

A DATA GOVERNANCE MATURITY ASSESSMENT: A CASE STUDY OF SAUDI ARABIA

Hmood Al-Dossari and Aisha Ali Sumaili

Department of Information Systems, King Saud University, Riyadh, Saudi Arabia

ABSTRACT

Nowadays, data has become important and influences the decision-making process on government and business sectors. Data governance strategy should not be underestimated because it increases the value of data and minimize data-related cost and risk. The data governance concept promotes the accomplishment of organizational objectives by developing and implementing an appropriate strategy for processing data in perfect and secure manner. This study aims to assess the maturity of data governance for Saudi sectors by design a framework and using it to measure whether the data governance have been applied or not. To do so, we have designed a questionnaire based on five criteria for assessing the current state of data governance implementation which are: policies and standards of data management, data quality, risk of poor data quality, cost of data correction, and data security. The questionnaire was then distributed to the employees in the IT department or who are related to data management or data security in Saudi sectors either government or private. The results show that approximately 48% of the respondents stated that they have a data governance committee in the sectors in which they work. Also, 55% of the respondents indicated that there are legislation and regulations for data governance in the sectors, as well as for making data available. Moreover, 42% from the respondents stated that their organizations have policies and procedures to enforce data management.

KEYWORDS

Data Governance, Maturity Assessment, Saudi Arabia

1. INTRODUCTION

Nowadays many organizations realize the importance of data and its great impact in improving their performance, increase data quality, and enhance competitiveness. The growing of digital data inside and outside the organizations increases possibilities of data availability and accessibility. Therefore, the organizations must be aware of the need for the right use of their data and to protect from unauthorized access. So, it is important for all organizations either public or private which process data especially personal data to have clear policies and laws to deal with their data in an appropriate way by conducting and adhering to data governance [1]. To design and apply the data governance effectively, these five data domains must be considered by the organizations: data principles, data access, data quality, metadata, and data lifecycle [2]. In this study, we assess the maturity of data governance for Saudi sectors by design a framework and using it to measure whether the data governance have been applied or not. To do so, we have designed a questionnaire based on five criteria for assessing the current state of data governance implementation which are: policies and standards of data management, data quality, risk of poor data quality, cost of data correction, and data security.

The remaining of this study is organized in the following structure, Section 2 provides the background of the research then Section 3 illustrate the main related works and studies. Section 4

presents several maturity assessment models and the proposed methodology introduced in Section 5. Experimental studies and results are shown in Section 6. The limitations of the study are highlighted in Section 5 and conclusion is provided in Section 7.

2. BACKGROUND

Data governance is the comprehensive management of usability, availability, security, and quality of data both inside and outside the organization. It includes establishing policies, standards, processes and frameworks to ensure the right use and effective data protection [3].

The word governance means 'to steer', governance in the generic meaning is the practice of control and authority and it is related to laws and regulations. The data in an organization must have rules and regulations governing it, it must be known who is authorized to access it and who is the owners of each dataset. There are many definitions of data governance, one of them was introduced by Ning Zhang in [1] that the data governance means the management of overall of data security, availability, integrity and usability in an organization. Rene Abraham et al., (2019) [6] considered data governance as the framework for managing data as a strategic organizational asset. DAMA international defined data governance in [5] as "the exercise of authority, control and shared decision-making (planning, monitoring, and enforcement) over the management of data assets". Olivia Benfeldt (2017) [7], defined data governance as the processes at the company level that define responsibilities and rights of decision-making that achieved organizational aims to enable data treatment in desirable behaviour as an asset in an organization. According to the Privacy Technical Assistance Center (PTAC) [8], data governance is the organizational approach to data management formalized as a set of procedures and policies that comprise the full data life cycle.

The American Health Information Association (AHIMA) explained data governance as follows: "data governance is a sub-part of information governance that develops for the assurance of data quality and execute the data needs planning in concert with the strategic information needs of the organization" [4]. The data governance involves data mapping, data modelling, data audit, data architecture, data quality controls, and data quality management [4]. The main reason for the emergence of data governance is the transformation of organizations from traditional paper transaction to electronic transactions. So, nowadays with the rapid transformation from paper to electronic transactions the data governance is a must (i.e. it is not optional).

Data governance and data management are often used interchangeably, however, there are entirely difference between them [2]. While data management is data logistics, data governance is the strategy of data. Data governance is a business strategy that is bigger and more comprehensive than data management. It is referring to what decisions related to the effective management of an organization must be made and it also defines the relation between organization users and data, which means who can access each dataset and why. On the other hand, data management is considered as an IT practice that aims to allow organizations to organize and control their data resources so that it is reliable, accessible and timely available. It is referring to describe the method that organization users actual handling with data, means where the users access data and how they are processing it [2]. The key data governance purpose is to implement effective management of data and ensuring the high quality of data.

3. RELATED WORK

There are many studies and researches that discussed the importance of data governance and its implementation in various sectors. In this section we will review several studies on assessment of data governance six countries.

3.2.1. A Case Study of South Korea

In 2018 Hee Yeong and June-Suh proposed a framework for big data governance to analyse the case of National Pension Service (NPS) in South Korea [9]. The NPS is the fourth largest pension organization in the world and it is forecasted to discover data of national pension by Korean government with its open policy that called 'Government 3.0'. The NPS vision is National Pension for a Successful Life to operate the services under two departments which are the department of information protection and the department of personal information protection to prevent cyber infringement.

The researchers realized that to develop an efficient data governance system, the system should be connected with IT governance, corporate governance, and information technology architecture. So, they proposed a framework for big data governance in order to create a new analysis based on data along with organization goals. The framework uses some strategies to achieve those goals such as: defining data responsibility, personal data protection and preserving quality of data. The data processing is controlled by the independent audit authority of the IT department and monitor procedures and guidelines that should be maintained in the system audit in that framework. The key goal of a big data governance framework as clarified by authors is to avoid failures in organization policies through the successful implementation of big data projects. The researchers in that paper examined the actual state of NPS in big data governance framework perspective and they noticed that it is difficult to study big data governance based on actual cases. Finally, they advised public sectors to provide big data services in order to improve the quality of life of an individual through considering a big data governance framework.

3.2.2. A Case Study of China

Another framework for big data governance was proposed in 2019 by QUAN LI et al. [10] for regional health information networks (RHINs) in China. The researchers focused on practical activities related to using big data in RHINs to resolve the main issues of data utilization and improve the quality of healthcare services.

The framework that proposed contains three domains and 12 elements. The domains are: drive domain that defines whether the framework can run, support domain that defines the fast of its run, and capability domain that defines how far it can run. The researchers presented the framework based on Chinese practice which provides worthy references for RHINs. They discussed the governance needs for each domain in their framework. For the driving domain, they recommended that the governance must include four areas: strategy planning governance, open transaction governance, industry support governance, and governance of laws and regulation. For the support domain, the researchers highlighted that the governance of the healthcare industry should involve resources planning of healthcare big data, privacy security protection, and standard system. For the capability domain, the governance of the healthcare life cycle must involve healthcare big data organization, collection, usage, storage, analysis, and process.

3.2.3. A Case Study of Bureau

In 2019 Dwitama Heryadi et al. [11] conducted a study to assess the maturity level of the IT Bureau within the Audit Board of State which has many tasks closely with data such as database administrator, database monitor, database consumer, etc. The researchers stated that the Audit Board can reach to some data governance maturity level of DAMA through adjust and reorganize their strategy based on the assessment. Their study used the Stanford maturity model which contains two main components: the first one is foundation components that measured data

governance and development of program resources with three subjects which are awareness, formalization, and metadata. The second component is project components which focus on measuring the effectiveness of data governance concepts with three subjects which are: data quality, stewardship, and master data.

Their results showed that the division leaders have got higher scores on most components dimensions except two: stewardship and master data components. They also found that the stewardship is still weak, and this sector requires important improvements for maturity level increment. Some of the improvements they suggested are determine the coordinative and operational side of stewardship, increasing the control of stewardship processes, and reform the data governance people's perspective on the role of the stewardship. Finally, they ended up by that the IT Bureau of Audit Board is at the level of 2.63 according to Stanford data governance maturity model.

3.2.4. A Case Study of USA

The status of data governance implementation across tier- one Universities in the United States was investigated by Cary and Hsa in 2018 [12]. They examine the status of data governance and its relationship with IT governance across 30 tier-one university websites. Their study was based on rankings from Carnegie Classification of Institutions of Higher Education which is a recognized ranking system to classify the higher education in the U.S.

The researchers followed a certain methodology by selecting 30 tier-one research universities based on their business structure and geographic location. Then, their websites were investigated to determine the information regarding data governance practice for that university and used as an indicator of their awareness of data governance. A data governance checklist was developed which contains seven main criteria that fit of higher education context. The seven criteria are: data quality, data access, data governance body, data stewardship, data security, metadata documentation, and business process integration.

The results showed that most universities have procedures of data access at their institution, three-quarters of the universities presented a data governance as a committee or council that enhance practice of data governance at their institution, and more than half of the universities implemented data governance principle in data quality and data stewardship at their institution. They concluded that most higher education institutions in the USA have awareness of data governance and in progress to formulate procedures and policies for their institutions.

3.2.5. A Case Study of Indonesia

Alivia Yulfitri in [13] proposed an operational model of data governance for a government agency X in Jakarta. The researcher used Data Management Body of Knowledge (DMBOK) framework [14] which has 10 functions of data management, but the researcher focused only on 3 functions: data governance, metadata management, and data quality.

The study assessed the organization, human resources, and business processes and concluded the results as follow: for organization, there is no unit for managing the activities of data governance. For business processes, there are no standards and policies of data governance although there are some activities for data management but are not formalized. For human resources, the employees did not carry out data governance activities in their works and only IT staff are performed these activities. As a conclusion, the researcher recommended the need for the development of the operational model of data governance to enhance the implementation of data governance in an efficient manner.

4. MATURITY ASSESSMENT MODELS

In this section, we will discuss three popular models that can be used for data governance maturity assessment.

4.1. Capability Maturity Model (CMM)

It is a methodology was developed by the Software Engineering Institute (SEI) in the mid-1980s. It provides a framework to analyse and assess the approaches and procedures that the organization is following from a data governance perspective. This model can be used to enhance organizations' processes and structures. The CMM model consists of five levels where each level has specific characteristics of processes that used to categorize the organization according to their data management ways [15].

Level 1: Initial - the processes at this level are characterized as undocumented, uncontrolled, disorganized, not measured, and ad hoc. The level 1 organizations don't have formalized procedures or strict rules for data management and their data exist in multiple files with various formats. Roughly 30% to 50% of organizations work on the level as mentioned in [15].

Level 2: Repeatable - repeated, determined to track schedule, cost and functionality are processes characteristics of this level and the processes of data management are described as reactive. The organizations at level 2 follow the governance program but they should to institutionalize that program and they will start data governance practice on specific data types to prepare a report for a business unit. Level 1 organizations can move to level 2 by restructuring data elements and they must start to involve best practices for data governance. Robert in [15] stated the 5% to 20% of organizations approximately operate at this level.

Level 3: Defined - at this level, the processes are defined as a consistent group of documents and defined standards. The organizations established a data governance program as a main component of the data usage, and they implemented that program in a regular manner to assess their data quality. The interaction among data governance and functions of project management and use appropriate tools to maintain data governance are essential to success organization at this level. The organizations that work on this level roughly 10% to 15% [15].

Level 4: managed - the processes are described as measured, controlled, and organized. At this level, the organization manages metadata, and this will help the data governance team to index and preserve metadata and they allow IT staffs to access the data that stored inside organizations. Organizations at level 4 often audits the data to measure the quality of production data and top management should be handling data as they handle other assets. Roughly 5% to 10% of organizations work at this level [15].

Level 5: Optimized - at this level, the processes concentrate on improving performance continually. The practices that developed from level 1 to level 4 are used by organizations to improve data quality and accessibility. The key goal of level 5 organizations is continuous improvement of data governance. A few organizations operate at this level, roughly less than 5% [15].

4.2. Gartner Maturity Model

Gartner was introduced at the end of 2018 by defining Enterprise Information Management (EIM) as an organizational adherence to secure and ensure enterprise information accuracy. The Gartner model consists of six stages of maturity [16].

Stage 0- Unaware: at this stage, there is no awareness of any data governance activities. There is no ownership, security, or any system defined for data in the organization. There is a lack of definition of standards for data gathering and management and strategic decisions are often made without enough information.

Stage 1 - Aware: at this stage, business and IT leaders start to understand and acknowledge the value of information and enterprise information management. There is awareness about issues regarding inconsistent information and importance of data quality.

Stage 2 - Reactive: at this stage, sharing of information takes place between the teams. Metrics are collected and used in order to focus on information and data protection. There are some policies for data management, however, the level of adherence to the information management is still low.

Stage 3 - Proactive: at this stage, information management system is accepted and adopted. The roles of governance become formalized and a part of every project. There is a compliance with procedures and policies of information management

Stage 4 - Managed: at this stage, enterprise information management standards and policies are well understood and implemented. The information assets are classified and the governance body is considered to determine best practices and solve cross-functional information issues.

Stage 5 - Effective: at this stage, the enterprise information management strategies are used to reduce risks and improving productivity. The organization has reached its goal in terms of information management. The EIM organization is organized very well and controls the processes of information management across the organization [17].

4.3. IBM Maturity Model

Introduced in 2007, this data governance model addresses a total of 11 domains involving data risk management, organizational structure and awareness, data quality management, data architecture and other domains [18]. This model consists of a total of five levels. The model assesses each one of the eleven data governance categories individually. This can help the organization to begin with the one that suits its needs and objectives. Most organizations when they begin data governance are at level 2 and only a few organizations can fulfil the level four and five.

Level (1) Initial: there is little to no awareness of the importance of data and there are no set standards for managing data. The estimated scheduled and budget of a data project is often overtaken.

Level (2) Managed: the importance of data and how it can benefit the organization is realized. Data starts to be viewed as an asset in the organization. However, there is a need for a set of data management tools and processes in place.

Level (3) Defined: data regulation and management guidelines are defined and integrated with the organization processes. There is a better use of technology to manage data and data management practices are widely implemented throughout the organization.

Level (4) Quantitatively managed: at this level, all the projects follow the data governance guidelines and principles. Data models are documented and made available throughout the organization. Measurable quality goals are set for each project and data process and maintenance. The performance of the business operations is continuously measured against the set goals.

Level (5) Optimized: at this level, data governance becomes an enterprise-wide effort that improves productivity and efficacy. Operations are more comfortable to navigate through and are streamlined.

5. METHODOLOGY

Motivated by the different maturity models discussed in the previous sections, we used a hybrid model from CMM and Gartner models to assess the maturity of Saudi organization in terms of data governance. The hybrid model includes the following maturity levels:

level 1- Unaware: there is no data governance, no formal principles or standards for data management and data is not secure. The processes at this level are characterized as undocumented and uncontrolled.

level 2 - Aware: there is awareness about issues regarding inconsistent information and increasing data quality and some awareness for data governance strategies. The processes of data management are characterized as reactive and determined to track schedule, cost, and functionality.

Level 3 - Defined: data governance is considered part of all projects that develop, and the roles of governance becomes formalized. The processes at this level are a consistent group of documents and defined standards and there is whole compliance with procedures and policies of information management.

Level 4 - Managed: the information is seen as critical; information assets are classified and the metrics for data management are determined. The processes at this level are described as measured, controlled, and organized. The governance body is put and determine best practices to resolve quality issues.

Level 5 - Optimized: information management is considered a competitive feature for enterprises. The processes at this level focus on continual improvement of data governance strategy and implement formalized procedures to make sure that information is accurate and consistent.

The flow chart in Figure 1 shows the sequential steps of the proposed methodology. The methodology begins with designing a questionnaire based on five criteria for assessing the current state of data governance implementation which are: policies and standards of data management, data quality, risk of poor data quality, cost of data correction, and data security. The questionnaire was distributed to a sample of IT employees who are working on a department responsible for data management or data security in Saudi private and public sectors.

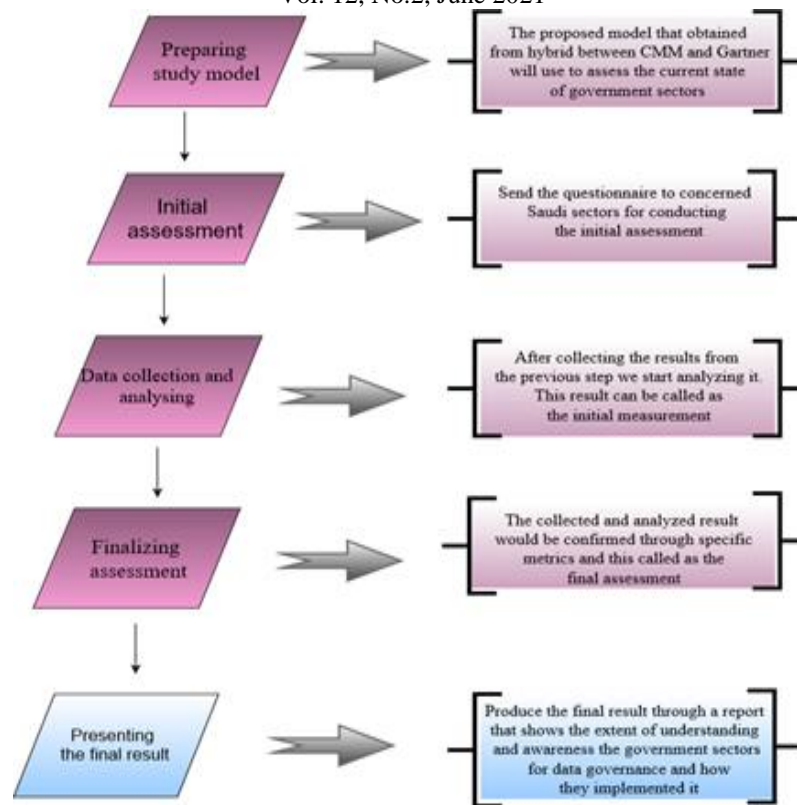


Fig. 1 Data Governance Assessment Methodology

The questionnaire consists of 49 questions organized in five sections each measures one of the five criteria mentioned above. For analysis purposes, we used the Statistical Package for the Social Sciences (SPSS) [20] as a data analysis technique to analyse the data collected by the questionnaire in order to determine the extent of applying data governance strategy in Saudi sectors and to test the proposed maturity model.

Data collection in this study was conducted through design an online questionnaire. The targeted participants were IT employees in private and government sectors in Saudi Arabia. The questionnaire has been created based on 5 criteria previously mentioned and implemented using Google Forms. It consists of 49 questions from which 39 are single-choice and 10 open questions. The participants of the survey were asked to answer group of questions by choosing which answer can describe the state of data governance in their organizations.

The designed questionnaire has 43 questions distributed into five sections. The first section has 11 questions about assessing the policies and procedures of Saudi sector to manage their data. The second section has 9 questions to measure the quality of data, while the third section aims to assess the risk of poor quality. The fifth section consists of 5 questions to measure the required cost and effort for data correction. The last section has 10 questions to assess data security. The questionnaire was distributed to various employees from private and public sectors in Saudi Arabia and 31 responses was received.

The participation may seem to some extent acceptable compared to total number of entities in Saudi Arabia. The other reason is that some sectors may not participate due to their lack of experience and knowledge of data governance.

6. EXPERIMENTAL STUDY

6.1. Characteristic of Dataset

The descriptive statistics are shown in Table 1. It represents the main characteristics of the participants in three dimensions: the sector they work in, type of the sector, and its domain. The participation from the government sector reached 77% from different sectors such as ministries, hospitals and universities. They work in different domains, education (6), health (2), commercial (5), civil (10), and others. To test the reliability, we calculate Cronbach's Alpha and obtained 0.896 indicating that the participants' responses were consistent and reliable.

Table 1: Margin specifications

	Characteristics	Frequency	Percent
Sector	Governmental	24	77.42%
	Private	7	22.58%
Type	Ministry	5	16.13%
	Agency	5	16.13%
	Authority	6	19.35%
	Public intuition	1	3.23%
	Hospital	1	3.23%
	University	4	12.9%
	Others	9	29.03%
	Domain	Education	6
Health		2	6.45%
Commercial		5	16.13%
Civil		10	32.26%
Other		8	25.81%

The questionnaire was based on five criteria as illustrated in Table 2. The mean for each criteria was calculated and shown in the table. According to the ranges of Likert scale [19]: (1.00 - 1.66) consider as bad, (1.67 - 2.33) represents good result, and (2.34 - 3.00) is very good. From the results on Table 2, it is obvious that the Saudi sectors have policies and procedures for data management as well as good data quality. The required cost for data correction is appropriate and the risks of poor quality is low. However, there might be some unusual issues with data security.

Table 2: Margin specifications

Criteria	Mean
1. Policies and standards of data management	1.8333
2. Data quality	1.8871
3. Risk of poor data quality	1.9000
4. Cost of data correction	1.8333
5. Data security	1.6607

6.2. Results

Based on the different statistical analysis that conducted on the respondents' answers we are reached to the following results. Although approximately 58% of the respondents as illustrated in

Figure 2 stated that either there is no or they are not sure of the existence of a data governance committee in the sectors in which they work, 55% of them indicated that there are legislation and regulations for data governance in the sectors (Figure 3), as well as for making data available (Figure 4). Moreover, as shown in Figure 5, 42% from the respondents stated that their organizations have policies and procedures to enforce data management.

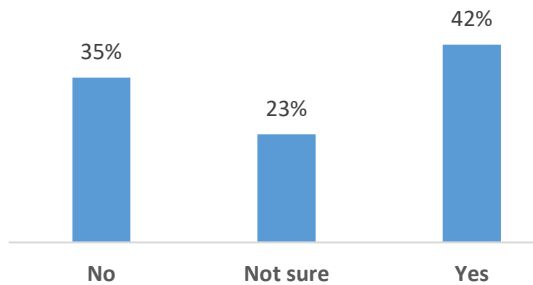


Fig. 2 Do you have a data governance committee?

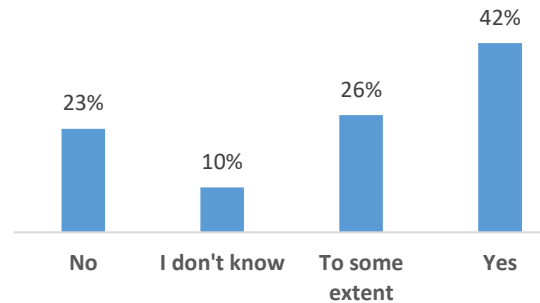


Fig. 5 Do you have policies and procedures to enforce data management?

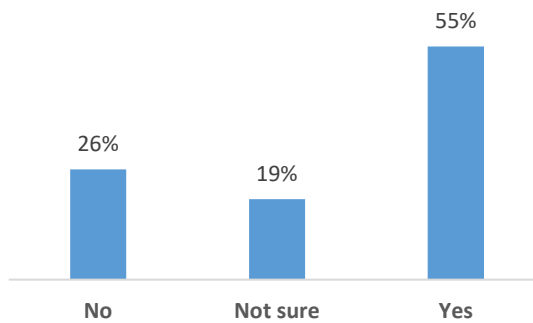


Fig. 3 Is there any regulations or legislation that governing the data?

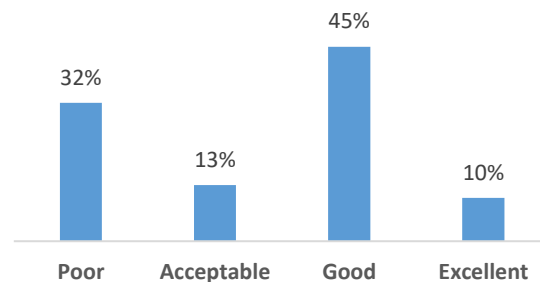


Fig. 6 How do you assess awareness level of data governance strategy in your organization?

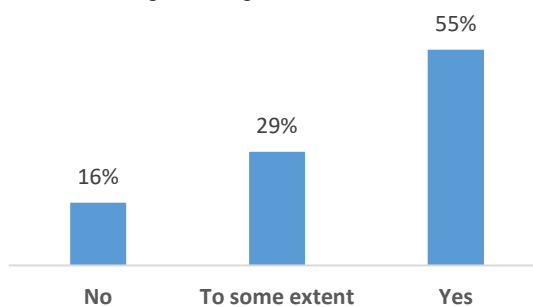


Fig. 4 Do you have clear policies for data openness?

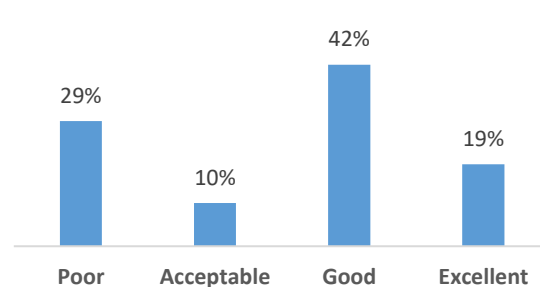


Fig. 7 To what extent your organization applying data governance strategy?

With regard to data governance strategy, 55% of respondents as shown in Figure 6 mentioned that the level of awareness on the importance of data governance is at least good. In terms of implementation, it is obvious that 61% of the participants (Figure 7) demonstrated that the organizations in which they work are applying a data governance strategy to control and manage their data.

7. DISCUSSION

In the data collection process, we have faced two main challenges. First, some participants are working on the same sector so it is hard for us to decide which answers are reflected the state of the sector. Second, it is hard for us to determine whether the participant is working on a department that is responsible for handling data in the sector or not.

This study might be enhanced by conducting direct observations and interviews with employees directly on site. Such data collection technique will help the researchers to collect more data with high confidence. That is, the duplication of responses will not occur and the researcher will target the suitable employees.

8. CONCLUSIONS

In this paper, we assessed the maturity of data governance in Saudi public sectors. To do so, we designed and developed a framework to measure whether the data governance strategies have been applied or not. Our results show that the Saudi public sectors have policies and procedures for data management. They also have good data quality and data correction cost is appropriate and acceptable. However, there may be some issues that relevant to data security. As a future work, we may consider expanding the scope of this study to cover other sectors in Saudi Arabia such as non-profit sectors.

ACKNOWLEDGEMENTS

The authors would like to express their cordial thanks to all participants in the questionnaire from Saudi public sectors.

REFERENCES

- [1] N. Zhang, Q. Yuan. "An overview of data governance", Economics Paper, December, 2016.
- [2] V. Khatri, C.V. Brown." Designing data governance". Communications of the ACM, 53, V1, pp. 148–152, 2010.
- [3] R. Mahanti, "Data Governance Implementation: Critical Success Factors," Software Quality Professional, Vol. 20 No. 4, pp. 4-21, 2018.
- [4] S. Bennett. "What is information governance and how does it differ from data governance?". Governance Directions, 69(8), 462–467, 2017.
- [5] R. Egelstaff and M. Well. "Data Governance Frameworks and Change Management," in Health Information Governance in a Digital Environment, 2013.
- [6] R. Abraham, J. Schneider, J. vom Brocke. "Data governance: A conceptual framework, structured review, and research agenda," In International Journal of Information Management 49, pp. 424-438, 2019.
- [7] O. B. Nielsen, "A Comprehensive Review of Data Governance Literature," 2017.
- [8] P. T. A. Center, "Data governance and stewardship," Vols. PTAC-IB-1, 2015.
- [9] H. Y. Kim and J.-S. Cho. "Data governance framework for big data implementation with NPS Case Analysis in Korea," Journal of Business and Retail Management Research (JBRMR), Vol. 12 Issue 3, 2018.
- [10] Q. L. LAN, N. ZENG , L. YOU, J. YIN, X. ZHOU and . Q. MENG, "A Framework for Big Data Governance to Advance RHINs: A Case Study of China," Vols. 7, 2019.
- [11] D. H. Kurniawan , Y. Ruldeviyani , M. R. Adrian , S. Handayani , M. R. Pohan and R. K. T , "Data Governance Maturity Assessment: A Case Study in IT Bureau of Audit Board," in 2019 International Conference on Information Management and Technology (ICIMTech), 2019.
- [12] C. K. Jim and H. C. Chang , "The current state of data governance in higher education," 2018.

- [13] A. Yulfitri. "Modeling operational model of data governance in government- case study: government agency X in Jakarta," in 2016 International Conference on Information Technology Systems and Innovation (ICITSI) Bandung - Bali, October 24 – 27, 2016.
- [14] DAMA. DAMA-DMBOK2, "Data Management Body of Knowledge," 2017.
- [15] R. S. Seiner, "A Data Governance Maturity Model," Vols. [online] Available from <http://tdan.com/a-data-governance-maturity-model/16702>, 2017.
- [16] D. Proença and . J. Borbinha, "Maturity Models for Data and Information Management A State of the Art," 2018
- [17] G. Firican, "Gartner Data governance maturity model" Vols. [online] Available from <https://www.lightsondata.com/data-governance-maturity-models-gartner/>, 2018.
- [18] G. Firican, "IBM Data governance maturity model," Vols. [online] Available from <https://www.lightsondata.com/data-governance-maturity-models-ibm>, 2018.
- [19] I. E. Allen, C. A. Seaman, "Likert scales and data analyses," Quality Progress, 40(7), 64-65, 2007.
- [20] N. H. Nie. "SPSS: Statistical package for the social sciences," McGraw-Hill New York, Vol. 227, 1975.