Privacy Considerations Throughout the Data Life Cycle

James J. Pomykalski
pomykalski@susqu.edu
Sigmund Weis School of Business
Susquehanna University
Selinsgrove, PA 17870, USA

Abstract

Due to the most recent technological advances of the Internet-of-Things (IoT) and data analytics algorithms, the issues of data privacy has been at the forefront of many enterprise strategies. Data privacy has been seen, throughout the years as a legal, public policy and, most recently, a risk management issue. In this paper, we begin the discussion of data privacy at the operational level by addressing operational initiatives and activities to be implemented. More specifically, these initiatives and activities are discussed across a data life cycle model. The focus of this paper is to examine the characteristics of each data life cycle stage and propose a series of organizational data management processes to address data privacy.

Keywords: Privacy, Strategy, Data Management, Risk Management, Organizational Processes, Consumer Data

1. INTRODUCTION

Despite the long history on the discussion of the issue of the "right to privacy", the privacy issue still remains largely ill defined. Nissenbaum (2009), in one of the seminal volumes on privacy, discusses privacy through the lens of contextual integrity because of the changing nature and culturally varying use of the term (Nissenbaum, 2009). Brandeis and Warren (1890) offered the most succinct argument for the right to privacy in 1890 when, against a growing backdrop of technological advances, they argued that privacy is simply "the right to be left alone".

In today's business environment privacy, specifically data privacy, is still an issue for many consumers and enterprises. Some commentators have suggested that enterprises should "abandon privacy as an organizational framework" (Bamberger & Mulligan, 2015, p. 22); largely due the issue of contextual integrity (Nissenbaum, 2009). However, many enterprises are addressing privacy through the development of a data strategy, which is becoming a necessary part of an overall enterprise strategy. "Data was once critical to only a few back-office processes, such as payroll and accounting. Today it is central to any business, and the importance of managing it strategically is only growing." (DalleMule & Davenport, 2017, p. 121).

A data strategy "helps by ensuring that data is managed and used like an asset. It provides a common set of goals and objectives across projects to ensure data is used both effectively and efficiently" (SAS, Inc., 2018, p. 4). Specifically, DalleMule and Davenport's (2017) data strategy framework includes both a data offense and data defense component.

"Data defense and offense are differentiated by distinct business objectives and the activities designed to address them" (DalleMule & Davenport, 2017, p. 114). The data offense component deals with the use of data for the purpose of increasing profitability, revenue, and customer satisfaction, while data defense component focuses on minimizing downside risk for the enterprise. Data defense includes activities that ensure compliance with regulations and the integrity of the data; including privacy

(DalleMule & Davenport, 2017); this is the focus of this paper.

However, in order to execute an effective data strategy an enterprise must understand the path data takes within the enterprise. Chisholm (2015) has proposed a seven-stage data life cycle (DLC). The DLC attempts to characterize the activities performed on data as it moves through an enterprise. This paper examines the activities that take place in each of the phases of the DLC and proposes measures and activities necessary for an enterprise to manage data privacy effectively.

Before proceeding further, it is important to discuss the use of terms data and information. The author acknowledges that data and information have distinct meanings, namely that information is data placed in context. However, in this paper, especially in reference to the work of other scholars, the words data and information are used interchangeably. In addition, at some point during the data life cycle data does become information therefore the privacy of the data as well as the privacy of derived information are essentially synonymous.

The goal of this research is examine the characteristics of each phase of the Chisholm DLC model and propose a series of organizational data management processes, using known practices as well as existing regulations, to address the privacy of data within an enterprise.

This paper is structured in the following manner. In the next section, we briefly review the literature around data privacy in terms of public policy, legal, and risk management frameworks. In section three, the Chisholm model (Chisholm, 2015) is described in more detail. In sections four through nine, privacy concerns/issues across each of the seven stages of the DLC are discussed in more detail. Section 10 briefly addresses future work in this area, while the final section presents a summary and some conclusions from this work.

2. LITERATURE REVIEW

In the early 1970s, the privacy literature concentrated on privacy as a public policy issue; that citizens had the right to privacy. Regan has stated that privacy should be framed as "a common value, a public value, and a collective value" (Regan, 1995, p. 241). She further argued that society is better when privacy is considered. Nissenbaum (2009) argues that is a social good. She states that the use of analytics can breach the privacy of individuals and others, such as in the 23andMe (Belluz, 2014) or the Target case (Duhigg, 2012). The 23andMe case involves the matching of genetic testing material by police to privately developed DNA databases. The Target case related to the use of data to identify pregnant customers. Target used this information to send coupons to these customers; one customer was a 17-year-old girl whose father was not aware of her pregnancy. Both of these cases are discussed later in this paper.

The legal discussion of information privacy was raised even earlier than Regan (1995) and Nissebaum (2009). Following on the work of Brandeis and Warren (1890), privacy was defined further by Westin (1967) where he initiated the concept of "informational self-determination" which means that individuals should have the right to determine the extent of the use of their information. While this does not show privacy as a public policy issue, it does assert that individuals should control privacy.

The issue of privacy has become more concerning with the acceleration of information technologies (Shaw, 2009; Johnson & Miller, 2009). Shaw (2009) believes that technological advances have made people reconsider the public/private distinction and has led the legal profession to become more concerned as well. Solove (2005) created a taxonomy of activities that invade the privacy of individuals. This taxonomy includes 16 privacy harms ranging from collection through surveillance, information aggregation, insecurity of information, and disclosure and exposure. Zuboff (2019) updates and extends the discussion on nature of surveillance activities undertaken with information technology.

The underlying principles of informational self-determination has been enforced since European Union's General Data Protection Regulation (GDPR) went into effect in May 2018. One aspect of this regulation focuses on data protection as it relates to ensuring people can trust an organization to use their data fairly and responsibly; this is a practical level of the fundamental right to privacy (Information Commisioner's Office (ico.), n.d.). The GDPR is built on seven underlying principles of data protection:

1. Lawfulness, fairness and transparency—enterprises have identified appropriate bases for data collection and processing, the enterprise have considered how processing data will impact individuals, and that enterprises are open and honest inform the public on data usage,
2. Purpose limitation—that enterprise have clearly identified and documented the purpose for data collection and processing,
3. Data minimization—enterprises collect only the data they specify; that the data has a specified purpose,
4. Accuracy—enterprises must build in checks for data accuracy throughout the life cycle,
5. Storage limitation—enterprises must maintain and adhere to data collection and retention policies,
6. Integrity and confidentiality—enterprises must maintain strict security measures to protect the data, and
7. Accountability—requires enterprise to take responsibility for personal data in their possession; demonstrate and document compliance.

These underlying principles are key to the GDPR and the issue of data protection and privacy. Each of these principles have an impact on data across the data life cycle.

Finally, a third area that warrants discussion is the practical operations and policies used by enterprises; namely C-level executives like Chief Privacy Officers (CPO). Hilliman states that "data governance oversight [especially in terms of privacy and security] should exist in an interdisciplinary and accountable setting" (Hilliman, 2013, p. 136). Bamberger and Mulligan (2015) have focused their examination of enterprises safeguarding privacy on current CPOs both in the US and in Europe. These CPOs view their role as strategic; that they spend time attempting to integrate "privacy concerns throughout decision making about firms goals, products, and services ensuring a voice on privacy matters is heard at the [executive] table" (Bamberger & Mulligan, 2015, p. 78). Another vital part of the job of a CPO is to interact and understand the external environment in which their enterprise operates. They spend time interacting and building relationships with privacy regulators, including the Federal Trade Commission and privacy advocacy groups.

"To the extent privacy governance requires the dynamic, 'learning' approach that many [CPOs] described, privacy is increasingly framed as a part of the evolving practice of risk management" (Bamberger & Mulligan, 2015, p. 81). In this way, privacy governance is about setting up guidelines for operational managers to operate in, then the role of the privacy team is to monitor and audit. As one CPO said it, "my team is not responsible for compliance, they're responsible for enabling the compliance of the business" (Bamberger & Mulligan, 2015, p. 84). If they hear of potential violations.

3. CHISHOLM DATA LIFE CYCLE MODEL

A data life cycle model fits well into this strategic view of privacy. The data life cycle does not necessarily define all the specific processes involved in handling data, it does provide "high-level", i.e., strategic, understanding of the activities within that stage regarding enterprise data.

While the Chisholm (2015) model is not the only data life cycle model available (National Network of Libraries of Medicine (NNLM), nd), the seven stage model best fits this research effort for three primary reasons.

First, the Chisholm model has generic applicability and specifically addresses the issue of data governance; the other DLC models address the needs of handling library and/or (National Network of Libraries of Medicine (NNLM), nd) research data. Data governance, within an enterprise, consists of the development of "a system of decision rights and accountabilities for information-related processes" (Thomas, 2014, p. 3).

Second, the Chisholm model has clearly described the separation of the stages and the clarity of the specific activities. The model phases depict "logical dependencies and not actual data flows" (Chisholm, 2015, "Critique"); data are harder to capture and are informed by enterprise business processes.

Third, the model phases focus on specific data governance activities that are unique to each phase; these are shown in Figure 1 below.

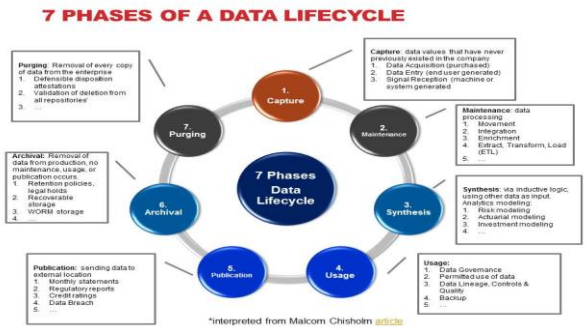

Figure 1: Data Governance in the Data Life Cycle

The focus of the Chisholm model is on the need to ensure sound data governance principles in every stage of the DLC. In the following sections, each phase of the Chisholm model—Figure 1—is introduced and the specific privacy concerns/issues in terms of data governance are addressed.

4. CAPTURE

The first stage, capture, deals with the initial, and original, entry of the data to the enterprise. Chisholm specifically mentions three primary means for data capture: data acquisition, data entry, and signal acquisition. Data acquisition involves the capture/acquisition of data by an enterprise from outside of the firewalls (Chisholm, 2015). This data capture can be done on a one-time or continuous basis.

In this phase, the enterprise has made a conscious effort to incorporate this data into their infrastructure; the data fits their business needs. The first three GDPR principles—lawfulness, fairness, and transparency; purpose limitation; and data minimization—are most applicable in this phase. The enterprise decides to collect particular data items based on identified appropriate bases for data collection and identifying and documenting the purpose for the data collection and usage. These criteria can be clearly shown in the documentation of data requirements and business need. In addition, adherence to privacy standards, set by management, can only be achieved through the establishment of clear priorities and formal policies. These policies should proactively define and align the rules for data collection provide on-going services to data stakeholders, and react to and resolve issues arising from non-compliance (Thomas, 2014). In addition, in terms of transparency of the data capture, enterprises should develop "guidelines that align the interests of companies and their customers, and ensure that both parties benefit from personal data collection" (Morey, Forbath, & Schoop, 2015, p. 100). A survey of 900 people from five different countries showed that consumers are largely unaware of the type of data that is captured routinely by firms (Morey, Forbath, & Schoop, 2015).

Privacy concerns, due to increased surveillance, have escalated especially since the late 1980s. In fact, in 1986 only about one percent of all data was digitized. However by 2013 nearly 98% of all data was in digital form (Zuboff, 2019); this increased the opportunity for many enterprises to collect data first and then postulate a purpose second. In her recent book, Zuboff (2019) suggests that this intensive digitization, along with increased use of information technologies, more powerful than those suggested by Shaw (2009) have led to this increase in "surveillance capitalism". Surveillance capitalism is "not technology; it is logic that imbues technology and commands it into action. Surveillance capitalism is a market form that is unimaginable outside the digital milieu" (Zuboff, 2019, p. 15).

5. MAINTENANCE & SYNTHESIS

The maintenance phase involves the storage, "integration, cleansing, enrichment, changed data capture, as well as the familiar extract-transform-load processes" (Chisholm, 2015, "Data Maintainence"). New data values are derived in this stage by using deductive logic. While in the synthesis phase, new data values are created by using, not deductive logic but inductive logic which requires the use of expert judgment, experience, and/or intuition as part of the logic. (Chisholm, 2015). The privacy issues and concerns of both of these phases are similar so the discussion covers both of these stages.

In these stages, captured data is combined with either other captured data or existing data from within the internal systems of the enterprise. The primary privacy concern at this stage is that of data anonymization or data pseudonymization. A lengthy discussion of the primary difference between an anonymization process and a pseudonymization, the activities in each process, and the risks—to security—associated with each process appears in a document published by the Information Commissioner's Office (Information Commissioner's Office (ico.), 2012). The

document provides guidance to enterprises that need to anonymize data, it also helps identify issues to consider ensuring effective anonymization of personal data and, finally, the document focuses on the legal tests required in the GDPR.

The data creation accomplished in these phases is the result of efforts to classify vast amounts to data into homogenous clusters or data found through insight creation techniques such as data mining; much of this latter work occurs in the synthesis phase. There is a higher risk of these new data elements containing personally identifiable data on specific individuals or groups making the need to understand the issues of data anonymization/pseudonymization important. Again, the establishment and adherence to policies regarding the storage of this data are of vital importance. In addition, the establishment of individual accountability to compliance with the policies needs clear documentation and enforcement throughout the enterprise.

This stage also requires a high degree of data security and many enterprises have experience in handling the integration, cleansing and enrichment of data through their efforts in creating a data warehouse. Data governance efforts that focus on data security, similar to those that focus on data privacy are also critical at this stage. The data governance efforts may be limited to only certain data—i.e., master data—and responsible data governance personnel will be accountable for:

- access management and security requirements,
- alignment of frameworks and data governance initiatives,
- assess risk and develop risk management plan,
- enforce regulatory, contractual, architectural compliance requirements,
- identify stakeholders, establish decision rights, and clarify accountabilities (Thomas, 2014).

Another issue that needs to be addressed in this stage, as well as in the next stage on data usage, is the development of the use of systems that automate data integration. Recent research has led to the development of machine learning/data mining systems that build in privacy preservation. Clifton et al. (2004) call for the need for further research, development, and use of privacy preserving systems especially in data integration and sharing efforts.

There are two recent examples of data synthesis efforts that draw concern. The first example involves the use of facial recognition software in US airports to enhance the boarding process (Funk, 2019). As passengers board their facial image is compared to images in a database of photos taken from visas, passports, and related immigration applications. While many believe that this increases security, many privacy concerns exist with facial recognition software systems. First, these systems do have a 99 percent accuracy rating for identifying white men, while the error rate for females and people with darker skin tones is as high as 35 percent (Funk, 2019). In this case, females and minorities have an increased likelihood of being targeted for additional screening measures. The second concern raised with the use of facial recognition is the use of the databases itself. Funk (2019) states "Americans should be concerned about whether images of their faces collected by this program will be used by companies and shared across different government agencies. Other data collected for immigration purposes—like social media details—can be shared with federal, state, and local agencies. If one government agency has a database with facial scans, it would be simple to share the data with others" (Funk, 2019, Para 10). In this case, it is not the data that is misused but the technology utilizing the data that could result in privacy violations (Funk, 2019).

In the second, more widely publicized case, an open source DNA database was used by police in California to track down the Golden State Killer (Molteni, 2018). The case involves questions of privacy because of the frequent use of genetic testing through firms like 23andMe and Ancestry. In this case, the police matched the killer's genetic profile, obtained through old crime scene samples, to samples in the open source DNA database. Once a pool of individuals was obtained they used other clues—sex, age, place of residence—to rule out suspects. Eventually their search found a single suspect, which was confirmed through matching his DNA to the crime scene samples (Molteni, 2018). While some people may be comfortable with this police tactic, others, especially privacy experts, are concerned.

6. USAGE

The usage stage of the DLC is described as "the application of data as information to tasks that

the enterprise needs to run and manage itself" (Chisholm, 2015, "Data Usage").

Chisholm (2015) goes on to point out that with any data there may be additional permissions that are needed through the data governance structure. Specifically, Chisholm (2015) refers to legal or regulatory restrictions imposed by outside agencies that restrict the use of the data to certain business processes.

The other issue left to consider at this stage is the use of privacy-preserving data models, namely machine learning/data mining, which will further protect the privacy of individuals in this stage. The current state of these privacy-preserving systems is outlined in a recent post by engineers at Dropout Labs (Mancuso, DeCoste, & Uhma, 2018).

In this report, the engineers discuss the impact of machine learning systems on privacy and the ongoing research and development in privacy-preserving systems. The authors view the development of these systems, especially open source systems, as positive and encourage enterprises to examine and utilize these systems.

7. PUBLICATION

Chisholm (2015) describes the data publication stage as the point in time where data is sent to a location outside of the enterprise. Data publication is the stage that represents the point in time where data is beyond recall or correction. This stage also covers the breach of data from internal systems.

Data publication is often the first time that consumers may react to lack of transparency in data capture, synthesis, and usage of their personal data. After the publication of the Duhigg article (2012) on Target's use of personal data to identify pregnant customers, in particular a minor female, Target made a public apology and quietly withdrew the program.

Data breaches cause enterprises major issues in that consumers often find that their personal data, much of which they did not know was being held by a firm, was compromised. Data breaches carry both legal and public relations implications, which cost enterprises in terms of not only dollars but also reputation, brand, consumer trust and, potentially, future sales (Bowers, 2011). Data breaches may also have implications more directly on consumers in the form of identity theft possibilities.

8. ARCHIVAL

Data archival is "the copying of data to an environment where it is stored in case it is needed again in an active production environment, and the removal of this data from all active production environments" (Chisholm, 2015, "Data Archival"). A data archive is a place to store data once its useful life has be exhausted. The data remains as part of the data infrastructure and can be restored if necessary, however, no further maintenance, usage or publication occurs.

While many of the same concerns about security and data anonymization/pseudonymization exist at this stage to those discussed in the maintenance and synthesis stages, one of the specific issues questions facing enterprises with data archival is the development of a data retention policy. The GDPR address the issue of data retention through the storage limitation principle, which states that for personal data enterprises should (1) not keep the data longer than needed, (2) be able to justify the length of retention for the data, (3) have a policy setting group set compliance standards that are well documented (Information Commissioner's Office (ico.), 2012).

9. PURGING

The final stage of the data life cycle is the purging of all instances of the data item from the enterprise and its systems (Chisholm, 2015). Data deletion is difficult for enterprises but, for much of the data in current systems, "the costs of keeping data are higher than you think, and the benefits are lower…there is a chance it will be harmful—like being lost in a breach or subpoenaed in a lawsuit" (Branscombe, 2019, Para 6). Branscombe (2019) goes to state that about a third of data, stored in current data centers, is likely redundant, obsolete or trivial and since it holds no business value; it should be purged.

Purging data can be cost-effective as well as risk reducing. Additional costs may be incurred due to additional anonymization/ pseudonymization processes. In addition, the risk of having the data lost and de-identified often outweigh the loss of this data. In fact, Joshua de Larios-Heiman, chair

of the California Lawyers Association Internet & Privacy Law Committee, warns that enterprises should think of this data as uranium rather than oil; as assets that could become toxic (Branscombe, 2019).

10. FUTURE WORK

This work has two natural extensions. First, the focus of future work can shift from looking at data governance activities that help ensure privacy to those that help ensure data quality. Data quality is still a part of the data defense strategy (DalleMule & Davenport, 2017) and is a natural extension of this work on data privacy. Data quality has been cited, as one of the key issues that enterprises face, especially with the use of analytics, in ensuring that the data that is being used is of the highest quality; as defined by the end users (Kwon, Lee, & Shin, B, 2014; Hazen, Boone, Ezell, & Jones-Farmer, 2014). Understanding how users view data quality in each of the different phases of the life cycle is important in understanding how data quality improvement is accomplished.

Second, further delineation of the data governance activities described in each of the seven phases of the DLC needs to be undertaken. This would include the use of current practices found within existing enterprises in managing privacy throughout the DLC. Subsequent work can focus on specific key phases in the DLC individually.

11. SUMMARY/CONCLUSIONS

This paper addressed an initial set of data governance activities and initiatives that need to be addressed within each stage of the data life cycle. While this is not the first work to begin to address these data governance activities, it is the first look at data governance with respect to privacy across particular stages of a data life cycle.

This work illustrates the number of issues and concerns that must be addressed when it comes to data privacy. While, in the US the legal oversight concerning data privacy is scattered because individual state laws are enforced. For example, California will have a new law in effect, the California Consumer Privacy Act (CCPA) (Californians for Consumer Privacy, n.d.), in January of 2020 to address data privacy. However, the recent enactment of the European Union's (EU) General Data Protection Regulation (GDPR) has come a long way in forcing US enterprises to address privacy more rigorously due to the fact that any enterprise doing business within the EU must adhered to these new standards.

The problem is that while some enterprises are dealing with data privacy as a risk management issue (Bamberger & Mulligan, 2015) there seems to be little progress being made on privacy as a public policy issue.

Overall, this paper focused on the examination of the characteristics of each data life cycle stage and the unique activities that need to be addressed in each stage with regard to data privacy.

12. ACKNOWLEDGEMENTS

The author would like to thank the anonymous editors of this paper for their thoughtful, constructive, and detailed comments and feedback. This version of the paper is not only more readable, but also provides a clearer and better-articulated understanding of privacy activities across the data life cycle.

13. REFERENCES

Bamberger, K. A., & Mulligan, D. K. (2015). *Privacy on the Ground: Driving Corporate Behavior in the Inited States and Eurpoe.* Cambridge, MA: The MIT Press.

Belluz, J. (2014, December 18). Genetic testing brings families together, and sometimes tears them apart. *Vox*, p. 2014. Retrieved June 21, 2019, from https://www.vox.com/2014/9/9/6107039/23andme-ancestry-dna-testing

Bowers, T. (2011). *Security as Business Risk: How Data Breaches Impact Bottom Lines.* Experian.

Brandeis, L., & Warren, S. (1890). The Right to Privacy. *Harvard Law Review, IV*(5).

Branscombe, M. (2019, July 3). Data Deletion: Your data Strategy' Greatest Defense. *CIO Magazine*. Retrieved July 24, 2019, from https://www.cio.com/article/3405129/da

ta-deletion-your-data-strategys-greatest-defense.html

Californians for Consumer Privacy. (n.d.). *About the Law*. Retrieved July 24, 2019, from CAPrivacy.org: https://www.caprivacy.org/about

Chisholm, M. (2015, July 9). Seven Phases of a Data Life Cycle. *Information Management*, p. n.p. Retrieved May 28, 2018, from https://www.information-management.com/news/7-phases-of-a-data-life-cycle

Clifton, C., Kantarcioglu, M., Doan, A., Schadow, G., Vaidya, J., Elmagarmid, A., & Suciu, D. (2004). Privacy-preserving data integration and sharing. *9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery* (pp. 19-26). Paris: ACM. Retrieved September 27, 2019

DalleMule, L., & Davenport, T. (2017, May-June). What's Your Data Strategy? *Harvard Business Review*, pp. 112-121.

Duhigg, C. (2012, February 16). How Companies Learn your Secrets. *New York Times*, p. 2012. Retrieved from https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

Funk, A. (2019, July 2). I Opted Out of Facial Recognition at the Airport—It Wasn't Easy. *Wired*. Retrieved July 22, 2019, from https://www.wired.com/story/opt-out-of-facial-recognition-at-the-airport/

Hazen, B., Boone, C., Ezell, J., & Jones-Farmer, L. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics, 154*, 72-80.

Hilliman, C. A. (2013). Data Privacy, Security, and Compliane through Data Governance. In N. Bhansali, *Data Governance: Creating Value from Information Assets* (First ed., pp. 125-148). Boca Raton, FL: CRC Press.

Information Commisioner's Office (ico.). (n.d.). Introduction to Data Protection. Wilmslow, Cheshire, UK. Retrieved June 26, 2019, from https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/

Information Commissioner's Office (ico.). (2012). *Anonymisation: Managing Data Protection Risk Code of Practice.* Wilmslow, Cheshire (UK): Information Commissioner's Office (ico.). Retrieved July 20, 2019, from https://ico.org.uk/media/for-organisations/documents/anonymisation-code.pdf

Johnson, D. G., & Miller, K. W. (2009). Information Flow, Privacy, and Surveillance. In D. G. Johnson, & K. W. Miller, *Computer Ethics* (pp. 81-108). Upper Saddle River, NJ:: Prentice Hall.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management, 34*(3), 387-394.

Mancuso, J., DeCoste, B., & Uhma, G. (2018, January 10). Privacy-Perserving Machine Learning 2018: A Year in Review. Retrieved September 27, 2019, from https://medium.com/dropoutlabs/privacy-preserving-machine-learning-2018-a-year-in-review-b6345a95ae0f

Molteni, M. (2018, April 27). The Creepy Genetics Behind the GOlden State Killer Case. *Wired*. Retrieved July 22, 2019, from https://www.wired.com/story/detectives-cracked-the-golden-state-killer-case-using-genetics/

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. *Harvard Business Review, 93*(5), pp. 96-105. Retrieved July 24, 2019

National Network of Libraries of Medicine (NNLM). (nd). *Data Life Cycles: Research Data Cycles and Guides*. Retrieved June 5, 2019, from National

Institute of Health: https://nnlm.gov/data/data-life-cycles

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of the Social Life.* Palo ALto, CA: Stanford University Press.

Regan, P. (1995). *Legislating Privacy: Technology, Social Vaues, and Public Policy.* Chapel Hill, NC: University of North Carolina Press.

SAS, Inc. (2018). *The Five Essential Components of a Data Strategy.* Cary, NC: SAS Inc.

Shaw, J. (2009, September). Exposed: The eroision of privacy in the Internet era. Cambridge, MA. Retrieved June 26, 2019, from https://harvardmagazine.com/2009/09/privacy-erosion-in-internet-era

Solove, D. J. (2005). A taxonomy of privacy. *University of Pennsylvania Law Review, 154*, 477-560.

Thomas, G. (2014, June 11). *The DGI Data Governance Framework.* Retrieved 2019, from The Data Governance Institute: http://www.datagovernance.com/dgi-data-governance-framework/

U.S. Geological Survey. (nd). *Data Management*. Retrieved June 5, 2019, from USGS.org: https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle

Westin, A. F. (1967). *Privacy and Freedom.* New York, NY: Atheneum.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Fronier of Power.* New York, NY: PublicAffairs.