INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING & COMPUTER SYSTEMS (IJSECS)
ISSN: 2289-8522 e-ISSN: 2180-0650
VOL. 6, ISSUE 1, 41 – 51
DOI: https://doi.org/10.15282/ijsecs.6.1.2020.5.0068

ORIGINAL ARTICLE

# Towards Data Privacy and Security Framework in Big Data Governance

Jacentha N.Maniam[1], Dalbir Singh[1*]

[1]Center for Software Technology and Management (SOFTAM), Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

**ABSTRACT** – Data privacy and security are among the most important aspects to be considered in implementing a data-driven system. A well generated big data contains a wealth of information that can reveal personal information that should be kept in private. Although various studies are widely documented, the methods and policies to ensure the privacy and security of data in big data governance remain unclear. Big data governance also seeks to protect the affected person to disclose new and critical personal information that is not intended for disclosure. Thus, this study focuses on data privacy and security issues in big data governance. The objective of this study aims to propose of big data governance framework that complements data privacy and security factors. This study uses a qualitative approach in the development of the research framework based on a systematic literature review and evaluated by experts to validate the framework. This study is expected to benefits the public and private sectors where the proposed framework could be applied as a guide to preventing any data leakage or misuse of big data. However, the study could be expanded in the future to include technological and legal perspective.

ARTICLE HISTORY
Received: 23 Apr 2020
Accepted: 29 May 2020

KEYWORDS
*Big Data Governance*
*Data Privacy*
*Data Security*

## INTRODUCTION

Structured and unstructured data generated by various organizations and social entities are referred to as big data. Human leaves digital footprints in almost every daily activity. Data generation occurs in a big scale every time someone is online, may it be using a smartphone equipped with GPS, or even communicating on social media and performing online shopping (Gregory & Halff, 2020). It shows a continuous advance in the flow of sensors, photography, text, and voice data that strengthened the use of industrial data (Bernard Marr, 2019). The development of big data is thriving in various public and private sectors as these include health, pharmacy, power, telecommunications, and transportation enterprises (Tallon, 2013). Big data opens a wide opportunity in every industry to accurately predict customer behaviour and improve performance efficiency as the industry begins to take steps in the field of big data (Elia, Polimeno, Solazzo, & Passiante, 2020).

Balancing the amount of personal data in the facilities and application of big data should be addressed to avoid the occurrence of data leakage issues (Panda Security MediaCenter, 2018). Moreover, the level of trust in data privacy and security protection so that it does not fall into the wrong hands still haunts the public. It is a responsibility for all individuals that work closely with big data to address any privacy and security breach to avoid any reputational, legal or financial issues to the industry (Alex Bekker, 2018).

It is known that big data and data privacy, as well as data security, are intertwined together. Even though there are various spaces for improvement in data privacy and security, but still a complete big data governance framework that focuses together data privacy and security is nowhere to be found. A framework that emphasizes on data privacy and security in big data governance is a worthy research topic that requires extensive investigation (Al-Badi, Tarhini, & Khan, 2018). A framework of big data that covers the aspects of data privacy and security can solve a variety of issues regarding data privacy and security, especially on data leaks in implementing a big data governance system (Jawwad A. Shamsi ; Muhammad Ali Khojaye, 2018). However, how are the specific factors regarding data privacy and security are related and formulates a big data governance framework that could provide optimum results? (Moghadam & Colomo-Palacios, 2018). Such fundamental understanding forms the foundation that differentiates the proposed study in the article from others.

Therefore, the purpose of this study is to investigate the continuity of data privacy and security in big data governance. A framework will be proposed to incorporate the details of privacy and security into the existing big data governance framework to make it more efficient and usable in existing systems. Finally, the expert opinion could be taken into account for verification and improvement purposes. In this article, there are three (3) main sections that focus on (a) privacy and security issues in big data governance, (b) methodology and (c) contribution of study that outlines relevant privacy and security factors that constructs the proposed big data governance framework.

*CORRESPONDING AUTHOR | Dalbir Singh | ✉ dalbir@ukm.edu.my

## BACKGROUND AND PREVIOUS STUDIES

The background section discusses related previous studies in detail. It comprises the following four sections that focus on big data governance and issues related to data privacy and security. It aims to provide a fundamental understanding of the related area and prepares for a critical review of privacy and security factors in the subsequent section.

### Elements of Big Data

Big data is a new way of processing information to improve the quality of decision making, understanding discovery and process escalation. Big data has yet to have a universal mutual definition in literature but it is usually classified as an entity, with its 5Vs - volume, velocity, variety, veracity, and value (Jenn Cano, 2014) (Dabab, Craven, Barham, & Gibson, 2018). Big data is also considered as a holistic information management approach, for acquiring, cleaning, integrating, storing and analyzing data from a variety of sources either internal or external, in which data can be structured or non-structured. It is to generate insights and analysis to support a decision (Randy Bean, 2016). Data depends on four data processing stages such as collecting, storing, analyzing and using the respective data. According to Gartner, the technology used in those stages are not suitable and does not cope up with current technology development (Gartner, 2019). Furthermore, since 2015, big data was not considered in the yearly hype cycle report as big data has entered a higher phase from its previous (Nancy Davis Kho, 2018). Big data is also associated with the need for coordination in data handling especially in data-intensive applications. Figure 1 shows the elements that are often associated with big data.

Usually, volume pointed to the data size. Volume relates to the enormous growth of data from various sources. A large amount of data are generated by organizations, individuals and sensors from various field. Data distributors are almost unable to monitor and supervise data that they actively or passively distribute. The data can likely predict a person's identity and behavior which ultimately leads to an individual's privacy leakage. Velocity refers to the high frequency of data. This element triggers issues related to information security and privacy that are more severe. Rapid and repetitive data require a non-relational database. Therefore, the design of the framework relating to the distribution of data should focus on privacy and security factors (NIST 2018) (NBDPWG - Security and Privacy Subgroup, 2018). Variety refers to the various types of data formats and data sources available (Ye et al., 2018). Data formats are grouped as unstructured data, semi-structured data, and structured data. There is a variety of data types as text, figures, and video. Veracity refers to the reliability of a data source, defect, noise and quality of the data type (Whitson, P., 2013). Data storage and retrieval focus more on data authorization. Moreover, this aspect also shows the level of effectiveness of the data to be analyzed. Value refers to the results obtained from a large data set. The bigger the value of the data, the higher the capability to attract the attention of hackers (Whitson, P., 2013). When a hacker successfully hacks a database, the hacker will able to obtain a huge amount of sensitive information which eventually reduces the data attacking cost. Thus, it will increase the chances for cyber-attacks to occur.

The size of data is referred to as volume. Volume usually compared with the expansive growth of data from various sources. A large amount of data are generated by organizations, individuals and sensors from various field. Data distributors are almost unable to monitor and supervise data that they actively or passively distribute. The data can likely predict a person's identity and behavior which ultimately leads to an individual's privacy leakage. Velocity refers to the high frequency of data. This element triggers issues related to information security and privacy more severely. Rapid and repetitive data require a non-relational database. Therefore, the design of the framework relating to the distribution of data should focus on privacy and security factors (NIST 2018)(NBDPWG - Security and Privacy Subgroup, 2018). Variety refers to the different range of formats and data sources (Ye et al., 2018). Data formats can be classified as unstructured data, semi-structured data, and structured data. There are a variety of data such as text, figures, and video. Veracity refers to the trustworthiness of a data source, noise, defect and quality of the data type (Whitson, P., 2013). Data storage and retrieval focus more on data authorization. Moreover, this aspect also shows the level of effectiveness of the data to be analyzed. Value refers to the results obtained from a large data set. The bigger the value of the data, the higher the capability to attract the attention of hackers (Whitson, P., 2013). When a hacker successfully hacks a database, the hacker will able to obtain a huge amount of sensitive information which eventually reduces the data attacking cost. Thus, it will increase the chances for cyber-attacks to occur.

There are many causes of data privacy and security issues to occur. One of many reasons, data leakages, and illegal data exposure occurs because of these elements of big data. When the volume of the data increases, it increases the risk of the data being misused and exposed as well. This has resulted in efforts to detect and mitigate data loss by conducting a detailed study of factors that should be emphasized in the field of big data. It can conclude that a high flow of big data could harm the database system if the database does not have sufficient features to store the data that has been transmitted. These data are easy to leak and personal information can be distributed illegally. Therefore, a framework that can serve as a guide in the big data governance which could solve these potential issues is vitally needed.

Big data is generated from many variations, data storage and maintenance play an important role in big data governance. However, issues related to leaks of data privacy and unauthorized transmission of data from unreliable sources are still prevalent.

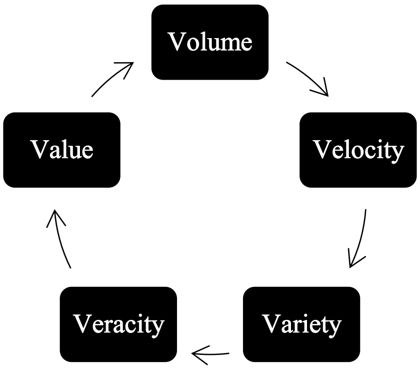

**Figure 1.** Elements of Big Data

## Big Data Governance

Advances in information and communication technologies are rapidly connecting the surrounding environment with enormous data quantity. This phenomenon has led to the emergence of advanced data technology among scientists and technologists around the world, particularly in the exploration of big data. According to Datameer (2015), the more advances the data technology is, the more important it is to adhere to corporate standards and older industry regulations. A robust governance system must have the capability to audit additional changes of data, detect data offspring and permit role-based data access to perform impact analysis. Datameer (2015) also suggested five pillars of strong big data governance such as data policies and standards, privacy and security, quality and consistency, compliance and regulatory, and lastly archiving and retention. These five factors play an important role in developing a complete data governance system (Datameer, 2015). Big data governance technology is evolving because it has many possibilities and unprecedented insights can be offered to an individual, organization, and society. However, some challenges need further exploration regarding the big data governance system.

## Privacy and Security in Big Data Governance

This section discusses the relationship between data privacy and security in big data governance. Issues and risks related to data privacy and security in big data governance are further discussed in this section. Data privacy and security has become one of the important factors for the current improvements in big data technology, particularly for data storage and data preparations. The significance of data privacy and security is growing in line with the development, accessibility, and utilization of big data.

According to the special edition of the National Institute of Standard and Technology (NIST) 2018, big data usage is looking forward to doubling up every two years by 40,000 exabytes by 2020. This number also estimates that more than one-third of data by 2020 can add value if analyzed in a certain way. A detailed study also mentions that less than one-third of data needs data privacy and security in 2010 but by 2020 more than 40 percent of data will require privacy and security (NBDPWG - Security and Privacy Subgroup, 2018). Nowadays, big data is gathered in the public cloud infrastructure built using a variety of operating systems and hardware built-in with computer analysis software. Big data privacy and security require a different perspective compared to the traditional system approaches.

According to Javier Salido (2010), four principles support organizations in choosing strategies and activities that safeguard the privacy and security of data assets (Javier Salido, 2010). The first principle is the adherence to policies and standards throughout the validity of private data. It also means that the data processing must obligate with the current rules and regulations. Moreover, it is important to shield the confidentiality of users concerning their preferences that permits individuals to re-establish the information as needed. The second principle lowers the likelihood of unapproved entry or illegal usage of classified data.

A well-managed information system should provide appropriate practical administrative protection, to make sure the confidentiality, integrity, and availability of data.  The third principle is to decrease the effect of important data loss. The system that protects information must allow appropriate protection such as anonymization that make sure the data confidentiality is protected from being lost or theft. Appropriate plans and actions of a data breach must be provided to all users who are related to the infringement response. The last principle is to state the control sets that should be documented to demonstrate its effectiveness. This helps to ensure the accountability and the compliance of the organization towards privacy and confidentiality principle. Moreover, data has to be demonstrated via splendid monitoring, controls, and auditing. A corporation must have processes for over-looking of non-compliance and clear alternatives, in case if such incidents occurs (Javier Salido, 2010).

The privacy and security of individuals applies to the initial generation of data who also include themselves in the second-generation of data in conjunction with existing data sets. Data usage is related to the accumulation of data from the recipient and using it for another purpose. Thus, higher data generation causes higher risk of data misuse or data leakage. Someone’s privacy is being infringed upon and disseminated on or off ground as a result of a lack of exposure on the significance of data privacy and security problems. This can be overcome by having a framework as guidance in managing big data which specifically focusses on factors that are an important part of preserving data privacy and security.

## Data Privacy and Security Issues in Big Data Governance

Multiple risks and issues have to be overcome in big data governance, especially in the context of data privacy and security. Big data aggregation process shows that data is often aggregated from many different sources and being used or shared by many actors for various purposes (Gloria González Fuster, 1989). The purpose is to perform illegal actions like infringing on an individual's privacy by leaking personal information. Attention should be given to avoid such issues like data redundancy, data leakages or illegal data sharing before its too late.

According to Matturdi, Zhou, Li, & Lin (2014), data privacy and security are among the top challenges that need to be focused on big data governance. Government agencies, the healthcare industry, researches in biomedical and large scale businesses invest a high amount of resources in aggregating and sharing personal data to leverage the big data (Matturdi, Zhou, Li, & Lin, 2014). "National Security Administrationregularly collects and analyses a large amount of personal data obtained from heterogeneous data sources such as telecommunications, Internet, and end-user of widespread business database including Microsoft, Yahoo, Google, Facebook, PalTalk, Youtube, Skype, AOL, and Apache"(James BamFord, 2012). Such facts show that big data is capable of threatening the privacy of users if it is not properly addressed. Thus, it shows that factors such as privacy, availability, and confidentiality play an important part to lower the big data risk from threatening the user's privacy.

Most of the information obtained from big data can provide insight for the individual. Personally, identifiable information is everywhere and sometimes it can be found in the most unexpected places. Usually, an organization underestimates the rules and legalization of data generation and processing even after knowing they have to deal with such an enormous amount of data. Users must have the right to know what data is being collected from them. Thus, the data privacy and security of user can be safeguarded in the big data governance (Felix Rosbach, 2019). Data privacy can be improved through encrypted data and avoiding third party from discovering the content of confidential information.

Additionally, when an individual's personal information is combined with a large set of external data, it enables the information to expand and resulted in obtaining new conclusions about the related individual. Usually, a conclusion or an inference upon an individual should not be leaked. It is kept secret by a data analyst from the data providers. Unfortunately, such confidential data and information always tend to leak and the source of the leakage remains unidentifiable. This happens due to weaker infrastructure security in the big data governance system.

Big data ownership is addressed as one of the factors of data privacy and security. Ownership is an attribute (which may or may not be visible to the user) that associates data with one or more entities that own or influence what can be done with the data. Database ownership allows create, read, update and deletion towards the data. Transparency in ownership enables trust and control for data owners as well as availability and utilities for enterprises and communities. Unfortunately, the transparency of data ownership is often being questionable among data users. An issue such as data leakage can be avoided from the beginning of the data privacy and security systems if it is properly managed although retention of data provenance enables tracking through the data life cycle and tracking data ownership changes (Gruschka, Mavroeidis, Vishi, & Jensen, 2019). Figure 2 shows the relationship between the elements of big data and the current challenges associated with privacy in big data governance. These issues have been repeated over and over again to make a realization that big data governance system still has room for improvement whereby privacy and security aspects can still be improved.

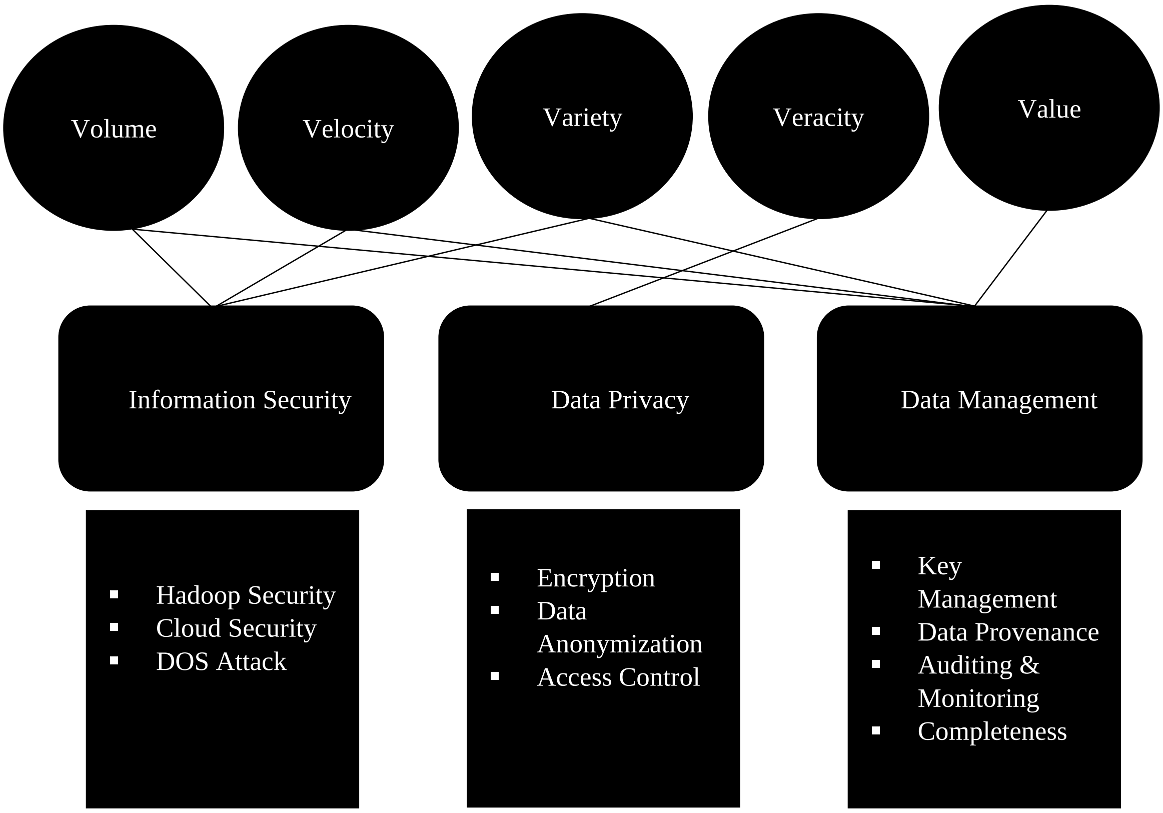

**Figure 2.** Data Privacy and Security Challenges in Big Data Governance

## METHODOLOGY

This section discusses methods and approaches regarding the literature review that has been used in this study. Systematic literature reviews (SLR) can accomplish the objective of the study. SLR is chosen as it is the most widely used approach to review and discover crucial findings from relevant past studies (Mendes, Wohlin, Felizardo, & Kalinowski, 2020). It also has been applied to a past study related to the area of big data with similar attributes (Cui, Kara, & Chan, 2020). This study captures all the information gained from the systematic literature review that focused on past studies conducted from the year of 2016 to 2020 on big data privacy and security issues in big data governance. All insights are written in English for academic databases such as Scopus, Science Direct, IEEE Xplore, ACM, Web of Science, Springer Link, JSTOR, and Emerald Insight. All published results are from academic studies such as journals, conferences, and academic reports. This literature review was conducted as mentioned by Kitchenham and Charter (Kitchenham & Charters, 2007).

### Article Selection Method

Overall fifteen (15) study materials have been chosen that consists of documents and journals that would analyze thirteen (13) factors including accuracy, privacy, integrity, communication, access control, infrastructure security, completeness, confidentiality, availability, storage, encryption, anonymization, and data provenance which are identified within their respective frameworks.

i. Specific documents and journals from the year of 2016 to 2020 were selected to identify emerging studies.
ii. Keyword sampling techniques have been used in related journals and documents searching. Keywords such as big data, big data governance, factors influencing big data, privacy and security in big data, the framework of data privacy and security, big data governance impacts and the importance of data privacy and security in big data governance have been chosen.
iii. Content analysis is expanded by identifying the importance of data privacy and security factors in big data governance. Comparisons are made between documents and journals that are found to address issues that must be directly linked with maintaining the privacy and security of data in big data governance.

The analysis was carried out with selecting five (5) relatable frameworks which would be analyzed to reveal the similarity and differences that could be adapted within the framework of the preliminary study. Data privacy and security factors combination technique has been used to identify the frameworks from the previous study. This analysis aims to propose of big data governance framework that complements data privacy and security factors. In summary, the flow chart of the SLR process and inclusion/exclusion criteria is illustrated in Figure 3.

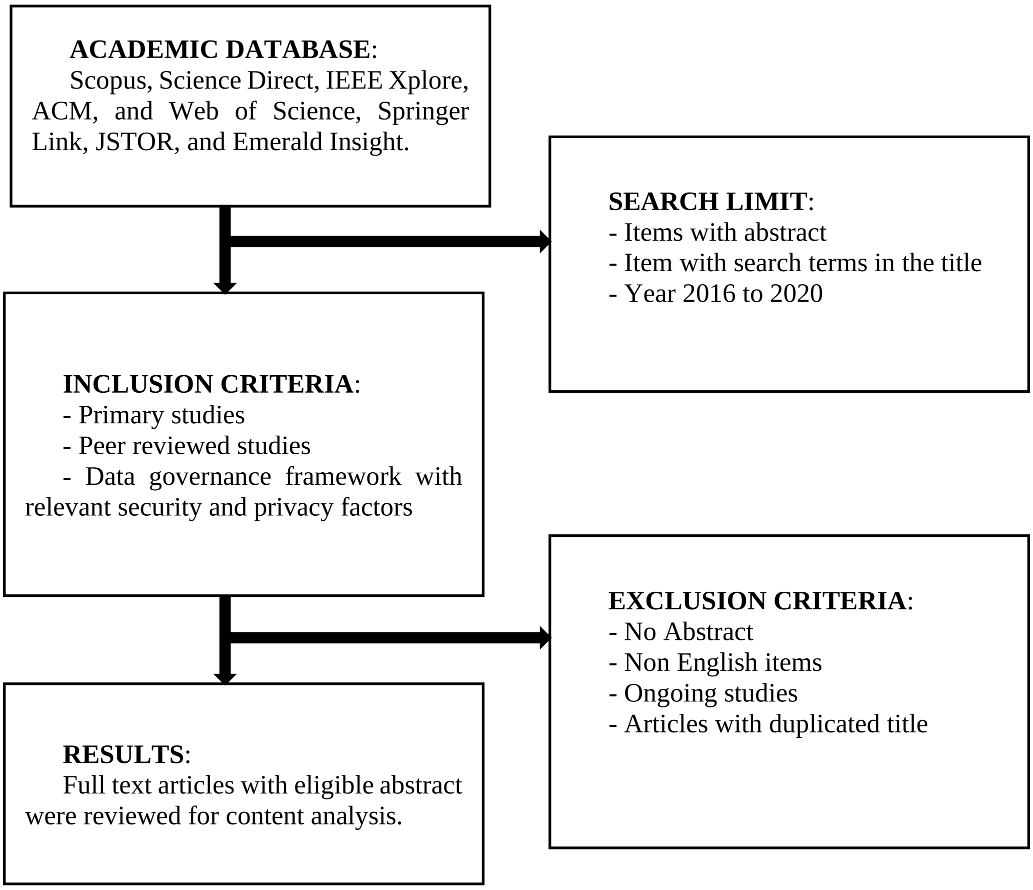
**ACADEMIC DATABASE**:
Scopus, Science Direct, IEEE Xplore, ACM, and Web of Science, Springer Link, JSTOR, and Emerald Insight.
**SEARCH LIMIT**:
- Items with abstract
- Item with search terms in the title
- Year 2016 to 2020
**INCLUSION CRITERIA**:
- Primary studies
- Peer reviewed studies
- Data governance framework with relevant security and privacy factors
**EXCLUSION CRITERIA**:
- No Abstract
- Non English items
- Ongoing studies
- Articles with duplicated title
**RESULTS**:
Full text articles with eligible abstract were reviewed for content analysis.

**Figure 3.** Flow chart of the SLR process and inclusion/exclusion criteria

## RESULT AND DISCUSSION

This section elaborates on the factors that are engaged in data privacy and security especially in big data governance. The proposed framework is discussed in this section based on the findings of the SLR for this study. Fifteen (15) articles on big data governance that focused on data privacy and security is chosen and reviewed to discover the factors related to the suggested framework. Table 1 shows the studies conducted by past researchers based on data privacy and security for big data.

**Table 1.** Model and Framework of Data Privacy and Security in Big Data Governance

| Title | Year | Outcome |
|---|---|---|
| Exploring Big Data Governance Frameworks | 2018 | Eight components of Big Data management are organizational structure, stakeholders, the scope of data, policies, and standards, optimizations, quality, data storage, communication and data management. |
| Exploratory Strategic Road-mapping Framework for Big Data Privacy Issues | 2018 | An exploratory framework using QFD techniques and technological analysis plan to address social, technology, resources and industry issues that are relevant to data privacy and security. |
| Towards a Data Governance Framework for Third Generation Platforms | 2019 | Proposed an initial schema for building third-party big data governance programs which was a concept tool that guides an organization to identify and design their services in line with the vision and mission of the 4.0 Industrial Revolution. |
| Data Governance Framework for Big Data Implementation with a Case of Korea | 2017 | The framework is divided into four sections including the objectives, strategies, components and technological infrastructure. |
| Factors influencing effective use of Big Data: A research framework | 2019 | The framework is divided into three categories. These include motivations, operations, and support mechanisms. Each of these categories is linked to big data so that the value can be measured. |

Table 2 shows data privacy and security factors in big data governance. According to Table 2, the majority of studies choose privacy as an important and must-have factor in big data governance. The availability factor is chosen in 8 studies followed by accuracy factors in 7 studies comparatively. It is found that infrastructure security also plays a vital role in protecting data privacy and security in big data governance. Completeness was also taken into consideration as only complete data able to have a higher value. Whereas factors such as control access, data storage, anonymization, encryption, and data provenance are chosen in 4 studies respectively. Data confidentiality and data integrity also have been discussed in 3 studies.

**Table 2.** Data Privacy and Security Factors in Big Data Governance

| | | Data Privacy and Security factors | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Title | Year | A | P | I | C | C.A | I.S | CO | A.V | S | E | AN | D.P | C.M |
| Exploring Big Data Governance Frameworks | 2018 | x | | | x | x | | | | x | | | | x |
| Information security governance in Big Data environments: A systematic mapping | 2018 | | x | | | | x | | x | | | | | |
| Understanding Privacy Violations in Big Data Systems | 2018 | | x | | | | | x | x | | | x | | x |
| Obstacles to Implementation of Information Security Governance | 2018 | x | x | | | | x | | | | | | x | |
| Exploratory Strategic Road-mapping Framework for Big Data Privacy Issues | 2018 | x | x | x | | | | | x | | x | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Privacy-Preserving Big Data Publishing | 2018 | | x | | | | x | | | | | x | | |
| NIST Big Data Interoperability Framework: Volume 4, Security and Privacy | 2018 | | x | x | | x | x | x | | | | | x | x |
| Towards a Data Governance Framework for Third Generation Platforms | 2019 | x | | | | | x | | x | | x | | | |
| Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR | 2018 | | x | | | | | | x | x | | x | | |
| Big Data and Government Governance | 2018 | | x | x | | x | | | | x | | | x | x |
| Security, privacy & efficiency of sustainable Cloud Computing for Big Data & IoT | 2018 | x | x | | x | | | | | x | | | | |
| Data governance framework for Big Data implementation with NPS Case Analysis in Korea | 2018 | x | | | | | | x | x | | x | | | |
| Efficient privacy preservation of Big Data for accurate data mining | 2019 | x | x | | | | x | | x | | | | | x |
| Achieving correlated differential privacy of Big Data publication | 2019 | | x | | | | | | x | | x | | | |
| Privacy-Preserving, Protection of Personal Data, and Big Data: A Review of the Colombia Case | 2019 | | | | | x | | | | | | x | x | |
| | Total | 7 | 11 | 3 | 2 | 4 | 6 | 3 | 8 | 4 | 4 | 4 | 4 | 5 |

A=Accuracy, P=Privacy, I= Intergrity, C=Communication, C.A=Control Access, I.S=Infrastructure Security, CO=Confendtiality, A.V=Availability, S=Data Storage, E=Encryption, A.N=Anonymization, D.P=Data Provenance, C.M=Completeness

Finally, the communication factor is described in only 2 studies which shows that it has less prominence towards protecting data privacy and security in big data governance. Figure 4 shows the average usage of factors chosen in the protection of big data privacy and security by comparing the literature review in a pie chart.

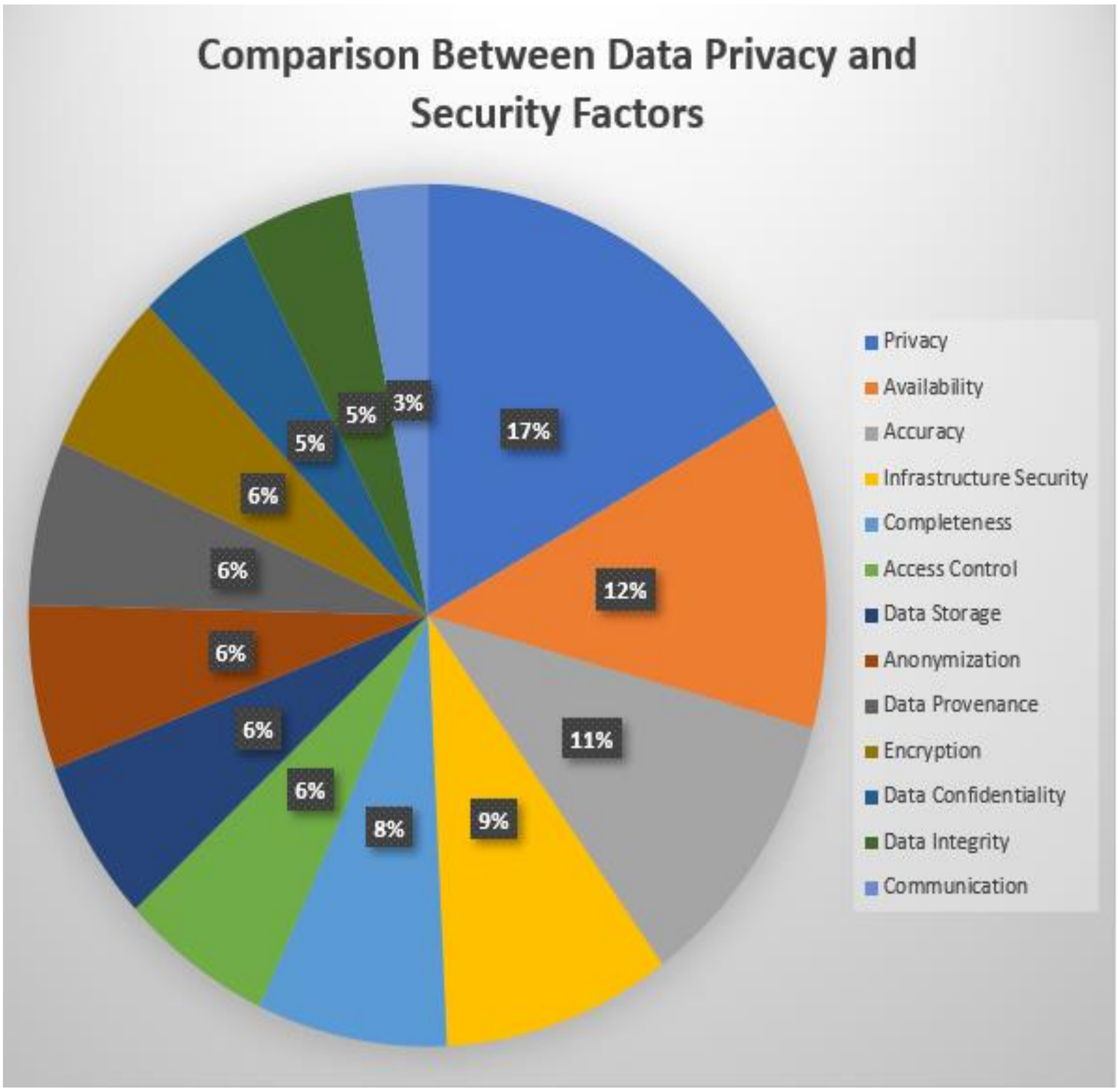

**Figure 4.** Comparison between Average Factor Usages

Figure 5 shows the proposed framework of data privacy and security in the big data governance system. The framework is initiated with the five (5) pillars of big data governance whereby privacy and security are one of the pillars in it. Privacy and security divided into three (3) parts such as confidentiality, integrity, and availability (CIA). This was adapted from the 'CIA' triad which is also known as the information security model. 'CIA' refers to confidentiality, integrity, and availability and it is very crucial in traditional data security or even cybersecurity. Researchers use the CIA security model in studies involving data privacy and security. For example, the same approach has been used by (Al-Far, Qusef, & Almajali, 2019) & (Prasad et al., 2011). Finally, the privacy and security of data affecting factors were identified through a literature review and have been classified according to their characteristics and criteria.

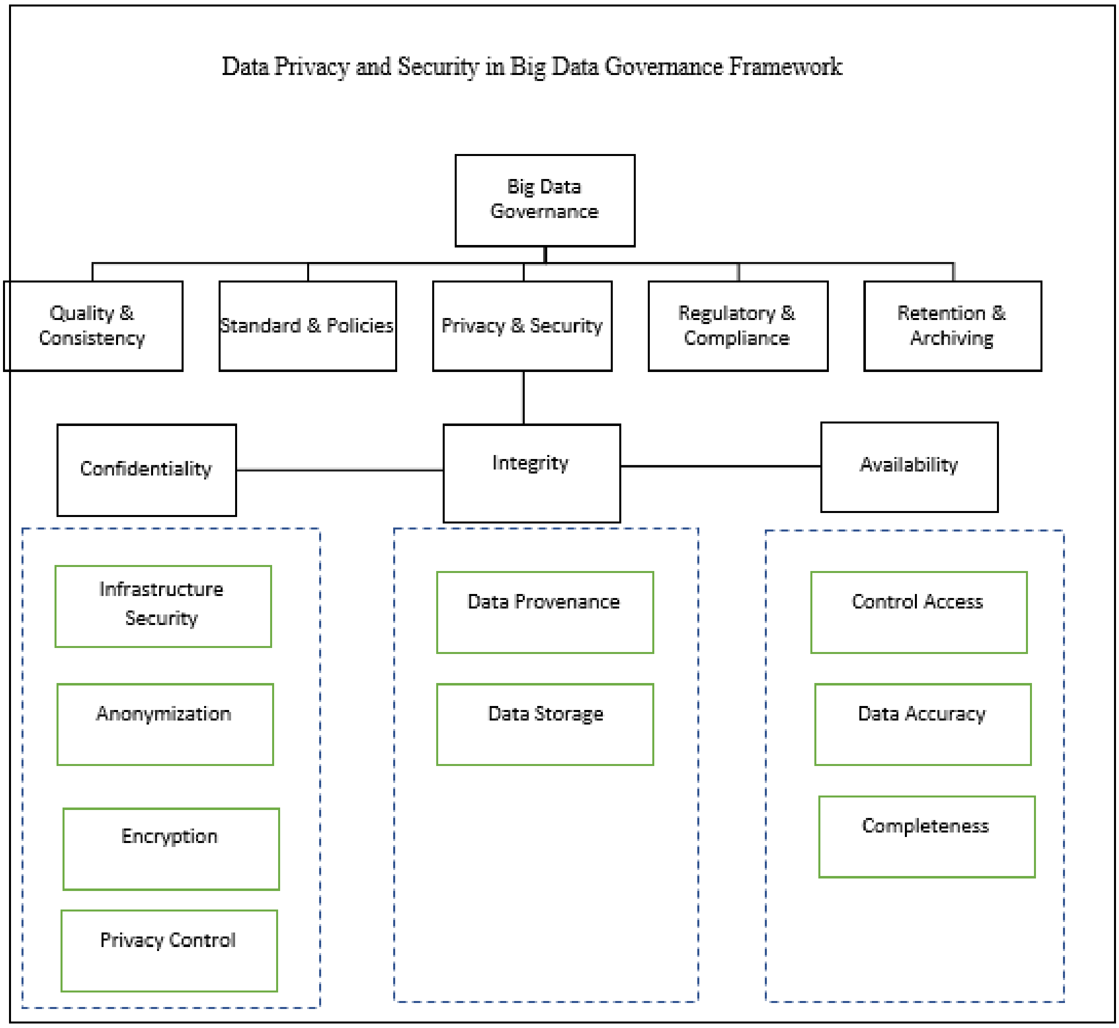

**Figure 5.** Proposed Framework for Data Privacy and Security in Big Data Governance

## CONCLUSION

While advanced data technology continuously has room for improvement in terms of data privacy and security, a framework that encompasses all aspects of detailed data privacy and security is yet to be discovered. A big data framework that covers features of data privacy and security can solve a variety of problems regarding privacy and data leakage in the implementation of big data governance systems. This study implies that it can help enterprise sectors in managing big data governance that prioritizes data privacy and security. The complexity of policies around big data governance is one of the reasons that makes a single analysis method not effective in conducting review of the overall scope of opportunities and risks in big data governance. Thus, more suggestions for additional studies should be continued to incorporate various analytical methods as an effort to understand and deepen the relationship and interaction aspects of data privacy and security in growing big data governance technology. Besides that, the study is not bounded by any legal procedures or technological platform used to govern big data resources in any organization. Such aspects should be considered in the adoption of the proposed framework. It may also open up new frontiers for future investigation.

## REFERENCES

Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring Big Data Governance Frameworks. *Procedia Computer Science*, *141*, 271–277. https://doi.org/https://doi.org/10.1016/j.procs.2018.10.181

Alex Bekker. (2018). Big data security: issues, challenges, concerns. Retrieved from https://www.scnsoft.com/blog/big-data-security-challenges

Al-Far, A., Qusef, A., & Almajali, S. (2019). Measuring Impact Score on Confidentiality, Integrity, and Availability Using Code Metrics. *ACIT 2018 - 19th International Arab Conference on Information Technology*, 1–9. https://doi.org/10.1109/ACIT.2018.8672678

Bernard Marr. (2019). What is Big Data? Retrieved from https://www.bernardmarr.com/default.asp?contentID=766

Cui, Y., Kara, S., & Chan, K. C. (2020). Manufacturing big data ecosystem: A systematic literature review. *Robotics and Computer-Integrated Manufacturing*, *62*, 101861. https://doi.org/https://doi.org/10.1016/j.rcim.2019.101861

Dabab, M., Craven, R., Barham, H., & Gibson, E. (2018). Exploratory strategic roadmapping framework for big data privacy issues. *PICMET 2018 - Portland International Conference on Management of Engineering and Technology: Managing Technological Entrepreneurship: The Engine for Economic Growth, Proceedings*, 1–9. https://doi.org/10.23919/PICMET.2018.8481834

Datameer. (2015). Datameer Big Data Governance, 1–8.

Elia, G., Polimeno, G., Solazzo, G., & Passiante, G. (2020). A multi-dimension framework for value creation through Big Data. *Industrial Marketing Management*. https://doi.org/https://doi.org/10.1016/j.indmarman.2020.03.015

Felix Rosbach. (2019). The Top 3 Big Data Security and Compliance Challenges of 2019.

Gartner. (2019). Gartner Glossary. Retrieved from https://www.gartner.com/it-glossary/big-data

Gloria González Fuster, A. S. (1989). Big Data and Smart Devices and Their Impact on Privacy. *Journal of Chemical Information and Modeling*, *53*, 160. https://doi.org/10.1017/CBO9781107415324.004

Gregory, A., & Halff, G. (2020). The damage done by big data-driven public relations. *Public Relations Review*, *46*(2), 101902. https://doi.org/https://doi.org/10.1016/j.pubrev.2020.101902

Gruschka, N., Mavroeidis, V., Vishi, K., & Jensen, M. (2019). Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 5027–5033. https://doi.org/10.1109/BigData.2018.8622621

James BamFord. (2012). The NSA Is Building the Country’s Biggest Spy Center (Watch What You Say) | WIRED.

Javier Salido. (2010). Data Governance for Privacy, Confidentiality and Compliance: A Holistic Approach. *Isaca Journal*, *6*, 1–7.

Jawwad A. Shamsi ; Muhammad Ali Khojaye. (2018). Understanding Privacy Violations in Big Data Systems. *IT Professional*, *20*(3), 73–81. https://doi.org/https://doi.org/10.1109/MITP.2018.032501750

Jenn Cano. (2014). The V’s of Big Data: Velocity, Volume, Value, Variety, and Veracity.

Kitchenham, B., & Charters, S. (2007). Source: " Guidelines for performing Systematic Literature Reviews in SE " , Kitchenham et al Guidelines for performing Systematic Literature Reviews in Software Engineering Source: " Guidelines for performing Systematic Literature Reviews i, 1–44. https://doi.org/10.1145/1134285.1134500

Matturdi, B., Zhou, X., Li, S., & Lin, F. (2014). Big Data security and privacy: A review. *China Communications*, *11*(14), 135–145. https://doi.org/10.1109/CC.2014.7085614

Mendes, E., Wohlin, C., Felizardo, K., & Kalinowski, M. (2020). When to update systematic literature reviews in software engineering. *Journal of Systems and Software*, *167*, 110607. https://doi.org/https://doi.org/10.1016/j.jss.2020.110607

Moghadam, R. S., & Colomo-Palacios, R. (2018). Information security governance in big data environments: A systematic mapping. *Procedia Computer Science*, *138*, 401–408. https://doi.org/https://doi.org/10.1016/j.procs.2018.10.057

Nancy Davis Kho. (2018). The State of Big Data 2018 - EContent Magazine. *Feb 19*.

NBDPWG - Security and Privacy Subgroup. (2018). NIST Special Publication 1500-4 Security and Privacy. *NIST Big Data Interoperability Framework*, *4*(June). https://doi.org/https://doi.org/10.6028/NIST.SP.1500-4r1

Panda Security MediaCenter. (2018). Data security in the age of Big Data. Retrieved from https://www.pandasecurity.com/mediacenter/security/big-data-implications/

Prasad, P., Ojha, B., Shahi, R. R., Lal, R., Vaish, A., & Goel, U. (2011). 3 Dimensional security in cloud computing. *ICCRD2011 - 2011 3rd International Conference on Computer Research and Development*, *3*, 198–201. https://doi.org/10.1109/ICCRD.2011.5764279

Randy Bean. (2016). Just Using Big Data Isn’t Enough Anymore.

Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, *46*(6), 32–38. https://doi.org/10.1109/MC.2013.155

Whitson, P., G. (2013). ISO/TEC JTC 1 Information Technology. *Salem Press Encyclopedia Of Science*.

Ye, H., Cheng, X., Yuan, M., Xu, L., Gao, J., & Cheng, C. (2018). A survey of security and privacy in big data, (December). https://doi.org/10.1109/ISCIT.2016.7751634

## ACKNOWLEDGEMENT

This work was supported by (a) Fundamental Research Grant Scheme (FRGS). Ministry of Education (MOE), Malaysia: FRGS/1/2016/ICT01/UKM/02/2, (b) Universiti Kebangsaan Malaysia Research Grant: GGP-2019-026.