UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF COMPUTING AND INFORMATICS

An Assessment of Data Governance at Kenya Health Professionals Regulatory Authorities

BY

Victor Elijah Were

Supervisor

Mr. Christopher Moturi

October, 2016

Submitted in partial fulfillment of the requirements for the Degree of Master of Science in Information Technology Management of the University of Nairobi

DECLARATION

I declare that this project report is my original work except where due references are cited. To the best of my knowledge, this it has not been submitted for any other award in any University.

Signature: ________________________________ Date: ___________________

Victor Elijah Were P54/79244/2015

APPROVAL

This project report has been submitted in partial fulfillment of the requirements of the Master of Science Degree in Information Technology Management of the University of Nairobi with my approval as the University supervisor.

Signature: ________________________________ Date: ___________________

Mr. Christopher Moturi
School of Computing and Informatics
University of Nairobi

ACKNOWLEDGEMENT

First of all I would like to thank God for giving me strength and courage to carry out this study.

Special thanks to my supervisor Mr. Christopher Moturi and other lecturers. Thank you to all the members of staff of the health professional regulatory authorities in Kenya for their cooperation and generosity of time.

To my wife, Moureen, thank you supporting me during this period. You support was special.

To all my classmates, who from time to time were sources of inspiration. May the Almighty God Bless you all abundantly.

ABSTRACT

There is an increase in adoption of information communication technology in the health care by most of the developing countries. This has resulted in existence of most of their healthcare data in electronic form as a paradigm shift from the more traditional manual form. There is need to derive value from these data by making more evidence based decisions and reporting one version of the truth. Research shows that data governance initiatives help to improve data quality.

Health workforce data is critical in health care workforce planning and subsequent health services delivery in any country. The study focused on assessing data governance at the existing Kenya Health Professional Regulatory authorities and proposing a data governance model that can be used to establish a data governance program at the authorities. It further sought to determine the drivers and barriers of data governance at the authorities. The study used data governance decisions areas based on Khatri and Brown, 2010.Qualitative and quantitative research methods were used in this study to collect data.

The study results identified maintenance of quality of data; customer satisfaction; data security and control; operational efficiency as the drivers of the authorities to adopting formal data governance. Similarly, the authorities are faced with lack of data governance awareness; inadequate management ownership and support as well as limited funding and resource allocations as barrier to data governance. In addition, the study proposed that for the authorities to increase their data governance they need to identify their data as an asset; initiate more data quality management mechanism to increase data quality; restrict access of their data by strengthening their data controls. Furthermore, they need to create awareness; increase management ownership and support; allocate funding and resources to the initiatives. Metadata and data lifecycle were found not to be significant factors of data governance at the authorities.

The finding of this study can be used to establish a data governance program for health regulatory authorities in Kenya. In order to evaluate the impact of the model, there is need to implement the model and conduct a longitudinal study to determine the model impact.

Key Word:

Data governance, data quality, healthcare delivery, healthcare systems, health regulation

TABLE OF CONTENT

ABBREVIATIONS & ACRONYMS

1. CMMI -Capability Maturity Model Institute
2. EA - Enterprise architecture
3. HRH – Human Resources of Health
4. ICT – Information Communication Technology
5. IT - Information Technology
6. MDM – Master Data Management
7. WHO – World Health Organization

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - INTRODUCTION

1.1 Background of Study

Information Communication Technology (ICT) sector has emerged as a steadily growing contributor to the economy of Kenya. The sector has outperformed all others in the recent past. This has been attributed to the expansion in infrastructure, favorable government policy and an innovative private and public sector. Similarly, Kenyans are now involved in the use, creation and development of technology. Organizations have also embraced technology as a business driver or enabler. With the advancement of technology more organizational data is available in digital form.

Organizations have discovered that their success depends on the quality of their information. They rely on this data to make significant decisions that can affect core business. Organizations believe that data can be a valuable asset (Logan, 2012).This is because the data can be used to make better decisions (Otto, 2011). Effective data governance is required with organizations adopting a data-driven strategy. Research shows that formal data governance programs help in increasing data quality (Cheong & Chang, 2007).

Healthcare provision remains one of the main worldwide challenge and hindrance to human capital growth. There are several significant milestones in the struggle for health care provision which provide a platform for healthcare sector planning and development in the country. These form standard benchmark for which healthcare progress is assessed. Healthcare workforce is one of the core building blocks of any health system. Currently, there are eight established health professional regulatory authorities in Kenya which are mandated to regulate the training and practice of the various health professional cadres. These authorities rely heavily on technology to perform their mandate and have most of their data in electronic format. The health professional regulation data is important to inform healthcare workforce planning in the country.

There are numerous definitions of data governance. Master Data Management (MDM) institute (http://www.tcdii.com/) defines data governance as the formal composition of people, processes, and technology with the aim of enabling organization leverage its data as an enterprise. Similarly, Data Governance Institute (http://www.datagovernance.com/) states that data governance is an organization of decision, rights, and accountabilities for information related processes according to agreed-upon models. Furthermore, Fu et al, (2011) defines data governance as an agreed set of processes to improved consistency, accuracy, security of data while reducing the cost of management. In this study the researcher adopted definition of data governance based on Fu et al(2011).

1.2 Problem Statement

Healthcare workforce data is valuable for healthcare provision planning which influences national policy in the healthcare sector and in turn the overall health services delivery in the country. The value of this data is directly associated with the quality of the data and the ability to access and analyze such data to, find new patterns, new meanings, new data relationships, and new knowledge (Hovenga, 2013).Cheong and Chang (2007) found that maintaining the data quality of organizational data is not effective without a formal data governance program. Currently, the health regulatory authorities have their data in electronic format. The existence of this data necessitates the need for adopting a formal data governance program in order to derive value and improve its quality. Furthermore, data governance has been listed an area that requires strengthening in the One Monitoring and Evaluation Framework for the Health Sector in Kenya, 2016.

1.3 Research Objectives

The goal of the study was to determine the current status of data governance and develop a model that can be used to establish a formal data governance program for the health professional regulatory authorities.

Specific Objectives

The specific objectives of this study are outlined below:

1. To determine the status of data governance at the Kenya health professional regulatory authorities.
2. To identify the drivers and barriers of data governance at authorities.
3. To develop a data governance model for the authorities.

1.4 Research Questions

The study sort to answer the following research questions:

1. How is data governed by the Kenya Health professional regulatory authorities?
2. What are the main drivers of data governance at the Kenya health professional regulatory authorities?
3. What are the factors that hinder data governance at these authorities?
4. What data governance domains should be considered in a data governance program for the authorities?

1.5 Scope of the Study

The research was conducted at eight established health professional regulatory authorities in Kenya. The authorities regulate nurses, medical doctors, pharmacists, clinical officers, nutritionist, radiographers, laboratory technologists, and public health officers among other healthcare workforce cadres. Furthermore, the study also evaluated any mechanisms that currently exist which sought to address data governance at the board e.g. policies.

1.6 Research Significance

The findings of study contribute to an understanding of data governance at the health regulatory authorities and deduce the drivers and barriers for adopting a formal data governance program. The research further develops a model that the authorities can use to establish a formal data governance program. Lastly, it contributes to the body of knowledge in data governance for health professional regulatory data.

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to provide an outline of existing research in the area of data governance models. The existing literature on the topic of data governance models includes journals articles, and research papers, and material published on the Internet – which covers the general research themes as well as the specific topics forming part of the study. The literature review has been organized into a discussion of the overview of data governance; discussion of existing data governance models; discussion of barriers and drivers of data governance.

2.2 Overview of Data Governance

Research has shown that formal data governance programs help in increasing data quality (Cheong & Chang, 2007) and that there is a positive correlation between an organization commitment to governing data, and its capacity to get value out of data assets (Economist Intelligence Unit, 2008). In the technology perspective, data needs to be stored appropriately while on the business perspective it needs to be interpreted to deduce its meaning. Therefore, a close collaboration of the business and IT forms a core component of data governance (Cheong & Chang, 2007).

Despite data governance not being a new topic there a few scientific publications on this topic. Otto (2011) conducted a scientific literature review and found out that there are only 33 existing scientific journals or conference proceedings on data governance, and the first one was published in 2005.

It is also essential to determine the link between data governance and IT governance. While data governance has been derived from IT governance, they both share the similar layer on the enterprise architecture. Despite, the similarity they have distinguishable differences. Whereas IT governance deals with IT assets (that is: applications and infrastructure), data governance focuses on data assets in order to transfer it into information (Khatri & Brown, 2010). We can conclude that IT and data governance are interconnected but independent disciplines (Kooper et al. 2011).

2.3 Data Governance in Health Regulatory Authorities

Though the use of health data has a significant potential to facilitate research in order to improve the quality of healthcare, and reduce its costs. Policy issues should be addressed for its full potential to be realized (Hripcsak, et al, 2012).Skilled and adequate supported healthcare workforce is required for the delivery of quality health care services. The workforce goal as stated by the World Health Organization is to deploy the right health care workforce with the right skill to meet the needs that are required. A data-driven decision making is required to inform health sector planning. Health managers and policy makers require a robust understanding of workforce dynamics. This includes a better understanding of the workforce supply pipeline and skill-mix required at the facility level. This will result in health promotion, prevention of disease, and an improved quality care. Kenya must also ensure that its workforce is well regulated and comprised of licensed professionals that are strategically deployed and equitably distributed at each level of care. The Kenya Health Policy 2012-2030 calls for adequate health information for evidence-based decision-making. This policy forms the basis for the 2012 Kenya Health Information Policy, whose goal is to strengthen the generation, validation, dissemination, and use of health data.

2.4 Data Governance Models

There are several data governance model developed by scientists and industry leaders in data governance. Kenyan regulatory authorities have simple organization structures typically with no ICT departments or very few individuals in the ICT department. Therefore, important to consider a data governance model that is a simple and adaptable to the current environment. The model should consider the key domains that are important in governing the health regulatory data. The study reviewed existing scientific and industrial models.

2.4.1 Data Governance Contingency Approach Model - Weber et al (2009)

The author developed an extensive contingency approach model for data governance that can be used to design data governance for an organization. The model identifies three components: namely data quality roles, decision areas, and responsibilities which form a responsibility assignment matrix. It further defines the different roles each component plays in the exiting organization structure.

The author extensively studied data governance and proposed a model that incorporates the multidimensionality of data governance. By showing contingencies in data governance design, the model can help to interpret data governance. Although the model is supposed to help organizations in designing their data governance, the author provided no means to assess its governance.

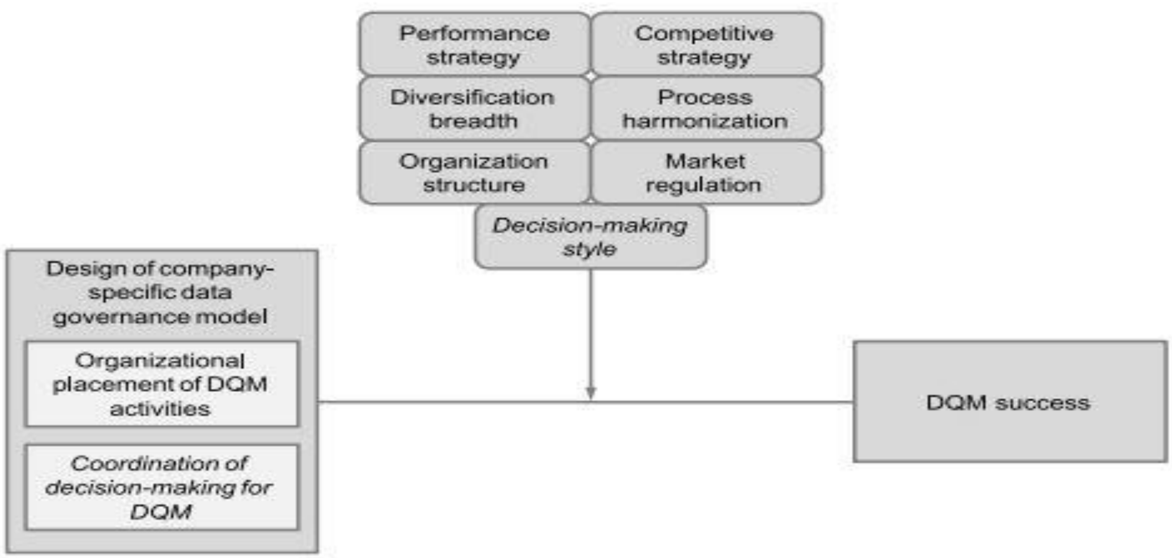


Figure 1: Contingency model for data governance, Weber et al. (2009)

2.4.2 Data Governance Structure Model - Cheong and Chang (2007)

The author identifies a data governance structure, based on a case study. The structure consists of several organizational bodies and their relation to each other. On an organization's strategic level, there should be a data governance council that is responsible for endorsing policies, aligning business and data initiatives, and reviewing budget submissions for data related projects. On the tactical level, data custodians and data stewards play a large role. On the lowest level, user group are involved. User groups consist of key data stakeholders from various divisions.

Cheong & Chang's model helps in understanding what data governance roles should operate on what organizational layer, but it does not provide a way to establish data governance. It is noteworthy that whereas. Weber et al (2009) promote a contingency approach to data governance, Cheong and Chang seem to promote that all organizations should adopt the same structures.

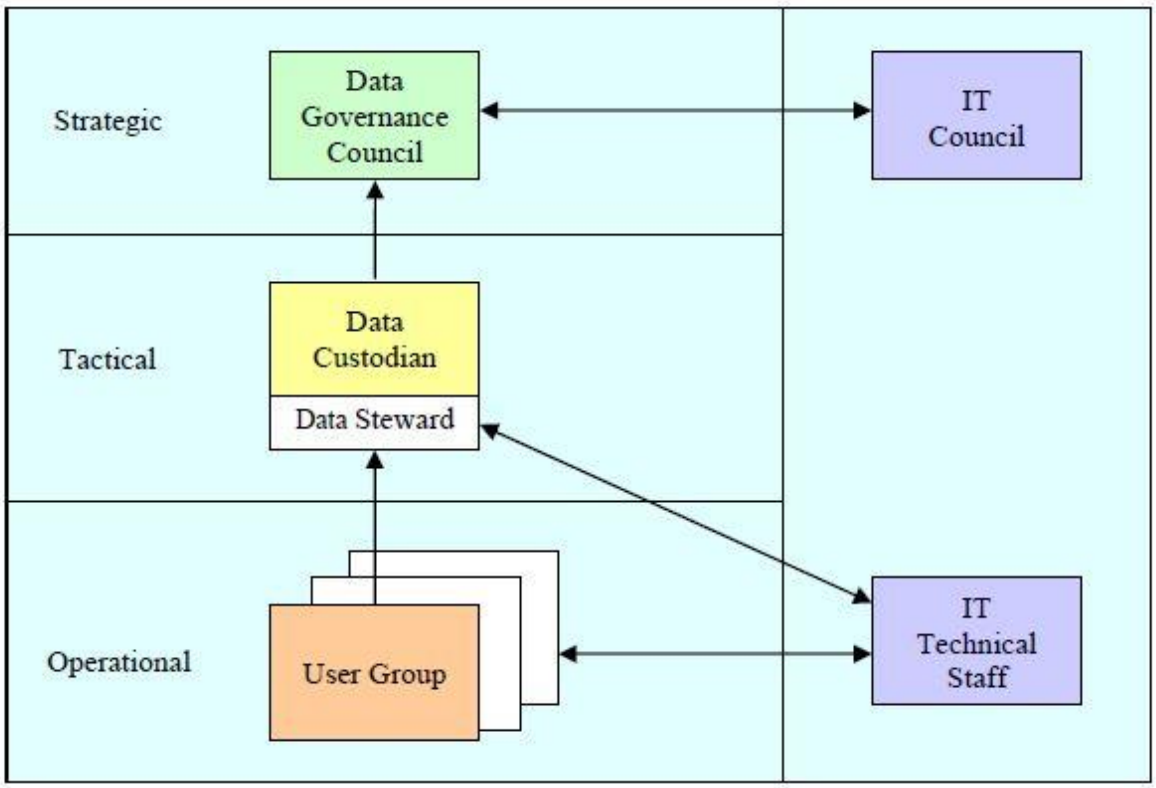


Figure 2: Data Governance Structure Model, Cheong and Chang, (2007)

2.4.3 IBM Data Governance Council Maturity Model – IBM (2007)

This model has been developed by IBM Council members through collaborative approaches. The members collaborated to form a recognizable means which organization can use to design their data governance programs. The model is based on Capability Maturity Model and provides a set of milestones that help organizations measure the way they govern its data. The model defines eleven data governance which are described in table 1 below. It further groups the domains as outcomes, enablers, core disciplines and supporting disciplines. The outcome requires the core disciplines, while the enablers support the core disciplines. Meanwhile, the supporting disciplines support the core disciplines. Like any CMMI-based maturity model, the model allows organizations to identify their current data governance maturity, determine their objectives and provide the activities that will move them to the next stage (Smith, 2015).

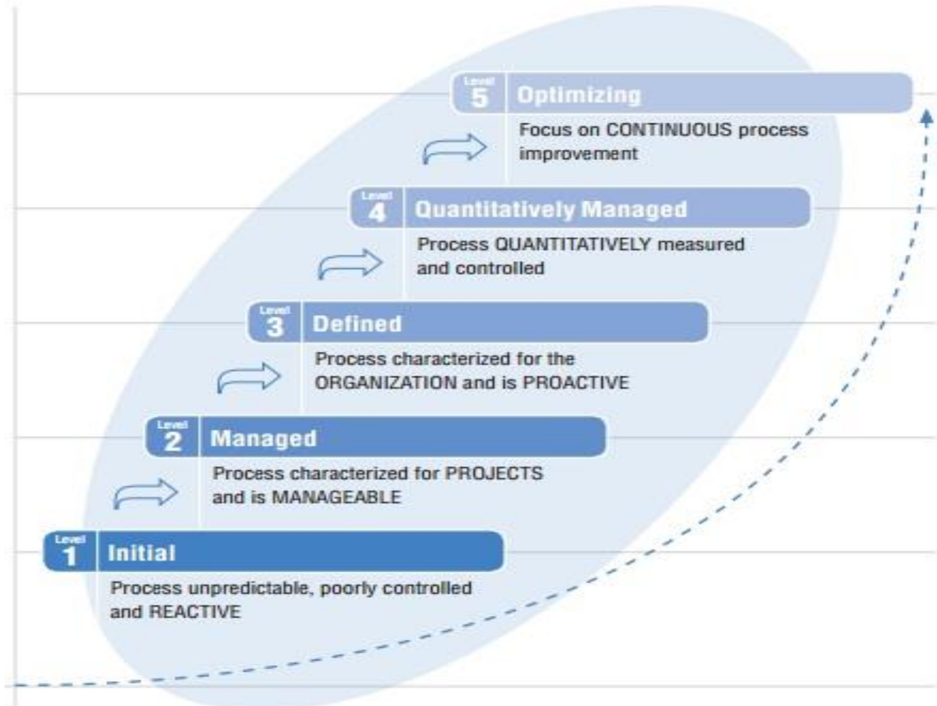


Figure 3: Data Governance Council Maturity Model, IBM (2007)

Table 1: IBM Data Governance Council Maturity Model 11 Domains

IBM Data Governance Council Maturity Model Domains	
Organizational Structures & Awareness	This domain recognizes the joint responsibility between IT and business as well as its role at the different levels of management
Stewardship	This domain offers quality control for the management of data.
Policy	This domain provides a formal documentation of the desired organizational behavior.
Value Creation	Describes how the data assets are measured to maximize its value.
Data Risk Management & Compliance	The methodology by which data risks are managed.
Information Security & Privacy	Describes the methods which an organization mitigates its data risks.
Data Architecture	Refers to the architectural design of the data systems its applications.
Data Quality Management	Describes the method which to measure data integrity
Classification & Metadata	Provides common semantic for data elements within an enterprises.
Information Lifecycle Management	Describes a formal approach to collection, use and destruction of data.
Audit Information, Logging & Reporting	The domain provides for means of measuring and evaluation of the data value, risks and efficacy of governance.

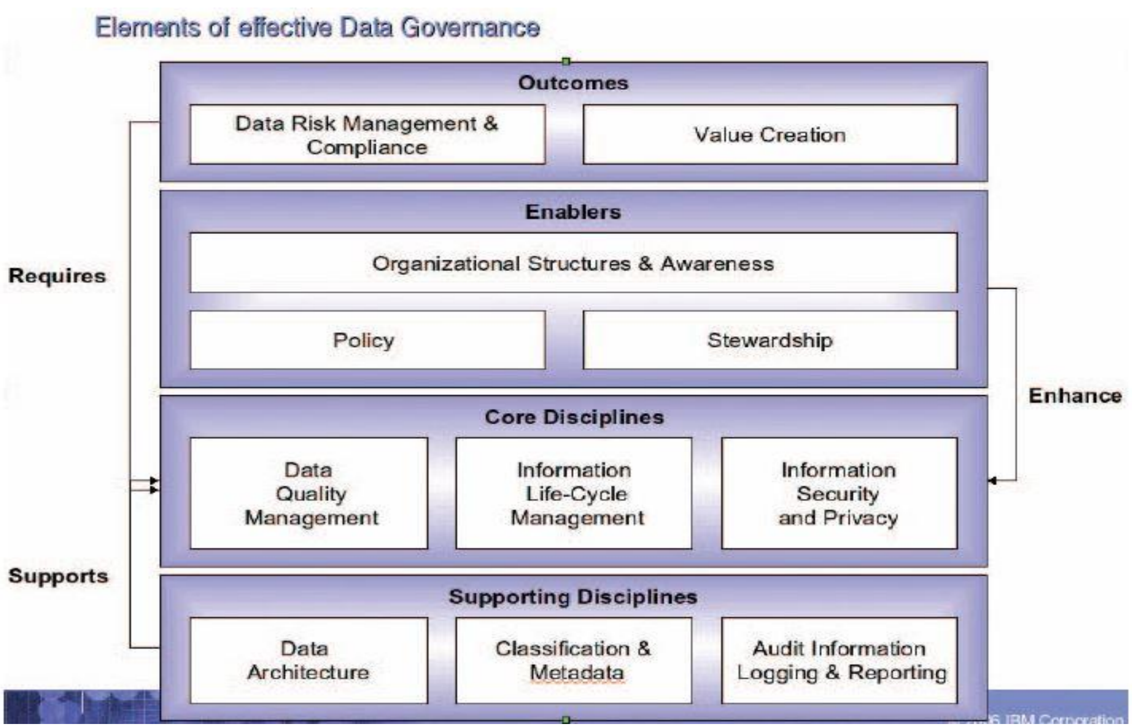


Figure 4: Data Governance Council Maturity Model, IBM (2007)

2.4.4 Kalido Data Governance Maturity Model - Chen (2010)

The author developed a model based on Capability Maturity Model a model, though unlike the CMMI it only has four stages. The four maturity stages include application-centric, enterprise repository-centric, policy-centric, and fully governed. These stages are based on the way organizations manage their data assets.

The application-centric stage defines a stage when organizations begin processing its data through systems which are developed to support the data transactions. In addition, some organizations govern data through data modeling. The second stage, enterprise repository-centric defines when organization starts to rely on data for decisions through data analysis. Hence lead to the organization thinking of data use on a broader perspective. The third stage, the policy centric stage is as a result of data being complex and large overtime. This leads to demand that requires different ways to manipulate by combination, manipulation, and storage. The final stage, fully governed stage when successful implementations of policy-centric data governance results in long term improvement of business performance. With time the scope of the data governance initiatives increase to cover all areas of data governance i.e. quality, security and lifecycle.

Despite, the elaborative nature of the model, it is rigid since it only gives indicators for organization, process and technology which are required to be aligned before moving to the next stage. Furthermore, it only looks at 3 domains i.e. organization, process and technology.

	Stage 1	Stage 2	Stage 3	Stage 4
Organization	Nothing	Silo'ed	Formed	Permanent
Process	Nothing	Informal	Defined	Optimized
Technology	Transaction	Data	Data Policy	Policy Driven
	Application-Centric	Enterprise Repository-Centric	Policy-Centric	Fully Governed

Figure 5: Kalido Data Governance Maturity Model, Chen (2010)

2.4.5 Data Governance Antecedents - Tallon etal (2013)

Tallon etal. (2013) studied data governance literature and conducted interviews with data professionals. Their final research model gives an overview of positive and negative antecedents for data governance, the composition of data governance, and positive and negative consequences of data governance on firm performance.

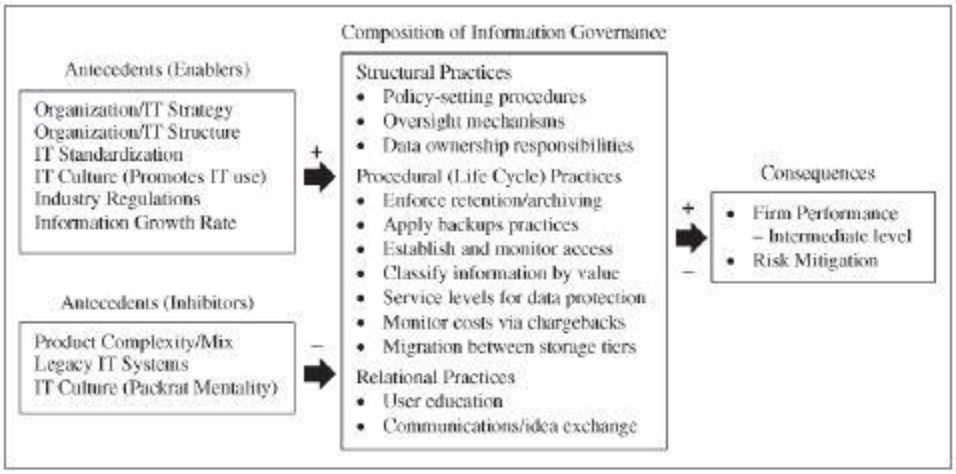


Figure 6: Data governance antecedents, Tallon etal (2013)

Whereas their research models give welcome insight into antecedents for, and composition and consequences of, data governance, again no way is provided to assess data governance.

2.4.6 Data Governance Decision Areas – Khatri & Brown (2010)

Khatri & Brown (2010) identifies five data governance decision areas i.e. data principles, data quality, metadata, data access, and data lifecycle. The decision areas have been derived from IT governance by Weill & Ross (2004).However, it is not explained how the derivation was done. The decision areas are interrelated but deal with a distinctive set of core issues. The author further points out that each decision area should be established along with identification of the decision makers. In addition, the author urges the need to establish data governance in close association with IT governance. The author only tested the model in a large insurance firm.

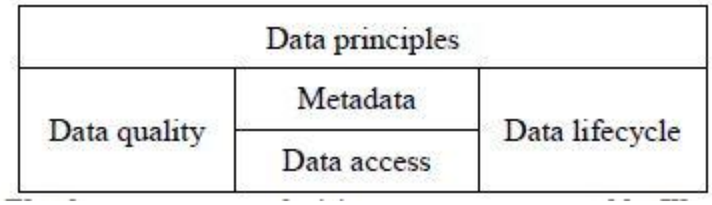


Figure 7: Data governance decision areas, Khatri and Brown (2010)

The model identifies critical domains that should be considered when developing a data governance program model for any organization.

2.5 Rationale for Choice of Approach

Table 2: Data Governance Models Comparison

Model	**Dimension**
Weber (2009) Data Governance Contingency Approach Model	The model is based on contingency approach. It identifies three components i.e. data quality
Cheong and Chang (2007) – Data Governance Structure Model	Cheong & Chang's model helps in understanding what data governance roles should operate on what organizational layer.
IBM (2007) - Data Governance Council Maturity Model	The model is based on Capability Maturity Model and provides a set of 11 domains to be considered.
Chen (2010) - Kalido Data Governance Maturity Model	The model is based on four maturity stages i.e. application-centric, enterprise repository-centric, policy-centric, and fully governed which are based on how organizations manage their data assets.
Khatri& Brown (2010) - Data governance decision areas	Identifies five data governance decision areas: data principles, data quality, metadata, data access, and data lifecycle. These decisions areas should be considered when designing an organization data governance program.
Tallon etal (2013) - Data governance antecedents	The model gives an overview of positive and negative antecedents for data governance, the composition of data governance, and positive and negative consequences of data governance on firm performance.

Capability maturity models have been reported to be expensive and difficult to apply for small and resource poor institutions, Duarte and Martins (2011). The models are time consuming and need excessive training for people to be in a position to implement resulting in lost time or focus on main areas of operation. The study therefore adopted Khartri and Brown (2010) who highlight the main decision areas that should be considered in designing data governance.

2.6 Data Governance Drivers & Barriers

Otto (2011) states the most common drivers of data governance are: is to guarantee compliance; allow decision-making; increase customer satisfaction; increases operational efficiency; achieve business integration; increase data quality.

Chalker (2014) states seven business drivers of data governance. Firstly, the author states that in order to maintain compliance, organization result in data quality and control procedures which indicate a need for improved data controls and accuracy. Secondly, in cases where there are fragmented approaches on a different process, organizations result in a need for centralized oversight control. The author furthers states that a need to increase operating effectiveness and reduce administrative costs may necessitate defining clear roles and responsibilities for data management with agreed measures and metrics to improve efficiencies and avoid errors. In addition, the author states that data quality efforts lack developed measures, tracking, and metrics which hinder quick and effective responses that address root causes rather than merely correcting errors. Other drivers of data governance include data error; data sources are not properly utilized to improve the efficiency of data origination and maintenance of data; difficulty meeting market demands for flexible, timely and relevant information and finally the inability to efficiently and accurately deploy data for external use.

Chalker (2014) further highlights several barriers to data governance which include lack of understanding of data governance. Most of the organizations do not have knowledge on what data governance means and what is entails. Organization also experiences lack of awareness on the governance mechanism with only a few of the individuals within an organization being aware of the data governance mechanisms. The author further, identifies lack of support, ownership, and limited resources being allocated by organization towards data governance programs. This is attributed to data governance being a lower priority in the organization.

2.7 Conceptual Model

A conceptual framework describes the relationship between different concepts that are applicable to the problem under investigation (Cavana, 2001). As stated in the previous chapter, the purpose of this study is to assess and develop a data governance program model for the Kenya health professional regulatory authorities. From the literature above, the study used Khatri and Brown (2010) to identify which data governance areas to be considered to design a data governca program model for the health regulatory authorities.

Khatri and Brown (2010) defined five data governance domains: data principles; data quality; metadata; data access; data lifecycle.

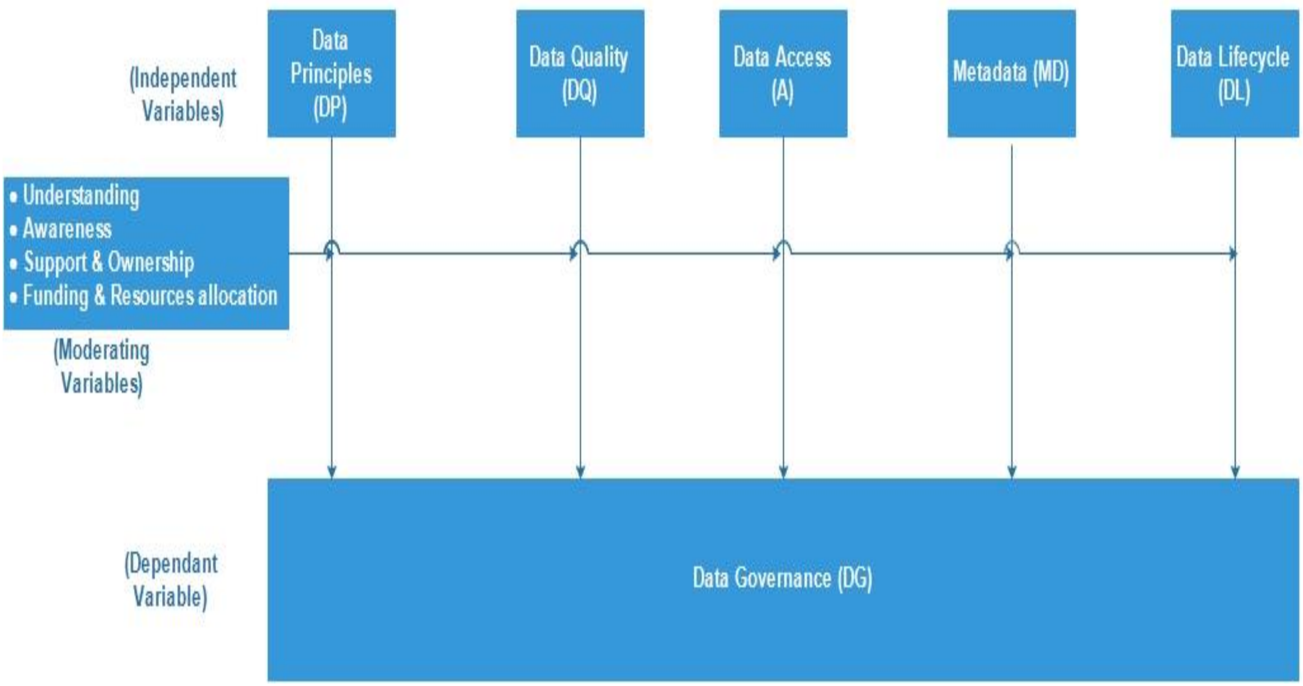


Figure 8: Conceptual Model, Khatri and Brown, 2010

2.8 Definition of Model Constructs

a) Data Principles

Khatri and Brown (2010) named their first areas as data principles. This sets the basis for the intended use of data. The domain sets the rules for the data as an enterprise wide asset and its appropriate policies and guidelines which are required. In addition, the domain considers data from external sources.

b) Data Quality

Wang et al (1996) argues that establishing and maintaining of data quality is critical in success of data governance. Data of high quality is critical in the ability to turn data into useful information. At least for the most important data sources, there should be a clear understanding of the level of data quality that is needed. To do so, a definition of what data quality means in these assets is advised. Such a definition will be different for each data asset. It opens the doors for being able to measure, monitor, and evaluate data quality. A data quality definition also allows for communicating these results to relevant stakeholders in both business and IT. Many software tools are available to support organizations in their data quality work. Data quality can be expressed using a variety of different attributes, such as accuracy, completeness, and relevance (Cheong & Chang, 2007). This makes data quality inherently multidimensional.

c) Metadata

In order to know desired levels of data quality, one should understand how it should be interpreted. Metadata provides descriptions about how data should be interpreted to use it as information. Metadata deals with data about data. Examples are when and by whom it was created, what another piece of data it may be based on, and what it means.

Just like data quality, metadata means something different in every different context (Duval et al. 2002). One may consider an appropriate level of data quality, which should have access to it, and when it should be deleted, also part of the metadata. Strictly speaking, this indeed is part of metadata, but for comprehensibility, we scope metadata to ‘interpretation’. Metadata can be a simple description in a document or technical in nature. Metadata helps in fostering a common understanding of the importance of data objects and facilitates communication about technical or business unit boundaries (Weber, 2009).

d) Data Access

Data access refers to which users should have access to certain data. It should be based on the definition of unacceptable uses of data within the organization, and compliance requirements for audibility, privacy, and availability (Khatri and Brown, 2010). Organizations can use international standards, such as ISO 27000 for information security, to derive their data access guidelines from (ISO 2012). Data is typically stored on media that are also physically accessible. Therefore, data access guidelines should also include physical data access (Khatri and Brown, 2010). Whereas many organizations struggle with rather vague and multidimensional concepts such as data quality and metadata, most organizations have specific security guidelines that include information security (PricewaterhouseCoopers, 2012).

e) Data Lifecycle

Managing data as a product with a lifecycle is one of the key principles of data quality management, which in turn inspired data governance. Just like a physical product, data typically goes through a number of stages. It is created, used, needs maintenance may be lost during an infrastructure crash, and will eventually need to be deleted or archived. Guidelines for data quality, metadata, and data access and data lifecycle should be specified to all these stages.

Guidelines for these stages play a key role in operationalizing the data principles into IT infrastructure, making data lifecycle the domain that has the closest connection to IT (Khatri and Brown, 2010).

CHAPTER 3 - METHODOLOGY

3.1 Introduction

This chapter outlines the methodology that was employed in carrying out the study. The goal of research methodology is to provide a systematic guide on how to ensure that the research is timely and consistent. Similarly, it seeks to ensure quality study results. The chapter covers a description of the research philosophy and design, target population, sample size, sampling procedure, data collection and data analysis procedures.

3.2 Research Philosophy

This refers to the belief about the way data about a phenomenon should be collected, examined and used. Broadly there are two major scientific research philosophies i.e. positivist also known as scientific and interpretivist (also known as anti-positivist) (Galliers, 1991).

a) Positivist Research Philosophy

This philosophy belongs to an epistemology which can be defined as a philosophy of knowing. In positivism studies, the researcher is free of the study and there is no room for personal interests within the study. Researchers warn that when one assumes a positivist approach to their study, then they are independent of their study hence their research is objective. Independent implies that one maintains minimal interaction with their research participants when carrying out your research (Wilson, 2010).

b) Interpretivist Research Philosophy

Livesey (2006) says that interpretivist methodology is specified by the collection of qualitative data and use of unstructured interviews as well as observations. Interpretivists believe that only through the subjective interpretation can the reality be fully understood. The main drawback associated with interpretivism relates to its subjective nature.

Discussion and Rationale for Choice of Approach

Different authors have recommended use of combination of both philosophies in order to achieve quality of the research (Kaplan and Duchon, 1988). The main aim was that the research to be undertaken should be both relevant to the research questions, as set out in Chapter One.

The main questions are to be answered by the research include; how data is governed by the Kenya Health professional regulatory authorities; the main drivers of data governance for the authorities; the critical factors that hinder data governance at the authorities; data governance domains to be considered when designing a data governance program at the authorities. To answer the questions adequately requires and objectivity approach hence the choice of positivist research philosophy.

3.3 Research Design

The research design guides the data collection and analysis. It provides the guide with which a research is conducted. Similarly, it provides the method for collection, measurement and analysis of data (Kothari, 2004).

In this study, the researcher employed both descriptive and explanatory research design. Descriptive research aims at describing characteristics of an individual or group (Kothari, 2004). Similarly, explanatory research identifies relationship between the factors to the research problem.

A descriptive research was selected in order to study the status of data governance at the Kenya Health Professional regulatory authorities. Meanwhile, the explanatory research design was used to determine the drivers and barriers to data governance at the authorities as well as domains that affect data governance are to be considered when establishing a data governance program at the authorities.

3.4 Target Population

A population refers to different elements that meet the minimum requirement to be included in the sample study. (Burns and Grove, 1993). The study population consisted of 65 employees of the health regulatory authorities whose work is in line with data initiatives at the authorities.

3.5 Sampling

Kumar & Phrommathed (2005) defines sampling as the procedure of choosing a few elements from bigger population elements. This becomes a basis of estimating the characteristic of the bigger population elements.

The study utilized a purposeful random sampling with the employees of the eight Kenya health professional regulatory authorities. A purposeful random sampling aims at identifying a population which provided a way of them not having prior information about the research outcome. The aim is to achieve reliable and credible results.

The respondent included people with knowledge of data governance in these organizations. This includes IT Manager, IT officers, head of departments, data managers, data coordinators, IT security managers, key process owner's users.

The subjects included in the sample were picked to meet defined criteria. The employees met should have worked with the organization for at least 6 months and have used any of the existing information system or data for more than 3 months.

Sample Size

Pande P. etal (2000) argue that it is important to keep the sample size manageable without affecting the quality of the results. This enabled the researcher to get reliable and detailed information while minimizing on the cost of time, finances and human resources (Mugenda and Mugenda, 2003).Yamane (1967) provides a simplified formula to calculate sample sizes.

$$n = \frac{N}{1 + N(e)^2}$$

Where **n** is the sample size, **N** is the population size**,** and **e** is the level of precision.

By using Yamane's formula of sample size with an error of 5% and with a confidence coefficient of 95%, the calculation from a population of 65 came up with 55 as the sample size. To take care of non-respondents the researcher raised the sample size to 64.

3.6 Data Collection

Data collection refers to the process by which the research collects information to answer the research question. Its forms a key component of the research study since inaccurate data collection can lead to inaccurate study results. In collecting the data the research considered data to be collected, who and how the data was to be collected.

A questionnaire was chosen as data collection instrument. Data was collected with the aid of questionnaires to assess of data governance at the authorities. Questionnaires were decided upon because of their high response rate and the less time required for administering. They also offered an opportunity to be less bias if presented in a consistent manner.

Questionnaires were personally distributed to selected respondents to complete. The respondents were given three weeks in which to answer the questionnaires before the researcher collect them.

3.7 Data Preparation & Analysis

a) Data Preparation

Once the completed questionnaires were collected the following steps were performed to prepare the data for analysis:

- The questionnaires were checked to eliminating unacceptable questionnaires ie incomplete ones, instructions not followed, little variance and missing pages.
- Data editing was performed to correct illegible, incomplete, inconsistent and ambiguous answers.
- Data coding to assign alpha or numeric codes to answers that do not already have them so that statistical techniques can be applied.
- The data transcription to transfer the data to an electronic format or database.
- Data cleaning was performed to review data for consistencies. Inconsistencies were from faulty logic, out of range or extreme values.

b) Data Analysis & Hypothesis Testing

Descriptive analysis was used to measure the percentages, measures of central tendency (mean, mode, median) and measures of variability (range, Standard deviation, and variance). Linear regression was used to test the relationship between independent variables (data access, data principles, metadata, data quality and data lifecycle) and dependent variable (data governance) (Saunders etal, 2011). Cronbach’s Alpha was used for reliability which is the degree to which the measure of a construct is dependable (Bhattacherjee, 2012). Construct validity was conducted to measure the extent to which a measure effectively represents the underlying construct that it is supposed to measure (Bhattacherjee, 2012).

The following hypothesis was tested as shown in the figure below.

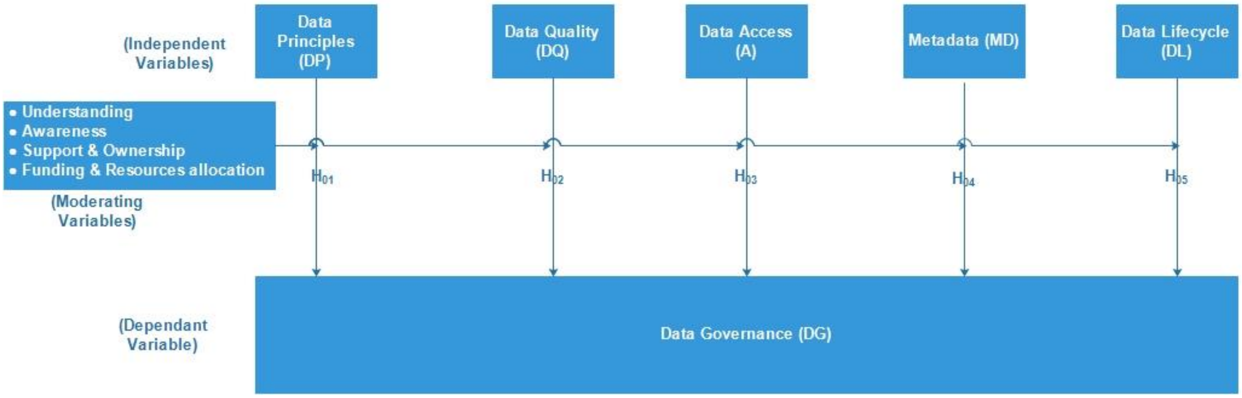


Figure 9: Operationalized Model through Hypothesis Testing

Source: Researcher

H_{01}: Data principles have a significant role in data governance at health professional regulatory authorities.

H_{02}: Quality of data has a significant role in data governance at health professional regulatory authorities.

H_{03}: Access of data has a significant role in data governance at health professional regulatory authorities.

H_{04}: Metadata information has a significant role in data governance at health professional regulatory authorities.

H_{05}: Data lifecycle has a significant role in data governance at health professional regulatory authorities.

3.8 Ethical Considerations

Res Honesty and integrity is required for any research process which is done to protect the rights of the study respondents. To term the study as ethical, an informed consent was sort before administration of the research tools to the respondents.

Burns & Grove (1993) defines informed consent as the potential subject's agreement to participate voluntarily in a study, which is reached after the subject understands the important information about the research study.

Respondents were informed about the purpose of the study and the methodology that would be used to collect the data. Similarly, they were assured on any risk that they might be subjected to in the event they are involved in the study. The respondents identity were kept anonymous and their response confidential throughout the study.

3.9 Limitations of the Methodology

Research quality is heavily dependent on the individual knowledge of the organization and the study could be easily influenced if they discuss the questions with other respondents of the study. To minimize this, the questionnaire was administered independently to the respondents. The study also used a purposive sampling to target only respondents with data governance knowledge at the organizations.

4 CHAPTER 4 –RESULTS & DISCUSSIONS

4.1 Introduction

This chapter presents the results and discussions of the results.

4.2 Response Rate

Field (2009) states that the sample size needed to achieve a certain level of power as the numbers of predictors vary depending on the level of effect and number of predictors. The study required a larger effect and had five predictors. A total of 64 questionnaires were sent out targeting at least 8 respondents per regulatory authority. In order to yield a high response rate the researcher did several follow-ups with the respondents. 61 questionnaires were returned which is 95.3 % response within four weeks with three follow-ups.

4.3 Demographic Characteristics

This section shows the results of the demographic characteristics of the respondents. Descriptive statistics was used to analyze the data and present the results.

4.3.1 Respondents Roles

The participants were coded according to their roles in the organization with regard to data. The roles include data user (DU), system administration (SA), system developer (SD), system champion (SC), data coordinator (DC) and project manager (PM). The figure 11 below shows that the top respondents were data users followed by system champions and system administrators. Therefore, most of the respondents understood the health regulation data and its difference facets.

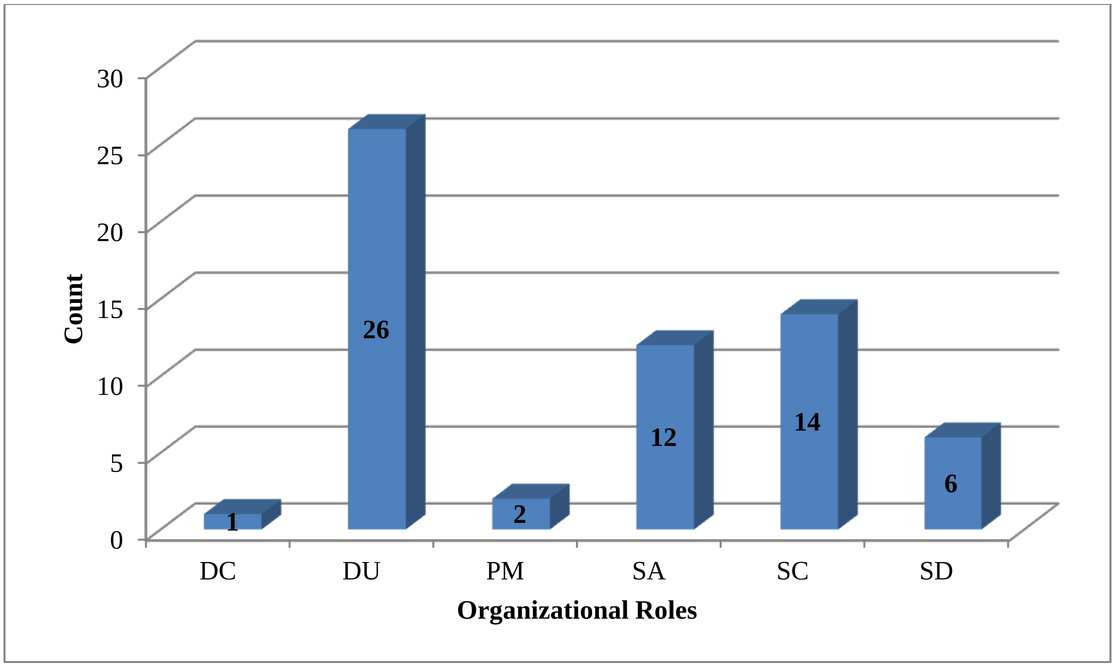


Figure 10: Respondents Roles

4.3.2 Years of Experience

Likewise, the years of experience in their current position indicates that respondents (23%) had worked for at least two years. The highest duration reported was 15 years by two respondents while the least duration 7 and 8 years reported by one respondent respectively. As a result, it is noted that most of the respondents had adequate experience to understand health regulation data.

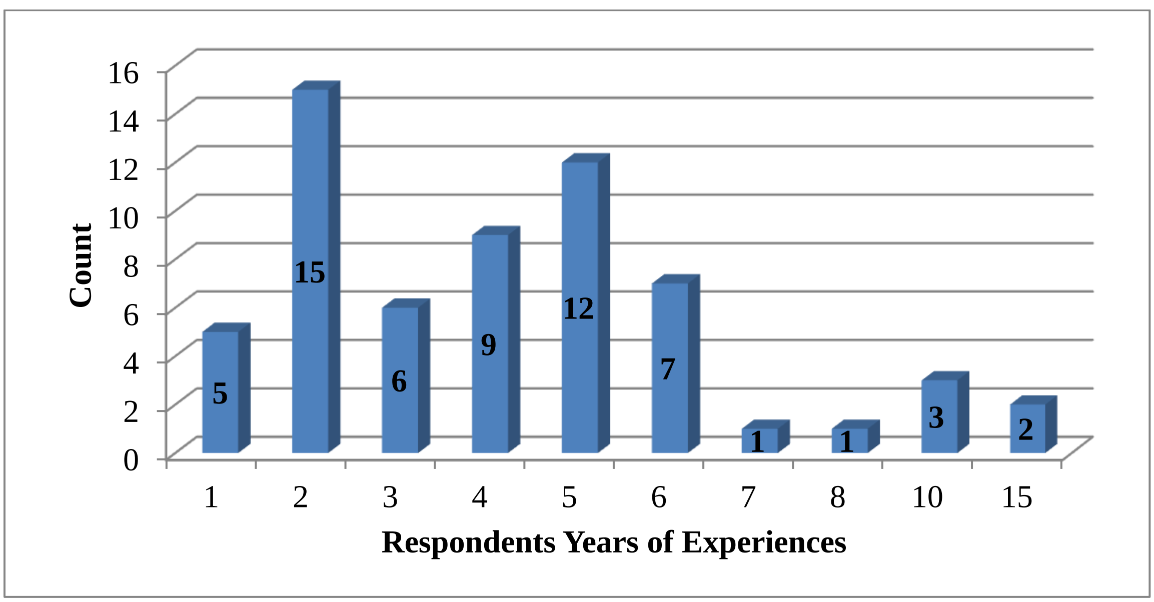


Figure 11: Participants years of experience

4.4 Reliability and Validity Analysis

4.4.1 Reliability

The study assessed whether instruments were reliable and valid to ensure the accuracy of the evaluation. Though research shows that reliability and validity being closely related, it should be noted that the two do not depend on each other (Nunnally and Bernstein, 1994). The study used Cronbach’s alpha to measure reliability. This is an index that provides a measure of the extent to which the data collected is consistent on a scale between 0 and 1. Internal consistency refers the extent to which all the items in a test represent the same construct and are therefore related to each other. Internal consistency is recommended to determine validity for a research study (Nunnally& Bernstein, 1994). On the other hand, reliability estimates the extent of measurement error in a test. The table4 below shows the alpha test of the research

Table 3: Cronbach's Alpha Index

Construct	Cronbach’s Alpha Index
Data Principles	0.804
Data Quality	0.704
Metadata	0.706
Data Access	0.705
Data lifecycle	0.853
Data Governance	0.800

As shown on table4 above the reliability coefficients for all the constructs were above the threshold (alpha<0.7).

4.4.2 Validity Analysis

To establish the convergent and discriminant validity of the test instrument, factor analysis was done. This is the statistical measure to analyze the interrelationship among a large number of variables and explain underlying dimensions. In factor analysis, if these items load together, it represents similar areas of concern.

Table 4: Correlation Matrix

Correlation Matrix[a]

		DP1	DP2	DP3	DP4	DQ5	DQ6	DQ7	DQ8	MD3	MD4	A5	A6	DL3	DL4	DG1	DG2
Correlation	DP1	1.000	.791	.813	.026	-.137	.000	-.033	-.214	-.177	-.067	-.089	-.236	-.022	.034	.349	.342
	DP2	.791	1.000	.874	.025	.000	.169	-.032	-.074	-.127	-.015	-.153	-.229	.113	.115	.390	.483
	DP3	.813	.874	1.000	.197	-.059	.130	-.013	-.124	-.216	.050	-.161	-.288	.100	.071	.500	.468
	DP4	.026	.025	.197	1.000	-.055	-.130	-.140	-.141	.046	.248	.170	.017	.268	.239	.154	.023
	DQ5	-.137	.000	-.059	-.055	1.000	.595	.079	.395	.060	-.024	-.076	.136	-.010	-.021	.327	.297
	DQ6	.000	.169	.130	-.130	.595	1.000	.285	.356	.077	.114	-.048	.057	-.131	-.185	.304	.341
	DQ7	-.033	-.032	-.013	-.140	.079	.285	1.000	.631	-.021	-.025	.364	.412	-.123	-.250	.087	.115
	DQ8	-.214	-.074	-.124	-.141	.395	.356	.631	1.000	-.023	-.058	.260	.610	-.060	-.229	.135	.195
	MD3	-.177	-.127	-.216	.046	.060	.077	-.021	-.023	1.000	.534	.088	.099	-.192	-.095	-.093	-.087
	MD4	-.067	-.015	.050	.248	-.024	.114	-.025	-.058	.534	1.000	.011	-.044	.070	.161	.083	.159
	A5	-.089	-.153	-.161	.170	-.076	-.048	.364	.260	.088	.011	1.000	.551	.037	.162	-.199	-.195
	A6	-.236	-.229	-.288	.017	.136	.057	.412	.610	.099	-.044	.551	1.000	-.125	-.015	-.101	-.116
	DL3	-.022	.113	.100	.268	-.010	-.131	-.123	-.060	-.192	.070	.037	-.125	1.000	.742	-.037	.076
	DL4	.034	.115	.071	.239	-.021	-.185	-.250	-.229	-.095	.161	.162	-.015	.742	1.000	-.066	-.025
	DG1	.349	.390	.500	.154	.327	.304	.087	.135	-.093	.083	-.199	-.101	-.037	-.066	1.000	.661
	DG2	.342	.483	.468	.023	.297	.341	.115	.195	-.087	.159	-.195	-.116	.076	-.025	.661	1.000
Sig. (1-tailed)	DP1		.000	.000	.423	.148	.500	.402	.050	.088	.306	.250	.035	.433	.399	.003	.004
	DP2	.000		.000	.425	.500	.099	.404	.288	.167	.455	.122	.039	.194	.191	.001	.000
	DP3	.000	.000		.065	.327	.162	.461	.172	.049	.351	.110	.013	.223	.296	.000	.000
	DP4	.423	.425	.065		.338	.161	.143	.141	.363	.028	.097	.449	.019	.033	.120	.432
	DQ5	.148	.500	.327	.338		.000	.274	.001	.323	.429	.282	.151	.469	.437	.005	.011
	DQ6	.500	.099	.162	.161	.000		.014	.003	.280	.192	.357	.332	.160	.078	.009	.004
	DQ7	.402	.404	.461	.143	.274	.014		.000	.438	.425	.002	.001	.174	.027	.255	.191
	DQ8	.050	.288	.172	.141	.001	.003	.000		.432	.329	.022	.000	.324	.039	.152	.068
	MD3	.088	.167	.049	.363	.323	.280	.438	.432		.000	.252	.227	.071	.235	.241	.253
	MD4	.306	.455	.351	.028	.429	.192	.425	.329	.000		.466	.369	.297	.110	.264	.113
	A5	.250	.122	.110	.097	.282	.357	.002	.022	.252	.466		.000	.390	.108	.064	.067
	A6	.035	.039	.013	.449	.151	.332	.001	.000	.227	.369	.000		.171	.455	.221	.188
	DL3	.433	.194	.223	.019	.469	.160	.174	.324	.071	.297	.390	.171		.000	.389	.281
	DL4	.399	.191	.296	.033	.437	.078	.027	.039	.235	.110	.108	.455	.000		.308	.424
	DG1	.003	.001	.000	.120	.005	.009	.255	.152	.241	.264	.064	.221	.389	.308		.000
	DG2	.004	.000	.000	.432	.011	.004	.191	.068	.253	.113	.067	.188	.281	.424	.000	

a. Determinant = .000

In our findings, we established the correlation of the items using the determinant, (It has to be greater than .00001, shown in table 4) the determinant here was .000 which is greater than 0.0001.

Table 5: KMO and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.611
Bartlett's Test of Sphericity	Approx. Chi-Square	477.493
	df	120
	Sig.	.000

On Kaiser-Meyer Olkin Measure we got a value of 0.611, any value above .6 and greater is said to be adequate. Meanwhile, a statistical significance of .000 was achieved compared to .001 or less which is acceptable.

4.5 Variable Analysis

4.5.1 Data Principles

87.5% of the organizations reported that both internal and external data sources have been identified as shown in table 7 below. Meanwhile, only 75% reported that the owners of the data sources have also been identified. In addition, all the regulatory authorities except one reported to rely on their data to make policy decisions.

Table 6: Summary of Regulatory Authorities Data Principles Status

	REG 1	REG 2	REG 3	REG 4	REG 5	REG 6	REG 7	REG 8	Summary
Data Sources Identified	I&E	I&E	I	I&E	I&E	I&E	I&E	I&E	Internal & external data sources –87.5% Internal data sources –12.5%
Data sources owners defined	Y	Y	N	Y	Y	N	Y	Y	Yes – 75% No – 25%
Data used to make policy decisions	Y	Y	N	Y	Y	Y	Y	Y	Yes – 87.5% No – 12.5%

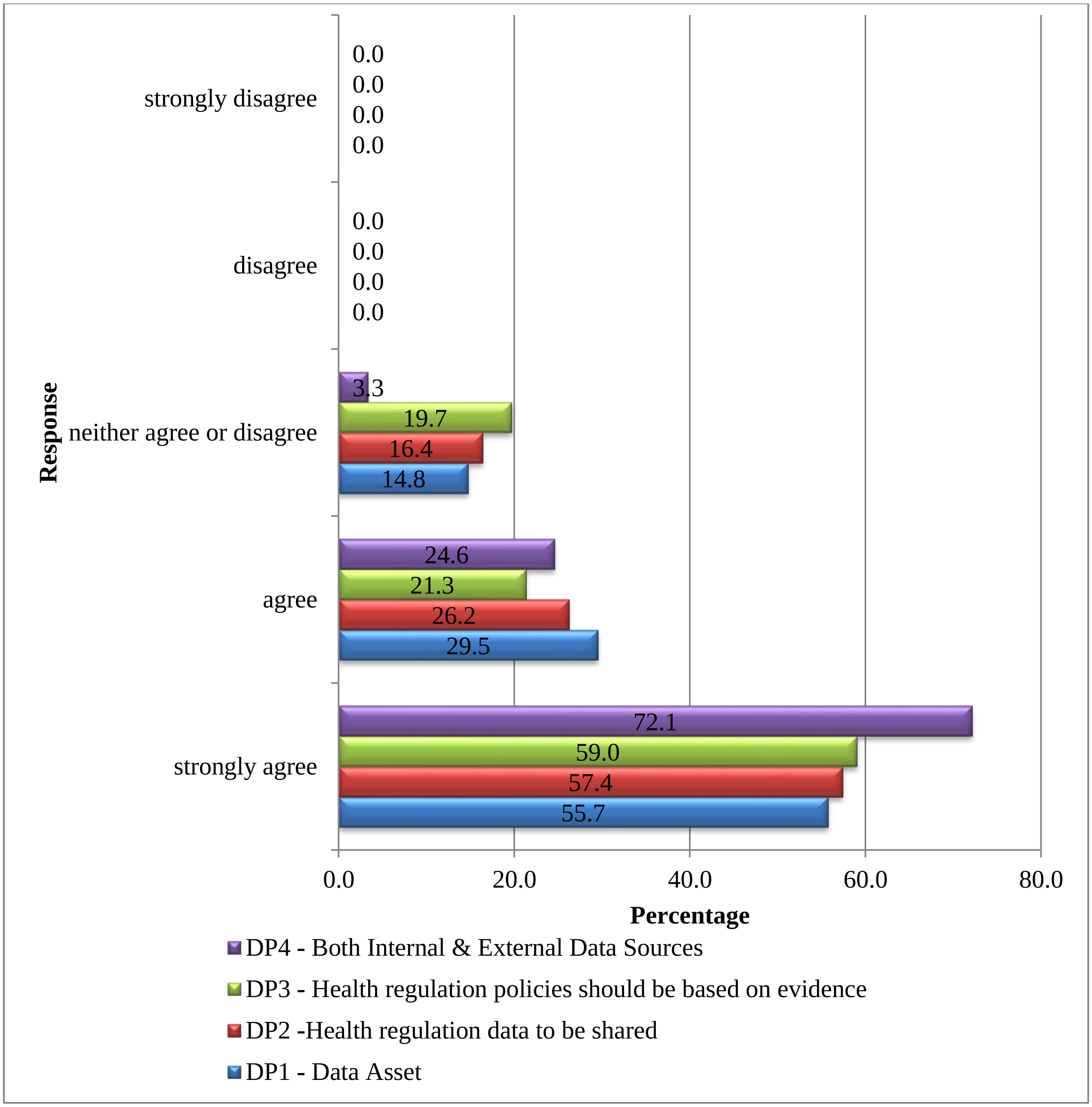


Figure 12: Data Sources Identification

Meanwhile, the majority of the respondents (96.7%) either strongly agreed or agreed that internal and external data sources should be identified in health regulation. A further 55.7% strongly agreed that health regulation data should be identified as an asset in health regulation. 80.3% of the respondents either agreed or strongly agreed that health regulation policies should be based on evidence. 83.6% either strongly agreed or agreed that there is need for the health regulation data to be shared by other health information systems.

4.5.2 Data Quality

50% of the organization reported having a formal data policy with a similar number reporting to have data stewards present. Table 8 shows that only the organizations that reported to have data policy present also reported to have data stewards. In addition, the six organizations reported that data cleaning is done as a day to day activity.

Table 7: Summary of Regulatory Authorities Data Quality Status

	REG 1	REG 2	REG 3	REG 4	REG 5	REG 6	REG 7	REG 8	Summary
Data Quality Policies Present	Y	N	N	Y	Y	N	Y	N	Y- 50% N- 50%
Frequency of data profiling	M	Q	Q	W	W	Q	W	M	Weekly – 37.5% Monthly – 25.0% Quarterly –37.5%
Frequency of data cleaning	D	D	D	D	D	W	D	W	Daily – 75.0% Weekly – 25.0%
Data Stewards Present	Y	N	N	Y	Y	N	Y	N	Yes – 50% No – 50%

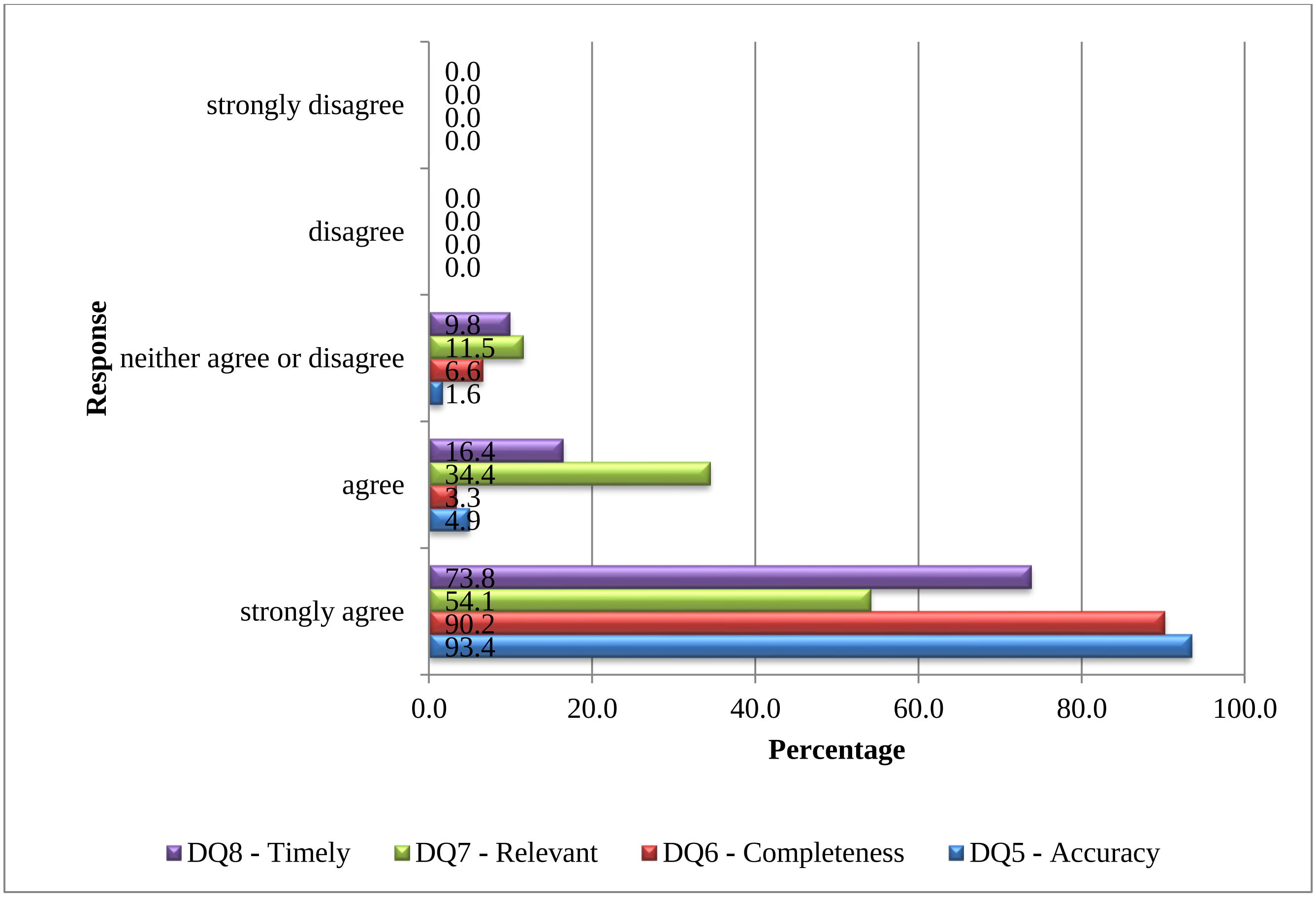


Figure 13: Health regulation data indicators

Similarly, most of the respondents (93.4%) strongly agreed that regulation data should be accurate while 90.2% strongly agreed that data should be complete. Furthermore, 54.1% respondents agreed that regulation data should be relevant. Lastly, 73.8% strongly agreed that regulation data should be timely.

Data accuracy, completeness, and relevance can only be achieved by constant monitoring of the current data and updating the data frequently. The result shows that despite most of the respondents strongly agreeing that health regulation data should be accurate, complete and relevant, there are still some health regulation authorities who perform data profiling on a monthly basis. This may result in the data not being accurate and current.

4.5.3 Metadata

On metadata, only two (25%) of the regulatory authorities reported that a formal data dictionary is present. Meanwhile, none of the authorities reported that the data dictionary is available for all to use.

Table 8: Summary of regulatory authorities' metadata status

	REG 1	REG 2	REG 3	REG 4	REG 5	REG 6	REG 7	REG 8	Summary
Formal data dictionary exist	N	N	N	N	Y	N	Y	N	Yes–25.0% No –75.0%
Data dictionary available for all	N	N	N	N	N	N	N	N	Yes – 0% No – 100%

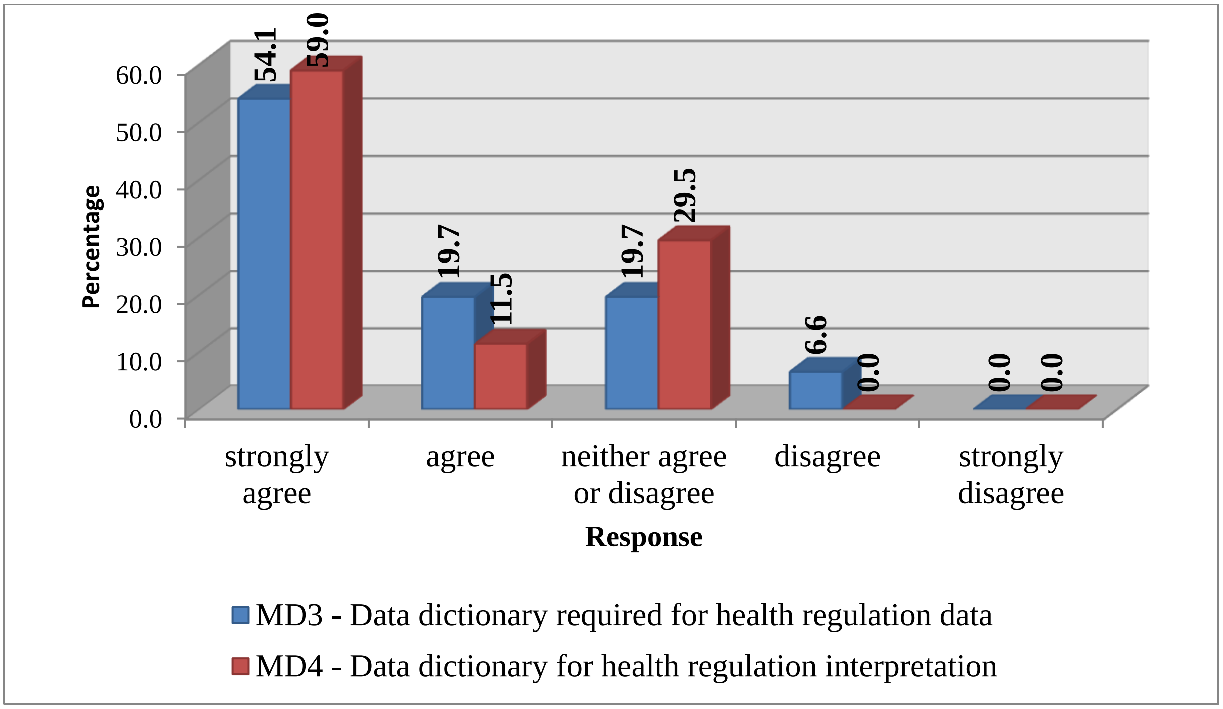


Figure 14: Health regulation metadata

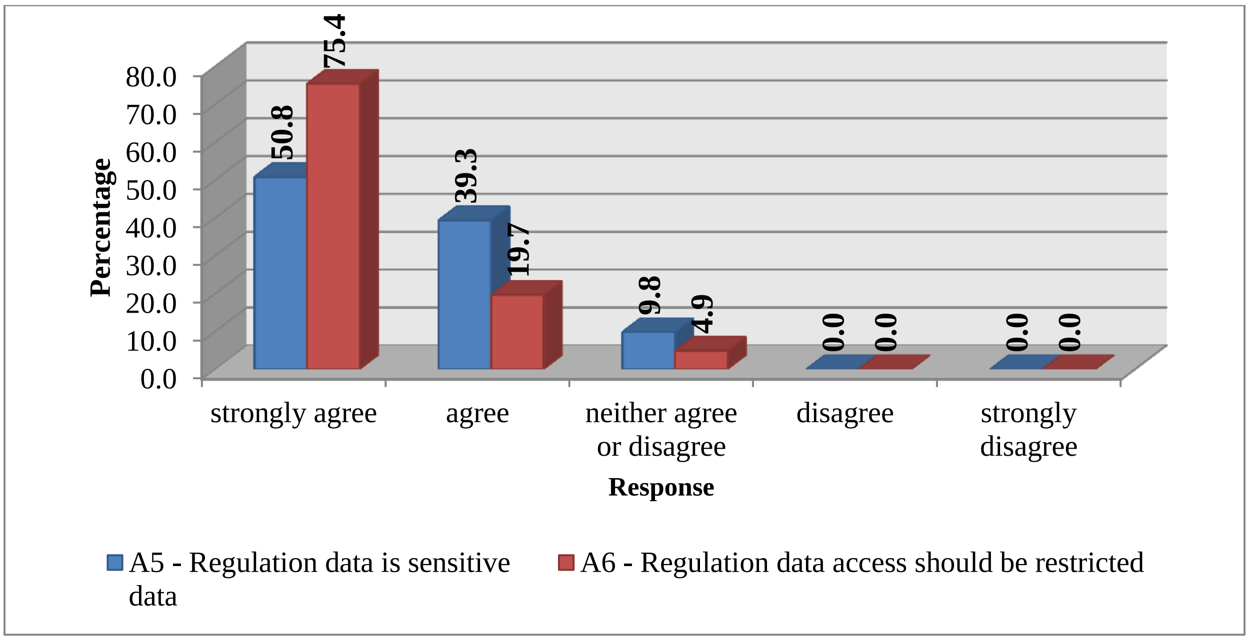
54.1% and 19.7% of the respondents strongly agreed or agreed respectively that data dictionary should be defined for all health regulation data elements. On the other hand, 59.0% and 11.5% strongly agreed or agreed respectively that a data dictionary is important for health regulation data interpretation.

4.5.4 Data Access

On data access, all the authorities reported to have tagged their sensitive data, restricted access and prevented unauthorized changes to their data. Meanwhile, three of the authorities reported that they don't scrutinize all data requested to protect personal information

Table 9: Summary of Regulatory Authorities Data Access Status

	REG 1	REG 2	REG 3	REG 4	REG 5	REG 6	REG 7	REG 8	Summary
Sensitive data tagged	Y	Y	Y	Y	Y	Y	Y	Y	Y- 100% N- 0%
Data access restricted	Y	Y	Y	Y	Y	Y	Y	Y	Y- 100% N- 0%
Unauthorized Database changes prevented	Y	Y	Y	Y	Y	Y	Y	Y	Y- 100% N- 0%
External data requested scrutinized	Y	N	N	Y	Y	N	Y	Y	Yes- 62.5% No – 37.5%



The data access domain results on the desired state showed that 50.8% strongly agreed that regulation health regulation data contains sensitive information while 75.4% strongly agreed that regulation data access should be restricted. In conclusion, 90.1% strongly agreed or agreed that health regulation data is sensitive with 95.1% strongly agreed or agreed that its access should be restricted.

The data further shows that data access is a strong requirement for health regulation data. Furthermore, most of the health regulation authorities reported having already data access control mechanism in place.

4.5.5 Data Lifecycle

None of the regulatory authority reported having a formal data retention policy. Only two of the regulatory authorities reported having ever discarded their data.

Table 10: Summary of Regulatory Authorities Data Lifecycle Status

	REG 1	REG 2	REG 3	REG 4	REG 5	REG 6	REG 7	REG 8	Summary
Data retention policy present	N	N	N	N	N	N	N	N	Yes–0.0% No –100.0%
Previous data discarded	Never	Never	Never	Never	More Than 10 years	Never	More Than 10 years	Never	Never – 75.0% More Than 10 years– 25.0%

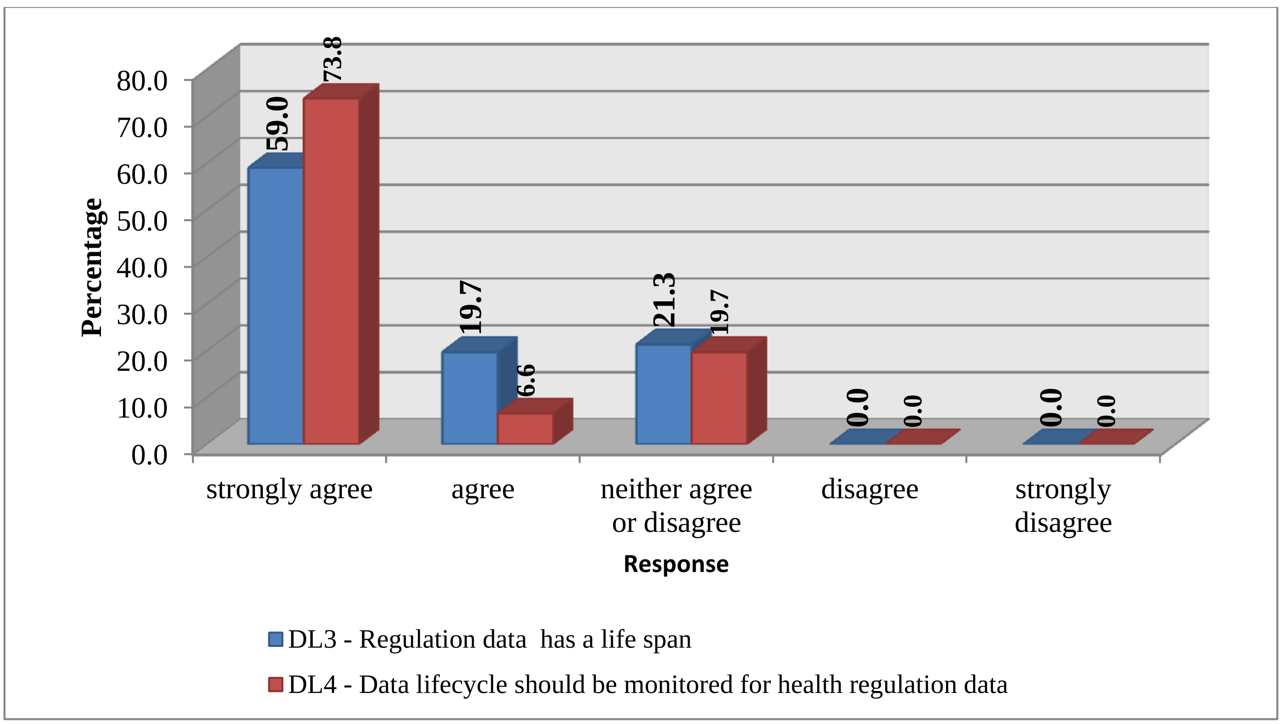


Figure 15: Data Lifecycle Graph

59.0% of the participant strongly agreed that health regulation data has a lifespan while 73.8% strongly agreed that health regulation data should lifecycle should be monitored throughout its lifecycle. Despite, all the regulatory authorities reporting that they don't have a data retention policy, most of the respondents agreed that there is aneed to monitor the lifecycle of the health regulation data from creation to disposal.

4.5.6 Data Governance

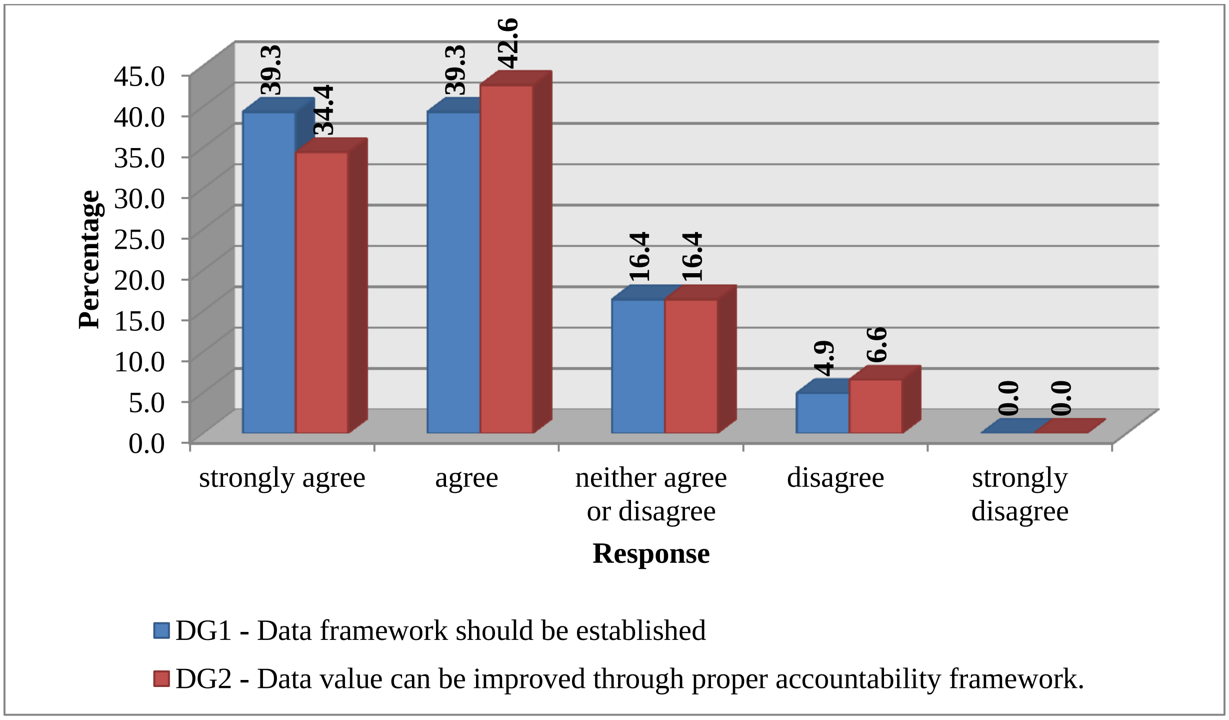


Figure 16: Data Governance Graph

On data governance, 78.6% participants either agreed or strongly agreed that health regulation data framework should be established while 77% of the participants either agreed or strongly agreed that health regulation data value can be improved through proper accountability framework.

4.6 Factors Affecting Data Governance

4.6.1 Drivers of Data Governance

The participants were required to agree on the drivers that would drive their organization to adopt formal data governance based on the existing possible drivers according to literature. The results are shown in table 12below. Most of the respondents strongly agreed that maintenance of data quality, customer satisfaction, and operational efficiency and to ensure data security were the most common drivers of their organization adopting data governance.

Table 11: Data Governance Drivers

	Description	Mean	Std. Deviation	Mode	N
D1	Ensure Compliance	3.44	1.232	5	61
D2	Maintain data quality	1.30	.527	1	61
D3	Achieve customer satisfaction	1.33	.507	1	61
D4	Ensure data security & control	1.59	.938	1	61
D5	Achieve operational efficiency	1.84	1.019	1	61
D6	Maintain competitive advantage	3.30	1.487	5	61

4.6.2 Barriers of Data Governance

The participants were required to identify the barriers that their organizations are facing to achieve formal data governance. The results are shown in table 13 below. Most of the respondents strongly agreed that lack of support & ownership and lack of resource were the most experienced barriers. In addition, lack of awareness on data governance was also agreed as a barrier.

Table 12: Data Governance Barriers

	Description	Mean	Std. Deviation	Mode	N
B1	Lack of data governance understanding	3.54	.787	4	61
B2	Lack of data governance awareness	2.08	1.215	1	61
B3	Lack of support & ownership	1.64	.775	1	61
B4	Lack of resource i.e. funding & human resources	1.64	.659	1	61

4.7 Hypothesis Testing

The table below shows the model summary and overall fit statistics. We found that the adjusted R^2 of our model is 0.460 with the $R^2 = .505$ that means that the linear regression explains 50.5% of the variance in the data.

Table 13: Model Summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.711[a]	.505	.460	.5725

a. Predictors: (Constant), Data Lifecycle, Metadata, Data Quality, Data Principles, Data Access

Table 14: ANOVA table

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.408	5	3.682	11.233	.000[b]
	Residual	18.026	55	.328		
	Total	36.434	60			

a. Dependent Variable: Data Governance

b. Predictors: (Constant), Data Lifecycle, Metadata, Data Quality, Data Principles, Data Access

Table 15: Coefficients[a]

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.443	.408		1.086	.282
	Data Principles	.730	.134	.530	5.436	.000
	Data Quality	.858	.196	.457	4.386	.000
	Metadata	-.001	.088	-.002	-.016	.987
	Data Access	-.474	.156	-.327	-3.036	.004
	Data Lifecycle	-.002	.094	-.002	-.021	.983

a. Dependent Variable: Data Governance

The table 15 above shows that data principles, data quality and data access are the factors that significantly impact data governance. Data access was the only negative predictor of data governance. While both data quality and data principles were positive predictors. Therefore as data quality increases data governance also increases meanwhile a decrease of access to data results to and increases to data governance. It also shows that an increase in data principles also results in an increase in data governance.

Below is a summary (Table 16) indicating acceptance/ rejection of the hypothesis regarding the main variables relating to the data governance program model for health regulation authorities. From the table, data principles, data quality and data access were accepted with a p-value of less than 0.0**5.**

H_{01}: Data principles have a significant role in data governance at health professional regulatory authorities.

H_{02}: Quality of data has a significant role in data governance at health professional regulatory authorities.

H_{03}: Access of data has a significant role in data governance at health professional regulatory authorities.

H_{04}: Metadata information has a significant role in data governance at health professional regulatory authorities.

H_{05}: Data lifecycle has a significant role in data governance at health professional regulatory authorities.

Table 16: Model Variable Test Summary

Hypothesis	Coefficient	t-statistic	p-value	Decision
H_{01}: Data principles have a significant role in data governance at health professional regulatory authorities.	**.530**	**5.436**	**.000**	**Accept**
H_{02}: Quality of data has a significant role in data governance at health professional regulatory authorities.	**.457**	**4.386**	**.000**	**Accept**
H_{03}: Access of data has a significant role in data governance at health professional regulatory authorities.	**-.327**	**-3.036**	**.004**	**Accept**
H_{04}: Metadata information has a significant role in data governance at health professional regulatory authorities.	-.002	-.016	.987	Reject
H_{05}: Data lifecycle has a significant role in data governance at health professional regulatory authorities.	-.002	-.021	.983	Reject

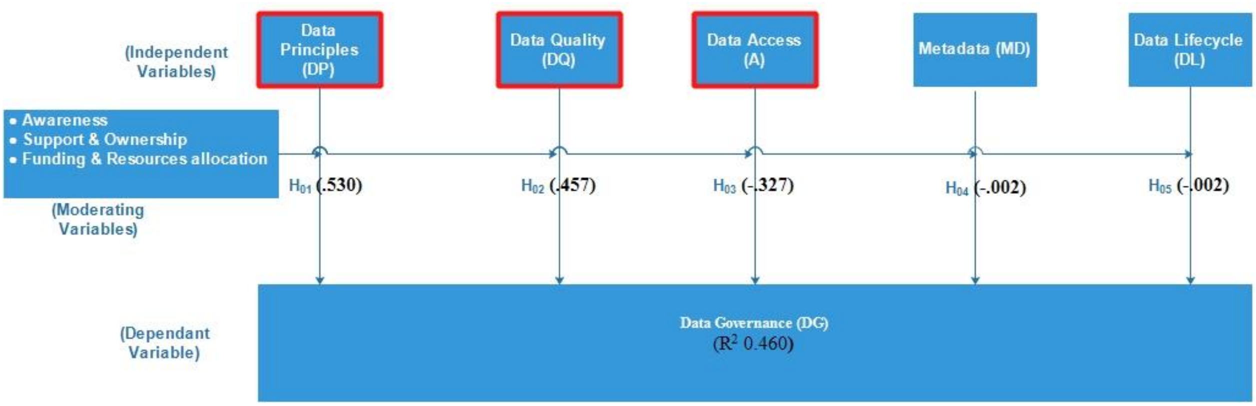


Figure 17: Summary of Hypothesis Test

5 CHAPTER 5 –CONCLUSION & RECOMMENDATION

5.1 Introduction

This chapter summarizes the findings of the study in relation to the objectives, literature review, and constructs. Suggestions for further areas of study are also captured as a way of filling the gaps identified in the study.

5.2 Study Achievement

The primary purpose of this study was to assess data governance status at the Kenya health regulatory authorities and propose a data governance program model for the authorities. Literature was synthesized and five variables namely data principles, data quality, metadata, data access and data lifecycle were identified. The identifiable relationships were captured through a proposed conceptual model in figure 8. This section relates the results presented in Chapter 4 to the research objectives, questions, and existing literature.

a) Research Objective 1 -Determine status of data governance at the Kenya health professional regulatory authorities.

Most of the health professional regulatory boards have identified both internal and external data source. This implies that health regulation data does not work in a silo and there is a need to consider also external data sources.

On data quality, some regulatory boards have formal data quality policy with formal data stewards. Data profiling frequency varies with some of the regulatory board performing it weekly while others monthly. Data cleaning is performed mostly as a daily activity. The difference in frequency in data profiling and data cleaning implies that some of the data issues are not identified immediately as soon as they occur so that they are cleaned.

On data access, the health regulatory authorities were performing well with all of them having tagged sensitive data; restricted access and implemented the mechanism of preventing authorized access to data. On the other hand, some of the authorities do n't scrutinize data request to prevent access to personal information.

Metadata was found to be new fold in the health regulatory authorities. Only two of the authorities were found to have a formal data dictionary and none have the data dictionary accessible to all to use. None of the regulatory authority was found to have a data retention policy. In addition, only two of the regulatory authorities had discarded data before.

b) Research Objective 2- Identify the drivers and barriers of data governance at authorities

The study results showed that there are four factors that would drive the health regulatory authorities into adopting formal data governance. The factors include: need to maintain the quality of data; achieve customer satisfaction; ensure data security and control; achieve operational efficiency.

Meanwhile, the study identified three factors that may hinder adoption of data governance by the health regulatory authorities. These factors include lack of awareness on data governance; lack of support and ownership; lack of resource i.e. funding and human resources.

c) Research Objective 3 - Develop a data governance program model for the authorities.

The study proposed a data governance program that would consider data principles, data quality and data access. It should also consider awareness, support, and ownership as well as resources as moderating factors. The regulatory authorities should sort to identify their data as an asset and improve its data quality. Data access should also be restricted. In addition, the authorities should build awareness, management support and ownership as well as allocate funding and required resources. This is shown in the figure below.

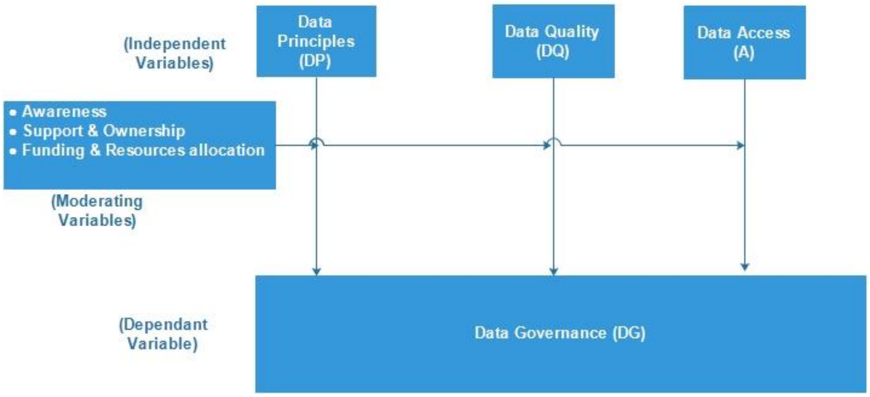


Figure 18: Data Governance Program Model for Health Regulatory Authorities in Kenya

5.3 Study Limitation

All studies have limitations which may vary from methodology, participants and to the procedure followed.

The first limitation of this study was from the participants. Some of the participants did not have abroad knowledge of some of the areas under scrutiny. The sample size of the participants was 61 which could be regarded as small. This is explained due to the lean staff that the health regulatory boards have.

5.4 Future Direction

As a future direction, there is a need to actually implement data governance model based on the proposed domains and evaluate its impact on data governance compared to the current status.

5.5 Conclusion & Recommendation

The study showed that metadata and data lifecycle had no significant effect on data governance at the Kenya health regulatory authorities. The authorities should focus on strengthening data principles, data quality and data access in order to strengthen their data governance. They should also focus on building awareness, ownership, and support as well as allocate resources in order which are the moderating factors of data governance at the organizations.

REFERENCES

1. Bennett, E. (2008). The future of enterprise information governance. *Economist Intelligence Unit*.
2. Burns, N., & Grove, S. K. (1993). Advanced statistical analyses. In *The practice of nursing research. Conduct, critique and utilisation*. WB Saunders Company Philadelphia.
3. Chen, W. (2010). Kalido Data Governance Maturity Model. Retrieved from www.gleanster.com/system/.../Kalido-Data-Governance-Maturity-Model-1-5.pdf
4. Cheong, L. K., & Chang, V. (2007). The need for data governance: a case study. *ACIS 2007 Proceedings*.
5. Clemmons, S., & Simon, S. J. (2001). Control and coordination in global ERP configuration. *Business Process Management Journal*, *7*(3), 205-215.
6. Collins, H. (2010). *Creative research: the theory and practice of research for the creative industries*. Ava Publishing.
7. Cooper, D. R., & Schindler, P. S. (2003). Business research methods.
8. Duarte, D. and Martins V. P. (2011). Towards a Maturity Model for Higher Education Institutions. A PhD project at University of Algarve Faro, Portugal available at ceur-ws.org/vol-731/05.pdf
9. Fu, X., Wojak, A., Neagu, D., Ridley, M. J., & Travis, K. (2011). Data governance in predictive toxicology: A review. J. Cheminformatics, 3, 24.
10. Galliers, R. D. (1991). Strategic information systems planning: myths, reality and guidelines for successful implementation. *European Journal of Information Systems*, *1*(1), 55-64.
11. Hovenga, E. J. (2013). Impact of data governance on a nation's healthcare system building blocks.
12. Hripcsak, G., Bloomrosen, M., FlatelyBrennan, P., Chute, C. G., Cimino, J., Detmer, D. E., & Keenan, G. M. (2014). Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. *Journal of the American Medical Informatics Association*, *21*(2), 204-211.

https://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf

13. Hüner, K. M., Ofner, M., & Otto, B. (2009, March). Towards a maturity model for corporate data quality management. In *Proceedings of the 2009 ACM symposium on Applied Computing* (pp. 231-238). ACM.
14. IBM, (2007). Data Governance Council Maturity Model: Retrieved from :
15. Kaplan, B., & Duchon, D. (1988). Combining qualitative and quantitative methods in information systems research: a case study. *MIS quarterly*, 571-586.
16. MOH (2016). One Monitoring and Evaluation Framework for the Health Sector in Kenya, 2016. Retrieved from http://www.healthdatacollaborative.org/fileadmin/uploads/hdc/Documents/Kenya_Health_Data_Collaborative_Conference_Report.pdf
17. MOH (2015). Data, System Governance and Change Management Framework. Division of Health Informatics, Monitoring and Evaluation, Kenya.
18. MOH. (2015). The Kenya Health Enterprise Architecture (KHEA): Version 1.0. Division of Health Informatics, Monitoring and Evaluation, Kenya.
19. Khatri, V., & Brown, C. V. (2010). Designing data governance.*Communications of the ACM*, *53*(1), 148-152.
20. Kooper, M. N., Maes, R., &Lindgreen, E. R. (2011). On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management*, *31*(3), 195-200.
21. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.
22. Kumar, S., &Phrommathed, P. (2005). *Research methodology* (pp. 43-50). Springer US.
23. Livesey, C. (2006). The relationship between positivism, interpretivism, and sociological research methods. *As Sociology for AQA*.
24. Logan, D., & Knox, R. E. (2010). Hype Cycle for Enterprise Information Management, 2010. *Gartner report*, (G00200503).
25. Matney, D., & Larson, D. (2004). The four components of BI governance.*Business Intelligence Journal*, *9*, 29-36.
26. Moturi, C., & Kahonge, V. W. (2015). Association Analysis of Pharmaceutical Imports in Kenya. International Journal of Computer Applications Technology and Research, 4(6).

27. Mugenda, O. M. &Mugenda AG (2003). *Research Methods, Qualitative and Quantitative Approaches*
28. Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability.*Psychometric theory*, *3*(1), 248-292.
29. Otto, B. (2011, June). Morphology of the organisation of data governance. In *ECIS* (Vol. 20, No. 1, p. 1).
30. Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises. *Information Systems Management*, *32*(3), 234-251.
31. Pande, P. S., Neuman, R. P., & Cavanagh, R. R. (2000). *The six sigma way*. McGraw-Hill,
32. Polit-O'Hara, D., & Beck, C. T. (2006). *Essentials of nursing research: Methods, appraisal, and utilization* (Vol. 1). Lippincott Williams & Wilkins.
33. Sekaran, U. (2006). *Research methods for business: A skill building approach*. John Wiley & Sons.
34. Selltiz, C., Wrightsman, L. S., Cook, S. W., Balch, G. I., Hofstetter, R., &Bickman, L. (1976). Research methods in social relations.
35. Thomas, R. L., Giles, S., Greenwalt, D., &Knecht, K. (2013). Enterprise Data in Healthcare.
36. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*,*12*(4), 5-33.
37. Watson, H. J., Fuller, C., &Ariyachandra, T. (2004). Data warehouse governance: best practices at Blue Cross and Blue Shield of North Carolina. *Decision Support Systems*, *38*(3), 435-450.
38. 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), 4.
39. Weber, R, The Rhetoric of Positivism Versus Interpretivism: A Personal View MIS Quarterly Vol. 28 No. 1, pp. iii-xii/March 2004
40. WHO. (2010). World Bank Economic Update 2010.
41. Yamane, T. (1967). Elementary sampling theory.

APPENDIX 1

STUDY QUESTIONINNAIRE

Health Professionals Regulatory Agencies Data Governance Assessment Questionnaire

Part A: Cover Letter

Dear Participant,

My name is Victor Elijah Were. I am a student at the University of Nairobi taking Masters of Information Technology Management. I am undertaking a study designed to assess data governance at the Kenya Health Regulatory Authorities. This study will enable development of assessment model for improving the data governance either in this and/or other institutions.

Data governance (DG) refers to the set of process that assist an organization improve it data consistency, accuracy, security while minimizing its cost. A comprehensive data governance program should include a governing body, procedure and plan on how the procedure will be executed.

Your participation is voluntary. You may choose to answer any question or discontinue participation at any point. There is no personal risk to you in responding to this questionnaire since your identify will remain anonymous. All the responses will be aggregated, summarized, and analyzed for the award of the Master's Degree. For inquiry about rights as a research participant, feel free to contact the University of Nairobi Offices in Kenya. If you have any questions related to the study or this questionnaire, please send an email to **victorelijah@students.uonbi.ac.ke**

I have read and understand the above information. I agree to participate in this study.

☐ Yes ☐ No

If No, Please may we know the reason: ____

Part B: Demographic data

In this section, the study shall basically focus on the general information regarding the respondent, their experience on data governance. In this case, choose a single option for all questions unless required to do otherwise.

1. How many years have you worked at the organization?

   Years ☐☐

2. Did you use any of the current systems at the organization?

   ☐ Yes ☐ No

3. Which of the following will best describe your role at the organization on information system?

   ☐System Administrator ☐ System Champion

   ☐ System User ☐ Project Manager

   ☐ Data User ☐ Data User

   ☐ System Developer ☐ Data administrator

   ☐ Other – Specify source: ______________________________________

4. What is the name of your organization? ____

5. Date of Interview: ____

Part C: Data Governance Drivers

In this section is concerned more on identifying data governance driver at the health professional regulatory authorities. Therefore, you need to select one option at a time unless you have further information or suggestions.

1. Based on your experience, please indicate your level of agreement with this statements in regard to reason which would drive your organization to adopt a formal data governance

	Strongly disagree	**disagree**	**Neither disagree nor agree**	**Agree**	**Strongly agree**
a. Ensure compliance with an existing regulatory framework	☐	☐	☐	☐	☐
b. Maintain data quality	☐	☐	☐	☐	☐
c. Ensure data security & security	☐	☐	☐	☐	☐
d. Increase organizational efficiency	☐	☐	☐	☐	☐
e. To increase customer satisfaction	☐	☐	☐	☐	☐
f. Meet competitive advantage	☐	☐	☐	☐	☐

Part D: Data Governance Barriers

1. Based on your experience, please indicate your level of agreement with this statements in regard to factors that hinder your organization into achieving a formal data governance

	Strongly disagree	**disagree**	**Neither disagree nor agree**	**Agree**	**Strongly agree**
a. Lack of understanding on data governance	☐	☐	☐	☐	☐
b. Lack of awareness on existence of data governance	☐	☐	☐	☐	☐
c. Lack of support & ownership on data governance	☐	☐	☐	☐	☐
d. Lack of resources i.e. their limited resources allocated towards data governance programs	☐	☐	☐	☐	☐

Part E: Data Principles

1. Based on your experience, please indicate your level of agreement with this statements in regard to data assets for health professional regulation

	Strongly disagree	disagree	Neither disagree nor agree	Agree	Strongly agree
a. Data is an important asset in health professional regulation	☐	☐	☐	☐	☐
b. Health regulatory data should be shared by other Health Information Systems	☐	☐	☐	☐	☐
c. Health professional regulatory is based on evidence based decisions	☐	☐	☐	☐	☐
d. Internal & external data sources should be considered as data assets for health regulation	☐	☐	☐	☐	☐

Part F: Data Quality

1. Based on your experience, please indicate your level of agreement with this statement with regard to data quality

	Strongly disagree	disagree	Neither disagree nor agree	Agree	Strongly agree
a. Regulation data should be accurate	☐	☐	☐	☐	☐
b. Regulation data should be complete	☐	☐	☐	☐	☐
c. Regulation data should be relevant	☐	☐	☐	☐	☐
d. Regulation data should be timely	☐	☐	☐	☐	☐

Part G: Metadata

1. Based on your experience, please indicate your level of agreement with this statements in regard to meta data on health professional regulation

	Strongly disagree	disagree	Neither disagree nor agree	Agree	Strongly agree
a. Data dictionary should be defined for all data elements	☐	☐	☐	☐	☐
b. Data dictionary is importance for health regulation data interpretation	☐	☐	☐	☐	☐

Part H: Data Access

This section assessed the need for knowledge about which users should have access to certain data in as part of data governance.

1. Based on your experience, please indicate your level of agreement with this statements in regard to data access on health professional regulation

	Strongly disagree	disagree	Neither disagree nor agree	Agree	Strongly agree
a. Regulation data is sensitive data	☐	☐	☐	☐	☐
b. Regulation data access should be restricted	☐	☐	☐	☐	☐

Part I: Data Lifecycle

This section assesses the need for knowledge about the need of managing data as a product with a lifecycle i.e. creations, storage and destroy data.

1. Based on your experience, please indicate your level of agreement with this statements in regard to data lifecycle on health professional regulation

	Strongly disagree	disagree	Neither disagree nor agree	Agree	Strongly agree
a. Regulation data has a life span	☐	☐	☐	☐	☐
b. Data lifecycle should be monitored for health regulation data	☐	☐	☐	☐	☐

Part I: Data Governance

This section assesses the importance of governing data for health regulation.

1. Based on your experience, please indicate your level of agreement with this statements in regard to data governance on health professional regulation

	Strongly disagree	disagree	Neither disagree nor agree	Agree	Strongly agree
a. In order to derive more value on the health regulation data, a data framework should be established	☐	☐	☐	☐	☐
b. Health regulation data value can be improved through proper accountability framework.	☐	☐	☐	☐	☐

Part J: Others

This section allows one to give any other important area that needs to be considered in data governance of health regulatory data

a) Please mention any other area that should be considered when governing health regulation data?

____

____

____

____

____

____

____

DATA GOVERNANCE ASSESSMENT TOOL

1. Name of the organization? ____
2. What data sources have been identified?
   - ☐ None
   - ☐ External data sources only
   - ☐ Internal data sources only
   - ☐ Internal & External data sources
3. Are the owners of the data sources defined?
   - ☐ No
   - ☐ Yes
4. Do you use your data to make policy related decisions?
   - ☐ No
   - ☐ Yes
5. Is there a formal data quality policy?
   - ☐ No
   - ☐ Yes
6. How often is data profiling and audit performed in the organization?
   - ☐ Never
   - ☐ Quarterly
   - ☐ Once a year
   - ☐ Monthly
   - ☐ Twice a year
   - ☐ Weekly
7. How often is data cleaning and monitoring done in the organization?
   - ☐ Never
   - ☐ Quarterly
   - ☐ Daily
   - ☐ Yearly
   - ☐ Monthly
   - ☐ Biyearly
   - ☐ Weekly
8. Are there data stewards who are in charge of data quality?
   - ☐ No
   - ☐ Yes
9. Is there a formal data dictionary in the organization?
   - ☐ No
   - ☐ Yes
10. Is the data dictionary available for all in the organization?
    - ☐ No
    - ☐ Yes
11. Has sensitive data been identified and databases which carry sensitive data tagged?
    - ☐ No
    - ☐ Yes
12. Has the data access been restricted to authorized business and IT users, and single login created, with all login activity traced?
    - ☐ No
    - ☐ Yes
13. Is there a process for database changes so that unauthorized changes are prevented?
    - ☐ No
    - ☐ Yes
14. Are all external data requests scrutinized, personal identification information protected as needed, and approved by the Privacy and Security work-group?
    - ☐ No
    - ☐ Yes
15. Has a data retention policy been defined for all your data?
    - ☐ Don't Know
    - ☐ Yes
    - ☐ No
16. When did your organization last discarded data?
    - ☐ Don't Know
    - ☐ 5-10 years ago
    - ☐ Never
    - ☐ More than 10 years ago
    - ☐ 1-5 years ago