Minnesota State University, Mankato

Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato

All Graduate Theses, Dissertations, and Other Capstone Projects

Graduate Theses, Dissertations, and Other Capstone Projects

2021

Proposed Data Governance Framework for Small and Medium Scale Enterprises (SMEs)

Rejoice Okoro
Minnesota State University, Mankato

Follow this and additional works at: https://cornerstone.lib.mnsu.edu/etds

 Part of the Databases and Information Systems Commons, Data Science Commons, and the Information Security Commons

Recommended Citation

Okoro, R. (2021). Proposed data governance framework for small and medium scale enterprises (SMEs) [Master's thesis, Minnesota State University, Mankato]. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. https://cornerstone.lib.mnsu.edu/etds/1126/

This Thesis is brought to you for free and open access by the Graduate Theses, Dissertations, and Other Capstone Projects at Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. It has been accepted for inclusion in All Graduate Theses, Dissertations, and Other Capstone Projects by an authorized administrator of Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato.

PROPOSED DATA GOVERNANCE FRAMEWORK FOR SMALL AND MEDIUM SCALE ENTERPRISES(SMES)

By

Rejoice Okoro

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

In

Data Science

Minnesota State University

Mankato, Minnesota

April 2021

4/5/2021

Proposed Data Governance Framework for Small and Medium Scale Enterprises (SMEs)

Rejoice Okoro

This thesis has been examined and approved by the following members of the student's committee.

_____
Advisor

_____
Committee Member

_____
Committee Member

Acknowledgment

I also want to deeply appreciate Dr. Christophe Veltsos, my thesis supervisor, without him this project would have not been such a success. Also, I want to specially acknowledge my family and friends for their constant support.

Contents

List of Tables iv

List of Figures iv

1.INTRODUCTION 1

2. REVIEW OF LITERATURE 3

2.1 Data governance 3

2.2 Key Concepts and Principles of International Data Governance 6

2.3 The Data Management Body of Knowledge 9

3. SMALL AND MEDIUM SCALE ENTERPRISES (SMEs) 12

3.1 General challenges and opportunities 12

3.2 SMEs responsibilities to data governance and privacy laws: 14

3.3 Data security and protection 15

4. DATA SCIENCE AND DATA GOVERNANCE 16

4.1 The intersection of data science and data governance 18

4.2 Master Data Management 20

4.3 Data decision domain 21

5. COMPARATIVE ANALYSIS BETWEEN DATA PROTECTION LAWS 23

5.1 General Data Protection Regulation (GDPR) 23

5.2 Personal Data Protection Act 2012 (PDPA) 24

5.3 California Consumer Privacy Acts of 2018 (CCPA) 25

6. DATA GOVERNANCE FRAMEWORK FOR SMEs 33

7. SUMMARY AND CONCLUSION 36

8. REFERENCES 36

List of Tables

Table 1. Key Requirements of EU GDPR 26
Table 2. Detailed Similarities between GDPR and CCPA 31
Table 3. Detailed Similarities between GDPR and PDPA 32

List of Figures

Figure 1 The core concept of data governance, data management, and data quality management based on Otto [14].....4
Figure 2 Relationship between data governance and data management [37].....8
Figure 3 The long list of key concepts and principles of data governance [32].....9
Figure 4 The DMBOK Guide Knowledge Area Wheel [29].....11
Figure 5 Adapted DMBOK Guide Knowledge Area Wheel [29].....11
Figure 6 Data governance and data science intersection [57].....19
Figure 7 Data governance domain [12] 22

Abstract

Data governance is not a one size fits all, instead, it should be an evolutionary process that can be started small and measurable along the way. This research aims at proposing a data governance framework by ensuring data management processes, data security and control are compliant with laws and policies. This article also presents the first results of a comparative analysis between three data privacy laws and outlines five components which together form a data governance framework for SMEs. The data governance model documents data quality roles and their type of interaction with data quality management activities exploring how data is perceived and applicable to SMEs providing best practices for proper data management which includes roles and responsibilities about the use of data for automated decision making, privacy, compliance to data laws, the intersection of data governance and data science in the digital era.

1. INTRODUCTION

Data governance affects both private and public organization and can also deliver tremendous benefits. Most large organizations are relatively advancing in terms of data governance; however, SMEs also recognize the potential of a data-driven organization and its alignment to data governance. The focus is to combine both the business and IT-related perspectives alongside their processes to reach a broader view, thus enabling them to recognize and achieve their maximum business potential soon. The problem is derived from the increased important for SMEs to focus more on their data assets, SMEs needs a specific data governance framework that fit its context which will aid better decision-making prospects because one size does not fit all. With data governance framework, companies can implement corporate-wide accountabilities for data, compliance with international privacy laws and data transparency in its processes regarding the usage of individual's data. A data governance framework will help companies structure and document their data quality accountabilities.

The purpose of this paper is to explore how SMEs perceive data and design a framework for data governance whilst investigating whether current data frameworks available apply to SMEs. The lack of attention given to SMEs by the data governance community is uneven given the enormous contribution of this sector to the economy at large. Also, due to current technological advancements such as cloud computing, SMEs will likely store and process a greater amount of data in the cloud. Data governance should be a universal approach relating to data accountability, which must fit all data aspects and meet organizational needs [1][2]. Previous literature reveals that over 58% of 200 companies surveyed recognized their data as a strategic asset [3]. Data governance is not a one size fits all, instead, it should be an evolutionary process that can be started small and measurable along the way.

SMEs are facing several data-related issues as the amount of data available increases daily; they are concerned with managing their data assets according to specific requirements. The need for data governance helps ensure they achieve data objectives by the development and implementation of a data strategy on both organizational and technical levels. However, SMEs are still facing challenges while achieving some considerable goals related to data governance. Data governance includes people, process, and technology aligned to ensure maximum protection and the use of data assets.

The various data protection laws ensure the appropriate use on how data is obtained, used, stored, and including the rights of individuals to control their data. These laws provide a legal framework in the use of data. The European Union (EU) has the General Data Protection Law, which was enforced since May 25th, 2018, The United States currently does not have a comprehensive privacy law but adopted a sectoral law which is the California Privacy Act (CCPA). The law governing data protection in Singapore is the Personal Data Protection Act 2012 (PDPA).

Forty-five percent [4] of participants within the global sphere do not have a data governance policy in place. Thus, data governance is far more important and requires more attention from both the public and private sectors [5]. Data governance has since been gaining popularity and it is also an emerging field in the data science field [7][8]. Data governance is also seen as a promising approach to maintain the quality and use of data assets of organizations [6]. Data has a huge influence on both the strategic and operational decisions of SMEs thus should therefore be treated as an asset [9]. Data assets can be managed more effectively by the adoption of an appropriate data governance framework.

With an appropriate framework in place, SMEs will be able to implement appropriate data quality management (DQM) which will bring together both the IT and business departments. Data governance defines roles and responsibilities for decision areas about data by establishing organization-wide standards for efficient data quality management, most importantly, by assuring compliance with laid down laws governing data use.[10] Most sources assume that data governance to be the universal approach, one size fits all.

The distribution of accountability for data management differs between companies. The contingency approach is different for organizations, this shows that each company needs a specific data governance framework that fits into a set of contextual factors. Moreover, several IT governance models exist an example is centralized and decentralized IT governance [11]. SMEs will continue to face significant challenges and pressure as the business environment is constantly changing as well as the technological sphere this will place enormous demands on them to control data assets and manage data effectively.

Therefore, SMEs will need standard framework practices to adopt which will help them to survive and succeed as compliance with laws tightens. However, as a literature review was conducted, initial investigations revealed how SMEs perceive data assets, they do not recognize data asset values and view data as being independent of its systems that support various business processes.

This paper examines, in section one, the introduction, in section two, definitions of data governance, Key Concepts, and principles of data governance, data decision domain, and ethical AI principles. Section three presents general challenges and opportunities faced by SMEs, data security and protection, and roles and responsibilities, section four presents the intersection between data governance and data science, section five discusses the comparative analysis between data protection laws. This paper concludes, in section six and seven by identifying the original contribution of our research to the body of knowledge and presents future avenues for investigation.

2. REVIEW OF LITERATURE

2.1 Data governance

Data can be defined as the set of characters that has no meaning unless it is converted into a specific type of usage which then converts data into information [12]. Data and information are sometimes used interchangeably. Data value is not represented in a company's financial records, but they are a vital part of every activity. Key data attributes include accessibility, availability, quality, consistency, auditability, and security. Misconceptions exist between governance and management. Governance focuses on effective management and the use of IT in alignment with its overall IT strategy whilst management focal point is on implementation of decisions [13]. The relationship between data management and data governance is based on a differential proposed by the International Organization for Standardization (ISO), data governance specifies which decisions are to be made in data management and who makes such decisions.

However, data management ensures these decisions are made and actions are taken place appropriately. The goal of SMEs is to generate the utmost value of their data assets. Organizations are constantly looking for how to generate data value, data governance measures, and monitors aspects in respect to data use, due to different organizational structures, there are numerous ways to implement data governance. Consequently, data governance has several definitions due to organizational complexity. Data governance depicts a structured framework for decision-making, rights, and responsibilities regarding the use of data in an enterprise [15]. Generally, data only generate value if it is being used and analyzed. Data quality refers to data "fitness for use" [16]. Data quality management (DQM) aims is to maximize data quality.

Data Management Body of Knowledge (DAMA) international [17] defines data quality management as a function of the measurement, evaluating, improving, and ensuring data fitness for use. Therefore, data quality management is a sub-level of data management which includes the essential function of planning, controlling, and provisioning of data asset [17]. The following figure shows the Core concept of data governance, data management, and data quality management.

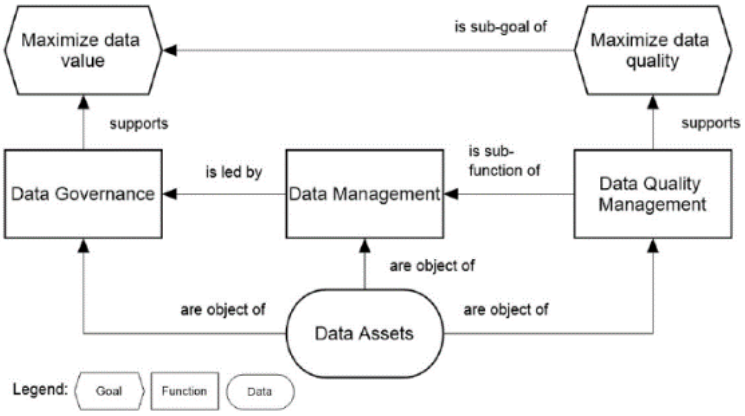

Figure 1 The core concept of data governance, data management, and data quality management based on Otto [14]

Data quality [18] is vital when discussing data governance, data quality refers to the completeness, correctness, uniqueness, referential integrity, and consistency across all data domain, data fitness for use is directly aligned with the six data quality dimensions of accessibility, availability, quality, consistency, auditability, and security.

Bias is inborn in data, and if a machine learns from biased historical data, it reinforces bias more. An immediate observation is that a learning algorithm is designed to pick up statistical patterns in training data. If the training data reflect existing societal biases against a minority, the algorithm is likely to incorporate these biases. This can lead to less favorable decisions for members of these minority groups. While automated decision systems have the potential to bring more efficiency, consistency, and fairness, it also opens the possibility of new forms of discrimination, which may be harder to identify and address. The impervious nature of machine learning algorithms and the many ways human biases can creep in, challenge "our ability to understand how and why a decision has been made" and our capacity of guaranteeing fundamental values of society, such as fairness, justice, and due process insights. Algorithmic discrimination may arise from different sources. First, input data into algorithmic decisions may be poorly weighted, leading to disparate impact. For example, as a form of indirect discrimination, overemphasis on zip code within predictive policing algorithms can lead to the association of low-income African American neighborhoods with areas of crime and as a result, the application of Fair, Transparent, and Accountable Algorithmic Decision-making Processes.

Secondly, discrimination can occur from the decision to use an algorithm itself. Categorization can be considered as a form of direct discrimination, whereby algorithms are used for disparate treatment. Third, algorithms can lead to discrimination because of the misuse of certain models in different contexts. Fourth, in a form of the feedback loop, biased training data can be used both as evidence for the use of algorithms and as proof of their effectiveness. The use of algorithmic data-driven decision processes may also result in individuals being denied opportunities based not on their actions but on the actions of others with whom they share some characteristics. The prediction quality, which is measured by any Lipschitz continuous loss function, whereby each attribute will contain a protected attribute such as race or gender, in respect to which seek to guarantee fairness.

Poor data quality is associated with an increase in complexity of the technological sphere which includes Customer Relationship Management (CRM), Supply Chain Management (SCM), and Enterprise Resource Planning (ERP) systems. The value of data assets depends on data quality of source data increase in data productivity is because of effective business intelligence which further leads to effective decision making [19]. Furthermore, this will also follow regulatory compliance by delivering accurate, complete, and timely data. Data quality effectiveness is harnessed on reports generated and decisions that are to be made based on the quality of data. The main issues regarding data quality are 1) Data is spread across various systems within organizations. 2) Data is collected and used by various levels in the organization 3) System development methodologies do not incorporate data quality assurance.

These issues mentioned can be addressed by having a master data management, this will help ensure data quality using an effective data governance framework. A standard data governance framework will help data managers by giving them the mandate to manage data quality as an enterprise data asset [20]. Data quality issues can be addressed by adopting a holistic approach that focuses on the people, process, and technology [21], data quality should be constantly measured in these organizations. Addressing data quality issues requires data to be governed [22], techniques and tools shape data. Therefore, data governance is simply the governance of people and technology.

There are various definitions of data governance, Newman, and Logan [22]; define data governance as “the collection of decision rights, processes, standards, policies and technologies required to manage, maintain and exploit information as an enterprise resource” [23]. Cohen [24] defines data governance as “the process by which a company manages the quantity, consistency, usability, security, and availability of data” [24]. Considering the above definitions, the need for a data governance framework will thus

enable enterprise-wide collaboration for, different levels in the organization to manage enterprise-wide data and thus aligning data-related assets to overall corporate objectives. Three most important questions, data governance must find answers for SMEs [25][26][27] include:

- What decisions are to be made regarding corporate data on an enterprise-wide level?
- Which roles are accountable in the decision-making process?
- How are the roles involved in the decision-making process?

These questions are all addressed in this paper. With regards to the first question, data governance-related decisions are built upon fundamental principles of data management which include data quality requirements, data quality measurements, metadata, data standards, data access requirements, and data lifecycle management [25]. The second question refers specifically to the roles involved in decision making in the data governance framework, based on previous research roles mentioned frequently includes data stewards, data committees, and data owners. Data stewards evaluate the business requirement and problems with data. They are also responsible for data handling from a business department [28] whereas data owners specify business requirements on data and data quality [9]. This term "data owner" has been criticized by some researchers because it suggests the data is owned by a certain business division or group which contradicts the point that data should be viewed as a company's assets. The central decision-making board is where the data committee becomes vital, specify principles for data usage throughout the organization. The third question refers to the linking of roles and responsibilities to decision areas. In SMEs, for example, the decision regarding data architecture could be assigned to a data committee while the executive power to make decisions can be assigned to a data steward.

2.2 Key Concepts and Principles of International Data Governance

For SMEs, data governance is not all about perceiving data assets as an important asset in the organization but also extending the value of these assets for the whole company by recognizing the cross-functional challenges that these businesses and their leaders are facing. An investment must be made for data governance to be successful and propel long-term support. Numerous concepts have been identified in past kinds of literature, key concepts of data governance are divided into common understanding, compliance, organization, and alignment. Data management needs to comply with the strategic, tactical, and operational policies of the organization.

Data governance is seen as the combination of IT and business aspects, this model excludes technical aspects. DMBOK includes major dimensions such as data strategy, data

integration and interoperability, data architecture and data warehousing, and business intelligence. Unique data governance cannot be used by all organizations. Moreover, these key concepts below are related to data governance related to public organizations.

The following questions should be answered when it comes to data governance for SMEs:

- How do SMEs define their corporate data that needs to be aligned through the organization?
- What are the various roles for the decision-making process in SMEs?
- How effective will data governance be for these SMEs?

Overall, the most important aspect of data governance includes people, process, and technology [34]. Seven goals of data governance that can be directly aligned to goals for SMEs. The universal goals for data governance framework include:

1. *Goal* – Enable better decision-making.
2. *Goal* – Reduce operational friction.
3. *Goal* – Protect the needs of data stakeholders.
4. *Goal* – Train management and staff to adopt common approaches to data issues.
5. *Goal* – Build standard, repeatable processes.
6. *Goal* – Reduce costs and increase effectiveness through coordination of efforts.
7. *Goal* – Ensure transparency of processes.

The standardization of data definition is of uttermost importance across SMEs and other goals they are trying to achieve depends on the focus of the data governance framework. Data governance for SMEs comes into existence because of data quality issues, data integrity issues, or the reusability of data. A set of data quality groups should be in charge to ensure better data quality, different types of data will be used in the early stage of iterations of a data governance framework, which will include sets of master data, sensitive data, acquired data, and data of interest to stakeholder.

Weber and Otto [1] identified seven factors which include competitive strategy, breadth of diversification, organizational structure, company's strategy, harmonization of process decision making, and market regulation these factors are likely to influence data governance. Data quality roles. Responsibilities and decision areas, these three important components create a responsibility assignment matrix. Khatri and brown [25] framework identified dimensions for the development of data governance strategy which includes data quality, metadata, data access, data lifecycle. Data is seen across these principles as a driving force and thus supports SMEs strategy [35].

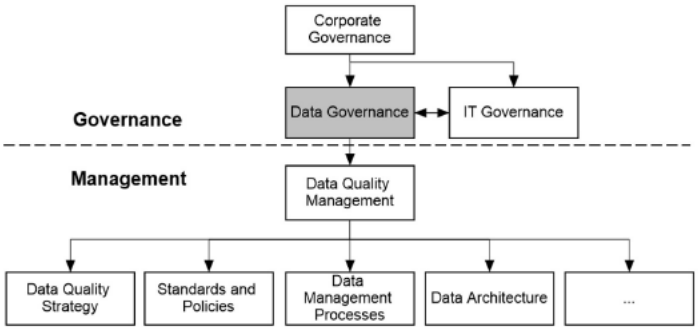

Figure 2 Relationship between data governance and data management [37]

For data to be managed efficiently, it should be combined with metadata [38]. Data governance is seen to be comprised by IT governance [39], this is perception is related to data management as a discipline of IT management [40]. The technological, IT investment, IT management function, IT infrastructure management are based on the analysis of IT governance research. Whilst data quality management deals with the organizational issues that are external to the scope of IT management which includes scope creep, sponsorship, political issues [41].

Data Quality Management (DQM) is focused on the collection, organization, storage, processing, and presentation of high-quality data, data quality management has both organizational and business relevance most often, the responsibility to handle and improve data quality is assigned to the IT departments, while some other companies implement data management. With data governance, it involves the implementation of data quality management accountabilities which incorporates both professionals from the business and IT department. If the implementation of data governance is not in place, then companies will be susceptible to a high risk of failure in implementing compliance and corporate-wide governance which will further decrease data asset value thus resulting in liability issues, data governance implementation should be made horizontally by the assimilation of data governance throughout the enterprise and support in the administration of data-related business processes. Furthermore, data governance helps to ensure that heterogeneous databases and their applications are integrated into a business environment by the support of the implementation of master data management, and this relates also to the concept of enterprise linked data.[42][43].

The dimensions are illustrated in the below diagram:

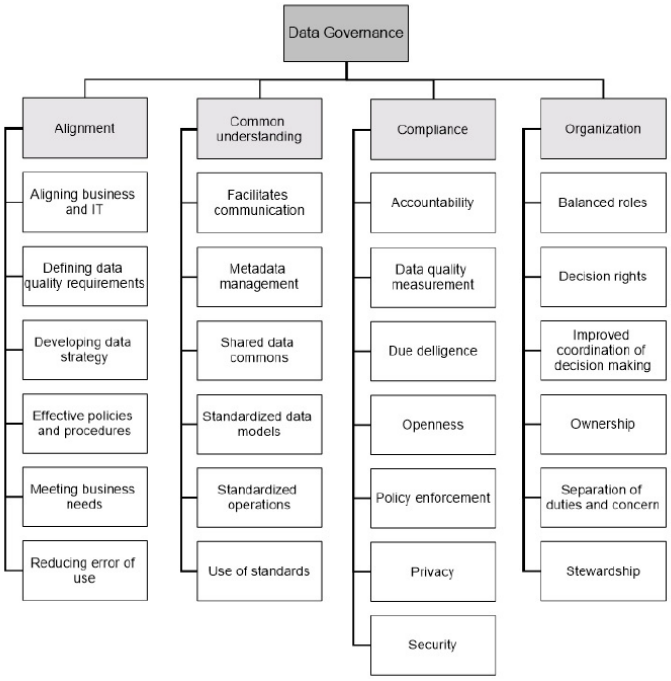

Figure 3 The long list of key concepts and principles of data governance [32]

2.3 The Data Management Body of Knowledge

Various data governance framework has been proposed in research, the first edition of the Data Management Body of Knowledge (DMBOK) was first published in 2008 and served as a guide in data governance research area. The second version was released in 2012 an expanded version thus further in the scope of security issues and cloud services to cope with the overall increase in the size of data. The first version focuses on ten data management functions which include data quality, metadata, data warehousing, and business intelligence, reference and master data, documents and contents, data architecture, data modeling, and design, data storage and operations, data security, data integration, and interoperability and data governance [29]. The dimensions above are a guiding force for data management.

The DMBOK framework provides a structure for data management thereby enabling organizations to see the importance of data assets. The basic and most important elements are the practices and procedures, technology, organization roles and responsibilities, and culture. The eleven data management knowledge includes:

Data integration and interoperability: This deals with data acquisition, extraction, and virtualization, how data can be integrated with other business functions.

Data security: Security of data assets is highly important as this will ensure privacy, confidentiality, and appropriate data integrity. Protecting data from unauthorized access and data corruption. This might include encryption, tokenization, and hashing.

Data storage and acquisition: Data assets should be protected and managed and have structured storage.

Data modeling and design: Analysis and design of data deal with building and appropriate testing of the data model. This is the first step in process of database design.

Data architecture: Policies, rules or standard that guides data collection, storage, and integration.

Documents and content: Enabling, storing, and protecting access to data contents which will be further used for analysis and decision making.

Reference and master data: Ensuring a 360-degree view of organization data assets, master data management is used for defining and managing critical data of an organization to provide a single point of reference while reference data is the analytical data that supports decision making.

Data warehousing and business intelligence: Used for data analysis and reporting, data warehouse stores historical data typically from various disparate sources and it is also a core part of business intelligence.

Metadata: Provides information about other data, this makes working with an instance of data easier.

Data quality: Data fitness for intended use, common traits found in data quality include accuracy, completes, reliability, relevance, and timeliness.

Data governance: Includes people, process, and technology aligned to ensure maximum protection and the use of data assets.

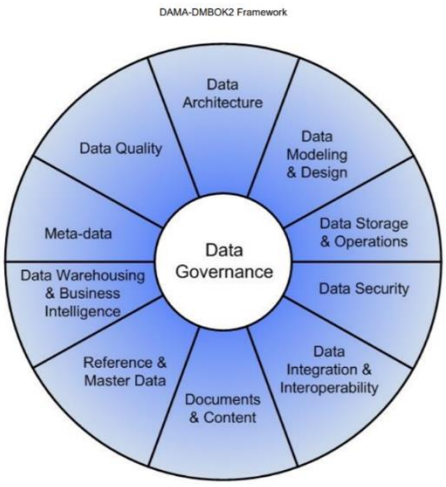

Figure 4 The DMBOK Guide Knowledge Area Wheel [29]

There are also environmental elements that provide a logical and consistent way that describes these knowledge areas above. There are seven elements which have additional descriptor: people, process, and technology.

The highlighted elements in the diagram below are included in the proposed data governance framework as these are the core and vital elements for SMEs adapted from the original DMBOK framework.

Figure 5 Adapted DMBOK Guide Knowledge Area Wheel [29]

For data governance to be effective in SMEs, data storage and operations need to be considered. Firstly, a data lake which is a company's massive collection of all data sources uses analytical applications which is scalable and can hold a huge amount of data in raw form, stores it until ready for further processing operations. Various data formats can be found in a data lake, elements have a unique identifier and metadata labels. Data lake works well with the vast amount of data, that has various raw data formats until further insights are derived from these data. In contrast, data lakes are data spaces that model the relationship between various data repositories, thus it helps to reconnect a variety of data formats that are stored in different locations [30].

Creating structured data deals with the concept of linked data which can be integrated into a flexible bottom-up approach [31]. Having SMEs' data under control with the help of linked data, which will enable them to be aware of processes and attributes. Compliance with laws and the incorporation of data governance will result in a massive result in the change process. To strengthen the quality of data, an ontology-linked data technique should be used. Cloud data governance is an emerging field, challenges in this field include the legal, business, and technical dimensions [32].

3. SMALL AND MEDIUM SCALE ENTERPRISES (SMEs)

3.1 General challenges and opportunities

European Union Commission [8] defines SMEs as a small business is a business with less than 50 employees with an annual turnover of less than €10 million or an annual balance sheet of less than €10 million [47]. The United States Small Business Administration (SBA) defines a small business as one that is independently owned and operated, is organized for profit [48]. Small businesses are important to a nation's economy. For small businesses, they are more likely to fail as compared to a larger business which makes them more focused on survival and incremental growth rate to avid failure [49].

Furthermore, they lack the resources, time and advanced technology, and expertise needed to leverage innovations and new business opportunities because of data analysis of data assets. The management structure is also different as they are mostly owned and managed by the same Individuals with few employees compared to a large business whose owners are mostly private shareholders with managed lead by C suite professionals [49]. Data governance is not a one size fits all, instead, it should be an evolutionary process that can be started small and measurable along the way.

Small businesses are constantly making decisions relating to their business, all decisions made are based on data available, there is an increase in data, business tore and process daily to gain value. There have been numerous technological advances, the business environment is constantly changing, and organizations are looking for ways to scale up to meet up these new requirements. These technologies are collecting even more data than before [44] and this also comes with associated risks such as privacy, issues, risk issues, and security issues. How can vital information be retrieved from these available data? How can SMEs trust the data that they have? All these answers lie in a formal data governance framework.

Research shows the need for a data governance framework specific for each business sector [45]. Few frameworks have been suggested in past literatures but targeted at large businesses. Business drivers in these large organizations are not like those in SMEs [46], making decisions daily based on available data. Thus, data governance is important to SMEs as compared to their larger counterparts. It will help ensure data security, compliance, privacy, quality, credibility, and usability of their data assets. SMEs are constantly seeking options for cost reduction due to limited resources, also, simultaneously, seeking to be more competitive in the market. Most SMEs challenges revolve around attracting new customers, and this is a problem big firms struggle with too.

However, the larger firm finds it quite easy to attract customers due to the historic success behind its brand, for greater recognition for SMEs, smart branding is required, and this will increase the ability to drive customer growth and while maintaining high-quality customer service. SMEs needs to be proactive by making cold calls, attending various business conferences, which will in return help channel SMEs business presence and maintaining profitability remains a challenge for also big players in the firm, number of things that can help increase profitability includes reducing cost, increasing turnover, increasing productivity and efficiency.

Data governance framework for SMEs should:

Be easy to implement: The degree of understanding and learning how to use innovative tools and the degree of difficulty associated with the usage is defined as complexity [50]. For SMEs, the ease of use and implementation is of great priority in any decisions to adopt a framework or tool as technology business decisions are, most times handled by a small group of employees. A framework is likely to be rejected by a business, once it is perceived to be complex and difficult to implement, Therefore, the data governance framework for SMEs should not be complex and easy to implement.

Low cost: The cost associated with implementation, maintenance, and training should be less expensive. An innovation is adopted quickly if it is perceived to be less expensive [51,

52]. During the adoption of a new solution, the cost is noted to be a major factor. SMEs are limited in terms of their resources as compared to larger organizations that have a larger pool of funding both external and Internal [53]. Therefore, the data governance framework for SMEs should be low in terms of the cost of implementation.

Management and internal structure of SMEs: The management structure of SMEs is different compared to larger businesses, there is no board of directors or shareholder and it is also limited [53]. Decision making in SMEs are implemented quickly. Often, during decision-making, there is no need for a board meeting with serval committees. Therefore, the management structure should be considered in the development of a framework.

3.2 SMEs responsibilities to data governance and privacy laws:
This includes:

1) **Data protection by design and default:** The controller should ensure appropriate technical measures are taken into consideration such as **pseudonymization** and appropriate organizational measures during the time of determining means of processing personal data and at the time of processing itself thus implementing data protection principles such as data minimization to protect rights of data subjects. Furthermore, only data meant for specific purposes should be processed.

2) **Data protection by impact assessment:** An assessment must be carried out if processing using new technologies may likely result in a risk to the rights and freedom of data subjects. Furthermore, the advice of a data protection officer should be sought during an assessment.

3) **Designation of a data protection officer:** There are several cases where the designation of a data protection officer is needed:

   a) When processing of special categories in such a case where personal data relates to criminal convictions is done on a large scale.

   b) When processing is done by a public authority with an exemption for courts that act in their judicial capacity.

   c) When processing requires regular monitoring of data subjects on a large scale.

3.3 Data security and protection

Compliance with all laid down regulatory laws and internal requirements is highly important for SMEs. Data stewards are held accountable to the following:

- Locating sensitive data across all systems.
- Alignment of SMEs to all regulatory initiatives, security/technological frameworks.
- Assessment of risk and related controls to risk management.
- Enforcement of regulatory requirement across all business units.

Personal Data breach notification to a supervisory authority:

In event of a data breach, the controller must notify authorities not later than 72 hours of becoming aware of it, except the data breach is not resulting in the risk to rights and freedom of data subjects. If the data breach is not made known within 72 hours, the controller must give reasons for the delay.

The notification shall contain:

a) A comprehensive description of a personal data breach states the number of data subjects affected.
b) Description of impact of a data breach.
c) Description of the measures of control taken.

Personal Data breach notification to data subjects:

Data breach notification must also be communicated to the data subject, if the case of the breach results in a high risk that will affect data subjects, the controller must communicate without delay to affected data subjects in an easy and plain language with measures taken to contain the data breach.

Five steps to protect your small business from cyberattacks.

According to Keeper Security's 2020 SMB Cyberthreat Study, 66% of senior decision-makers at small businesses still believe they are unlikely to be targeted by online criminals. Similarly, 6 in 10 has no digital defense plan in place whatsoever, underscoring the need for heightened industry awareness and education across the board. The five steps include:

1) **Encourage Security Conscious Workplace Environment:**
   The more your employees know about cyber-attacks and how to protect your data, the better off you will be. It may be as simple as reminding them not to open an attachment from people they do not know or expect, posting procedures for

encrypting personal or sensitive information so they do not forget, also requiring them to change their passwords regularly.

2) **Have data breach prevention tools including intrusion detection:**
   Ensure employees are monitoring the detection tools, it is important to not only try to prevent a breach but to make sure that if a breach occurs, the company is aware as soon as possible. If a breach occurs, there should be a clear protocol outlining which employees are part of the incident response team and their roles and responsibilities.

3) **Appoint a Data Processing Officer to ensure compliance:**
   Data privacy laws are in place now, this means companies of all sizes needs to act now and start putting in place robust standards and procedures to counter the cybersecurity threat or face the prospect of paying high costs in regulatory fines and thus cause reputational harm to their business.

4) **Data Encryption:**
   Full-disk encryption software is available from all major computer and mobile operating systems: use this to encrypt all the data you manage, and make sure all your company devices have this software activated and updated. Many cloud services offer data encryption features as well. Backing up or archiving business data is essential for recovery from cyber-attacks, theft of devices, or loss of equipment or media resulting from a flood or fire.

5) **Conduct Regular Risk Assessments:**
   Risk assessment involves identifying, analyzing, and evaluating risk and ensuring that you have picked appropriate cybersecurity controls to protect your business from cyber-attacks. Try a free or trial online risk assessment to get a sense of how your security measures match up to recommended practices.

4. DATA SCIENCE AND DATA GOVERNANCE

Data manipulation and use of various statistical methods are carried out in almost all facets of business today, the degeneration of data manipulation combined with the use of various stats method has become imperative that a strong data governance policy must be in place

to curb any data degeneration used to arrive at a data-driven decision. Data governance will play a key role in various data science practices to help ensure validity at any phase of data analysis, manipulation, or prediction thus preventing any misuse and corrupt scientific methods used to arrive at any decision. Another implication may be that decisions made are based on misused statistical methods, manipulated by humans to suit a specific business need.

Therefore, data governance plays a vital role, ensuring validity checks and balances are put in place so data in data science is not being manipulated or misused for the wrong reasons. The first point of intersection between data science and data governance is in big data which typically resides in all facets of our life and this tends to lose integrity during data transfers, major beaches mostly happen during data transfers. For all data-driven practices, data security should be a priority.

Data governance is intertwined with data science during:

1) Ensuring regularity compliance with laid down laws such as the EU GDPR.

2)Managing enterprise data from an end-to-end value chain.

3) Quality check of unstructured data.

In the next five to ten years, businesses will completely rely on data science, data analytics to make critical business decisions thus ensuring integrity and trust in data and analytics is of the highest priority. Secondly, another point of intersection is in the various market scandal, with the rising data scandals from top organizations like Facebook, data governance, protection, and security are becoming the topmost priority in the business world. Data governance platforms are becoming more prominent now as the goal of data governance solution is to maintain data at its highest level of quality while managing master data management too. This will also form the core of data ethics A code of conduct or ethics for data scientists, like the purpose of the Hippocratic Oath in guiding medical professionals.

Businesses are aware of data ethics to ensure fair and appropriate data usage, build trust with customers by proving themselves worthy by minimizing bias and social injustice, and most importantly, complying with regulations. Thirdly, the intersection between data science and data governance is the chief data officer; with the growth of automated advanced technologies, the power of data-driven activities is becoming more prominent.

Lastly, another point of intersection is building ethical machine learning models that are free of bias, algorithm silently structure our lives, algorithms make some life-changing decisions; the algorithm can determine whether someone is hired, promoted, offered a loan,

or provided housing as well as determine which political ads and news articles consumers see. Yet, the responsibility for algorithms in these important decisions is not clear. Bias is inborn in data, and if a machine learns from biased historical data, it reinforces bias more.

An immediate observation is that a learning algorithm is designed to pick up statistical patterns in training data. If the training data reflect existing societal biases against a minority, the algorithm is likely to incorporate these biases. This can lead to less favorable decisions for members of these minority groups. While automated decision systems have the potential to bring more efficiency, consistency, and fairness, it also opens the possibility of new forms of discrimination, which may be harder to identify and address.

The impervious nature of machine learning algorithms and the many ways human biases can creep in, challenge "our ability to understand how and why a decision has been made" and our capacity of guaranteeing fundamental values of society, such as fairness, justice, and due process [55][56]. However, addressing bias requires more than a technological fix but an understanding of the underlying structural inequalities [55][57]. A particularly challenging question is how to decouple legitimate information and sensitive information carried by the same variable, such as zip code. Looking in-depth into discrimination-aware data mining to measure the performance of algorithms. With, structured data governance in place, bias, and discrimination related to these will be curbed to a significant level.

4.1 The intersection of data science and data governance

Data governance ensures data management processes, data security and control are compliant with laws and policies, ensuring confidential features are removed during data processing or before data is shared or used legally.

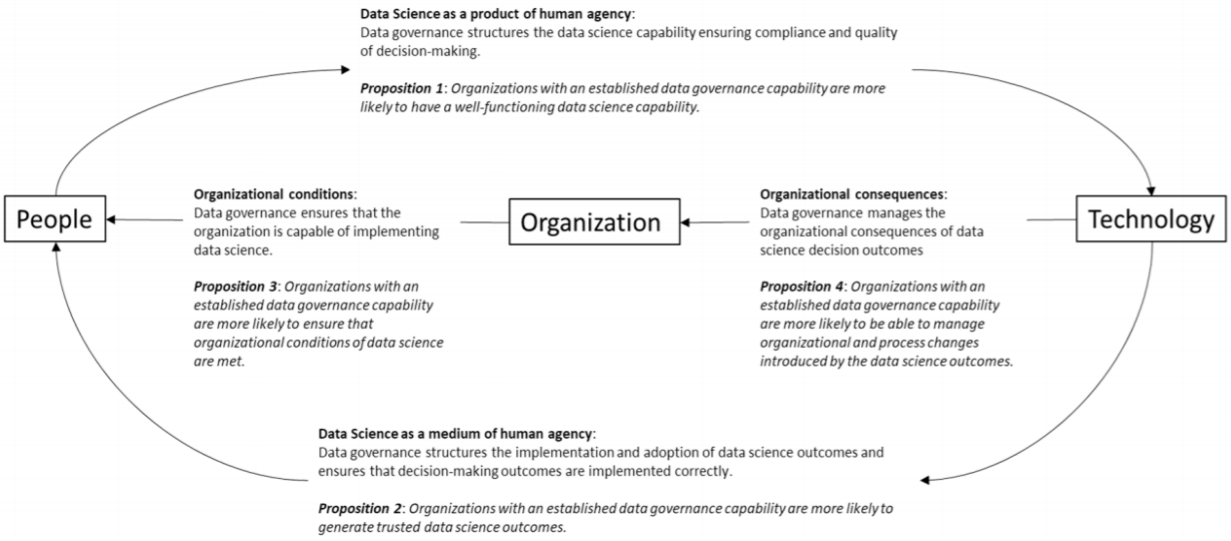

Figure 6 Data governance and data science intersection [57]

In this figure above, Orlikowski (1992) identifies four vital propositions as stated above. Technology here is referred to as data science, understanding the role data governance plays as a boundary condition for data science. Organizations with a data governance capability are better positioned for trusted data decisions, highly valuable for reducing risks in data science tasks or initiatives, creating trust in the outcome of data science decisions thereby influencing positively the use and acceptance of data science outcome. Data science as a product of human agency enables the structuring of data science capability thereby ensuring compliance and helps quantify decision making. First proposition states that organizations that establish a strong data governance capability show more success in terms of data science functions. Data Science is a medium of human agency.

Data governance is also beneficial as it structures decision-making outcomes and the implementation of data science outcomes, thereby the second proposition states, an established data governance presence in an organization, generates trusted data science outcomes. A relationship between technology, organization, and people ae establishes, for organizational conditions, data governance helps ensure the implementation of data science practices and thereby linking it to the third proposition that states, an established data governance structure in an organization ensure organizations conditions of data science are met, there is also an organizational consequence of data science which data governance helps manages. The last proposition explains the need for established data governance due to organizational and process changes introduced by data science outcomes.

Most importantly, an increasing number of decisions are made by various machine learning models, and with the world-embracing more of AI to drive decision-making using data.

Whether these algorithms are approving loans application, or determining the risk profile of a particular candidate, there is a broad interest to ensure data governance of all the systems are building accountable transparent, and fair algorithms. All data-driven organizations must reckon with data governance requirements. There is also legal compliance as discussed in the chapters above that must be taken into consideration as no organizations want to be pulled by the media for violating a law or perceived discrimination/bias. However, all these can be curbed by following a data governance framework. The EU law, CCPA, and PDPA provide a framework for transparency policy and the governance of automated decision-making systems. These laws help bridge the gap in the data management and data privacy space.

4.2 Master Data Management

Master data management enables the control of data showing a single system of record and has any changes replicated across the various systems in an automated manner. Master data management focus mainly is to create an accurate, timely, integrated, and complete set of master data to help manage and grow business. The single location in which the integrated set of master data is stored is called the master data system of record.

Master data management is classified into I) operational MDM and II) Analytic master data management.

Operational master data management deals with the integration of operational applications such as ERP or CRM and the supply chain management in data flow main upstream. Analytic Master data management such as data warehousing; customer data integration all forms the enterprise master data management. Maintaining and publishing all the organization master data is the function of an enterprise Master data management system. Master data management applications, master data store, master metadata store, and master data integration services all make up the main components of an enterprise Master data management.

Content and requirements for master data:

A successful Master data management solution requires heavily on following:

1. A data governance framework which, manages and integrates enterprise data into a master data environment.

2. Seamless facilitation of information extraction, sharing, and delivery.

3. Proper definitions and resolution involves the usage and meaning for entities, object relationships, and hierarchies.

4. Transparency involved in exposing services related to enterprise clients for accessing managing master data assets.

Successful master data management process involves:

- A proper definition of master data flow.
- Identification of sources and consumer master data.
- Master data model definition.
- Definition of a functional and operational characteristic of master data management tool.
- Create a master data element by merging source data.
- Creation and maintenance of business and technical metadata.
- Publish master data.

Overall, master data management should have more of a business focus instead of a technology focus or a mixture of both. The three elements that form the one master data include data, processes, and information system. They form the framework for successful Master data management implementation. Data, information systems, and processes form the framework which makes the core entities.

4.3 Data decision domain

IBM created a maturity model for data governance which includes eleven categories which include classification, metadata, data architecture, data quality management, information security, and privacy, data risk management and compliance, information lifecycle management, audit information, and logging and reporting [36].

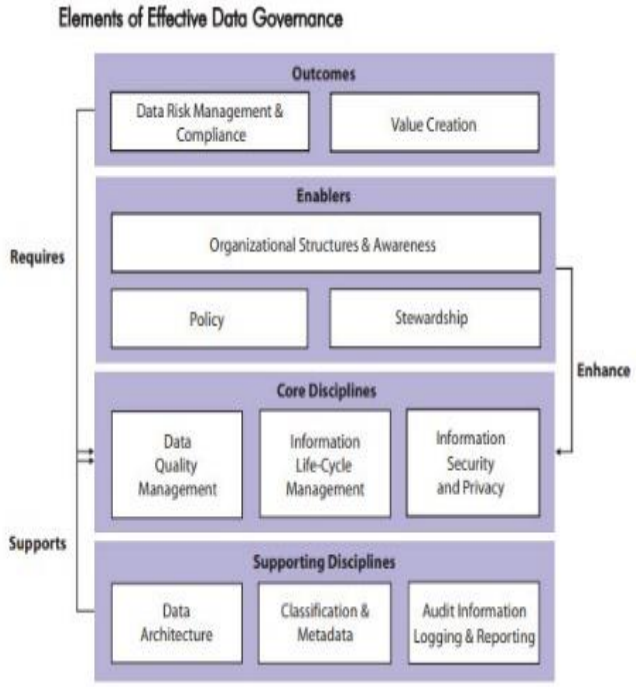

Figure 7 Data governance domain [12]

Data risk management and compliance: Identification of risk and ensuring it is avoided, accepted, or mitigated.

Value creation: Data assets are maximized to create value which will enable a business to maximize its value.

Organizational structural and awareness: Establishment of a structure with descriptive levels of responsibility for both the business and IT department

Policy: Set guidelines or rules which enable the decision-making process.

Stewardship: Ensuring data assets are in custodial care which will enable asset enhancement, risk management, and better organization control.

Information lifecycle management: Policy-based approach which involves data collection, use, retention, and deletion, to maximize its utility, lower cost, and minimize legal risks by ensuring data is stored securely and is not retained for longer than is needed. A strong information lifecycle management will help control data retention.

Information security and privacy: Privacy relates to rights you have to data assets and how it is used and information security, on the other hand, refer to how these data assets are protected.

Classification & metadata: Metadata contains an underlying definition or description of data.

Audit information, logging & reporting: This involves monitoring and taking into consideration the data value risk and efficacy of data governance.

5. COMPARATIVE ANALYSIS BETWEEN DATA PROTECTION LAWS

5.1 General Data Protection Regulation (GDPR)

The European Union (EU) is known to have the strongest data governance framework alongside rules and regulations introduced in the past years. The main piece of data regulation which is the General Data Protection Regulation (GDPR) turned the EU as a global driver in data protection, this was released in 2016. GDPR came into effect on May 25th, 2018. The most recent is the creation of a data strategy that aims to provide a single market point for personal data thus offering strong protection. The EU designed GDPR to harmonize all data privacy laws across its member countries thereby pricing greater protection and rights to individuals. This also affects how businesses handle information on their users and there is also a large fine to be paid if not in compliance with laid down laws. Furthermore, the cybersecurity act and open data directive, and the free flow of non-personal data was later released in 2019.

GDPR's seven principles are lawfulness, fairness, and transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, and confidentiality (security), and accountability.

- **Lawfulness, fairness, and transparency:** Ensures that data is fair, lawful, and transparent to its various data subjects.
- **Purpose Limitation:** Data must be processed legitimately, or its purpose explicitly specified for its data subjects at the point of collection.
- **Data Minimization:** Data should be collected specifically for its intended purpose.
- **Accuracy:** Personal data should be kept up to date.
- **Storage Limitation:** Data should be stored for as long as necessary for its intended purpose only.

- **Integrity and confidentiality:** Data processing should be done in such a way as to ensure confidentiality and integrity.
- **Accountability:** Data controller must ensure compliance to all laid down GDPR laws.

There are also several rights an individual can exercise, this includes:

- **Right to access:** The data subject shall have the right to access and obtain personal data, this also includes the purpose for which the data is used and processing information.
- **Right to rectification:** The data subject shall have the right to rectify inaccurate personal information.
- **Right to processing restriction:** The data subject shall have the right to restriction of processing where there is a contention in accuracy of personal data, the processing is unlawful, processing is no longer needed by the controller but required for the exercise or defense of a legal claim by the data subject.
- **Right to object:** The data subject shall have the right in the event of direct marketing or profiling or the processing of personal data under certain events.
- **Right to portability of data:** The data subject shall have the right to personal data provided to a controller in a readable format, right to transmit personal data from one controller to another where feasible.
- **Right to erasure:** The data subject shall have the right to obtain from a controller deletion of any personal data without any undue delay.
- **Right not to be a subject to automated decision making which includes profiling:** The data subject shall have the right not to be subject to automated decision making or profiling.

5.2 Personal Data Protection Act 2012 (PDPA)

Personal Data Protection Act 2012 (PDPA) provides a common baseline standard for the protection of personal data alongside existing laws to provide full protection. This helps signifies Singapore's commitment to the collection, disclosure, and use of personal data in this big data era as well as strengthening Singapore's position as the leading commercial hub. Most Importantly, the PDPA protection model has the right balance for the rights of an individual to protect its data against the collection roof an organization to use for legitimate and specific purposes only [54].

Key Requirements for PDPA:

- **Purpose Limitation:** Data should be collected and used for its intended purpose only.
- **Notification:** Data subjects should be aware in simple and clear terms what their data is used for during the collection, processing, and usage stages.
- **Consent:** Clear and concise consent should be given by data subjects before collecting and processing of their data.
- **Access and Correction:** When requested adequate information should be given in how data is used and disclosed and corrected if need be upon request.
- **Accuracy:** Data should be accurate and complete at all times.
- **Protection:** Data should be protected from all unauthorized users to avoid data breach form occurring.
- **Retention Limitation:** Retain data for its intended use only and delete when no longer needed.
- **Transfer limitation:** Follow all laid down laws regarding data transferability.
- **Openness:** Transparency and openness when communicating data practices or policies with the public.
- **Do-Not-Call Registry:** Do not send marketing messages to individuals who have registered in the National DNC registry.

5.3 California Consumer Privacy Acts of 2018 (CCPA)

The CCPA laws aim to guarantee strong protection for individuals regarding their data and to businesses that collect, use, and share consumer data in all ramifications, online or offline. The CCPA is known to be a significant legislative law guiding the country's providing stronger privacy, protection, and greater transparency, giving consumers more control over their data. California is known to be the fifth largest in the global economy. This has already taken effect as of 1 Jan. 2020. Californians have the right to know what type of personal data is collected from them, accessibility to the data, request for deletion, know if the personal data is shared with third-party organizations, opt-out of campaigns, equality in terms of service whether they choose not to exercise their privacy right. Third-party are also prohibited from selling the personal information of their consumers from ages 13-16, consent is required from their parents or guardian. Most Importantly, CCPA aims to increase transparency for California residents.

The four main California Resident Right:

- Right to know which personal data is collected by business how it is used and shared.
- Right to deletion of personal data.

- Right to non-discrimination for exercising CCPA rights.
- Right to opt-out of the sale of personal data.

Table 1. Key Requirements of EU GDPR

Key requirements of EU GDPR	
Application	EU GDPR applies to any organization that processes personal data of people in the EU, where the processing relates to an offer of goods/ services and /or relation to the monitoring of individuals in the EU. Furthermore, an appointment of EU representative applies if an organization is not established in the EU. However, an appointment is not required where processing done by the organization is occasional and is not a requirement of processing special segments of personal data.
Processing	Processing of personal data is deemed lawful when consent is given for the processing of personal data for a specific purpose by an individual.
Consent	Explicit consent must be given by individuals.
Individual's right	An individual has several rights which includes the right to access, rectification, processing restriction, object, portability of data, erasure, not to be a subject to automated decision making which includes profiling.
Data erasure	The data subject has the right to request to be forgotten, disseminated, and/or prevent third parties from processing their data.

Data portability	Data subjects should be able to receive personal data in a commonly used and machine-readable format.
Notification of data breach	Data breach notification must be communicated to the supervisory authority and data subjects within 72 hrs. of becoming aware of the breach.
Fines	Depending on the infringement, up to 10 million EURO or 2% of its annual turnover of the preceding financial year, whichever is higher, and up to 20 million EURO or 4% of its annual turnover of the preceding financial year, whichever is higher.
Data protection officers	There should be an appointment of a data protection officer for organizations whose core activities include data processing operations, monitoring data subjects on a large scale, processing of special categories of data which includes data related to criminal convicts.
Privacy by design	During the designing of a system, data protection must be included from the onset.

Like the GDPR, PDPA also extends to those who may not have any presence in Singapore. The only major similarity is in the technological definition. The Singapore Personal Data Protection Act 2012(PDPA) is a set of laws that governs how organizations collect, use, and disclosure the personal data of all individuals. PDPA came into full effect in 2014. PDPA allows only for collection and use of personal data for its defined intention or purpose only and thereby must inform individuals on the purpose and use of their data during collection. Before that, consent must have been obtained from individuals before processing personal data.

If requested by an individual, the organization must be able to provide information on how individuals' data has been used in previous years, if any mistake is found by an individual, a request for correction must be done. Organizations must ensure all personal data collected must be accurate and up to date and secure from unauthorized access. All personal data must be retained for business, legal purposes only, else destroy when no longer needed. A

data protection officer must be designated. Lastly, the establishment of a do-not-call registry, individuals registered under this registry should not send any form of marketing messages except clear consent has been obtained. Major changes being proposed to the PDPA, but during the time of this research, they are not yet passed and implemented. An organization that complies with these laws tends to build trust with its customers.

Individual rights include:

- Right to request access from an organization to personal data (few exceptions).
- Right to request rectification of an error or omission in personal data (few exceptions).
- An individual has no right to request an organization to delete personal data.
- Right to restrict processing.
- Right to withdraw consent.
- Right to file a complaint to relevant authorities.

Analysis of the three data privacy laws:

Scope

GDPR: Applies to all data controllers and processors in the EU, regardless of processing taking place within the EU. If the data processing activities relate to the provision of goods and services in the EU or activities of an individual in the EU.

PDPA: PDPA applies to any organization in Singapore processing personal data from anywhere, and organizations outside of Singapore processing personal data from individuals in Singapore. PDPA does not have a limit to its scope of extraterritorial jurisdiction.

CCPA: Applies to for-profit businesses that operate within California, this also gives more transparency and control to California residents.

Enforcement

GDPR: Date Made:14th April 2016, the implementation date was on 25th May 2018.

PDPA: DNC Registry came into effect on 2 January 2014 and the main data protection rules on 2 July 2014. Personal Data Protection Act (PDPA), which was passed in 2012 and became effective in 2014.

CCPA: Introduced on Jan 3rd, 2018 and passed on Jan 28, 2018, became effective Jan 1, 2020.

Personal Data

GDPR: Defines personal data as data relating to an identifiable natural person directly or indirectly which might include name, Identification number.

PDPA: Defines personal data as data about an individual, who can be identified and in which an organization has or likely might have access to.

CCPA: Defines personal data as data that relates to a particular consumer or household directly or indirectly, which might include browsing history, geolocation data.

Consent

GDPR: Does not allow for deemed consent at all. Consent must be explicitly and freely given before processing of data personal data. The GDPR sets a threshold of 16 years of age for consent.

PDPA: Ensure that the consent has been obtained from the individuals before collecting, using, or disclosure of personal data. Consent is not required under PDPA if data processing falls within deem consent. PDPA does not stipulate a minimum age of consent.

CCPA: Businesses must have opt-in consent to sell personal information of consumers under the age of 16 if businesses have "actual knowledge" that a consumer is under 16. For consumers under the age of 13, the child's parent or guardian must affirmatively authorize the sale of the child's personal information. They must enable consumers to exercise their rights to opt-out of the sale of their personal information or request its deletion.

Data Minimization

GDPR: Only personal data necessary for its intended purpose should be collected.

PDPA: Any personal data that is relevant to the purpose can be collected, a test of appropriateness reasonability that is judged reasonable by a person.

CCPA: Does not impose a data minimization mandate.

Penalties

GDPR: Less severe violations, 2% of global annual revenue or £10 million, whichever is higher, for especially severe violations, 4% of global annual revenue or £20 million, whichever is higher.

PDPA: Fines not exceeding S$5,000-10,000(Depending on the offense) or imprisonment up to 12 months, this is penalties for individuals in breach of policy. Fines not exceeding S$50,000-100,000 (Depends on the offense).

CCPA: $2,500 per record for each unintentional violation, $7,500 per record for each intentional violation.

Right of Individuals

GDPR: Right to access, right to rectification, right to processing restriction, right to object, right to portability of data, right to erasure, right not to be a subject to automated decision making which includes profiling.

PDPA: The right to give consent and withdraw consent at any given point in time with reasonable notice except it disrupts legal proceedings, right to request personal data from an organization's control, right to request an organization that has control over data to correct any inaccurate data, also subject to exemptions.

CCPA: The right to give consent and withdraw consent at any given point in time with reasonable notice except it disrupts legal proceedings, right to request personal data from an organization's control, right to request an organization that has control over data to correct any inaccurate data, also subject to exemptions.

Data Breach

GDPR: In case of a data breach, notification of a personal data breach to supervised authority, affected individuals not later than 72 hours.

PDPA: Protection of personal data under its control by preventing unauthorized access to the collection, use, and disclosure of personal information. There are no mandatory data breach reporting procedures enforced in PDPA. Organizations are also advised to notify the PDPC as soon as possible of any data breaches that may potentially cause public concern, particularly if the breach involves sensitive personal data, or where there is a risk of harm to some affected individuals.

CCPA: California consumers have the right to bring statutory damages into action if personal data is subject to the data breach, this right applies to certain kinds of breaches. I) If information is personal information, ii) personal info must not be encrypted/redacted iii) breach because of business violation to maintain reasonable security procedures.

Data Transfer

GDPR: Consent should be given before business transfer personal data to third parties.

PDPA: Protection of personal data under its control by preventing unauthorized access to the collection, use, and disclosure of personal information. There are no mandatory data breach reporting procedures enforced in PDPA. Organizations are also advised to notify the PDPC as soon as possible of any data breaches that may potentially cause public concern, particularly if the breach involves sensitive personal data, or where there is a risk of harm to some affected individuals.

CCPA: California consumers have the right to bring statutory damages into action if personal data is subject to a data breach, this right applies to certain kinds of breaches. I)If

information is personal information, ii) personal info must not be encrypted/redacted iii) breach because of business violation to maintain reasonable security procedures.

Data Protection Officer

GDPR: Appointment of a data protection officer who has expert knowledge of protection laws, they must be provided with adequate resources needed to perform their job duties well.

PDPA: No officials or officers need to be designated.

CCPA: Appointment of a data protection officer (DPO), to oversee compliance, rules, and regulations with PDPA during the development and implementation of processes when handling personal data.

Sale of Personal Information

GPDR: No restrictions under GDPR, however, compliance is required by data controllers in respect to data processing.

PDPA: Organization shall not use or disclose personal data about an individual except consent is given**.**

CCPA: Businesses must include the "Do Not Sell My Personal Information" link on their homepage. that "sell" and must give consumers the right to opt-out of that sale and must include be clearly labeled.

Table 2. Detailed Similarities between GDPR and CCPA

GDPR	CCPA
Right to data deletion: Both the GDPR and the CCPA allows an individual to request the right to deletion of their personal information. **Children:** Both GDPR and CCPA have similar protection for children. GDPR specifies protection when the processing of children's data for social services while CCPA provides a specific rule regarding children's data regarding "selling" personal information, not limited only for social services. **Pseudonymization:** Both GDPR and CCPA have a similar definition of the term "pseudonymization" which is the processing of personal data in such a manner that the personal data can no longer be attributed to an identifiable person without the use of	

additional information, by putting in place technical and organizational measures which keep the additional information needed for identification separately. **Right of data subjects:** Both the GDPR and the CCPA establish a right of access, which allows individuals to have full visibility of the data an organization holds about them they can obtain details about the data being processed, but also copies of the data items themselves. **Right to opt-out:** Both the GDPR and the CCPA give the right to an individual to ask organizations to cease the processing, and selling respectively, of their data. CCPA requires a "do not sell my information" link while GDPR requires a withdrawal consent. **Data Portability:** Both the GDPR and the CCPA allow for data portability which is a part of the right to access. Data portability is regarded as the right to access in CCPA, while the GDPR provides for a separate and distinctive right. **Non-Compliance:** Both the GDPR and the CCPA states clearly the monetary penalties associated with non-compliance. However, the nature of the penalties, the amount, and the procedure to be followed differ quite significantly. **Supervisory Authority:** Both the GDPR and the CCPA assist organizations in understanding and complying with the laws by providing for an authority to supervise the application of the law. **Judicial Remedies:** Both the GDPR and the CCPA allows for individuals to seek damages for privacy violations, they also both laws allow for the class or collective actions to be brought against organizations.

Table 3. Detailed Similarities between GDPR and PDPA

GDPR	PDPA
Personal Data: Both define personal data as information directly or indirectly relating g to an individual, GDPR has a more detailed definition of personal data and distinguishes between different categories of personal data. **Consent:** the GDPR and the PDPA provide data subjects and individuals with the right to withdraw consent to the processing of their data. However, the GDPR provides data subjects with the right to object to the processing of their data, whereas the PDPA does not provide such a right. **Data Controller:** Both have similar definitions in terms of the concept of who a data controller is. Data controller bob duties include informing on the purpose and the use of	

data subject personal data and providing the right to withdraw consent. However, PDPA does not provide the request for erasure or deletion of personal data.

Non- Compliance: Both have a designated supervisory authority that outlines monetary penalties and corrective procedures in case of noncompliance. GDPR has a higher penalty compared to PDPA.

Accountability: Both recognize the fundamental privacy principle of an organization's accountability to the protection of personal data under its possession, which includes the designation of a data protection officer.

Data Protection Officer (DPO): Both recognize the appointment of a DPO. However, the GDPR has an extensive definition, qualities, and expertise of a DPO but the PDPA does not.

Personal Data Transfer: Both provides an extensive restriction on the transfer of personal data to another country or an organization stating explicitly legal grounds where transfers can be legally done.

Rights of data subjects: However, the Personal Data Protection (Amendment) Bill would introduce a data portability obligation like the right to data portability under the GDPR, which would require organizations, at the request of the individuals, to share the individual's data to another organization, in a machine-readable format.

6. DATA GOVERNANCE FRAMEWORK FOR SMEs

The data governance structure will form a basis for a transparent decision-making process and fosters accountability, lack of clear roles in business affects management and data quality. Data governance has different implementations depending on the business objectives, every business thus needs a framework to help with the implementation of data governance.

Framework includes:

Data Security and Privacy as Priority:

A data governance strategy must have a strong linkage between privacy and data security, any data retained may pose a risk of the data breach and the way forward is to limit the amount of data collected to only what is essential for use and deleting data once they are no more critical for business use. Data retained must be properly secured to avoid data

breach and secured from misuse from insiders. Data security should also be of utmost propriety when in transit with a strong authentication mechanism in place for control and verification. Customer-sensitive data should only be made available when they require strong verification access.

Data architecture which includes policies, rules or standard that guides data collection, storage, and integration should be strictly monitored so access to them can be audited in case of a data breach. Data privacy and security forms a strong foundation for a data governance strategy.

Data Impact Assessment:

Data should be properly pseudonymized or anonymized to reduce the risk and sensitivity. Ensure only data there are appropriate for use should be collected and retained. Personally, identifiable data must be stored properly analyzed and non-sensitive data which can also be unique enough to a person should be stored safely for processing. Data governance strategy must consider and identify the risk about data that could be identifiable and should consider implementing differential privacy to safely store and process data. Two impact assessment namely Data Protection Impact Assessments (DPIAs) under the General Data Protection Regulation (GDPR) and PIA (Privacy Impact Assessment), are two different but distinctive assessment which will help to access various data processing activities with an organization and show high-low level risks to rights, compliance, and privacy. PIA is used to enforce privacy by design to ensure data privacy policies and laws are all enforced to avoid a data breach and fines. Most importantly, this assessment also helps in risk identification and mitigation.

DPIAs specifically to GDPR is related to identification and mitigation of risk related to the usage and processing of personal data. Organization can use DPIA to show compliance in all aspect related to data privacy and laws. DPIAs document provides a proof that an organization has evaluated all necessary risks related to all data process usage and taken all necessary steps to mitigate them. This paper highlights the importance of data governance for SMEs and a data framework to follow that will help guide SMEs. Each assessment gives a holistic view of how compliant and comprehensive view of privacy risks and data related risks and helps ensure each is addresses and mitigated.

Designation of Data Review Board:

A data review board should be set up which includes professionals from data science, legal, compliance, marketing, and security. The diversification of such a board will ensure a

diverse output with regards to that will give insights valuable to the data science process and the enterprise. The board's core focus should be on examining all necessary details from data storage and acquisition which includes data assets should be protected and managed and have structured storage, data collection, cleaning, analysis, processing, and use, attempting to predict how data usage will affect the company's reputation, the company's risk of a data breach, and the company's risk of legal noncompliance. This board should also engage the knowledge of experts who may be able to foresee relevant non-compliance or potential breaches that might occur.

Avoidance of Unfairness or Bias:

Bias can be inherent in data at all levels of data analysis (data can be culled from a non-representative sample; a particular segment which represents only a subset of a population or skews out a subset of minorities or user of lower-income status, rather data culled should be a representation of diversification. Data can also be skewed in a way that wrongfully represents a particular group more than others and lastly by human errors which might include, defining groups wrongly, handling missing values incorrectly amongst others.

Data scientist selects a particular machine learning technique for prediction or analysis based on the context of the problem, Systematic bias should be noted as it can be inherent in data or chosen method used and can cause the machine to predict a particular set of groups differently and in some cases poses bias. Context related to how data is considered is also of utmost importance as patterns inherent in such data may be aggravated or reversed in the case of Simpson's Paradox. To further implement an effective data strategy, all systems should be tested for bias during initial development up until testing. Data can also be audited to reduce bias. This will help enhance good data quality for data warehousing and business intelligence which will be used for data analysis and reporting, data warehouse which stores historical data typically from various disparate sources and it is also a core part of business intelligence.

Data Governance, Ethics and GDPR:

Offering statements of data responsibility, access to third-party rules and transparency helps to further advance ethical data governance, transparency, fairness accountability, and responsibility. The GDPR which became effective on 25 May 2018, widely applied to data processing of EU citizens whether performed in the EU or by any foreign entity. All organizations must take priority and consider how to build and apply responsible data

governance strategies in the coming years, as many data governance strategies are becoming legal requirements around the globe.

7. SUMMARY AND CONCLUSION

Based on the framework, the key components that should be in place in an SME before a data governance framework can be adapted. The components include data security and privacy as priority, risk of data identifiability, designation of data review board, avoidance of unfairness or bias, and data governance ethics and GDPR.

Some of the components discussed includes a detailed assessment of the SME to determine whether there are skilled personnel to ensure the correct, sustained, and measurable implementation of the IT governance processes in the enterprise; a consideration of the size of the enterprise to ensure it has the resources it needs to successfully manage the technical and human resources required to actualize the processes of an adapted data governance framework. The Data framework for SMEs is a critical role within any organization that creates and consumes data and enables organizations to align policy with the reality of how individuals leverage data. As more companies move towards utilizing data for automation and creating recommendations to leadership, the role of data governance framework is crucial to provide prior experience, consistent advisement and use of best practices across the organization. The Data framework for SMEs should be the point of accountability for ensuring that policies for data are well researched, align with industry best practices and are successfully deployed.

8. REFERENCES

[1] Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all–A contingency approach to data governance. Journal of Data and Information Quality (JDIQ), 1(1), 1–27.

[2] Wende, K. (2007). A model for data governance -Organising accountabilities for data quality management. Paper presented at the 18th Australasian Conference on Information Systems.

[3] Pierce, E., Dismute W., & Yonke C. (2008). Industry report: The state of information and data governance: Understanding how organizations govern their information and data assets. Baltimore, MD: International Association for Information and Data Quality. [Google Scholar]

[4] Holt, V., Ramage, M., Kear, K., & Heap, N. (2015). The usage of best practices and procedures in the database community. Information Systems, 49, 163–181. 10.1016/j.is.2014.12.004

[5] Fisher, T. (2006). Fourth Quarter. Data monitoring: Add controls to your data governance and compliance programs. Business Intelligence Journal, 11, 51–57.

[6] Otto, B. (2011). A morphology of the organization of data governance.

[7] Cheong, L. K., & Chang, V. (2007). The need for data governance: A case study. In, Proceedings of the 18th Australasian Conference on Information Systems, Toowoomba (Australia), 2007- 12-06, Pp. 999-1008.

[8] Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53, 148–152. doi:10.1145/1629175.1629210.

[9] BROWN, C. V. (1997). Examining the emergence of hybrid IS governance solutions: Evidence from a single case site. Inf. Syst. Res. 8, 1, 69–94

[10] Ali M Al-Khouri (2012). Data ownership: who owns "my data". Int. J. Management. Inf. Technology, 2(1):1–8.

[13] Vijay Khatri and Carol V Brown (2010). Designing data governance. Communications of the ACM, 53(1):148–152.

[14] Boris Otto (2011). Organizing data governance: findings from the telecommunications industry and consequences for large service providers.

[15] Khatri V, Brown CV (2010) Designing data governance. Communications of the ACM53(1):148–152.

[16] Wang RY (1998) A product perspective on total data quality management. Communications of the ACM 41(2):58–65.

[17] DAMA (2009) The DAMA guide to the data management body of knowledge. Technics publications, Bradley Beach, NJ.

[18] Bair, J (2004). Practical Data Quality: Sophistication Levels.

[19] Olson, J (2003). Data Quality: The Accuracy Dimension. Published by Morgan Kaufmann Publishers, USA.

[20] Russom, P (2006). Taking Data to the Enterprise Through Data Governance. TWDI Report Series, March, viewed 16 July 2006.

[21] Friedman, T (2006). Key Issues for Data Management and Integration.

[22] Thomas G (2006). Alpha Males and Data Disaster. Published by Brass Cannon Press, USA.

[23] Newman, D. & Logan, D. (2006). Governance Is an Essential Building Block for Enterprise Information System. Gartner Research, May.

[24] Cohen, R (2006). BI Strategy: What's in a Name? Data Governance Roles, Responsibilities and Results Factors. DM Review.

[25] Khatri V, Brown CV (2010) Designing data governance. Communications of the ACM 53(1):148–152 Krüger W.

[26] Pierce E, Dismute WS, Yonke CL (2008) The state of information and data governance– understanding how organizations govern their information and data assets. IAIDQ and UALR-IQ Schulte-Zurhausen.

[27] Weber K, Ofner M (2008) Case study Ciba – organizing master data management. BE HSG/CC CDQ/11, Institute of Information Management, University of St. Gallen.

[28] Loshin D (2008) Master data management. Kaufmann, Burlington.

[29] Cupola, S Earley, and D Henderson (2014). Dama-dmbok2 framework.

[30] Michael Franklin, Alon Halevy, and David Maier (2005). From databases to dataspaces: a new abstraction for information management. ACM Sigmod Record.

[31] David Wood, Marsha Zaidman, Luke Ruth, and Michael Hausenblas (2014). Linked Data. Manning Publications Co.

[32] Paul Brous, Marijn Janssen, and Paulien Herder (2016). Coordinating data driven decision-making in public asset management organizations: A quasi-experiment for assessing the impact of data governance on asset management decision making. In Conference on e-Business, e-Services and e-Society, pages 573–583.

[33] Majid Al-Ruithe, Elhadj Benkhelifa, and Khawar Hameed (2016). Key dimensions for cloud data governance. In Future Internet of Things and Cloud (FiCloud), 2016 IEEE 4th International Conference on, pages 379–386.

[34] Gwen Thomas (2012). The data governance framework. The Data Governance Institute, Orlando, FL (USA).

[35] Vijay Khatri and Carol V Brown (2010). Designing data governance. Communications of the ACM, 53(1):148–152.

[36] IBM Data Governance Council Maturity Council, IDG (2007). The IBM data governance council maturity model: Building a roadmap for effective data governance.

[37] Kristin Wende and Boris Otto (2007). A contingency approach to data governance. 2007.

[38] Javier Nogueras-Iso, F Javier Zarazaga-Soria, Javier Lacasta, Rubén Béjar, and Pedro R Muro-Medrano (2004). Metadata standard interoperability: application in the geographic information domain. Computers, environment, and urban systems, 28(6):611–634.

[39] Dyché, J., Levy, E. (2006), Customer Data Integration. John Wiley & Sons. Hoboken, New Jersey.

[40] Sambamurthy, V., Zmud, R. W., Arrangements for Information Technology Governance: A Theory of Multiple Contingencies. MIS Quarterly, 23 (2) 1999, pp. 261-290.

[41] Wang, R. Y., Lee, Y. W., Pipino, L. L., Strong, D. M. (1998), Manage Your Information as a Product. Sloan Management Review, 39 (4).

[42] Murtha Baca. Introduction to metadata (2008). Getty Publications.

[43] Stefhan van Helvoirt and Hans Weigand (2015). Operationalizing data governance via multi-level metadata management. In Conference on e-Business, e-Services and e-Society, pages 160–172. Springer.

[44] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz (2014), "Big data, analytics and the path from insights to value," MIT Sloan Management Review, vol. 21.

[45] J. Harper (2013), "2014 Trends in Data Governance," in Data Articles | Data Science, Business Intelligence, & More, ed: Dataversity.

[46] M. Levy and P. Powell (2005), Strategies for Growth in SMEs: The Role of Information and Information Systems, First ed. Oxford: Elsevier Limited.

[47] E. U. Commission, "Commission recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises," Official Journal of the European Union, L, vol. 124, pp. 36-41.

[48] SBA (2009). Federal Small Business Programs and the Small Business Act of 1953. Available: http://www.sba.gov/.

[49] D. J. Storey and F. J. Greene (2010), Small Business and Entrepreneurship. Essex: Pearson.

[50] E. M. Rogers (2011), Diffusion of innovations: Simon and Schuster.

[51] L. G. Tornatzky and K. J. Klein (1982), "Innovation characteristics and innovation adoption-implementation: a meta-analysis of findings," Characteristics of innovations, their acceptance and execution: An interim analysis of the results found, p. 28.

[52] G. Premkumar and M. Roberts (2000), "Adoption of new information technologies in rural small businesses," Omega, vol. 27, pp. 467-484.

[53] D. J. Storey and F. J. Greene (2010), Small Business and Entrepreneurship. Essex: Pearson.

[54] Personal Data Protection Commission website <https://www.pdc.gov.sg/legislation-and-guidelines/overview> (accessed 3 August 2020).

[55] United Nations. (2018). World Economic and Social Survey [Online]. Retrieved December 3rd from https://www.un.org/development/desa/dpad/document_gem/wess-report/ .

[56] Martin, K.(2018). Ethical Implications and Accountability of Algorithms. Journal of Business Ethics. Retrieved December 2nd from http://dx.doi.org.

[57] Campolo, A., Sanfilippo, M., Whittaker, M., Crawford, K. (2017). AI Now 2017 Report. [Online] AI Now Institute, 3(4): 398–404.