Privacy: The Cost of Convenience

Prabesh Dhakal[1]

[1]*Registration Number: 3032362, Email: prabesh.dhakal@stud.leuphana.de*

Abstract

This article presents a novel framework to discuss data privacy as it relates to professionals that work with digital data technologies. In addition, recent data privacy cases and three recent changes in the data privacy landscape are discussed. Some recommendations that are aimed towards data professionals are made. Finally, a need for a deeper study in the area of changes in the data privacy landscape and the consequences of these changes to professionals that work with data is also highlighted.

Keywords: data privacy, big data, technology, ethics, responsibility

1 Introduction

As the world embraces internet connected devices and services, data has become ubiquitous. Similarly, following the increase in computation capacity of computers and research in the field of data analytics and artificial intelligence, individuals, companies, and governments have been eager to leverage the available data and the advances in computing and analysis to achieve their respective goals.(Gandomi and Haider 2015; Chen, Chiang, and Storey 2012) This has resulted to many jobs that deal with some form of digital data technologies. Consequently, the professionals that are able to deal with the data and data related technologies are in high demand (Hammerbacher and Segaran 2009). Given that the field of information technology is in constant change, this paves way to many challenges. One such challenge that relates specifically to data is data privacy. This paper seeks to discuss a subset of this challenge: the cost of convenient action with regards to data privacy with a main focus on the role of data experts.

In order to build the argument that the data privacy landscape is changing and data experts must take actions that prioritize convenience, definitions of terminologies used in this paper are set in Section 3. These definitions fulfill two purposes: (i) they clarify how these terms have been used in this paper, and (ii) they allow provide insight into of how these terms are discussed and understood in the data privacy related literature. Further, a *key-vault* paradigm that aims to enable a clearer discussion about data exerts' role in the context of data infrastructure, data owners, and data users is proposed in Section 4. Additionally, Section 5 presents an assessment of the data privacy landscape is made from two perspectives: (i) privacy infringement from different actors in the landscape (individuals, companies, and the government), and (ii) new developments concerning data privacy. Finally, recommendations for data experts as the key audience are made in Section 6. However these recommendations could apply to individuals that are not strictly "data experts" as defined in this paper.

2 Methods

First, a literature review was conducted with the aim of gathering research papers, reports, and news articles relevant to the topics of data privacy and security. While more preference was given to recent literature related to matters such as data breaches and latest trends, more established literature was consulted to establish definitions of key concepts.

Then, key data privacy and security infringements as well as key changes taking place with regards to the topic of data privacy were selected to be highlighted in the paper. The selection

Data Privacy and Ethics (WiSe2019/20)

Submitted to:
Prof. Dr. Andreas Möller

Submission Date:
March 15, 2020

process was motivated by the recency and scale of the events being discussed. Similarly, the role of individuals in the data privacy landscape and how that relates to the issue of privacy were investigated based on a *key-vault* paradigm with focus placed particularly on the role of data experts in the privacy landscape.

Finally, recommendations were outlined based on the literature and the author's experience.

3 Defining Privacy, Cost, and Convenience

Key terminologies that are relevant to this paper are discussed in this section.

3.1 Privacy

Privacy is a multifaceted concept that does not have one clear definition as it encompasses a wide range of concepts such as freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations (Solove 2008). The fact that there is no set single comprehensive definition for privacy, the necessity of clarifying why privacy infringements are harmful is often overlooked when talking about privacy invasions (Solove 2015). Solove (2015) further states that the concept of privacy is historically contingent because it is a product of norms, activities, and legal protections. Solove (2008) also maintains that given the breadth of meanings that the term "privacy" carries, specifying the context in which it is used is crucial. Hence, the need for specifying on which definition of privacy this paper is based on was recognized.

The most apt definition of privacy in the context of this paper comes from Cambridge University Press (2008). It states that [privacy] "is the right that someone has to keep their personal life or personal information secret or known only to a small group of people". Further, the term privacy, used interchangeably with "data privacy" in this paper, applies to personal or group data in the sphere of the internet and world wide web, regardless of the data generation, storage, and transmission mediums involved. Further, which data is private depends on what the owner of the data perceives to be private.

Despite narrowing down the definition of privacy and providing a clear context of how the term is being used in this paper, a clear answer to "Who owns the (private) data?" is still lacking. However, answers to this question is not discussed in detail in this paper as it is out of the scope of this paper.

3.2 Data and Information

Zins (2007) states that data, information, and knowledge are human artifacts. The term "data" usually refers to facts that have been recorded in some way. The nature of data that are collected and the methods employed for their storage and transmission have changed over time. This means that everything from ancient hieroglyphs to the bits and bytes stored on a modern electronic data storage medium all count as data. While there is a general consensus on the meaning of data, the definition for the term "information" still seems to be contextual.(Machlup 2014; Zins 2007; Buckland 1991) Whereas Zins (2007) distinguishes between data, information, and knowledge, Buckland (1991) states that data falls under a special case of information, *information-as-thing*, because they are regarded as being informative. Machlup (2014) states that the line separating information and knowledge is so thin that distinguishing them is irrelevant.

In this paper, the approach of Buckland (1991) was adopted for the definition of data; since things like private data, behaviors, etc. are informative, they themselves constitute information.

Personally Identifiable Information

Whereas data can mean a wide array of things, one of the main types of data, or information, that this paper focuses on is "personally identifiable information". Personally identifiable informa-

tion is information or data that can be used to distinguish or trace an individual's identity (Arefi, Alexander, and Crandall 2018). Examples of personally identifiable information are the last names of a company's employees, or the date of birth of a student in a class room at school, etc.

For the purpose of this paper, data constitutes electronic records of peoples' private information, behaviors, and possessions such as photographs, audio recordings, and videos. The data or information in question could be personally identifiable, or could be of economic interest to some entity or group even if not personally identifiable.

3.3 Cost in the Context of Privacy

Although the notion of cost is most commonly applied in financial or economic sense, the term has additional meanings. It can be used to describe something of value that has to be traded away in order to receive something else. Similarly, the term "cost" could also be applied as an alternative to "loss" (Cambridge University Press 2008).

When discussing data privacy, the meaning of *cost* depends on the context. With regards to big data, the vast amounts of collected data have substantial economic value (Acquisti, Taylor, and Wagman 2016). The situation of a person giving away their email address for a discount on an online retailer platform is different from someone having their information hacked and leaked online because their password was compromised. However, in either case one deals with an object, namely personal data, that carries some value whenever one refers to the concept of cost. Hence, the object or concept of value for the purpose of this paper is *data*, and the cost incurred range from emotional toll and trust to money and individual privacy itself.

3.4 Convenience

The meaning for this term is derived from Cambridge University Press (2008) and which defines convenience as "the fact that something is suitable for your purposes and causes no difficulty for your schedule or plans or the fact of something being easy to do or get to."

Actions that constitute convenience in the of data privacy are termed as "convenience oriented actions" in this paper and are discussed in further detail in Section 4.

4 People, Data, and Privacy

Now that the key terms are defined, this section discusses how human beings, data, and privacy interrelate to each other using the *key-vault* paradigm that was briefly mentioned above. Further, the position of *data experts* in this paradigm is discussed.

If one considers data, and by extension privacy, as the object of value that needs to be stored in a secure manner, one can consider the storage medium as a *vault*, much like the bank vaults that store money. These vaults, then, are accessed by certain *key owners* who are certified to do so, either because they are the owners of the data, or because the owners of the data have provided them permission to access the data. An example for the latter case could be an online service platform that has been authorised to use the phone number of an individual for two factor authentication. Furthermore, the *key* used to access this vault could be a password, or biometric information such as the fingerprint scan, etc. Finally, there are certain individuals or companies that have a special role of acting as *vault makers and maintainers*. These are the individuals or companies that are responsible for the design, implementation, upkeep, and improvements of the data vaults.

This analogy is visually summarised with a simple diagram in Figure 1.

4.1 Where do data experts factor in?

Finally, *data experts*, individuals that access data for various purposes on behalf of the key owners or the vault makers, lie at the crossroads of the three key entities that were mentioned above (1). Specifically with regards to professionals that work in the field of data science and machine learning, data experts perform three main activities: (1) access the vault (the data), (2) train algorithms and generate insights, and (3) deploy results based on step 2. These activities constitute the main focus of the paper.

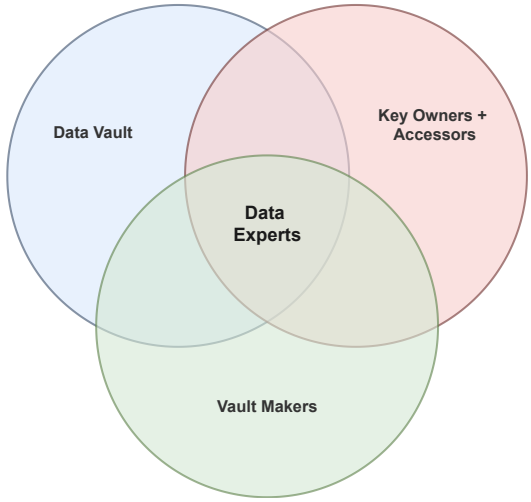

Figure 1. A Venn diagram representation of the "key valult" paradigm in Section 4 which highlights the three key components: *data vault, key owners and vault accessors, and vault makers*. Further, it shows that *data experts* lie at the intersection of all the three.

4.2 Convenience Oriented Actions

Although the term "convenience" was defined in Section 3, further clarification on what convenience oriented actions specifically mean is required. Thereby each of the three actions that *data experts* engage in on their professional and personal lives will be taken into account.

The term "convenient action" is a broad concept whose meaning depends on the context, for the purpose of this paper the term is used as an action or activity that is convenient to perform. This means that a convenient action is either easy to do or is a suitable action that causes no difficulty for schedule or plans at the time that the action needs to be performed. Some examples of convenience oriented actions of data experts for each activity that they may perform are laid out below.

When accessing the vault (or the data repository), the expert may choose to access any data from a vault without considering the legality or credibility of the source of the data. Similarly, when it comes to training machine learning algorithms or generating insights based on the data, the expert might place focus on getting the results without considering the legitimacy of the data, the code that they are using or the potential privacy violation that might result from their algorithms. Finally, when it comes to deploying the results of their analyses, they may deploy results without considering the consequences of their results or the security of their method of deployment; for instance, they may deploy an analysis with sensitive information on an unsecured public server.

Although these examples seem to be created ad-hoc, these are concrete cases where individuals or companies have made an offence against data privacy regulations by failing to employ more rigorous practices. The link presented in Footnote 2 contains a list of data privacy related incidents that were penalized in the EU after the enforcement of General Data Protection Regulation (GDPR).

5 Assessment of the Data Privacy Landscape

5.1 Recent Cases of Privacy Infringements

In the following, privacy infringement performed by entities at three different levels-individual, companies, governments-is presented.

Privacy Violation by Individuals

The first example of privacy violation is taken from the German state of Baden-Württemberg where the state fined a police officer for the first time ever in Germany. A police officer was fined for using the license plate number of a private individual via the central traffic information system of the Federal Motor Transport Authority to obtain a mobile phone number to contact the person without any consent or official intention. (Pressestelle, LfDI Baden-Württemberg 2019)

Privacy Violation by Companies

One of the most prominent examples of privacy violation performed by a company in the decade following 2010 is a case involving Facebook, the biggest social media platform at the time. Facebook settled a legal case with the Federal Trade Commission (FTC) in the US on July 24, 2019 in which FTC accused Facebook of misrepresenting the control that the users had over their personal information and failing to implement a reasonable program to ensure consumers' privacy. A penalty of 5 billion US dollars was imposed onto Facebook by FTC. (Federal Trade Commission (C) 2020)

On the same year, Facebook was involved in another scandal together with a British data analytics company named Cambridge Analytica. FTC alleged that Cambridge Analytica, along with a mobile application developer, used an app that harvested millions of Facebook users' private information for voter profiling and targeting. Cambridge Analytica settled the lawsuit from FTC. (Federal Trade Commission (C) 2020)

Yet another example of data privacy violation from a private company is the case involving Equifax, a credit scoring company in the US. FTC alleged that Equifax failed to secure customers' personal information that was stored in its network, failed to patch software vulnerabilities, failed to segment its database servers, and further stored Social Security Numbers in unencrypted plain text file format. Equifax settled for a sum of 575 million US dollars. (Federal Trade Commission (C) 2020; Federal Trade Commission (B) 2019; Federal Trade Commission (A) 2018)

Privacy Violation by Governments

Although governments are seen as the guardians of public interest, they are not free from actions that violate data privacy of private individuals. One of the biggest revelations came from Edward Snowden in 2013. Snowden revealed that the National Security Agency (NSA) implemented a mass surveillance program in collaboration with other governments' secret service agencies, including Germany and the U.K., and even private multinational companies like Google and Microsoft in order to collect data on private citizens with the aim of preventing terrorism.[1].

More recently, it was revealed that more than four hundred police departments across the US partnered with a doorbell and camera company owned by Amazon named Ring. The cameras are sold to homeowners as video security devices and are always connected to the cloud service that Ring offers. The police departments were given access to the video recordings of neighborhoods of their jurisdiction without a consent from the owners of the cameras. The police departments stated that this was done in order to increase the safety of the neighborhoods, however citizens were concerned that this move from the local governing body might give way to mass surveillance from the government. (Harwell 2019)

1. Details of the Snowden revelation has been covered comprehensively here: https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded

These examples make clear that data privacy infringements do not take place at one particular level of the society. Privacy violations have been committed by individuals, profit oriented companies, and governments for various reasons. While in many cases, the privacy violations are motivated by profit orientation (in case of private companies), or greater control over a society's security level (in case of the governments), there are also many examples where the violations were committed not in the least due to a preference to convenience oriented actions by individuals, companies, and even governments. There have been many other cases where individuals and companies have failed to comply to the legal data access, processing, and storage standards and have faced legal actions from regulatory agencies.[2]

5.2 Key Changes in the Data Privacy Landscape

The data privacy landscape is evolving in many ways. These changes in the data privacy landscape have far reaching consequences that not only affect data experts but also the general public in many frontiers (Manyika *et al.* 2011). This section presents three key global changes with regards to data privacy.

Increasing political discourse about data privacy

As the number of ways data can be generated, accessed, and stored has increased in recent years, there has been an increase in the political discourse about data privacy as evidenced by the fact that protecting privacy in big data is a fast growing research area (Mehmood *et al.* 2016). Even though the technology itself remains politically neutral, big data, in combination with Artificial Intelligence (AI), has been said to pose an unprecedented challenge to democracy (Körner 2019). In the wake of the Snowden revelations, Facebook and Cambridge Analytica scandals, and myriad other data privacy and security infringement cases, it is clear that the issues related to data privacy have achieved higher prominence in the public discourse.

Rapidly changing technology

Technology that deals with data generation, collection, access, and transmission is changing rapidly. There is an increase in easy access to fast internet and electronic devices that make spontaneous and carefree uploading of user-generated content extremely easy (Smith *et al.* 2012). As a result, the rate of data generation is rising everyday which has exacerbated the challenge of secure data storage, access, and usage (Mehmood *et al.* 2016).

At the same time, the field of data security and privacy is an active research area where new studies are published every day. Concepts such as *privacy preserving methods*, *differential privacy*, *identity based anonymisation*, *k-anonymity*, and *l-diversity*, etc. are taking momentum not only in the research arena, but also in commercial applications (Jain, Gyanchandani, and Khare 2016). These new developments could enable better handling of data by companies and data experts in terms of privacy and security.

Strengthening legal landscape

In response to the proliferation of data collection and transmission technologies (discussed above), governments across the world have updated and deployed more strict data privacy related regulations. One of the most significant events on this front was the adoption of the General Data Protection Regulation by the European Union (EU) in 2016, which was enforced in 2018. This move forced all companies handling EU residents' personal data to review and revise their organisational and technical privacy protection measures and develop new policies that ensure the compliance

2. An updated list of GDPR violation from individuals and companies can be found at https://enforcementtracker.com/

with the regulation (Tikkinen-Piri, Rohunen, and Markkula 2018). In the United States of America, the Federal Trade Commission (FTC) is responsible for the enforcement of data privacy and has been actively pursuing legal action against companies and individuals that infringe on data privacy and security regulations. Some recent examples of FTC's actions against offenders have been highlighted above. Specifically, the state of California in the United States of America has been pushing for stronger privacy laws and has passed the *California Consumer Privacy Act* that went into effect on January 01, 2020. (State of California, Department of Justice 2019; Kari 2019) Further, updating data privacy laws is also a highly prioritized issue in Canada (Government of Canada, Department of Justice 2019).

6 Recommendations

In light of the recent developments in the global data privacy landscape, there are several steps that individuals, specifically *data experts*, can take in order to ensure that they are not violating any privacy laws. Three of them are highlighted in this section. These recommendations are by no means meant to be exhaustive and are meant to serve as general guidelines instead of legal advice:

1. Cultivate interest in discussions about data privacy.
2. Update oneself on the data privacy laws for one's region of operation.
3. Adopt privacy oriented practices in one's workflow whenever possible.

With regards to the examples of *convenience oriented actions* that were highlighted in Section 4, some actions that are *not* convenience oriented could be the following:

When accessing data vaults, a data expert verifies the legality, authenticity, and the usefulness of the source of the data. Similarly, when training algorithms or generating insights using the data that they have, the data expert validates the security of the software or code that they use to work with the data. Finally, in the phase of deployment of results of the previous step, the expert ensures that the software and hardware environments where algorithm trained on the data or insights generated using the data are deployed are secure to the best of their capacity, and the expert also puts special care to ensure that publishing the results do not infringe on anyone's privacy.

7 Conclusion

The main aim of this paper was to illustrate that actors in the data privacy landscape can engage in actions that are oriented towards convenience, thereby infringing on the privacy of their stakeholders or the privacy laws of the legal region that they operate in.

A new paradigm named *key-vault* paradigm was proposed and used as a lens through which the role of data experts could be understood. The paradigm places data experts at the intersection of three different components that are related to how individuals interact with information technology systems: data storage and transmission infrastructure (vaults), access credentials and roles (key), and creators of the systems that store and transmit the data (vault makers).

Further, key changes in the data privacy landscape were discussed. This discussion formed the basis of the recommendations that were made. The recommendations with data experts as the target audience, emphasized the importance of keeping oneself up-to-date with the data privacy laws and regulations, active participation in data privacy related discussions, and active adoption of privacy oriented practices in one's professional and personal workflow.

Limitations of the study. This project serves as a starting point of a larger discussion in the field of data privacy and how it relates to data experts. Given the scope of the project and the concomitant time constraints, only a small cross-section of the larger issue could be investigated and discussed. Hence, there is a need for a more structured review of the research area which provides more comprehensive insights into the topic of data privacy in a changing technological,

legal, and societal landscape. Furthermore, how the changes relates to the roles which data experts take at an individual level, at companies or governing bodies that they work at needs to be further elaborated.

References

Acquisti, A., C. Taylor, and L. Wagman. 2016. "The economics of privacy." *Journal of economic Literature* 54 (2): 442–92.

Arefi, M. N., G. Alexander, and J. R. Crandall. 2018. "PIITracker: Automatic Tracking of Personally Identifiable Information in Windows." In *Proceedings of the 11th European Workshop on Systems Security.* EuroSec'18. Porto, Portugal: Association for Computing Machinery. ISBN: 9781450356527. doi:10.1145/3193111.3193114. https://doi.org/10.1145/3193111.3193114.

Buckland, M. K. 1991. "Information as thing." *Journal of the American Society for Information Science* 42 (5): 351–360. doi:10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199106%2942%3A5%3C351%3A%3AAID-ASI5%3E3.0.CO%3B2-3.

Cambridge University Press. 2008. *Cambridge Academic Content Dictionary Reference Book with CD-ROM.* Cambridge University Press. ISBN: 9780521691963. https://books.google.de/books?id=pqlRO2jdI2gC.

Chen, H., R. H. Chiang, and V. C. Storey. 2012. "Business intelligence and analytics: From big data to big impact." *MIS quarterly:* 1165–1188.

Federal Trade Commission (A). 2018. *Privacy & Data Security: Update 2017.* https://www.ftc.gov/reports/privacy-data-security-update-2017-overview-commissions-enforcement-policy-initiatives.

Federal Trade Commission (B). 2019. *Privacy & Data Security: Update 2018.* https://www.ftc.gov/reports/privacy-data-security-update-2018.

Federal Trade Commission (C). 2020. *Privacy & Data Security: Update 2019.* https://www.ftc.gov/reports/privacy-data-security-update-2019.

Gandomi, A., and M. Haider. 2015. "Beyond the hype: Big data concepts, methods, and analytics." *International Journal of Information Management* 35 (2): 137–144. ISSN: 0268-4012. doi:https://doi.org/10.1016/j.ijinfomgt.2014.10.007. http://www.sciencedirect.com/science/article/pii/S0268401214001066.

Government of Canada, Department of Justice. 2019. "Modernizing Canada's *Privacy Act*." Accessed: 2020-02-02. https://www.justice.gc.ca/eng/csj-sjc/pa-lprp/modern.html.

Hammerbacher, J., and T. Segaran. 2009. "Information platforms and the rise of the data scientist." *Beautiful data: the stories behind elegant data solutions:* 73–84.

Harwell, D. 2019. "Doorbell-camera firm Ring has partnered with 400 police forces, extending surveillance concerns." Accessed: 2019-12-02. https://www.washingtonpost.com/technology/2019/08/28/doorbell-camera-firm-ring-has-partnered-with-police-forces-extending-surveillance-reach.

Jain, P., M. Gyanchandani, and N. Khare. 2016. "Big data privacy: a technological perspective and review." *Journal of Big Data* 3 (1): 25. ISSN: 2196-1115. doi:10.1186/s40537-016-0059-y. https://doi.org/10.1186/s40537-016-0059-y.

Kari, P. 2019. "California's groundbreaking privacy law takes effect in January. What does it do?" Accessed: 2020-02-02. https:%5C/%5C/www.theguardian.com%5C/us-news%5C/2019%5C/dec%5C/30%5C/california%5C-consumer%5C-privacy%5C-act%5C-what%5C-does%5C-it%5C-do.

Körner, D. B., Kevin. 2019. "Digital politics: AI, big data and the future of democracy." *Deutsche Bank Research. EU Monitor.*

Machlup, F. 2014. *Knowledge: Its Creation, Distribution and Economic Significance, Volume I: Knowledge and Knowledge Production.* Princeton Legacy Library. Princeton University Press. ISBN: 9781400856008. https://books.google.de/books?id=DtT%5C_AwAAQBAJ.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. 2011. "Big data: The next frontier for innovation, competition, and productivity." *Technical report, McKinsey Global Institute.*

Mehmood, A., I. Natgunanathan, Y. Xiang, G. Hua, and S. Guo. 2016. "Protection of Big Data Privacy." *IEEE Access* 4:1821–1834. ISSN: 2169-3536. doi:10.1109/ACCESS.2016.2558446.

Pressestelle, LfDI Baden-Württemberg. 2019. "LfDI Baden-Württemberg verhängt erstes Bußgeld gegen Polizeibeamten." Accessed: 2019-12-02. https://www.baden-wuerttemberg.datenschutz.de/lfdi-baden-wuerttemberg-verhaengt-erstes-bussgeld-gegen-polizeibeamten/.

Smith, M., C. Szongott, B. Henne, and G. von Voigt. 2012. "Big data privacy issues in public social media." In *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*. doi:10.1109/DEST.2012.6227909.

Solove, D. J. 2008. *Understanding Privacy.* Understanding privacy, Bd. 10. Harvard University Press. ISBN: 9780674027725. https://books.google.de/books?id=XU5-AAAAMAAJ.

———. 2015. "The meaning and value of privacy." In *Social Dimensions of Privacy: Interdisciplinary Perspectives,* edited by B. Roessler and D. Mokrosinska, 71–82. Cambridge University Press. doi:10.1017/CBO9781107280557.005.

State of California, Department of Justice. 2019. "California Consumer Privacy Act (CCPA)." Accessed: 2020-02-02. https:%5C/%5C/oag.ca.gov%5C/privacy%5C/ccpa.

Tikkinen-Piri, C., A. Rohunen, and J. Markkula. 2018. "EU General Data Protection Regulation: Changes and implications for personal data collecting companies." *Computer Law & Security Review* 34 (1): 134–153. ISSN: 0267-3649. doi:https://doi.org/10.1016/j.clsr.2017.05.015. http://www.sciencedirect.com/science/article/pii/S0267364917301966.

Zins, C. 2007. "Conceptual approaches for defining data, information, and knowledge." *Journal of the American Society for Information Science and Technology* 58 (4): 479–493. doi:10.1002/asi.20508. https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20508.

Declaration of Authorship

I hereby declare that I am the author of the project report that I am submitting. The report is entirely my own original work except where otherwise indicated. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

_____

Prabesh Dhakal, March 15, 2020