. L. (2016). *Ctrl + Z: The Right to Be Forgotten*. New York, New York: New York University Press.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good Practices. In *2013 Sixth International Conference on Contemporary Computing (IC3)* (pp. 404–409). IEEE. https://doi.org/10.1109/IC3.2013.6612229

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. Yonkers-on-Hudson, NY: World Book Company.

Kepner, J., Gadepally, V., Michaleas, P., Schear, N., Varia, M., Yerukhimovich, A., & Cunningham, R. K. (2014). Computing on masked data: a high performance method for improving big data veracity. In *2014 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1–6). IEEE. https://doi.org/10.1109/HPEC.2014.7040946

Kirk, M. (2014). *United States of Secrets*. United States: PBS. Retrieved from https://www.pbs.org/wgbh/frontline/film/united-states-of-secrets/

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data and Society*, *3*(1), 1–10. https://doi.org/10.1177/2053951716631130

Klein, A. (2017, September 28). The real problem with credit reports is the astounding number of errors. Retrieved from https://www.brookings.edu/research/the-real-problem-with-credit-reports-is-the-astounding-number-of-errors/

Kornstein, S. (2015, June 29). The Rise of Mobile Phones: 20 Years of Global Adoption. Retrieved from https://blog.cartesian.com/the-rise-of-mobile-phones-20-years-of-global-adoption

Krämer, J., Wiewiorra, L., & Weinhardt, C. (2013). Net neutrality: A progress report. *Telecommunications Policy*, *37*(9), 794–813. https://doi.org/10.1016/j.telpol.2012.08.005

Kuempel, A. (2016). The invisible middlemen: A critique and call for reform of the data broker industry. *Northwestern Journal of International Law and Business*, *36*(1), 207–234.

Kurwa, R. (2019). Building the Digitally Gated Community: The Case of Nextdoor. *Surveillance & Society*, *17*(1/2), 111–117. https://doi.org/10.24908/ss.v17i1/2.12927

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group Research Note*, *6*(70).

Lashmar, P. (2017). No More Sources?: The impact of Snowden's revelations on journalists and their confidential sources. *Journalism Practice*, *11*(6), 665–688. https://doi.org/10.1080/17512786.2016.1179587

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, *40*(2), 133–146. https://doi.org/10.1016/S0378-7206(02)00043-5

Lewis, M. (2019). *Don't F**k With Cats: Hunting An Internet Killer*. United States: Netflix.

Li, L. B. (2018). Data Privacy in the Cyber Age: Recommendations for Regulating Doxing and Swatting. *Federal Communications Law Journal*, *70*, 317.

Lowry, P. B., Dinev, T., & Willison, R. (2017). Why security and privacy research lies at the centre of the information systems (IS) artefact: proposing a bold research agenda. *European Journal of Information Systems*, *26*(6), 546–563. https://doi.org/10.1057/s41303-017-0066-x

Lucas, G. (2005). *Star Wars: Episode III - Revenge of the Sith*. United States: Twentieth Century Fox.

Lum, C., Hibdon, J., Cave, B., Koper, C. S., & Merola, L. (2011). License plate reader (LPR) police patrols in crime hot spots: an experimental evaluation in two adjacent jurisdictions. *Journal of Experimental Criminology*, *7*(4), 321–345. https://doi.org/10.1007/s11292-011-9133-9

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Masden, C. A., Grevet, C., Grinter, R. E., Gilbert, E., & Edwards, W. K. (2014). Tensions in Scaling-up Community Social Media: A Multi-Neighborhood Study of Nextdoor. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 3239–3248). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2557319

Mason, R. O. (1986). Four ethical issues of the information age. *MIS Quarterly*, *10*(1), 5–12.

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for*

the Information Society, *4*(3), 543–568.
McLeod, A., Savage, A., & Simkin, M. G. (2018). The Ethics of Predatory Journals. *Journal of Business Ethics*, *153*(1), 121–131. https://doi.org/10.1007/s10551-016-3419-9
Medine, D., Brand, R., Cook, E. C., Dempsey, J., & Wald, P. (2014a). *Report on the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act*. Washington, D.C. Retrieved from https://www.pclob.gov/library/702-Report.pdf
Medine, D., Brand, R., Cook, E. C., Dempsey, J., & Wald, P. (2014b). *Report on the Telephone Records Program Conducted under Section 215 of the USA PATRIOT and on the Operations of the Foreign Intelligence Surveillance Court*. Washington, D.C.
Millar, S. A., & Marshall, T. P. (2019, April 24). State Data Breach Notification Laws – Overview of Requirements for Responding to a Data Breach – Updated April 2019. Retrieved from https://www.natlawreview.com/article/state-data-breach-notification-laws-overview-requirements-responding-to-data-3
Milne, G. R., Culnan, M. J., & Greene, H. (2006). A Longitudinal Assessment of Online Privacy Notice Readability. *Journal of Public Policy & Marketing*, *25*(2), 238–249. https://doi.org/10.1509/jppm.25.2.238
Mühle, A., Grüner, A., Gayvoronskaya, T., & Meinel, C. (2018). A survey on essential components of a self-sovereign identity. *Computer Science Review*, *30*, 80–86. https://doi.org/10.1016/j.cosrev.2018.10.002
Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *Proceedings - IEEE Symposium on Security and Privacy*, 111–125. https://doi.org/10.1109/SP.2008.33
National Law Enforcement and Corrections Technology Center. (2010). *The Results Are In: Automatic License Plate Reader Technology Leads to Success*. Research Triangle Park, NC.
Newkirk, D. (2018). "Apple: Good Business, Poor Citizen": A Practitioner's Response. *Journal of Business Ethics*, *151*(1), 13–16. https://doi.org/10.1007/s10551-016-3397-y
Nextdoor. (2019, December 5). Privacy Policy. Retrieved from https://legal.nextdoor.com/us-privacy-policy-2020/
Nickelodeon Consumer Privacy Litigation, 827 F.3d 262, 293 (3rd Cir., 2016)
Nicks, D. (2016, May 24). Mark Zuckerberg Bought Four Houses Just to Tear Them Down. Retrieved from https://money.com/mark-zuckerberg-houses/
Norman, J. (2016). Taking the Sting Out of the Stingray: The Dangers of Cell-Site Simulator Use and the Role of the Federal Communications Commission in Protecting Privacy & Security. *Federal Communications Law Journal*, *68*(1), 139-0_8.
Noyes, K. (2015, June 25). Scott McNealy on privacy: You still don't have any. Retrieved from https://www.pcworld.com/article/2941052/scott-mcnealy-on-privacy-you-still-dont-have-any.html
O'Neil, C. (2016). *Weapons of Math Destruction*. New York, New York: Crown.
O'Neill, K. (2019, January 15). Facebook's "10 Year Challenge" Is Just a Harmless Meme—Right? Retrieved from https://www.wired.com/story/facebook-10-year-meme-challenge/
Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, *57*(6), 1701–1777.
Otto, P. N., Anton, A. I., & Baumer, D. L. (2007). The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information. *IEEE Security & Privacy Magazine*, *5*(5), 15–23. https://doi.org/10.1109/msp.2007.126
Ozer, M. M. (2010). *Assessing the Effectiveness of the Cincinnati Police Department's Automatic License Plate Reader System within the Framework of Intelligence-Led Policing and Crime Prevention Theory*. University of Cincinnati.
Pavlou, P. A. (2011). State of the information privacy literature: where are we now and where should we go. *MIS Quarterly*, *35*(4), 977–988.
Pell, S. K., & Soghoian, C. (2014). Your secret stingray's no secret anymore: The vanishing government monopoly over cell phone surveillance and its impact on national security and consumer privacy. *Harvard Journal of Law & Technology*, *28*(1), 1–76.
Peslak, A. R. (2006). PAPA revisited: A current empirical study of the Mason framework. *Journal of Computer Information Systems*, *46*(3), 117–123. https://doi.org/10.1080/08874417.2006.11645905
Peterson, A., Yahr, E., & Warrick, J. (2014, September 1). Leaks of nude celebrity photos raise concerns about security of the cloud. Retrieved from https://www.washingtonpost.com/politics/leaks-of-nude-celebrity-photos-raise-concerns-about-security-of-the-cloud/2014/09/01/59dcd37e-3219-11e4-8f02-03c644b2d7d0_story.html
Power, D. J. (2016). "Big Brother" can watch us. *Journal of Decision Systems*, *25*(sup1), 578–588. https://doi.org/10.1080/12460125.2016.1187420
President's Council of Advisors on Science and Technology. (2014). *Big Data and Privacy: A Technological Perspective*. Washington, D.C.

Privacy Rights Clearninghouse. (2014, October 1). Brind Your Own Device (BYOD)…at Your Own Risk. Retrieved from https://privacyrights.org/consumer-guides/bring-your-own-device-byod-your-own-risk

Rahimian, R., & Kelly, A. (2019, December 15). The Decade Tech Lost Its Way. Retrieved from https://www.nytimes.com/interactive/2019/12/15/technology/decade-in-tech.html

Reding, V. (2012). The EU Data Protection Reform 2012 : Making Europe the Standard Setter for Modern Data Protection Rules in the Digital Age. In *Innovation Conference Digital, Life, Design* (pp. 1–6). Munich, Germany. Retrieved from http://europa.eu/rapid/press-release_SPEECH-12-26_en.htm

Richards, N. M. (2012). The Dangers of Surveillance. *Harvard Law Review*, *126*(7), 1934–1965.

Richards, N. M., & King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review*, *49*(2), 393–432.

Riederer, C., Kim, Y., Chaintreau, A., Korula, N., & Lattanzi, S. (2016). Linking Users Across Domains with Location Data. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16* (pp. 707–719). New York, New York, USA: ACM Press. https://doi.org/10.1145/2872427.2883002

Ring under fire over weakness in video device security. (2020). *Network Security*, *2020*(1), 1–2. https://doi.org/10.1016/S1353-4858(20)30001-5

Rosen, J. (2012). The Right to be Forgotten. *Stanford Law Review*, *64*, 88–92.

Rush, B. L. Data Accountability and Trust Act (2019). Washington, D.C.: U.S. Congress.

Sanchez-Garrido, B. (2016). Social media's criminal element. *Risk Management*, *63*(1), 8–10.

Schneier, B. (2009, December 9). My Reaction to Eric Schmidt. Retrieved from https://www.schneier.com/blog/archives/2009/12/my_reaction_to.html

Shim, J. P., Mittleman, D., Welke, R., French, A. M., & Guo, J. C. (2013). Bring Your Own Device (BYOD): Current status, issues, and future directions. In *19th Americas Conference on Information Systems, AMCIS 2013 - Hyperconnected World: Anything, Anywhere, Anytime* (Vol. 1, pp. 595–596).

Slipke, D. (2015). Court document reveals more about Sentinel, OK, bomb threat. Retrieved from https://oklahoman.com/article/5386857/court-document-reveals-more-about-sentinel-ok-bomb-threat

Smith, H. J., Dinev, T., & Xu, H. (2011). Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly*, *35*(4), 989–1015. https://doi.org/10.2307/41409970

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, *20*(2), 167–196. https://doi.org/10.2307/249477

Smolan, R., & Erwitt, J. (2012). *The Human Face of Big Data*. Sausalito, California: Against All Odds Productions.

Snow, J. (2018). Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots. Retrieved May 23, 2019, from https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

Solove, D. (2015, November 13). The Growing Problems with the Sectoral Approach to Privacy Law. Retrieved from https://teachprivacy.com/problems-sectoral-approach-privacy-law

Solove, D. J. (2007). "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review*, *44*(May), 1–23. https://doi.org/10.2139/ssrn.998565

Sprenger, P. (1999, January 26). Sun on Privacy: "Get Over It." Retrieved from https://www.wired.com/1999/01/sun-on-privacy-get-over-it/

Stewart, K. a., & Segars, A. H. (2002). An Empirical Examination of the Concern for Information Privacy Instrument. *Information Systems Research*, *13*(1), 36–49. https://doi.org/10.1287/isre.13.1.36.97

Stoycheff, E. (2016). Under surveillance: Examining Facebook's spiral of silence effects in the wake of NSA internet monitoring. *Journalism & Mass Communication Quarterly*, *93*(2), 296–311. https://doi.org/10.1177/1077699016630255

Streitfeld, D. (2013, March 12). Google Concedes That Drive-By Prying Violated Privacy. Retrieved from https://www.nytimes.com/2013/03/13/technology/google-pays-fine-over-street-view-privacy-breach.html

Sweeney, L. (2000). *Simple demographics often identify people uniquely* (Data Privacy No. 3). Pittsburg, PA. Retrieved from http://dataprivacylab.org/projects/identifiability/paper1.pdf

Thompson, R. M. (2014). *The Fourth Amendment Third-Party Doctrine*. Washington, D.C.

Tobin, A., & Reed, D. (2017). *The Inevitable Rise of Self-Sovereign Identity: A white paper from the Sovrin Foundation*. Retrieved from https://sovrin.org/wp-content/uploads/2017/06/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf

Tschider, C. A. (2015). Experimenting with Privacy: Driving Efficiency Through a State-Informed Federal Data Breach Notification and Data Protection Law. *Tulane Journal of Technology and Intellectual Property*, *18*, 45–81.

U.S. Privacy Protection Study Commission. (1977). *Personal Privacy in an Information Society*.

Vaas, L. (2013, March 17). Hackers launch DDoS attack on security blogger's site, send SWAT team to his home. Retrieved from https://nakedsecurity.sophos.com/2013/03/17/swat-ddos-brian-krebs/

Vaas, L. (2016, July 15). Serial swatter, stalker and doxer Mir Islam given 2 years prison. Retrieved from https://nakedsecurity.sophos.com/2016/07/15/serial-swatter-stalker-and-doxer-mir-islam-given-2-years-prison/

Van Rijmenam, M. (2014). *Think Bigger: Developing a Successful Big Data Strategy for Your Business*. New York, New York: AMACOM.

Walker, R. K. (2012). The Right to Be Forgotten. *Hastings Law Journal*, *64*(2), 257–286.

Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review*, *4*(5), 193–220.

West, E. (2019). Amazon: Surveillance as a service. *Surveillance and Society*, *17*(1–2), 27–33. https://doi.org/10.24908/ss.v17i1/2.13008

Whang, S. E. (2012). *Data Analytics: Integration and Privacy*. Stanford University.

Wheelan, C. (2013). *Naked Statistics: Stripping the Dread from the Data*. New York, New York: WW Norton & Company.

Zmud, J., Wagner, J., Moran, M., & George, J. P. (2016). *License Plate Reader Technology: Transportation Uses and Privacy Risks*. College Station, Texas. Retrieved from https://scholarship.law.tamu.edu/facscholar%0Ahttps://scholarship.law.tamu.edu/facscholar/923

Author Biographies

Jacob A. Young is an assistant professor of management information systems in the Foster College of Business and the Director of the Center for Cybersecurity at Bradley University. He earned his D.B.A. in Computer Information Systems from Louisiana Tech University. Dr. Young conducts research on privacy, security, and anonymity issues related to information systems with a primary focus on anonymous whistleblowing systems. He serves as the Senior Advisor on Cybersecurity at the National Whistleblower Center in Washington, D.C. His work has been published in *AIS Transactions on Human-Computer Interaction*, *Communications of the Association for Information Systems*, *Journal of Information Security Education*, the *Journal of the Midwest Association for Information Systems*, the *DePaul Business & Commercial Law Journal*, and other journals and conference proceedings.

Tyler J. Smith is an assistant professor of business law in the Foster College of Business at Bradley University. He earned his J.D. from Indiana University Robert H. McKinney School of Law and his LL.M. from Notre Dame Law School. He conducts legal research on information privacy, constitutional law, and labor-management relations. His work has been published in several journals, such as the *Fordham International Law Journal*, the *New York International Law Review*, and the *Indiana International & Comparative Law Review*.

Haoran (Shawn) Zheng is an assistant professor of management information systems in the Foster College of Business at Bradley University. He earned his Ph.D. in Information Systems and Business Analytics from Chapman Graduate School at Florida International University. Dr. Zheng's conducts research in adoption, integration, and assimilation of e-health systems, organizational changes and modern privacy issues related to big data and analytics.

This page intentionally left blank

Journal of the Midwest Association for Information Systems

Volume 2020 | Issue 2 Article 4

Date: 07-31-2020

Advancing Technological State-of-the-Art for GDPR Compliance: Considering Technology Solutions for Data Protection Issues in the Sharing Economy

Gail L. Maunula
University of Turku, gaimau@utu.fi

Abstract

Technology provides solutions that help create and drive growth in the Sharing Economy (SE). From the initial technologies that transformed the age-old practice of sharing into a digital disruption of traditional industries, to mobile-app technology delivering streamlined functionality to users, technology has paved the way. As the business model develops, it is no wonder the SE looks to developers to technologically capitalize on new trends and solve emerging challenges. One such challenge is the need to secure data and comply with data protection laws, such as the EU's General Data Protection Regulation (GDPR). The ongoing struggle to achieve and maintain compliance under the GDPR keeps companies reanalyzing business practices. Often, new compliance gaps emerge, as demonstrated in this research, wherein analysis of SE processing activities uncovers potential privacy, security and data protection concerns related to the platform's disclosure of personal data to end-users. This article invites those from tech fields into the inner workings of theoretical legal research, fostering a symbiotic relationship between the law, technology and industry. This research acknowledges that the path to GDPR compliance meanders through the fields of technology, and juxtaposes this reality with the results of close scrutiny of SE data processing practices to suggest future research paths.

Keywords: General Data Protection Regulation, Sharing Economy, personal data protection, legal compliance, service provider obligations

DOI: 10.17705/3jmwa.000060

1. Introduction

Personal data is at the heart of Sharing Economy (SE) business models. The growth and success of the industry are closely tied to the ability to collect copious amounts of personal data and extract the greatest value from them (Richter, 2019). Personal data allow users, facilitated by Sharing Economy Platforms (SEPs), to gain the trust needed to break down the walls of information asymmetry and non-familiarity that comes with the absence of face-to-face transactions (Seigneur, 2009). Were it not for these personal data, it would be difficult to garner the trust necessary to encourage these consumer-to-consumer interactions (Lutz et al. 2017) that predominantly involve access to physical goods and services (Dakhlia et al. 2016).

However, these personal data would be of little value without the supportive technologies that allow SEPs to collect, store and extract their value. These technologies engender the capabilities needed to fully capitalize on the benefits of personal data (Madsen 1992; Cohen and Kietzman 2014). The SE is not only built on these technological capabilities (Einav et al. 2016), but also driven by the development of new tech tools that extend these capabilities (Krivenchuk and Smutny 2019).

But technology goes a step beyond offering data capture, extraction and analysis capabilities. These technologies form the data privacy and security structures that help SEPs meet legal compliance obligations for the protection of personal data (GDPR Report). Despite the notion that technology challenges the protection of personal data (Seigneur 2009) and exacerbates the struggle to create effective data protection legal regimes (Weiner 2004), both the SE industry and data protection regulators look to technology to meet the demands of compliance, thereby looking to the exact infrastructures that create certain challenges to simultaneously solve those challenges (Room et al. 2018).

When the valuable personal data disseminated through SE transactions originate from users within the European Union (EU), the General Data Protection Regulation (GDPR) is triggered. Personal data protection under the GDPR is achieved by giving data subjects control over their personal data through a set of individual rights (GDPR, Articles 12-21) and imposing fines and penalties for non-compliance (GDPR, Article 84). The GDPR protects the personal data of those considered European data subjects in the hands of an enterprise, regardless of its location. Casting a wide net, this territorial scope places any business offering goods and services or monitoring the behavior of EU data subjects under the GDPR's purview (GDPR, Article 3).

In 2016, the GDPR entered into force as a direct response to technological developments, from the fields of information technology (IT) and information and communications technology (ICT), that challenge the privacy and security of personal data (van den Hoven et al. 2019). Many business enterprises across the globe, including popular SEPs, scrambled to achieve compliance before the GDPR applicability date of 25 May 2018, and many still struggle to achieve compliance (Irwin 2019). Indeed, balancing the affordances of available security and privacy technologies and the quest to achieve profitability (Flavián and Guinalíu 2006) is a constant process. Nevertheless, the GDPR acknowledges that modern technology is a vital pathway to data protection (Tankard 2016).

Two years on, however, there is a need for further research that focuses less on the initial compliance momentum and more on the compliance trajectory. For technology to meet the future compliance needs of any data-driven industry, it is incumbent upon legal researchers to communicate the deeper implications of the GDPR. To date, multidisciplinary research related to data privacy and security technologies, the SE industry and the GDPR focuses on meeting these initial compliance challenges (Urban et al. 2018; Lutz et al. 2017; GDPR Report 2017). These studies presume a platform positioning under the GDPR served by the present industry interpretation of the flow of data during an SE transaction and the requisite legal obligations under this interpretation. However, we are at a stage where researcher attention can be directed toward more detailed scrutiny and interpretation of implications of the GDPR in the SE. As this research shows, a GDPR-focused reinterpretation of this data flow signals the need for tools that effectively protect and secure data beyond the digital platform.

This article investigates a development at the intersection of SE business models, supportive privacy and security technologies and GDPR compliance. More directly, the article analyzes the flow of personal data inherent to SE transactions that results in a weakness in personal data protection for some end-users that may trigger a GDPR response. This data protection weakness arises when SE personal data provided to the SEP is disclosed to its service providers. Personal data disclosure to service providers is integral to the SE business model, so it is not a new function. What is new, is the analysis of weaknesses in data protection that closer scrutiny of this practice reveals and the potential obligations of service providers under the GDPR. This paper signals to tech industries that they should ready themselves for a significant shift in how SEP functions may evolve upon answering these legal questions regarding disclosure of data to service providers and offer an opportunity for technological development from

the perspective of data protection legal requirements (Room et al. 2018). The goal of this combination of theoretical legal research and practical research is to engage tech developers in the early stages of this legal discovery, to identify those technologies that will, once again, bridge the gaps toward compliance.

Following this Introduction, Section 2 explores the nature and handling of personal data in the SE and juxtaposes the industry's data processing realities with the stringent compliance demands of the GDPR. Section 2 offers a concrete example of a data privacy and security scenario that reveals the extension of data protection obligations to service providers. The Section also illuminates two critical data compliance red flags and frames the solution around the position, influence, and technological capabilities of the SEP. Section 3 analyzes the role that technology does and could play in support of the data protection compliance efforts of SE data processors, and considers the way forward for technology development in light of revelations discussed throughout this text. The paper culminates with a Conclusion and Considerations for Future Research that acknowledges the integral part privacy and security-enhancing technologies will play in achieving data protection legal compliance for the SE and other industries that similarly handle personal data. Combined, these elements provide novel research intended to spark future ICT research and achieve the data protection goals envisioned by the GDPR.

2. Sharing Economy Personal Data and its Protection

With modern technology, SEPs open a global forum for sharing that would be nearly impossible in the analog world (Codagnone and Martens 2016). SEPs leverage this technology to encourage and facilitate exchanges between individuals with an abundance of a resource, with those in temporary need of that resource (Horton and Zeckhauser 2016). The SE has social, environmental and economic benefits that have propelled its popularity, growing the industry rapidly since its emergence in 2008 (Kim et al. 2015; Heinrichs, 2013; Botsman and Rogers 2010). Already in 2015, 44% of US adults were familiar with the SE, with 25% utilizing the SE as a consumer or provider (Consumer Intelligence Series 2015). Additionally, revenues from the SE are expected to grow from $15 billion in 2015, to $335 billion in 2025 (Hawksworth and Vaughn 2014).

SEPs seek to facilitate the efficient connection between two end-user groups: service providers (offering services or resources) and end-consumers (utilizing services or resources). Both end-user groups are consumers of the SEP, which provides the technical and social infrastructure used to form their consumer-to-consumer (C2C) relationship (see Figure1). The formation of this triangular relationship requires intimate personal data (Teubner and Flath 2019) that fosters the trust necessary to conclude a transaction that occurs, in part, outside the auspices of the digital space. This unique conglomeration of digital- and physical-world contact heightens the need for reliable technical tools that collect and utilize the appropriate types and amounts of data to facilitate safe transactions efficiently.

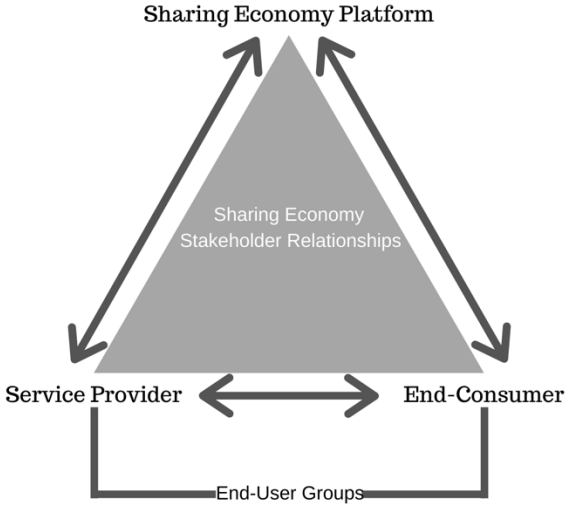


Figure 1. Sharing Economy Stakeholder Relationships

The choice to engage with a particular service provider, or the choice to offer your goods or services to a particular end-consumer, are decisions made after scrutinizing platform provided personal data collected from each user. Personal data under the GDPR is "any information relating to an identified or identifiable natural person" (GDPR, Article 4(1)).

This broad definition of personal data captures much of the data individuals provide in order to engage on SEPs. As research shows, the willingness to engage with a particular end-user depends on the richness of personal data provided (Ert, Fleischer and Magen 2016). The SEP must, therefore, espouse the technical tools necessary to assuage any hesitancy to provide this essential personal data while maintaining compliance with regulatory guidelines.

Unlike earlier peer-to-peer business models, such as eBay or Etsy, the trust between SE end-users expands beyond assurances of product quality to assurances of physical safety and responsible stewardship over another's personal property (Teubner and Flath 2019; Ranzini et al. 2017). Digitally supporting trust is challenging for any organization (Gefen et al. 2008). Still, the social exchange and real-world, physical encounter that comes along with SE transactions make trust even more vital (Williamson 1993). However, as challenging as it may be, pursuing the personal data governance mandated by the GDPR can assist in promoting the all-important trust factor necessary for successful SE interaction by demonstrating good business data practices (Zhang et al. 2020).

End-users eagerly provide their personal data supplied to the SEP, and research shows that they are more than willing to exchange their privacy for enjoying the benefits of participation in the SE (Lutz et al. 2017). Even though the choice is made to offset privacy, the duty to secure and protect personal data remains. This fact is the starting point for the remainder of this research. In Section 2.1, we go on to describe the nature of SE personal data further and track their movement to illustrate potential issues in meeting this duty to secure user personal data.

2.1 The Movement of Personal Data in the Sharing Economy

Tracking the movement of personal data is an important practice that not only allows an enterprise to clearly assess the amount and various types of personal data they govern, but also enables an accurate assessment of the appropriateness of systems used to protect this data. The GDPR recognizes the value in this process and requires some companies to keep internal records, called records of processing activities, intended "to support an analysis of the implications of any processing whether existing or planned, facilitates the factual assessment of the risk of the processing activities performed by a controller or processor on individuals' rights, and the identification and implementation of appropriate security measures to safeguard personal data" (WP29 Opinion). The information that follows is not intended to analyze the SE processing activities with the depth required of a GDPR mandated records of processing activities. Rather, the information gives a general overview of the SE personal data landscape, which still proves useful in evaluating the implications of processing activities and considering privacy and security measures.

Gateway personal data is the first necessitation for personal data in a SE transaction (Ranzani et al. 2017). This data is required to gain entry to the platform community and seek or offer goods and services. As Figure 2 depicts, the end-consumer and service provider approach the platform and provide this necessary preliminary data. Much of this data is similar for both the end-consumer and service provider; however, depending on the SEP's service offering, some of this data may be different. For example, on the Uber platform, only the service provider is required to supply data about their automobile and driving record. Some categories of gateway personal data include platform presence data (e.g., name, email address, profile information), personal property information (e.g., property address, property photos and descriptions), background and identity verification (e.g., copies of government-issued I.D.), payment information, location information (e.g., physical address, real-time GPS location) and device information (e.g., brand, battery life, and screen resolution). Gateway personal data goes a long way in shaping the trust and dual-safety SEPs attempt to digitally generate and lay the foundation for appropriate matchmaking (Ranzini et al. 2017).

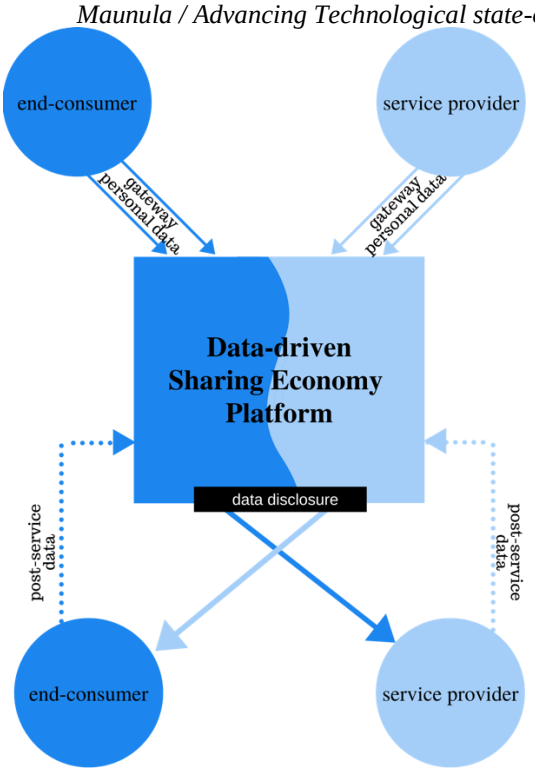


Figure 2. The flow of data in the Sharing Economy: Platform data disclosure to end-users

Figure 2 also shows that the personal data supplied to the SEP is what forms the very essence of the platform. The SEP decides which personal data is expected from each user to encourage the type of service transaction envisioned by the platform, be that on-demand travel, or lodging. The SEP assumes data controllership (Airbnb Privacy Policy), which includes the responsibilities and obligations set out by the GDPR to secure and protect this data (discussed further in section 2.3).

Once end-users supply the personal data necessary to gain access to the SEP, the next set of personal data directly contribute to the rendering of services. Here, the SEP decides which personal data it must disclose (see Figure 2), allowing the end-consumer and service provider to become transactional partners, and ultimately complete the transaction (Richter 2019). Some personal data are disclosed immediately for all platform participants to see, for example, platform presence data (as described above). Some personal data are disclosed to potential transacting partners; for example, real-time GPS location. And still, other data are disclosed to confirmed transactional partners; for example, physical addresses. SEPs do not themselves engage in the rendering of services or the provision of goods. It has been notably stated that, "Uber, the world's largest taxi company, owns no vehicles" and "Airbnb, the world's largest accommodation provider, owns no real estate" (Goodwin 2015). SEPs simply serve as matchmakers for their participants and facilitators of platform-level services, such as receiving and distributing payments.

At the time of services rendering, the utility of the data collected in an SEP's role as a matchmaker is put to the test by the end-consumer and the service provider. The proper execution of services at this point relies on the SEP identifying and passing on data already gathered for its purposes (facilitating matchmaking) to a service provider for their purposes (rendering services or providing goods).

After services are rendered, the service provider and the end-consumer return to the platform to share post-service data, as seen in Figure 2. These data include ratings and reviews and feedback data. Feedback data allows the service provider and end-consumer to rate the effectiveness of the SEP. Rating and review data, on the other hand, is personal data that enables service providers and end-consumers to evaluate the quality of the service interaction and provide future platform participants, and the SEP, with information about that interaction. Rating and reviews greatly enhance the necessary trust between service providers and end-consumers that allows participants to help one another make informed choices and steers the SEP in structuring a better platform experience for more appropriate matchmaking (Möhlmann and Geissinger 2018).

This movement of personal data is the hallmark of the SE and inherent to the business model. Users have come to expect the disclosure of their data and offset privacy concerns to receive the proven benefits of engaging through the SE (Lutz et al. 2017). However, legal compliance analysts must consider the ramifications of this data movement beyond the privacy trade-offs and scrutinize the security and protection concerns this data movement raises.

2.2 Legal Requirements for Personal Data Protection

When the personal data supplied to the SEP originate from an individual in the EU, the GDPR is triggered. In 2016, the European Commission set out to strengthen legal rules for data protection by updating the 1995 Data Protection Directive (DPD), the catalyst being the DPD's inability to address personal data protection and security in the face of emergent technologies (Room et al. 2018). The resulting GDPR has become a global gold standard in data protection with a territorial scope that captures business outside the EU, including those SEPs headquartered in the United States and elsewhere (Li and Yu, 2019; GDPR, Article 3(1-2)).

Personal data protection, as mandated by the GDPR, pivots on the designation of a data controller. The data controller determines the purpose and means - the 'why' and 'how' - for data processing (GDPR, Article 4(7)). Thus, the data controller is the entity responsible for the decisions concerning which technologies to employ for its data processing purposes. The data controller bears responsibility for compliance with the GDPR and must ensure and be able to demonstrate that all processing is done lawfully and in accordance with the provisions of the GDPR (GDPR, Article 5(2)).

Data controllers may engage competent data processors that assist in selecting appropriate technology and carrying out processing activities (GDPR, Article 28). However, this data processor can only process data by the instructions of the data controller (GDPR, Article 29) provided through a written contract known as a data processing agreement (GDPR, Article 28(3) and Article 29). As long as a data processor's actions fall within the confines of the data processing agreement, the data controller remains responsible for personal data and legally accountable for all technology choices, and any falters in security and privacy these choices bring.

This being said, a significant change from the DPD to the GDPR is that it places compliance obligations on data processors as well as data controllers. Data processors can be subject to the same hefty fines - up to 4% or 20,000,000 EUR of global annual revenue, whichever is greater (GDPR, Article 83) - for failing to meet compliance obligations applicable to the controller. While demonstrating adherence to the contractual obligations laid out in the data processing agreement can mitigate data processor penalties, the need to pay even a portion of these penalties may prove financially disastrous to some companies.

The GDPR grants individuals control over their personal data through a set of individual rights known as data subject rights (GDPR, Articles 16-21), the exercise of which should be enabled by the data controller (GDPR, Article 12(2)). Data subject rights include the rights to access, rectification, erasure, restriction of processing, data portability, and objection to processing. Data controllers must provide the transparency and tools that inform data subjects of all data collected concerning them, and the ability to access this data as envisioned by this set of rights. The data controller must provide any information pertaining to these rights in a "concise, transparent, intelligible and easily accessible form using clear and plain language…" (GDPR, Article 12(1)).

The GDPR bars any processing of personal data without establishing a legal basis to do so (GDPR, Article 6). The six lawful bases established by the GDPR include consent, the performance of a contract, compliance with a legal obligation to which the controller is subject, to protect vital interests of the data subject or other natural person, performance of a task carried out in the public interest, or legitimate interest (GDPR, Article 6(1)(a-f)). The collection, use and disclosure of personal data in the SE function primarily under the lawful bases of consent and performance of a contract (see, for example, Uber and Airbnb privacy policies).

Moreover, all processing must be done in observance of six principles. Those six principles demand that any personal data processing embrace (GDPR, Article 5(1) (a-f)):

1. lawfulness, fairness and integrity
2. a purpose limitation
3. data minimization
4. accuracy
5. storage limitations

6. with integrity and confidentiality

In the case of these principles, the data controller, once again, stands in the position of accountability and is responsible for compliance (GDPR, Article 5(2)).

Ensuring data subjects' access and control over their data, requiring a legal basis for all of their processing and demand for adherence to data processing principles are measures meant to make the GDPR a strong instrument for the protection of personal data, but also a challenge for compliance. These measures should also form the core of any technological structures to achieve compliance.

2.3 Data Protection Red Flags in the Sharing Economy Context

This Section merges the information in Section 2.1, regarding the movement of data in the SE, with the information in Section 2.2, regarding the lawful processing of personal data under the GDPR. The enmeshing of these two realities for the SE creates an entanglement that makes it easier to see irregularities that illuminate two critical GDPR compliance red flags. To deepen the understanding of the privacy and safety issues discussed, this Section begins with a real-world example of the data protection issues that can arise in the SE. Consider the following occurrence:

Denver, Colorado, USA – On 26 March 2015, Gerald Montgomery, a 51-year-old Uber driver, picked up a local woman for a trip to the airport. After successfully delivering her to the destination, Mr. Montgomery returned to her home address to burglarize the property. Mr. Montgomery was positively identified by his passenger's roommate, who was home at the time and scared him off. This identification was made possible by a screenshot the passenger took of her Uber receipt, which included his photograph.

Uber responded with the following statement: "[Uber] takes rider safety very seriously, and upon learning about this incident, we reached out [to] the rider. We immediately removed the driver's access to the Uber platform, pending an investigation. We continue to be in contact with the rider and will assist the authorities in whatever way we can." (Roberts 2015)

The data flow in the event recounted above mimics the depiction in Figure 2. The Uber rider (end-consumer) supplied her gateway personal data to the Uber platform, allowing them (through contract and consent) to use this data to efficiently match her with an available driver (service provider). The rider's personal data is protected while she is on the platform, with Uber serving as data controller. The Uber platform then disclosed necessary data to the passenger and driver, allowing them to connect to perform the service. At this stage, Uber removes itself from data controllership, as they are not involved in the service-rendering phase in any way. However, the incident shows the ease with which a service provider can use (process) personal data supplied by the platform for nefarious reasons. While banning the service provider from the platform may be appropriate business responses, what should be the proper data protection response? And, more importantly, what should be the appropriate data security support to curtail such abuses?

As Section 2.1 and the scenario above explain, personal data in the SE context is originally collected and used by the platform, who serves as the data controller (see Figure 2). The SEP then decides which data it must pass on to end-users (both service providers and end-consumers) to complete the service transaction. The reception and use of personal data by these end-users are acts of data processing. Thus, the first red flag raised by the data processing activity in the SE is the specific implication of the service provider as a data processor created by this disclosure and use of personal data.

The GDPR defines data processing as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" (GDPR, Article 4(2)). Although both end-users can be described as data processors, in general, it is the processing activity of the service provider that falls under GDPR obligations. This regulatory capture is due to the fact that service providers reap an economic benefit as a result of their data processing activities. Recital 18 of the GDPR explicitly states that exclusion under Article 2(2)(c) can only be enjoyed by those "with no connection to a professional or commercial activity." Article 2(2)(c) of the GDPR provides an exclusion from GDPR provisions for those that process data "in the course of a purely personal or household activity." End-consumers that merely pay for and receive goods and services through the SE fall under this exclusion, however, service providers do not.

Now, clearly classified as data processors beholden to GDPR obligations, a threat to privacy and security arises in the fact that service providers are not considered, addressed, guided, directed, nor contracted as data processors by SEPs (see

Uber and Airbnb Privacy Policies). As described in Section 2.2, data processors must only operate under the contractual stipulations of a data processing agreement. This agreement is a vital link that maintains a connection between the data controller and the data processor, ensuring that data is processed in line with the data controller's wishes and an equivalent level of data security. However, the recognition, and thus guidance, of service providers as data processors acting upon instructions of the SEP creates a relationship counter to the SEP's business function as a mere intermediary providing a matchmaking function. The SEP reliance upon GDPR consent and contract fulfillment mechanisms do not prove adequate as a means of data protection, considering the potential of harm that could befall personal data in the hands of a service provider.

The second red flag emerges as a consequence of the first. The SEP extraction from GDPR obligations over data processed during the service rendering phase causes a break in the consistency of data protection by allowing the SEP to abandon data controllership temporarily. As Section 2.2 (and Figure 2) points out, the recursiveness of the post-service rating and review data brings end-consumer back to the platform, reigniting SEP data controllership. Rating and review data is, in fact, personal data (Golbeck 2016) subject to protection under the GDPR, and is a part of the data disclosed by the SEP. Further, rating and review data has a bearing on the digital reputation of SE users, in particular, the service provider that stands to lose the economic benefits participation brings (Hawlitschek et al. 2016). Consequently, the GDPR places special protection over data that may "give rise to… damage to the reputation" (GDPR, Recital 75). Rating and review data may fall under this special category, requiring special security measures to minimize the risks to the rights and freedoms of data subjects. The implication created by this recursivity is that the SEP re-engages as the data controller, as end-users are on the platform to provide and benefit from this valuable data.

As Figure 2 shows, the SEP is irreducibly involved in handling, management and distribution of data. As such, the SEP remains the data controller at every stage. The recursive quality of rating and review data means that the purpose for the data has not changed – to assist in the completion of the SE transaction. Thus, rating and review data strengthens the SEPs line of involvement throughout the movement of data. This has consequences on GDPR obligations and sculpts a new data controller - data processor relationship between the service provider and SEP; a relationship where the SEP must exhibit greater responsibility over the entire SE data flow.

In-depth legal analysis of the interplay between legal obligations and platform interests in the processing activities in the SE raises red flags about the security and protection of personal data. It stands to reason that this socio-technical industry would look to technology for possible solutions to strengthen privacy, security and, thus, GDPR compliance. The shift will be from considering tech tools for security and privacy from a SEP compliance perspective to a service provider as data processor perspective.

3. Technology's Supportive Role in Protecting Sharing Economy Personal Data

Although the GDPR is technology neutral, meaning that it governs manual or automated data processing (GDPR, Recital 15), throughout the text of the Regulation, it is clear that there is a focus on what technology can do for achieving compliance and the challenges of remaining compliant while instituting new technologies. For example, the GDPR is built on a concept called Privacy by Design and Default (GDPR, Article 25). This concept means that organizations must always use "appropriate technical and organizational measures" to ensure they meet GDPR requirements (GDPR, Recital 78). In other words, data protection should be integrated at every stage of product or service development, from creation to implementation (by design), while adopting the strictest privacy standards (by default) (GDPR, Article 25). Technology is the clear way to adhere to these concepts.

Article 32 and Recital 83 of the GDPR directly deal with security of processing. The GDPR requires that companies consider the current "state of the art" when determining if security measures are appropriate. Although this ambiguous statement raises many questions, it makes clear that lawmakers expect industries to devise security strategies that "continuously evolve in line with anticipated advances in technology" (ElectricIQ 2018). Further, the concepts of GDPR's privacy by default and the state-of-the-art requirement must be viewed as a unit. If simultaneously, how the technology is arranged and the state-of-the-art must be applied, it naturally leads to an understanding that technology has to be at the forefront of any discussion. The communication between data-driven industries, such as that between the SE and ICT developers about their functions and their legal compliance, becomes an ongoing conversation.

As analyses of processing activities are meant to reveal, the analysis in this research shows the type, movement and gaps in protection of personal data. Inherent in the SEP business model is the need to collect data of various types that are disclosed to service providers to complete a transaction. During the course of this transaction, new data is compiled regarding the quality of services and the technologies that enable them. This recursive flow of data implicates the service provider as a processor of data under the GDPR, and maintains the role of the SEP as controller of this data. As data

controllers, the SEP will look to the same technologies that allow this collection, disclosure, and recursive nature of data to restructure security and privacy methodologies to meet the challenges of this implication.

3.1 A Call for Customized Technology for Developing Sharing Economy Data Protection Challenges

Customizing information security technologies for these developing compliance concern in the SE requires thinking on a nuanced level. Developers must move from considering general protections of all types of information from any unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, and availability (Kissel 2013), to considering the more specific protection of information related to an identified or identifiable natural person. Additionally, the tools must move from the broader compliance support of the quite capable SEP, to the targeted compliance support of service providers under specific processing conditions (Guamán 2016).

Current tools that enable the SEP and secure its valuable data are configured with the needs and capabilities of the SEP in mind. Envisioning service providers as data processors under the direction of the SEP (the data controller) demand new or reconfigured technologies that secure personal data while on the leg of its journey that takes it through the hands of the service provider.

In the SE context, disclosure of personal data extends to intermediaries and third-party data processors, such as payment processors, background check and identity verification providers, cloud storage providers, and marketing partners (Uber Privacy Policy). These categories of data processors that process on behalf of the SEP include business that are typically better equipped to secure data. The GDPR requires that data controllers "shall use only processors providing sufficient guarantees to implement appropriate technical and organizational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of that data subject" (GDPR, Article 28(1)).

This research shows that the requirement outlined in Article 28 to only use competent data processors must now extend to the service provider. A typical SE service provider, for example an Uber driver or an Airbnb host offering a room, is not necessarily equipped to guarantee appropriate "technical and organizational measures". However, the ongoing involvement of the SEP and its ability to implement appropriate technical measures may alleviate demands placed upon an ill-equipped data processing SE service provider. Established in the EU Court of Justice Google-Spain 2013 case, the balanced approach to protecting personal data insists that parties consider where "effective and complete control" would lie. It stands to reason that in the context of the SE, the Court would draw the line of effective and complete control around the SEP and its technological capabilities and contractual positioning over the flow of data.

An adequate system of extending controls from the platform to service providers may best be delivered through mobile devices and applications. The prolific use of mobile-apps to execute SE transactions creates a significant entryway into creating new compliance strategies. Mobile technology has already perfected the SEP's capability to communicate with its users, allow users to communicate with one another, acquire consent, verify identity, collect payments and encourage the robust rating and review systems that garner trust. Adapting technologies, with their dynamic user interfaces, for the mobile app environment can efficiently extend the SEP's data controller compliance obligations through the activities of the service provider's data processing. The goal is not only to provide tools that are easy to navigate, but those that actually achieve compliance under these unique circumstances.

Currently, there are tools on the market that could be configured for this purpose. For example, a company called MangoApps has developed a screen capture program called TinyTake which can blur specific parts of a screenshot to hide certain personal data (TinyTake 2019). This could prevent users from capturing data for later, unauthorized use. But again, this requires a nuanced approach, because this personal data, in large part, is necessary. The goal would be to prevent unnecessary screenshots that may lead to nefarious behavior.

Other possible solutions could allow the platform greater control over this data. For example, many service providers using the Uber platform utilize dashcams to ensure their physical safety and protect against any false claims. However, recent violations of the use of these cameras have allowed individuals to misuse this footage, which is personal data of their riders. In a recent case in St. Louis, Missouri, a passenger using the Uber ride-sharing platform, discovered that footage taken by her driver's dashcam was being livestreamed via a paid platform. The rider was made aware of the use of the dashcam and told the driver told her it was for his personal safety and security (Heffernen 2018). The Uber passenger agreed to relinquish her privacy for that effect, but, privacy notwithstanding, the misuse of her data would constitute a violation of EU personal data protection laws. A system where the platform controls the dashcam would better protect riders in these situations. Controls at the platform level could collect the data, manage retention time and ensure proper deletion, assigning contentions over any other nefarious dashcam use to other areas of the law (e.g., criminal law, tort law).

Recent attention concerning information privacy in the SE has focused on the use of blockchain technology. Blockchain allows peer-to-peer data transmission that bypasses a centralized server (De Philippi, 2017; Tumasjan and Beutel, 2019). Blockchain is problematic in the sense that the implication of further decentralization would remove the need for an SEP altogether, robbing the business model of its function. Beyond having structural problems, the blockchain is also problematic under the GDPR. The immutability of data, a strong characteristic of blockchain technologies, does not allow the proper exercise of data subject access rights, especially the right to rectification and deletion of data. However, blockchain could be a solution if the tool is tailored for use by the platform and ensures data subject rights.

These types of tailored solutions for security and privacy may exist, but need to be re-packaged and delivered from the perspective of data protection law and the consideration of ill-equipped service providers as processors under those laws.

4. Conclusion and Considerations for Future Research

Data-driven industries, including the SE, will continue to scrutinize their business practices in light of the complexities of achieving and maintaining GDPR compliance. This requires continually balancing the goals of the business model with innovative technological affordances. It can be said that the almost paradoxical push and pull between technology and legal compliance has dramatically influenced the SEP's relationship to the GDPR. This new relationship offsets this delicate balance of business and technology. Any solution for recalibrating security and privacy systems when business model analyses raise new GDPR compliance issues must be fashioned out of the dependency on technology.

As has been argued in this paper, a fresh legal analysis of processing activities in SE business models raises two GDPR compliance red flags pulls the service provider under the compliance canopy of the GDPR. The far-reaching implications of service provider data protection obligations realigns their relationship with the SEP and the technical tools already used to achieve privacy and security in the SE. This theoretical legal research will allow the industries of internet technologies, internet communication technologies, and information systems to sculpt the perfect solution. But research such as this is essential to clearly define the issues so that these technological remedies will be appropriate. As this paper reveals, disruption may just be knocking on the door of technology with regards to the sharing economy, because the SE business model reveals weaknesses in personal data protection that are not among the pool of readily available digital tools to pull from.

The major contributions in this research are, admittedly, heavily theoretical and legal. This is not meant to diminish the multidisciplinary balance called for throughout this paper. The intent of this paper, however, is to spark the interest in IT and ICT research to enhance data privacy and security for a specific legal challenge. With this legal analysis in mind, future research should be developed in the areas of mobile application interfaces, SEP supported blockchain and features that ensure appropriate data minimization and purpose and storage limitations as personal data lies in the hands of service providers.

Placing the SEP in control of the entire SE data flow is a game changer for the relationship between all SE stakeholders: SEPs, end-users and regulators. This new relationship demands quantitative and qualitative IT and ICT. Methodologies such as business case study can support a clear vision for platform actions, and technology gap analysis alongside legal gap analysis study ensures the way forward strengthens compliance, security and protection for all involved.

5. References

"Airbnb Privacy Policy." https://www.airbnb.com/terms/privacy_policy.

Dakhila, Sami, Andrés Davila, and Barry Cumbie. "Trust, but Verify: The Role of ICTs in the Sharing Economy." *Information and Communication Technologies in Organizations and Society: Past, Present and Future Issues.* Edited by Francesca Ricciardi, and Antoine Harfouche. Springer, Switzerland, 2016.

De Filippi, Primavera. "What Blockchain Means for the Sharing Economy." *Harvard Business Review*, 2017.

ElectricIQ. *GDPR and 'State of the Art' Security.* 2018. https://medium.com/@eclecticiq/gdpr-and-state-of-the-art-security-a5c07c04aeeb.

Gefen, David, Izak Benbasat, and Paul A. Pavlou. "A Research Agenda for Trust in Online Environments." *Journal of Management Information Systems*, vol. 24, no. 4, 2008. pp. 275-286. *JSTOR*. http://www.jstor.org/stable/40398920.

Goodwin, Tom. *The Battle is for the Customer Interface,* 2015. http://social.techcrunch.com/2015/03/03/in-the-age-of-disintermediation-the-battle-is-all-for-the-customer-interface/.

Guamán, Danny. *Privacy Vs. Data Protection Vs. Information Security – Software and Services Engineering,* 2016. https://blogs.upm.es/sse/2016/11/01/privacy-vs-data-protection-vs-information-security/.

Hawlitschek, F., T. Teubner, and C. Weinhardt. "Trust in the Sharing Economy ." *Die Unternehmung – Swiss Journal of Business Research and Practice*, no. 70(1), 2016. https://www.nomos-elibrary.de/10.5771/0042-059X-2016-1-26.pdf.

Heffernan, E. "Uber Evaluating Policies in Response to Story on St. Louis Driver's Secret Livestream." July 24, 2018. https://www.stltoday.com/news/local/metro/uber-evaluating-policies-in-response-to-story-on-st-louis/article_7c8e4558-ff49-54c0-8e8c-a6e79b954325.html.

Irwin, Luke. "Organisations Struggling to Meet GDPR Requirements, with Poor Planning and Lack of Awareness to Blame." 2019. https://www.itgovernance.co.uk/blog/organisations-struggling-to-meet-gdpr-requirements-with-poor-planning-and-lack-of-awareness-to-blame.

Kissel, Richard L. *Glossary of Key Information Security Terms.* 2013. https://www.nist.gov/publications/glossary-key-information-security-terms-1.

Li, He, Lu Yu, and Wu He. "The Impact of GDPR on Global Technology Development." *Journal of Global Information Technology Management*, vol. 22, no. 1, 2019. pp. 1-6. https://doi.org/10.1080/1097198X.2019.1569186, doi:10.1080/1097198X.2019.1569186.

Madsen, Wayne. *Handbook of Personal Data Protection.* Palgrave Macmillan UK. 1992. https://www.palgrave.com/gp/book/9781349128082.

Richter, Heiko, and Peter R. Slowinski. "The Data Sharing Economy: On the Emergence of New Intermediaries." *IIC - International Review of Intellectual Property and Competition Law*, vol. 50, no. 1, 2019. pp. 4-29, https://doi.org/10.1007/s40319-018-00777-7, doi:10.1007/s40319-018-00777-7.

Room, Stewart, Peter Almond, and Kayleigh Clark. *Technology's Role in Data Protection - the Missing Link in GDPR Transformation.* PwC, 2018.

Seigneur, Jean-Marc. "Social Trust of Virtual Identities." *Computing with Social Trust.* Springer London, London, 2009.

Tankard, Colin. *What the GDPR Means for Businesses.* 2016. http://www.sciencedirect.com/science/article/pii/S1353485816300563, doi:https://doi.org/10.1016/S1353-4858(16)30056-3.

TinyTake. "Best Free Windows Screen Capture & Video Recording Software." 2019, https://tinytake.com/.

"Uber Privacy Policy.", https://www.uber.com/global/en/privacy/notice/#data.

van den Hoven, Wolter, Jeroen, Blaauw, Martijn, Pieters, and Martijn Warnier. "Privacy and Information Technology." *The Stanford Encyclopedia of Philosophy.* Edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. 2019. https://plato.stanford.edu/archives/win2019/entries/it-privacy/.

Wiener, Jonathan B. *The Regulation of Technology, and the Technology of Regulation.* vol. 26. 2004.

Williamson, Oliver. "Calculativeness, Trust, and Economic Organization." *Journal of Law and Economics*, vol. 36, no. 1, 1993. pp. 453-486, https://EconPapers.repec.org/RePEc:ucp:jlawec:v:36:y:1993:i:1:p:453-86.

WP29 Opinion. *'Article 29 Working Party, 'Opinion 1/2010 on the Concepts of "Controller" and "Processor"' (WP 169, 16 February 2010).*

Author Biography

Gail L. Maunula is a doctoral researcher at the University of Turku Faculty of Law in Turku, Finland. Her research centers around the impact of human and fundamental rights issues on the Sharing Economy from a European socio-legal perspective, including the rights to data protection and privacy. Gail also holds a professional certification in information privacy for the European sector (CIPP/E). Gail further explores the wide range of issues arising from data protection and privacy as a member of the Digital Disruption of Industry Research Project (DDI), a multidisciplinary research consortium studying the impact of digitalization on Finnish society through the lens of industry.

Journal of the Midwest Association for Information Systems

Volume 2020 | Issue 2 Article 5

Date: 07-31-2020

Alignment of Coursework with Knowledge Requirements: A Textbook Content Analysis

Mark Weiser
Oklahoma State University, weiser@okstate.edu

Andy Bowman
Oklahoma State University, andy.bowman@okstate.edu

Abstract

Every information systems professional has a role to play in security. Analysts must consider security in their analyses and designs; programmers think through logic flaws that create vulnerabilities; and database managers need to provide appropriate access without exposing sensitive information to bad actors. Other disciplines also recognize the importance of employees having a respect for security and a broad understanding of concepts that enable it. Universities prepare students for careers across different domains; and the increasingly important formation of security knowledge falls to IS faculty. This study first examines relevant job postings to determine the knowledge, skills, and abilities most sought after by employers; then uses those results in a content analysis of current information security textbooks to indicate the degree to which employer-demanded concepts are covered in university-deployed teaching materials. The overall results of this study found that coverage of terms associated with security knowledge areas demanded by the marketplace is mixed among six leading textbooks, ranging from near complete coverage to just over half of the topics.

Keywords: *Content Analysis; Information Security Education; Textbook Analysis; Curriculum*

DOI: 10.17705/3jmwa.000061

1. Introduction

There is an extreme shortage of cybersecurity skills available in the U.S. workforce. In recent annual surveys by the Enterprise Strategy Group (ESG), in each of the last five years an increasing number of organizations reported "a problematic shortage of cybersecurity skills," with well over 50% in the most recent survey (Oltsik, 2019). U.S. News ranks Information Security Analysts 5th in all technology jobs, 19th in STEM, and 38th among the 100 best jobs in any field (U.S. News & World Report, 2020). Several job titles that rank even higher in surveys have information security knowledge as a basic requirement. CNNMoney/Payscale categorizes positions slightly differently but puts Information Assurance Analyst at number 5 among all job types (Braverman, 2017). Both rankings cite a median salary in excess of $98,000 with a bachelor's level education, compared to an overall IS salary average of $81,000 based on these same sources.

It's clear that demand for security knowledge far exceeds the number of appropriately-trained graduates with relevant four-year degrees. Many suggest that in order to bridge this gap, industry must turn to applicants with non-traditional backgrounds and even provide training to transition to this field (Zadelhoff, 2017). There is some question, however, if the general information security coursework provided by universities even covers the basic body of knowledge needed in entry-level positions.

It has become an expectation that IS programs incorporate security as a core component, as a separate course, or as modules embedded in multiple offerings across the curriculum. Other disciplines have recognized the importance of information security in their own career fields and have begun to offer required or optional coursework to build security awareness before entering the job market (Weiser & Conn, 2017). There is no agreement, however, about appropriate topics, depth, or scope for either the security practitioner or those for whom a broad understanding is sufficient. Because academia strives to place graduates in the most competitive positions, industry advisory boards comprised of prospective employers help educators shape topics. Recruiters, however, vary in opinion about the best direction in a constantly evolving field. Some employers turn to professional security certifications for affirmation that a job candidate has appropriate knowledge; but certifying bodies and training companies abound and do not agree on the set of desirable knowledge, skills, and abilities. The IS security field lacks a single unifying coordinating body upon which industry and academics can agree. Without accepted common topics and metrics like accountants have from the AICPA (AICPA, 2020), it is difficult to assess the degree to which information security curriculum offered in higher education, or tested by security certifications, actually matches the needs of the workforce.

This study analyzes relevant textbooks that provide a broad overview of information security; as indicated on the author's statements, publisher's marketing material, or the introduction to the book. These books are most often intended for second- or third-year bachelor's degree candidates who may then study information technology more broadly, or explore more depth in security during their upper-level courses. Although the same texts could be used for more advanced study, our scope is to explore the alignment of IT security knowledge for entry-level positions with the preparation that higher education provides.

Although a textbook certainly does not equate to the content of a course, it is a reasonable assumption that the book significantly influences the topics covered in the course. Content analysis of textbooks has been used for a variety of purposes across all fields of study. It can be used to gauge how long an emerging topic takes to appear in mainstream texts (Laksmana & Tietz, 2008); for comparative analyses of readability (Bargate, 2012); to indicate levels to which specific topics are covered; or even to identify different approaches to the same topics in different books (Foxman and Easterling, 1999; Fisher and Southey, 2005). Most commonly, however, textbook content analysis seeks to determine coverage of selected topics. Bracken and Urbancic (1999) analyzed ethics coverage in Accounting books, DeSensi & Jurs (2017) evaluated the presentation of psychological disorder stigma in Psychology texts, and Polikoff (2015) assessed the extent to which Mathematics textbooks meet common core standards.

To accommodate for the lack of an official knowledge list, this study systematically examines major job listing websites for terms most closely associated with information security knowledge. Because some positions referenced major professional certifications in lieu of, or in addition to, enumerating knowledge, skills, and abilities (KSAs) for candidates; content from those certifications was used to augment the list of KSAs. The superset of those words and phrases was then applied in an analysis of tables of contents from six major information security textbooks for inclusion of the concepts most sought by employers.

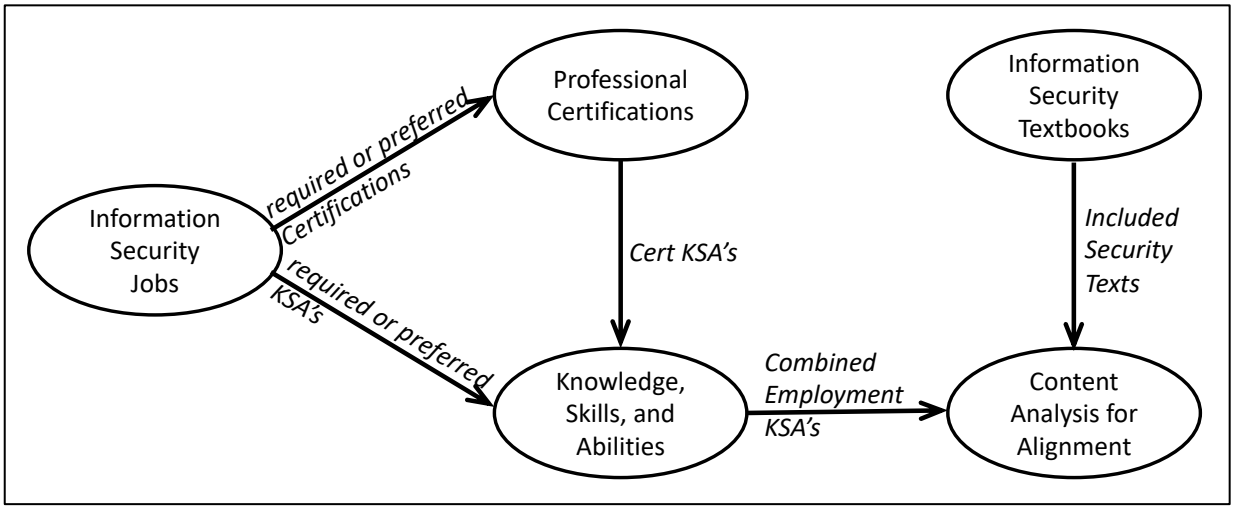
2. General Methodology

Content analysis is a systematic technique for representing longer passages of text into fewer content categories based on explicit coding rules (Berelson, 1952; Krippendorff, 1980). Holsti (1969) expands its application to include "any technique for making inferences by objectively and systematically identifying specified characteristics of messages." The most common application of content analysis is a word-frequency count, under the assumption "that the words that are mentioned most often are the words that reflect the greatest concern" (Stemler, 2000).

The lack of an accepted topic set for information systems security, as one may find for accounting or mathematics domains; however, means that we must first establish a set of knowledge objectives before gauging whether each is covered. A lack of common terminology and sub-topics further confounds this effort, so we employed the approach shown in Figure 1 to determine a non-redundant set of knowledge, skills, and abilities (KSAs) and then assess the presence of each in leading textbooks. We specifically:

1. searched relevant listings in multiple major job boards for KSAs and professional certifications that are required or preferred by prospective employers;
2. identified KSAs included in required or preferred professional certifications;
3. identified current textbooks with broad coverage in information security that could be taken by students without technical pre-requisites; and
4. analyzed textbook tables of contents to determine inclusions for job-driven KSAs.

Figure 1. Developing KSA's and Textbooks for Analysis



3. Task Methodology and Results

The scope of this study is to evaluate the alignment of security KSAs in entry-level positions with the content of textbooks used in IT security survey courses. To accomplish this, it was necessary to apply content analysis in three steps, each with a specific, objective and systematic approach. The first identified terms that represent knowledge skills and abilities required in relevant entry-level job postings. Because some postings referenced one or more security certifications rather than enumerating terms, we systematically determined representative KSA terms from those certifications. Finally, the most prevalent term from those two steps were used to objectively evaluate the content alignment in leading textbooks. Additional details of each process and the results follow.

3.1. Identification of KSA's and Certifications from Public Position Postings

Based on a Silk Road study involving the source of hired talent within corporations, ten top job search sites were identified (Research and Markets, 2016; Silk Road, 2016). Documentation of the query method employed by each site was reviewed and practical test queries were run to ensure that our systematic study process yielded objective and consistent results without redundancy that would skew results. Websites that require users to login were eliminated because the information gathered during the account creation process (skills, education, and location) was used to tailor results to the user. The biased results lead to a narrowed sampling of job listings that was not representative of entry-

level positions across the country. Other sites were excluded due to limiting the number of results per search, searches returning only local job listings, or use of semantic searches that lead to the number of results increasing for every additional word in the search, rather than refining the results. The table below details the reasons for excluding or including each site.

Table 1. Website Elimination Reasons

Website	Eliminated	Date Searched	Reason Eliminated
Glassdoor	Y	1/27/2020	Required login to search job listings
CareerBuilder	Y	1/27/2020	Limited search results to 5,000
Monster	Y	1/28/2020	More than three terms leads to increasingly large results
Simply Hired	Y	1/28/2020	More than two key terms leads to increasingly large results
Dice	Y	1/28/2020	Same jobs posted on other sites, introducing redundancy in data
Snagajob	Y	1/29/2020	Location based searches required
Indeed	N	1/27/2020	
Monster	N	1/29/2020	
Zip Recruiter	N	1/30/2020	
Google for Jobs	Y	1/27/2020	Focus on job postings

The first search used the term "Information Security," without any location preference. From among many relevant security-related searches, this term was selected because it returned the highest number of results of which at least 70% were identifiable as jobs within the domain of interest. A Boolean "AND" was used between terms on sites where necessary to return positions that included all terms in any section of the position description.

Each listing was reviewed in the order returned from the search and was first assessed for relevance to the domain and applicability for the purpose of this study. A position announcement was excluded from further review if:

- The job was not primarily in information security: eliminated secretarial jobs which required maintaining "privacy" of records, janitorial positions that listed a need "securing the site," and other references that fell outside of the scope of this study.
- More than two years of work experience was required: the scope of this study includes positions for which undergraduate study was the main preparation.
- A top-secret or higher clearance was required: the majority of these positions required specific experience such as prior military or government service which put it outside of the scope of this work.
- It was a temporary position, such as an internship or contract-based work: many of these occur before undergraduate study is complete.

If not eliminated, the position listing was reviewed in detail and KSAs and/or professional certifications that were required or preferred were recorded. Every job listing that fit the criteria was checked against all previously recorded job listings to ensure that no redundant postings were included. This process was repeated until KSAs and certificates from five non-eliminated position listings were recorded. Similar or highly related terms were consolidated, for instance "cloud", "cloud security", and "cloud architecture" were grouped into "cloud" and all subsequent cloud references were considered equivalent.

Because the different search engines vary significantly in size, we normalized the data by dividing the number of positions returned from each search by the number of positions returned by the search for only "information security" alone. This ratio for each KSA is recorded in Table 2 and ordered from highest to lowest by the sum of the ratios from all three included job search sites. The top twenty terms that were not direct references to names of professional certifications were retained to be used in further analysis as representative of knowledge areas that are most important in the field of information security.

Table 2. Top KSA terms from relevant position listings

Information Security…	Indeed	LinkedIn	Zip	Score
Risk	.1926	.0927	.1505	**.436**
Cloud	.0966	.1391	.1593	**.395**
Windows	.1036	.1060	.0822	**.292**
Networking	.0654	.0861	.0836	**.235**
Database	.0686	.0199	.1267	**.215**
Linux	.0621	.0795	.0639	**.205**
Mobile	.0574	.0596	.0632	**.180**
Vulnerability	.0441	.0728	.0329	**.150**
Web Application	.0673	.0331	.0405	**.141**
Active Directory	.0238	.0331	.0443	**.101**
Firewall	.0339	.0265	.0400	**.100**
Scripting	.0388	.0066	.0488	**.094**
Unix	.0248	.0397	.0294	**.094**
Security Standards	.0175	.0265	.0313	**.075**
Coding	.0297	.0063	.0315	**.068**
Patch	.0185	.0132	.0305	**.062**
Encryption	.0132	.0265	.0167	**.056**
Intrusion Detection/Prevention	.0120	.0265	.0165	**.055**
API	.0134	.0132	.0234	**.050**
SIEM	.0103	.0265	.0124	**.049**

Each job posting had a unique structure and employed varying terminology, so subjectivity in the analysis was necessary to convey results with practical importance that still retain objective comparable numeric measures. In some cases, terms were combined during the process. For instance, "encryption" and "decryption" either appeared together in most of the reviewed position descriptions or knowledge of both areas was implied. Based on the reviewed sets, the number of unique positions that included either or both terms were extrapolated to form the total for that combined topic. Similarly, "intrusion detection" and "intrusion prevention" were combined. "Scripting" and "coding;" however, were left as separate terms because the context of many reviewed positions appeared to use "coding" for more complex programming of full applications, whereas "scripting" in the security domain often referenced shorter segments to tie together processes and automate tasks. The use was inconsistent and each was sufficiently prevalent that, whether the terms were treated separately or collectively, both would be represented in our final results.

Many position descriptions included a preference that candidates had one of several professional certifications, with fewer having them as a requirement. Within the 50 most frequent terms during the job posting search, 13 were references to certifications in the IS security industry, including the five top certifications according to the ISSA (Oltsik, 2019). These certifications are awarded by professional organizations for individuals who can take a test proving knowledge and/or competence in information security areas. A review of the official website for each certification allowed us to delve deeper into the topics covered.

Appending the certification name to the term "information security" in the same way as was done previously, we recorded the number of positions returned by the job search sites and scored them in the same way described in our KSA discovery process. The results are in Table 3 and are ordered with the certification that is most referenced by position listings (CISSP) to the least (GCED). It should be noted that Security+ appeared in many position descriptions; however, it was not included in the table because the highly varied ways that the search engines processed a "+" prevented a comparable metric to be computed for that certification. There were also references to categories of certifications, such as "GIAC" which indicates the group of all SANS certifications, so the collective term GIAC was omitted in favor of individual certification names.

Table 3. Top certifications from relevant position listings

Information Security…	Cert Org	Indeed	LinkedIn	Zip	Score
CISSP	$(ISC)^2$	.0349	.0530	.0345	**.122**
CISA	ISACA	.0154	.0331	.0176	**.066**
CISM	ISACA	.0121	.0331	.0009	**.046**
CEH	EC-Council	.0074	.0132	.0005	**.021**
GSEC	$(ISC)^2$	.0060	.0132	.0004	**.020**
SSCP	$(ISC)^2$	.0063	.0066	.0003	**.013**
GCIH	GIAC	.0047	.0066	.0003	**.012**
CASP	CompTIA	.0047	.0066	.0003	**.012**
OSCP	Offensive-Sec	.0027	.0066	.0002	**.010**
GCIA	GIAC	.0020	.0057	.0002	**.008**
GPEN	GIAC	.0016	.0051	.0002	**.007**
GCED	GIAC	.0021	.0046	.0001	**.007**

3.2. Identification of Additional KSA's From Certifications found in Public Position Postings

Some employers listed required or preferred professional certifications in lieu of enumerating KSAs expected of a successful candidate, others stated desirable knowledge areas, and some included both. The implication is that a company listing professional certifications is seeking the knowledge demonstrated by a candidate to obtain that certification. In order to capture additional KSAs which were not revealed by the term search in Table 2, we reviewed the publicly available topic list for each certificate in Table 3, in addition to Security+, which was omitted from the table because of search anomalies described previously. The topic list for each certification, as presented on the official website for each ((ISC)2, 2020; CompTIA, 2020, EC-Council, 2020; ISACA, 2020; GCIA, 2020; Offensive Security, 2020) was reviewed for the presence of each of the terms that were drawn from our job search analysis.

Table 4 shows which topics appear in the knowledge list for each certification. Certifications are listed across the top are in order of their prevalence in position listings, as determined above. The "Job Appearance Score" shown for each is the taken from Table 3. Terms are sorted by the number of certifications in which each appears, as recorded in the right-hand column. Only terms that appeared in five or more certifications are shown. Six KSA terms that were not captured in our earlier job postings process appeared in at least five of the listed professional certifications, and those appear at the bottom of the table under the double line. "Penetration," for example, did not receive a sufficiently high score to appear in the results of our earlier process, but was identified by ten of the 13 professional certifications to which employers referred when searching for candidates. It is therefore assumed to be an important term that represents a KSA sought in future employees and is incorporated in our textbook content analysis. Similarly, "forensic tools", "control", "log", "auditing", and "client side / server side" (collectively) were included as relevant important terms in security education.

Table 4: KSA coverage indicated by certification topic list

Certification	CISSP	CISA	CISM	CEH	GSEC	SSCP	GCIH	CASP	OSCP	GCIA	GPEN	GCED	Sec+	count
Job Appearance Score	.122	.066	.046	.021	.020	.013	.012	.012	.010	.008	.007	.007		
Intrusion Det/Prev	X	X	X	X	X	X	X	X	X	X	X	X	X	13
Vulnerability	X		X	X	X	X	X	X	X	X	X	X	X	12
Networking	X			X	X		X	X	X	X	X	X	X	10
Malware	X				X	X	X	X	X	X	X	X	X	10
Firewall	X				X	X	X	X	X	X		X		8
Packet				X		X	X	X	X	X	X	X		8
Risk	X	X	X		X	X	X	X			X		X	9
Cryptography/Encrypt	X	X		X	X	X		X	X			X	X	9
Coding	X				X	X		X	X		X	X	X	8
SIEM	X	X	X		X	X	X					X	X	8
Web Application	X	X			X	X	X	X	X		X			8
Identity Management	X	X			X	X		X	X		X			7
Security Standards	X	X	X		X			X				X		6
Endpoint	X	X			X	X		X	X					6
Cloud	X				X	X		X				X		5
Database	X	X				X		X			X			5
Linux					X	X	X	X	X					5
Mobile	X	X			X			X	X					5
Penetration	X	X		X	X	X	X		X		X	X	X	10
Forensic Tools	X	X		X			X	X	X	X				7
Control	X	X		X	X		X	X	X					7
Log	X			X	X		X	X		X				6
Auditing	X	X		X	X	X		X						6
Client side / Server Side					X	X	X	X	X					5

3.3. Identification of Leading Relevant Textbooks

This analysis focuses on textbooks that are actively being used in overview courses in Information Security and are neither highly technical nor have significant pre-requisites. The authors first reviewed leading academic publishers' catalogs for textbooks in this domain which are currently being marketed in higher education settings. 15 titles were identified. For each book, the publisher's marketing material and authors' notes were reviewed to glean the type of class and student for which it was intended. Textbooks that were for advanced students, focused on a specific sub-domain of security or technology, or were published before 2016 were eliminated from further analysis. As a confirmation of the final selected set, we located 20 syllabi from general information security courses taught in the most recent academic year; each at a comprehensive 4-year university. None of the courses employed a textbook that was not on our original list, although two used texts that we had eliminated; one because it was an older edition, and the other because it focused on a more specific sub-domain of the field, putting it outside of this study's scope.

The remaining six textbooks had copyright dates of 2016 or newer and spanned four publishers and five authors. (Ciampa, 2016; Easttom, 2019; Smith 2019; Vacca, 2017; Whitman and Mattord, 2017; Whitman and Mattord, 2018). These six books are the subject of a content analysis for alignment to KSAs identified in position announcements.

3.4. Content Analysis of Textbooks for Alignment to KSA's from Current Position Postings

A detailed table of contents (TOC) was obtained for each of the six textbooks to be studied. For some publishers, the full text of the most current edition was available and the combined detailed outlines of each chapter was used. For others, a detailed TOC that spanned the entire textbook was accessible. In each case, the reviewed material included the highest-level subject and at least two levels of sub-topics below. This level of detail gave the authors of this study sufficient granularity to determine subject areas that were important enough or covered in adequate detail, without using the index which would have references to terms that may have simply been used or defined in the text.

Table 5. KSA Presence in TOC of Textbooks

KSA Term	T1	T2	T3	T4	T5	T6	#
Active Directory	x	x	x	x	x	x	6
Auditing	x	x	x	x	x	x	6
Cryptography/Encryption	x	x	x	x	x	x	6
Firewall	x	x	x	x	x	x	6
Intrusion Det/Prev	x	x	x	x	x	x	6
Log	x	x	x	x	x	x	6
Networking	x	x	x	x	x	x	6
Risk	x	x	x	x	x	x	6
Security Standards	x	x	x	x	x	x	6
Vulnerability	x	x	x	x	x	x	6
Patch	x	x	x	x	x	x	6
Client side / Server Side	x	x	x	x	x		5
Control	x	x	x	x	x		5
Endpoint	x	x	x	x		x	5
Forensic Tools	x	x	x	x	x		5
Identity Management	x	x	x	x	x		5
Mobile	x	x	x	x		x	5
Packet	x	x	x	x	x		5
Web Application	x	x	x	x		x	5
Malware	x		x	x		x	4
Penetration	x		x	x			3
Scripting	x		x			x	3
Windows	x	x	x				3
Database	x	x					2
Linux	x	x					2
SIEM	x	x					2
Unix	x	x					2
Cloud		x					1
Coding	x						1
API							0
	28	25	23	21	16	16	

For each KSA term in the superset from Table 2 and Table 4, a search on the detailed TOC was performed. The presence or absence of each term, or a close derivation of it, was then recorded. The results of this analysis are presented in Table 5. The first column shows the KSA term as discovered either in position descriptions on major job search sites, or in detailed content descriptions of certifications that appeared in job descriptions. The "T" columns indicate the presence of a term in one of the six textbooks. Textbook 1 (T1) the Smith (2019) textbook, T2 stands for the Vacca (2017) textbook, T3 indicates the Easttom (2019), T4 represents the Whitman and Mattord (2017) Principles book, T5 is the Whitman and Mattord (2018) Management book, and T6 is the Ciampa (2016) book. The last column shows the number of textbooks in which each term was located and the last row indicates the number of these terms that appears in that book. KSA terms are ordered from top to bottom beginning with the terms that appear in all textbooks. Texts are ordered from left to right beginning with the one which includes the most KSA areas.

Coverage of the areas is inconsistent across the textbooks, as indicated by the tables of contents. The first 10 KSA terms (33%) appeared in all six of the textbooks and 19 of them (63%) were addressed in at least four of the six texts. The remaining 10 KSA terms (33%) appeared in half or fewer of the reviewed textbooks, including "API" which was not explicitly mentioned in any table of contents from the textbooks.

4. Discussion, Limitations, and Conclusions

Although academia strives to educate students with knowledge, skills, and abilities that are demanded by employers; faculty perceptions of relevant topics may not align with industry requirements. In rapidly evolving fields like information security this problem can be particularly acute; with new topics constantly emerging, others becoming less important; and some transforming to the extent that they require entirely different approaches to understand. Because of this reality, instructors rely on textbooks authors to heavily influence the propriety of topics covered. Textbook authors, however, also face the same impediments in this dynamic field and are constrained by lengthy revision cycles of traditional publishers.

Because of the rapidly changing nature of information security, it might be expected that inclusion of knowledge topics currently demanded by the job market is strongly related to the publication date of the textbook. Indeed, the book with the most recent publication date (Smith, 2019) did include the most terms, however, it lacked "cloud" within the table of contents. This area has received growing attention in the IS community, particularly with respect to security, yet the only textbook that treated it with sufficient weight to appear in the table of contents was the second oldest of the textbooks (Vacca, 2017). In fact, a full-text search of two of the textbooks and their indices did confirm a lack of coverage in those books. It is important to recognize possible omissions such as these to ensure graduates who have taken even an overview course in information security understand the high-level issues.

An important contribution of this study is to draw attention to potential misalignment between areas of knowledge that are important in the information security industry, but that may not appear in popular textbooks intended to broadly cover the field. Instructors may use supplemental materials to cover any omissions; determine that the general security knowledge is sufficiently applicable to a specific omitted term; relegate coverage to a future course; or simply favor other topics. With limited resources, trade-offs may be necessary and this study assists in identifying potential problematic areas.

The authors performed a content analysis on textbooks as a surrogate for class content, because textbooks are more readily available than detailed syllabi or course outlines. The implicit assumption is that the content of a textbook has a significant influence over any course in which it is used. In many cases this is true; however, the authors recognize and accept that there are differences between formal materials and knowledge acquired within any course. Even if a topic has been identified by this study, some areas could be minimized or omitted in an introductory textbook with the expectations that later courses will cover them in more depth. It is also likely that there are so many potential subject areas in this field, that authors intentionally focus on a subset of areas to build a general awareness and competence.

It is also an unfortunate reality that in excess of 60% of university students do not purchase at least one of their required textbooks in a given term (Hilton, 2016; Martin et al., 2017). Students cite high cost and lack of alignment with exams and job aspirations. It can certainly be argued that students do not have a firm basis to know requirements for future jobs, so the burden falls to instructors to assure that selected textbooks and other content aligns well with actual job requirements to better justify the benefit that investment in a textbook has to students' career potential.

The process this study employed to determine the most important current and relevant knowledge, skills, and abilities is itself a contribution. Any highly dynamic academic discipline that lacks an official governing body to specify and revise expected knowledge requirements could benefit from this approach. In fact, the authors of this study recognize that it is highly likely that the specific terms identified may not appear if the study were replicated in the future. That does not diminish the value of the outcomes but highlights the need for a rigorous approach to assist in topical alignment between industry demands and academic offerings. A benefit of this simple cyclical method is that there appears to be a high level of convergence in terms within as few as 30 position announcements from two non-overlapping job search engines, making it very practical to replicate or apply elsewhere.

While professional certification review was an intermediate step in this study, we included the full results table because it contributes value in its own right. Out of the certifications explored, two of the top three certifications, CISA and CISM focus on management aspects of information security but lack direct mentions to many of the listed KSAs both are considered among the most important certifications to achieve for an information security job (Oltsik, 2019). This highlights the importance of both the technical and managerial aspects of security, as well as potential career paths for those able to manage projects and organizations, but who may not have the depth and breadth of technical detail.

It may also appear that using jobs specifically in the information security field to derive a set of KSAs for a general survey course is inappropriate. The study of job postings, however, was done to extract a set of the most critical issues facing companies today. The analysis of the textbooks, however, only revealed the presence of topical coverage, rather

than a measurement of depth. It is reasonable that an awareness-level coverage of the most pressing topics in a field be included, or that the instructor makes an informed decision to omit one or more areas.

Summing weighted results from multiple job sites could be considered overly simplistic. There is not an accepted method to determine unspecified inclusion terms for a textbook content analysis and other approaches were considered for this study. For instance, terms extracted from certification details could then have been multiplied by the Job Appearance Score, incorporating how often a certificate is referenced in a job announcement. The approach we used was more inclusive, significantly easier to understand, and met the guidelines for content analysis.

There are a few limitations in the content analysis itself that should be noted. First, this analysis only reviews six popular information security books. Although each is used by multiple universities for general coverage of information security, they do not represent the entire set of such textbooks. The study authors did, in fact, identify some university offerings that used other books. Second, a comprehensive search of the full text was not performed. The content analysis was done on a detailed table of contents for each book and the level of detail varied between authors and publishers. It is very likely that some of these terms appear within the full text, but the concept was not prevalent enough to appear in the table of contents. Third, although the study authors were diligent in the reviews and have sufficient experience in the domain to identify similar terminology to not unnecessarily exclude a text from the table, it is possible that some instances of KSA terms went undetected, or terminology was not similar enough to be recognized.

Finally, there is a need for a study more closely relating actual course content to information security positions. This study used textbook content as a surrogate for course coverage. A more direct review of detailed syllabi and course outlines would better reveal the degree to which higher education is addressing the needs of industry, creating the impetus for change and continuous improvement across the discipline.

4. References

(ISC)2. (2020) (ISC)2 Information Security Certifications. Retrieved from https://www.isc2.org/Certifications

AICPA. (2020). Association of International Certified Professional Accountants. Retrieved from https://aicpa.org

Bargate, K. (2012). The readability of managerial accounting and financial management textbooks. Meditari Accountancy Research, 20(1), 4-20.

Berelson, B. (1952). Content Analysis in Communication Research. Glencoe, Ill: Free Press.

Bracken, R. M., & Urbancic, F. R. (1999). Ethics content in introductory accounting textbooks: An analysis and review. Journal of Education for Business, 74(5), 279-284.

Braverman, B. (2017, May 9). Best Jobs in America. Retrieved from https://money.cnn.com/gallery/pf/2017/01/05/best-jobs-2017/5.html

Ciampa, M. (2016). Security Awareness: Applying Practical Security in Your World (5th ed.). Boston, MA: Cengage Learning.

CompTIA. (2020). CompTIA Security+. Retrieved from https://comptia.org/certifications/security

DeSensi, V. L., & Jurs, B. S. (2017). Coverage of Psychological Disorder Stigma in Introductory Psychology. North American Journal of Psychology, 19(3).

Easttom, C. (2019). Computer Security Fundamentals (4th ed.). London, UK: Pearson.

EC-Council. (2020). Certified Ethical Hacker. Retrieved from https://cert.eccouncil.org/certified-ethical-hacker.html

Fisher, C. D., & Southey, G. (2005). International human resource management in the introductory HRM course. The International Journal of Human Resource Management, 16(4), 599-614.

Foxman, E., & Easterling, D. (1999). The representation of diversity in marketing principles texts: An exploratory analysis. Journal of Education for Business, 74(5), 285-288.

GIAC. (2020). Cybersecurity Certifications. Retrieved from https://giac.org/certifications

Hilton, J. (2016). Open educational resources and college textbook choices: a review of research on efficacy and perceptions. Educational Technology Research and Development, 64(4), 573-590.

Holsti, O. (1969). Content Analysis for the Social Sciences and Humanities. Reading, MA: Addison-Wesley.

ISACA. (2020) ISACA CREDENTIALS. Retrieved from https://www.isaca.org/credentialing

Krippendorff, K. (1980). Content Analysis: An Introduction to Its Methodology. Newbury Park, CA: Sage.

Laksmana, I., & Tietz, W. (2008). Temporal, cross-sectional, and time-lag analyses of managerial and cost accounting textbooks. Accounting Education: an international journal, 17(3), 291-312.

Martin, M., Belikov, O., Hilton, J., Wiley, D., & Fischer, L. (2017). Analysis of student and faculty perceptions of textbook costs in higher education. Open Praxis, 9(1), 79-91.

Polikoff, M. S. (2015). How well aligned are textbooks to the common core standards in mathematics?. American Educational Research Journal, 52(6), 1185-1211.

Offensive Security. (2020). Course Overview. Retrieved from https://www.offensive-security.com/pwk-oscp/

Oltsik, J. (2019). The life and times of cybersecurity professionals. ESG and ISSA: Research Report. Retrieved from https://www.esg-global.com/hubfs/pdf/ESG-ISSA-Research-Report-Life-of-Cybersecurity-Professionals-Apr-2019.pdf

Research and Markets (2016). United States Online Recruitment Market 2016 – 2020 with LinkedIn, CareerBuilder, Monster & Indeed Dominating. PR Newswire.

Silk Road. (2016). Top Sources of Hire 2016. Retrieved from http://hr1.silkroad.com/source-of-hire-report-download

Smith, R. E. (2019). Elementary Information Security (3rd ed.). Burlington, MA: Jones and Bartlett Learning.

Stemler, Steve (2000) "An overview of content analysis," Practical Assessment, Research, and Evaluation, 7(17).

US News & World Report. (2020). Information Security Analyst Overview. Retrieved from https://money.usnews.com/careers/best-jobs/information-security-analyst

Vacca, J. R. (2017). Computer and Information Security Handbook (3rd ed.). Burlington, MA: Morgan Kaufmann Publishers.

Whitman, M. E., Mattord, H. J. (2018). Management of Information Security (6th ed.). Boston, MA: Cengage Learning.

Whitman, M. E., Mattord, H. J. (2017). Principles of Information Security (6th ed.). Boston, MA: Cengage Learning.

Weiser, M., & Conn, C. (2017). Into the breach: Integrating cybersecurity into the business curriculum. BizEd, 16(1), 36-41.

Zadelhoff, M. V. (2017). Cybersecurity has a serious talent shortage. Here's how to fix it. Harvard Business Review. Retrieved from https://hbr.org/2017/05/cybersecurity-has-a-serious-talent-shortage-heres-how-to-fix-it

Author Biographies

Mark Weiser is a Professor of Management Science and Information Systems at Oklahoma State University. He has published in leading journals and proceedings, focusing on the areas of upper-layer network protocols, security, forensics, and technology-supported teaching. Dr. Weiser was founding director of the Center for Telecommunications and Network Security and principle investigator for funded projects from DoD, NSA, AFOSR, NSF, and multiple private agencies.

Andrew Bowman is a PhD Student studying Management Science and Information Systems at Oklahoma State University. His research interests include Digital Piracy, Digital Activism, and Consequences of Blockchain Technology and other Decentralized Applications.