# Table of Contents

# CHAPTER ONE: INTRODUCTION

## 1.0 Background and Context

In an era defined by rapid technological advancement and increasing reliance on data-driven decision-making, the efficiency of data management in has become an important feature of an organizational success across various sectors, particularly public administration and large corporations. The effective management of data holds great importance for facilitating transparent governance, enhancing service delivery, and fostering socio-economic development such as informed decision-making (Eurini et al., 2023) within the public sector. Data management is defined as the process of collecting, storing, and protecting data within an entity to promote efficiency and decision-making.

The Gambia, a nation in West Africa celebrated for its rich cultural heritage and beautiful landscapes has recently undergone a period of significant change and improvement. Since the country's transition to democracy in 2017, there has been a renewed focus on promoting good governance, accountability, and transparency. This dedication is evident in both the previous National Development Plan (NDP) 2018-2021/2022 and the new Recovery-Focused National Development Plan (RF-NDP) 2023-2027. As part of this broader reform agenda, public institutions in The Gambia are increasingly acknowledging the important role of data in driving decision-making based on evidence and in achieving development goals.

Despite the recognition of its importance, data management practices within public institutions in The Gambia are often characterized by challenges and inefficiencies. Historically, these institutions have relied on manual processes and outdated systems for data collection, storage,

and analysis. Fragmented databases, inadequate infrastructure, and limited capacity have impeded efforts to harness the full potential of data for governance and development purposes.

The Gambian public sector operates in a dynamic environment characterized by diverse stakeholders, limited resources, and evolving regulatory frameworks. Against this backdrop, effective data management presents both opportunities and challenges. Access to accurate and timely data can empower decision-makers, enhance transparency, and improve service delivery to citizens. However, inadequate data management practices, including data silos, inconsistent data quality, and limited capacity, can hinder public institutions' ability to fulfill their mandates effectively.

The Gambian public sector operates in a dynamic environment with diverse stakeholders, limited resources, and evolving regulatory frameworks, presenting both challenges and opportunities for effective data management. With a diverse population across urban and rural areas, there's a need for robust data management systems to ensure equitable access to quality services and resources for all citizens. Furthermore, The Gambia's participation in regional and international initiatives like the Sustainable Development Goals (SDGs) and the African Union's Agenda 2063 emphasizes the critical role of accurate and timely data in monitoring progress and shaping policy interventions.

In this context, there is a pressing need to examine and address the challenges faced by public institutions regarding inadequate data management practices, including data silos, inconsistent data quality, and limited capacity, which hinder public institutions' ability to fulfill their mandates effectively and managing their data assets effectively. By understanding the underlying factors influencing data management practices and identifying opportunities for improvement, stakeholders can develop tailored strategies and interventions to enhance governance outcomes, improve service delivery, and promote inclusive development.

The study is structured as follows: chapter 2 provides insights on relevant literature on data management and practices within public institutions in the Gambia, chapter three discusses methodology used…

## 1.1 Problem Statement

Recognizing the growing importance of data management, it's evident that public institutions in The Gambia are facing challenges in this domain. Effective data management within these institutions is crucial for informed decision-making, efficient resource allocation, and transparent governance. However, despite the recognized importance of data management, there are opportunities to address the challenges being faced by public institutions in The Gambia. One of the primary challenges is the prevalence of fragmented and outdated data management systems. Many institutions currently rely on manual processes and disparate databases, leading to inefficiencies, errors, and difficulties in data integration. Furthermore, there is a need to address inadequate infrastructure and limited technological capacity, which pose significant obstacles to effective data management. By investing in the necessary resources, both in terms of technology and skilled personnel, public institutions can implement and maintain modern data management systems, leading to optimal utilization of available data and ensuring the quality and reliability of generated information. Additionally, addressing challenges related to data security and privacy is crucial. Strengthening data protection measures will mitigate risks such as unauthorized access, data breaches, and loss of confidentiality. This will not only enhance public trust but also ensure compliance with legal and regulatory requirements governing data management practices.

## 1.2 Research Questions

The study aimed to answer the following research questions through its research objectives.

1. What are the current existing data management practices within the public institutions in The Gambia?

2. What are the main challenges and obstacles that hinder effective data management at these institutions?

3. What strategies and policies can be employed in the Gambian public sector to strengthen the data management capacities?

4. How might data management enhancement contribute to improved governance effectiveness and service delivery outcomes in the public sector?

## 1.3 Objectives of the Study

The overall objective of this study is to evaluate the effectiveness of data management in public institutions in The Gambia that serve as a mechanism for the public sector for transparency and informed decision-making. These objectives are outlined for the study as follows:

1. To examine the present state of data management practices within the selected public institutions by reviewing current literature.

2. To evaluate the current challenges and obstacles that impedes efficient data management in these institutions.

3. Assess how data management practices impact the level of governance, service delivery, and institutional performance.

4. Explore and design various strategies and best practices that can be implemented to improve and enhance data management skills in the public sector of The Gambia.

## 1.4 Hypothesis

There are three hypotheses that have been established to identify the relationship between effective data management and service delivery and also, hypothesis on the impact of data management on Institutions performance within the public sector of the Gambia. These entire hypotheses are drawn from the theoretical framework of this research.

$H_0$: The impact of effective data management in service delivery.

$H_1$: Effective data management does not impact institutional performance.

$H_2$: Effective data management does impact institutional performance.

## 1.5 Scope of study

This study explores current data management practices across different public institutions in The Gambia, including government ministries. It examines the diversity of the data types, sources, and formats managed by these institutions. The scope includes an analysis of existing data governance frameworks within public institutions, focusing on policies, procedures, and organizational structures governing data management. This involves assessing the clarity of roles and responsibilities, data ownership, accountability mechanisms, and compliance with regulatory requirements. This study investigated how data are collected, stored, and managed within public institutions, including the use of digital systems, databases, and information management tools. It examines issues related to data quality, integrity, security, and privacy. This aspect of the study explores the processes involved in processing and analyzing data within public institutions. It will assess the availability of analytical tools and techniques as well as the capacity of staff to conduct data analysis and interpretation effectively.

## 1.6 Significance of the study

Effective data management is of utmost importance for the development and progress of any public institution and The Gambia is no exception to this phenomenon. This study contributes to the existing knowledge by providing empirical evidence into the state of data management practices in the country. It examines how data are collected, stored, processed, and utilized within public institutions. By understanding these processes, policymakers can make informed decisions at both strategic and operational levels (Obatolu, 2021), leading to better governance practices, transparency, and accountability. Effective data management enables the efficient

allocation of resources and targeted interventions, resulting in enhanced service delivery across various sectors such as healthcare, education, and infrastructure. The improvement of the data efficiency practices in sectors, these institutions can streamline activities, identify areas of development, and eventually address needs and improve service delivery to the citizenry through effective policies and programs.

Through investing in data management practices, entities can strengthen a country's overall data infrastructure. This offers numerous benefits to public institutions and also opens opportunities for the private sector, academia, and civil society to foster collaboration. This not only fosters innovation but also to improved levels of economic growth. By analyzing the gaps in data management practices, Policymakers can formulate interventions from the gaps in skills, technology and governance frameworks and provide valuable insights on capacity building and reforms for institutional overall performance. This has the potential to lead to more transparency and accountability within the public sector, facilitating more efficient public services and subsequently leading to more public trust and well-being of the nation.

## 1.7 Thesis Structure

This research will be done in eight chapters and arranged as follows for the first three chapters;

1. Chapter one introduces the background and context of the thesis, highlighting the significances of effective data management in the public sector of the Gambia and observing the limited studies in this field. The research objectives, hypothesis, the research scope, significant of study, and research questions are all captured in this chapter. Finally, it also emphases the important contributions to knowledge done by this study.
2. Chapter two introduces the key literature on effective data management which are relevant to the research problem identified. Grounded on the methodical review carried out, present data management structures are identified. Significantly, the exposure and weak areas of former frameworks are addressed to help in bringing advanced research study context.

3. Chapter three brings an overview of the methodology used for this study. First, it explained the research design employed and also explained the nature of targeted populations and the types of research instruments to be use were also discussed and analyzed.

# CHAPTER TWO: LITERATURE REVIEW

## 2.1 THEORITICAL REVIEW

### 2.1.0 Introduction

#### 2.1.1 Data Lifecycle Management (DLM)

Data lifecycle management refers to the whole process of managing data throughout the life cycle from data collection to deletion. It contains series of phases during its cycle such as data creation, storage, processing, usage, and deletion and emphasizes on how the phases are governed by policies that maximizes the significance of data during each stage of the lifecycle. DLM strategies such as data governance policies, data security measures and compliance with regulations safeguard business data assets and enhance innovation and growth when implemented (Chia, 2023).

This theory supports and guarantees users of longer accessibility of data using policies in an organization at a suitable time. It also helps businesses to minimize risk by mapping business tasks with data maps to maintain redundancy and improve consistency (Kumar & Banyal, 2010). However, DML has challenges like resource allocation and identifying methods to correct data storage, usage, and management.

Khatri and Brown (2010) classified data lifecycle management as the fifth domain and concerns the movement of data in the whole institution. The data lifecycle help organization the requirements to comprehend how data is obtained from different contributors within and outside

of the institution and its movement from dynamic use to its final archival is an important segment of a productive data regulation program (Allen Chamberlain, 2013)

Rouse (2010) also recommended that all data occupying in large datasets should be instantly reachable depending on its longevity. Rouse further explained that since data in more recent version is often accessed in bigger frequency, it should hence allow information technology branches to amend their storage and retrieval architecture to manage these explored requirements. The maximum data lifecycle that regulate the movement of data applying rule sets (Breaux and Alspaugh, 2011). The rule set can include mappings to command origins requirement for data management standards grounded on the context of the data. In utilizing the approach hinted by Breaux and Alspaugh (2011) virtual licenses containing rule sets are integrated inside data sets to give controls against non-compliant utilize immediately when the data is assigned to its final archival.

### 2.1.2 Information Lifecycle Management

Information lifecycle management is the management of Organization's information and associated metadata throughout the lifecycle from creation to disposal. This theory consists of strategies and technology to improve the control of information and minimize the cost and risk of data within the organization. ILM takes a policy-based approach to handling data and facilitates storage where data will be categorizing into various types. New Data that will be used frequently are stored on faster and more expensive storage Medias for easy access while data that are less complex are stored in slower and cheaper media (Robert, 2022).

According to (Malak, 2023) effective information lifecycle management helps organizations in increasing employee productivity and making better business decisions.

### 2.1.3 Data Governance

Data governance theory emphasize on establishing policies, processes and standard for managing data assets effectively. These processes determine data owners, data security measures, data usage and intended uses for the data and ensures data is trustworthy and consistent (IBM, 2023). Data governance ensures that quality data is available throughout the lifecycle and the process

which are implemented are in line with the organization's business goal. However, the theory identifies who takes an action, from which data, in which situation and what method to use.

Data governance improved the quality of data, minimize risk of errors and data management cost, increase the accessibility of data within the organization and ensures rules are set regarding the access to data and adhered to (Team, 2021). This theory guide businesses in decision making regarding their competitors and how to improve revenue. Despite all the benefits, organizational alignment is the biggest challenge most organizations face as stakeholders' lack awareness on what the data assets are and what their respective definition and formats should be.

According to Alhajj, Ridley, and Rifaie, (2009) decision making administrations for data quality and the entire data management program moderate in a role – base, hierarchical framework. Also, Allen Chamberlain, (2013) explained that different models are showed in the finding; a popular theme is a model that includes executive donors and representations from all over the institution. Information technology heads should collaborate with institution to obtain sponsorship and regulate a data management council which will include stakeholders from information technology and the business department. This organization body is mainly to be responsible of determining data management policies and functioning base on the data quality structure

Storey, Dewan and Freimer (2012) propose that it is important to hand on brief data quality policies in the entire institution and structure what data quality feature should be readily recognizable in data sets. Furthermore, they argue that people should be highly rewarded to carryout sound data quality framework and remunerated for their productivity.

The formal ideas of data governance are to improve the value of data assets as data governance control data assets and optimize its importance with the help of quality control (Otto, 2011). Other studies validated the concepts of Otto (2011a), although argued that as a result of its importance it should be managed and control its quality, supervised and protected (Khatri el al, 2010; Rifaie et al 2009; were et al 2017). Wende (2007) explained that data governance should guarantee standard information products are transferred to the user.

The diagram below (Figure 1) indicates Otto's portrayal of his standpoint on data governance and its correlated themes

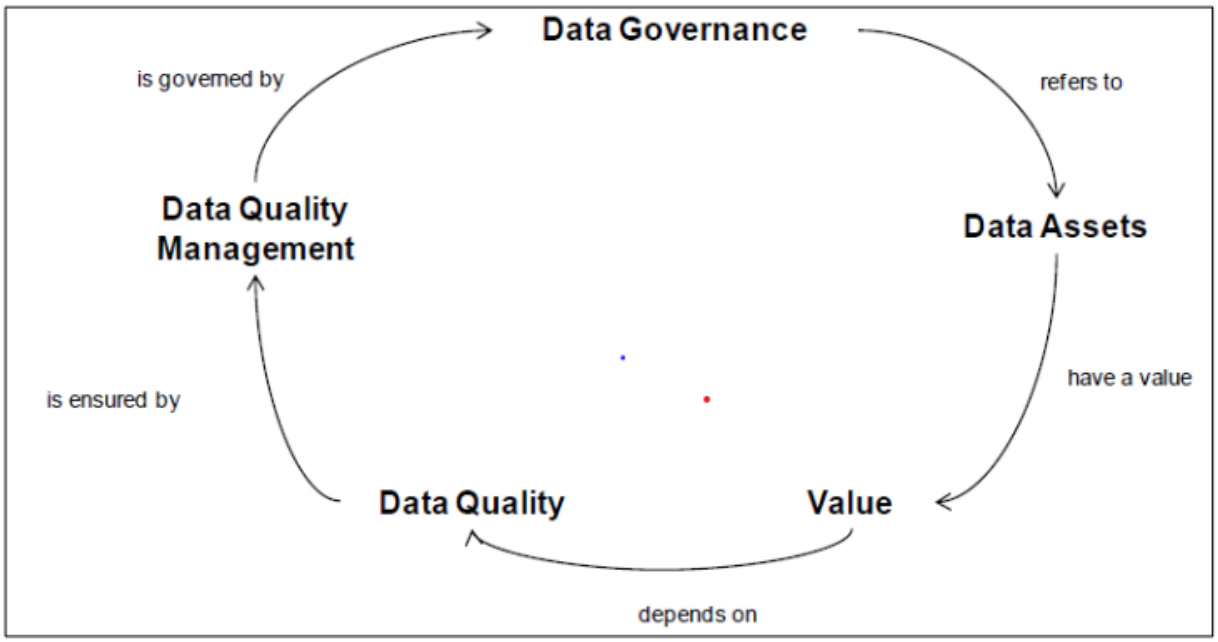


*Figure 1: Data Governance and Related concepts (Adapted from Otto, B. 2011a)*

### 2.1.4 Data Quality Management (DQM)

Data quality management refers to the practices and principles use to strengthen data integrity, usefulness, consistency, and accuracy. These principles require having the right people in place that uses advance technology to improve data quality of an organization. DQM make use of tools such as data cleaning, monitoring, profiling, parsing, matching, standardization, and enrichment to improve organization's data trustworthiness (Foote, 2022).

Data quality management positively impacts organization in making decisions about product development and marketing, corporate strategy, and stakeholder relationship. It also manages resources and maintains the business integrity (Onyegbula, 2023). Thus, receiving and accessing different data from different sources takes time to control and determine quality data from it.

Brown and Khatri (2010) recommend that institutions should provide standards to make sure data is current, complete and accurate. The consideration of data as an assets show a shift into what is termed as information economy (Parssian, Sarkar and Jacob, 2004). In the rising economy, the progress of an institution relies in huge part to the standard of its data. Research indicates that data quality issues can lead to unexpected cost, employee displeasure and negatively affect development and that quality standards lower financial of disorganized data quality (Allen Chamberlain, 2013). The expenses are assumed at 0.08 to 0.12 of the institution revenue. Information technology management should therefore regulate standards regarding data quality (Allen Chamberlain, 2013).

Oterle, Otto, and Weber, (2009) argue that while there is no unique solution for all institutions to apply a common data quality program, some information technology management frameworks are considering to include centralized, decentralized, and also federated model of the institution and describe the positive impacts of data management in linking information technology goals to the targets of businesses as a fundamental aspect of the management structure.

### 2.1.5 Data Integration

Data integration refers to the combination of data from various sources to a unified format that can be used by organization or individuals for meaningful decisions and analytical purposes. It has series of stages that ranges from identifying the source of the data to assessing the data and analyzing it. Data can be integrated using ELT (Extract Load and Transform), ETL (Extract, Transform Load), API (Application Programming Interface) and other types of data integration to harmonize data from different systems and formats (IBM, 2023).

This theory benefits businesses and organization in various ways. It increases business efficiency, improve business intelligence, reduce data silos and advanced analytics with a complete picture of financial risks and other processes (Simplilearn, 2023)

### 2.1.6 Information Retrieval

Information retrieval theories focus on identifying and extracting important data from a large database. Data is retrieve based on the query detected by users or applications. It plays a very important role in different sectors in the modern world from academia and ecommerce to healthcare and agriculture where data is needed the most. Ecommerce uses this theory to identify useful products based on customer's preferences and establish marketing strategies to overcome competitors.

Moreover, for the past year's information retrieval has never been perfect because of few challenges it encounters. Some data uses natural languages, making it difficult to interpret queries which affect the indexing and the evaluation process. Sometimes retrieval systems find it difficult to find the meaning of contents due to the gap between textual representation and human understanding (Elastic, 2023).

### 2.1.7 Metadata Management

Coleman, Hughes, and Perry (2009) posit that metadata is an important component in proper data management and recommend institutions to capitalize common coding structures for obtaining metadata and enhancing reporting. Also, Khatri and Brown (2010) recommend that metadata "provides a mechanism for a concise and consistent description of the data, thereby helping interpret the meaning". Alternatively, metadata gives a common means for institutions to actualize production searching and analyzing of various data posit by (Allen Chamberlain, 2013).

Furthermore, Allen Chamberlain (2013) proposes that data management authorities should direct a unique strategy in labeling information assets to realize the following to facilitate the following;

1. Upgraded search abilities and indexing of data assets
2. Automated distribution of information grounded on user conduct.
3. Instantly recording of what the data assets have been spread outside of the institution and inside the enterprise.
4. Clarity in where the data inhabits in its lifecycle
5. Improved efficiencies in deciding the genesis of the data assets.

Allen Chamberlain (2013) further advises that Information technology councils should also expect that the strategy in managing metadata will change in the long run. Management structures aimed to take a pre-emptive approach in steering the unpreventable changes to how an institution governances its metadata due to transitioning business activities and objectivities (Khatri and Brown, 2010). Irrespective of the unique methods used to acquire metadata, research conducted by Hua, Huang, and Yen (2010) recommend that data storage facility system should be internet – reachable and coding framework should maximize adaptable mark – up language (XML).

## EMPIRICAL REVIEW

### 2.2.1 Data Security & Privacy

Data security theory comprises of principles that protect digital information from theft, corruptions, or unauthorized access to information. It has tools that safeguard data through the process of data controls, encryption, masking, editing of sensitive information, auditing and ensure compliance with regulations (Fortinet, 2023). However, for a business to grow faster it needs a strong data security to maintain a clean reputation and have competitive edge against others.

Recently, organizations found it challenging to secure their data properly because of three major challenges. Having an insider, or former employee who has access to steal or corrupt data for personal use. Technical misconfiguration is another threat which can expose confidential data sets ad having a third-party in an organization.

Individuals are well informed that originations will protect their privacy but has a limited trust in technology. Encryption technology is a familiar attribute in all websites especially institution sites which help to protect information privacy, supported by a composition of separate unique identifiers, like, a password or key (Rajeev Shrestha, 2009). Privacy is a significant factor that impacts users' reason to embrace e – based transactions (Gerrard & Cunningham, 2003). Rajeev Shrestha (2009) argue that customers are indifferent in the trust ability of e – services' privacy

rules; trust is a major factor in customer's readiness to involved in an online transaction of funds and private sensitive details.

Hoehle et al (2012) posits that customers' expectation for purchasing products and services in financial institutions has many factors and security is one of those related factors. Security can be a great factor that influenced customer's expectations of using internet. Zhou (2012) explains that network servers are easily targeted by hackers and virus, which creates doubt and uncertainty in customers and institutions therefore, has a negative effect on their willingness to transact using devices.

### 2.2.2 Data Quality Strategy

Alhajj, Ridley, and Rifaie (2009) posit that data management schemes should bring clear understanding on how an effective data management maps in a grand design to an institution and should explain data quality goals. Data quality management should provide the following strategy components;

a. The procedure of evaluating how data quality decrease financial risk
b. A business case showing policies and strategies
c. The main purpose of data quality in an institution
d. A stockroom vault of all data quality rules and procedures

### 2.2.3 Data Access

Institutions ought to effectively regulate data accessibility resources in order to control risk and conformity (Khatri and Brown, 2010) and they highlighted the access guidelines that recognized principles for organizations to manage what information should be accessed and by who. Therefore, data structure should determine the worthy of data for organizational intelligence and instruct security controls for large data management sets. Also, effective data management help the institutions to understand insights with regards to their processes and customers attitudes, the intrinsic values of the datasets might be mishandled for villainous means without relevant, clear access regulated standards and methods (Allen Chamberlain, 2013).

Also, Ashford (2013) posits that information technology leaders and data regulation councils should revise safety programs to indemnify big volumes and data gathering into an organization. This aim to ensure that methods for gathering, analyzing, and explaining meanings from data are firmly paired with organizational administration risk and compliance projects. Institution leaders in large data warehouse, security and analytics should recommended that information technology pioneers change their strategies to security and convene real – time requirements on large data system (Ashford, 2013).

Allen Chamberlain (2019) advises that data management structures should secure the numerous digital channels from which information gathered is protected and well reliable. Furthermore, large data management and governance need clear explanation roles and a responsible framework to authorize and notify data custodians on an appropriate data handling method from the start of gathering to distribution within and outside of the institution (Madsen, 2012).

### 2.2.4 Data Principles

Organizations are obliged to create data principles "to which data is an enterprise wide asset, and thus what specific polices, standards and guidelines are appropriate" (Khatri and Brown 2010). They argued that success at administering large data set depends on first understanding that data is a key asset to an organization. By reaching the requirement for institutional data in the form of enterprise – wide asset, the data becomes important to the organization rules, control standards, and control routines (Allen Chamberlain, 2013). The recognition that data is an asset to be utilized in updated business intelligence entails so much of the same compliance habits carried out to different types of assets, which include human resources, tangible assets, and monetary assets or capital assets (Allen Chamberlain, 2013).

For managing data as an asset, McGlinchey (2013) advises institutions to form regulations that;

a. Explain what the planned goals of large data analytics are meant for in an organization.
b. Describe data management as the role of all stakeholders in both the information technology and the business side. They advise the establishment of a responsibility

framework for making sure that data management rules are frequently modified and applied.

Also, Allen Chamberlain (2013) recommends that data management principles should plainly explain how verdicts pertaining to data management rules and procedures are determined. Based on research conducted by Rifaie, Alhajj, and Ridley (2009) explained that such principles warrant the formulation of institutional framework to;

a. Establish data management regulations and procedures and authorize enforcements.
b. Control risk sustained by the setbacks of the data management program.
c. Sanction the capital spends for large data allocation such as technology and recruitment.
d. Convey data management decisions and bring clarity in the establishment and supervision of data management regulations in the organization.

## 2.3 Literature Gap

The diverse implementations, theories and evidences of an achievement has been unique (related to single moment within an institution), has solely dealt with exhibiting a particular area of focus (concentrating on solving a particular issue instead of an organizational widespread) and has neither been reproducible nor able to advance success to a different area of the institution. The present literature provides such evidence as stated below;

a. There is only few journal management scholarship which solves the difficulties of using data management tools – or, yet still, that explains the promise and chances of novel theories and practices that data management can bring about (George, Haas, Pentland, 2014)
b. In spite of the capacity innate in data driven institutional development applications to submit completely new method of visualizations for institutions, there are factors of potential blocks to use this methods as well. This captured three separate issues; capabilities, mind-sets, and ethics.

c. The domain of data – sharing agreements remains unofficial, disorganized, enforced manually, and connected to separate transactions (Koutropis & Leiponen, 2013).
d. The phenomenon of data management is one that has less focus by academics. Additionally, the uniqueness of data is arguable with possibly no overall strong description available (Calverd, 2014)
e. Transactions of such manual approaches and remote events represent an important road block to the market in data – social science in general and research management (George, Haas, Pentland, 2014).
f. Diversity, complication and diversity in arrangement describe data concepts and teachings (Motta et al., 2016)
g. "There is no singular big data theory, methodology, or value structured" and "just about anyone can enter the consulting space and engage in big data activity"
h. There are more difficulties which still occur with data and such issues have to do with heterogeneity, discrepancy and fragmentation, fluctuation scales, timelessness, confidentiality and data possession, and analytics and combination (Jagadish et al., 2014).
i. Data is a stage and as such when data is large and complex it demands more sophisticated processing machines to match it, and can possibly provide institutional insights and source of importance which lesser scale of data handling cannot (Mayer – Schonberger and Cukier, 2013).
j. The definition of data consists of different discrepancies. According to Church and Dutta (2013), Young (2014), (George, Haas, Pentland, 2014) incorporated the difference definition of data and are presented across different literatures and the confusion that exist in data definition.

# CHAPTER THREE: METHODOLOGY

## 3.0 Introduction

Effective data management is crucial for public institutions to efficiently utilize resources, make informed decisions, and improve service delivery. This methodology outlines the approach to

studying data management practices within various public institutions in The Gambia, including the Ministry of Finance and the Ministry of Agriculture. The research will involve the use of a quantitative method employing self-administered questionnaires structured for primary data collection. A Likert scale was utilized to gauge respondents' perceptions, and data analysis was conducted using the Stata 16 software.

## 3.1 Research Design

Base on the study by Sekaran and Bougie (2016), research design is planning for data collection, measurement, reviewing data, and intends to find answers to your research questions. Taking into account the research framework assist the researcher to discover an objective of the studies. There are six different types of research framework identified and they include; experiments, grounded theory, case studies, and research action.

This research is aligned for knowledge contribution in the field of data management in institutions; its potential utilization is geared to be a realizing by the government stakeholders and industrial experts. Also, it’s regarded noteworthy that the fundamental principle of the conceptual framework designs will analyze responses from number of experts, thus establishing the rationale into the problems it claim to tackle. The investigation responses and reports are to be presented and will be include some of the Gambia’s leading experts who are daily using data in the country’s public sector.

## 3.2 Target Population

In research, "population" refers to the entire group of individuals or units that the researcher is interested in studying. For this study, the population comprised employees working within different public institutions in The Gambia, including the Ministry of Finance, Ministry of

Agriculture, Ministry of Health, and other relevant departments and agencies. These individuals are involved in various aspects of data management in their respective institutions.

The target population for this study comprises employees directly involved in data management processes within public institutions in The Gambia. This includes data analysts, information officers, IT specialists, and other personnel responsible for data collection, storage, analysis, and utilization. A purposive sampling technique will be utilized to select participants from each institution, ensuring representation from different departments and levels of expertise.

## 3.4 Research Instruments

Kotler (2000) explained in a finding that four (4) different tools can be used to gather information/data and in a primary data collection method, furthermore, the instruments include surveys, focus group, experiments, and participant's observation. Nonetheless, this research follows a survey instruments to reach these inquiries as the researcher want to get answers and will do different private interviews.

### *3.4.0 Interviews*

Interview is one of the methods that can be carried out to gather information/data from different personals that possessed knowledge with regards to a particular topic. However, this study will do several interviews to gather information from public sector data management experts.

### *3.4.1 Questionnaires*

As the researcher will use primary data and concentration will be on questionnaires then will be used to gather specific data from respondents, this study will distribute questionnaires in three sections. Accordingly, section A will be the demography section and with seven questions; age, gender, monthly income, marital status, department, and experience. Additionally, section B will contain general questions regarding effective data management as part of the independent variables in the study and the researcher will use a five (5) point Likert scale in both section B and C will be level star from 1 as strongly agree to 5 strongly disagree. Furthermore, the majority of the questionnaires will be distributed by hand to various government departments and some will be sent by email in the form of word file and Google documents to be filled. The questionnaires contain few merits and demerits which are mentioned below;

#### 3.4.1.0 Merits

a. The answers are always fully unveiling because there is no name or signature.
b. The respondents will be able to fill questions with full comfort and relax
c. The questionnaires can be used in places where the researcher cannot reach with the use of a new method such as Google form or email.

#### 3.4.1.1 Demerits

a. It needs more time for respondents to return the questionnaires.
b. Few questions might not be clear, therefore, the respondents might give wrong answers.
c. All questions not to be answered by respondents will leave blank questions.

The researcher has employed methods to minimize the mentioned issues such as to leave questions blank, explaining the unclear questions for respondents to understand properly and will be filled the particular respondents and also double check all questions to verify if there are any question that was not answered.

## 3.5 Data Collection and Data Source

Primary data will be collected through a self-administered questionnaire. Primary data refers to original data collected firsthand by the researcher for a specific research purpose. In this study, primary data were collected directly from participants within the target institutions. These data will provide firsthand insights into the perceptions and practices of data management within public institutions in The Gambia. Questionnaires will be distributed to select participants within the target institutions. The questionnaire will be structured to collect information on various aspects of data management practices, including data collection, storage, analysis techniques, and utilization for decision-making.

The questionnaire consisted of closed-ended questions, with respondents providing their level of agreement or disagreement on a Likert scale. The Likert scale is commonly used in surveys and questionnaires to measure respondents' attitudes, perceptions, or opinions towards a given topic.

It typically consists of a series of statements to which respondents indicate their level of agreement or disagreement on a predefined scale, typically 1 to 5 e.g. 1 indicating "Strongly Agree" and 5 indicating 'Strongly Disagree". In this study, Likert-scale items were included in the questionnaire to assess participants' perceptions of various aspects of data management practices within their institutions. This will be shared electronically using the Google Form that consisted of numerous questions regarding data management practices such as data security, data privacy and others for individuals responsible with data handling at the studied public institutions to ensure quality of responses.

In this study, a quantitative method will be utilized to systematically measure and analyzes data management practices within public institutions in The Gambia. Quantitative research involves collecting and analyzing numerical data to understand phenomena and relationships. This approach allows for the quantification of responses, providing statistical insights into the prevalence, patterns, and correlations related to data management practices.

## 3.6 Source of Data

There are different methods of data source a researcher can use to collect data and the following are few examples;

a. Surveys can be used as a structured questionnaire due to embracing survey associated with large number of variables delegated with samples of respondents.
b. Qualitative research is important with a huge or smaller number of data purposive sample, with the help of different kind of methods example interview, observations, and focus in groups.
c. The use of experiments and quasi – experiment are vital because they usually include a research design that help strong causal inferences.

### *3.6.0 Primary data*

Hox and Boeije (2005) proposed that a primary data is that data which were collected for a particular problem and find the new data and it can further be used for another study, as they

stated before with regards that research can attract different solution to questions for example asking about attitudes, feelings, experience, and behavior regarding a problem in a population. Nevertheless, this study will collect a primary data through questions and personal interviews with officials in most public institutions in the Gambia.

### *3.6.1 Secondary data*

According to Ticharwa (2017), a secondary data can be explained as a data that was collected and summarized prior and the data is already presented for use in any purpose as needed. Secondary data is outdated and its accuracy is not recognized but it's cheaper and save more time than the primary data.

## 3.7 Data Analysis

The primary data will collected and they will be entered into the Stata 16 software for analysis. Stata is a powerful statistical software package widely used for data analysis in research settings. Descriptive statistics, such as frequencies, means, and percentages, will be calculated to summarize the responses to each Likert-scale item. Inferential statistical techniques can also be employed to examine the relationships between different variables related to data management practices.

After gathering the data, we inputted it into the Stata 16 software for analysis. Stata is a widely utilized statistical software package for analyzing data in research. We computed descriptive statistics, including frequencies, means, and percentages, to summarize the responses to each Likert-scale item. Furthermore, we can utilize inferential statistical methods to explore the connections between various variables associated with data management practices.

### 1.8 Reliability & Validity

The validity of the instruments, which are interviews and questionnaires in this case, will be determined by ensuring that all the questions contained in the questionnaire and the interview are in line with the study's overall research questions and objectives. The questionnaire will validated through pilot testing by some respondents to test their understanding and interpretation of the questions to ensure that such questions bear some meaning and the comments and/or queries raised by those respondents will be used to fine-tune the questionnaires.

In order to ensure reliability of the instrument, the questionnaire will be verified by the researcher's supervisor with better understanding of scientific research methods. For purposes of clarity, the questionnaire will further be pilot tested and consequently fine-tuned. In order to ensure logical completeness and consistency of responses, data editing will be carried out each day by the researcher and identified mistakes and data gaps will be rectified as soon as possible.


Chia, A. (2023). Data Lifecycle Management: A Complete Guide. *splunk*, 1-9.

Elastic. (2023). What is information retrieval? *Elastic*, 1-10.

Foote, K. D. (2022). Data Quality Management 101. *DATAVERSITY*, 1-5.

Fortinet. (2023). What Is Data Security? *Fortinet*, 1-5.

IBM. (2023). What is data governance. *IBM*, 1-6.

IBM. (2023). What is data integration? *IBM*, 1-5.

Kumar, R., & Banyal, R. K. (2020). Data Life Cycle Management in Big Data Analytics. *Procedia Computer science*, 369-370.

Malak, H. A. (2023). What is Information Lifecycle Management (ILM)? *The ECM Consultant*, 1-5.

Onyegbula , B. (2023). What is Data Quality Management? DQM Explained. *Splunk*, 1-2.

Robert, S. (2022). Information lifecycle management (ILM). *Techtarget*, 1.

Simplilearn. (2023). What is Data Integration - Popular Methods And Applications. *simplilearn*, 1-5.

Team, C. (2021). Data Governance. *CFI*, 1.

Chia, Y. K., Hong, P., Bing, L., & Poria, S. (2023). Instructeval: Towards holistic evaluation of instruction-tuned large language models. *arXiv preprint arXiv:2306.04757*.

Rahul, K., Banyal, R. K., & Arora, N. (2023). A systematic review on big data applications and scope for industrial processing and healthcare sectors. *Journal of Big Data*, *10*(1), 133.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148-152.

Chamberlain, A. (2013). *Using aspects of data governance frameworks to manage big data as an asset* (Doctoral dissertation, University of Oregon).

Rouse, P., Harrison, J., & Chen, L. (2010). Data envelopment analysis: a practical tool to measure performance. *Australian Accounting Review*, *20*(2), 165-177.

Breaux, T. D., & Alspaugh, T. A. (2011, August). Governance and accountability in the new data ecology. In *2011 Fourth International Workshop on Requirements Engineering and Law* (pp. 5-14). IEEE.

Breaux, T. D., & Alspaugh, T. A. (2011, August). Governance and accountability in the new data ecology. In *2011 Fourth International Workshop on Requirements Engineering and Law* (pp. 5-14). IEEE.

da Silva, M. D. M., Gamatié, A., Sassatelli, G., Poss, M., & Robert, M. (2022). Optimization of data and energy migrations in mini data centers for carbon-neutral computing. *IEEE Transactions on Sustainable Computing*, *8*(1), 68-81.

Sheikh, A., Malik, A., & Adnan, R. (2023). Evolution of research data management in academic libraries: A review of the literature. *Information Development*, 02666669231157405.

Rifaie, M., Alhajj, R., & Ridley, M. (2009, December). Data governance strategy: A key issue in building enterprise data warehouse. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (pp. 587-591).

Storey, V. C., Dewan, R. M., & Freimer, M. (2012). Data quality: Setting organizational policies. *Decision Support Systems*, *54*(1), 434-442.

Parssian, A., Sarkar, S., & Jacob, V. S. (2004). Assessing data quality for information products: impact of selection, projection, and Cartesian product. *Management Science*, *50*(7), 967-982.

Karremann, I., Zwierlein, C., & Groote, I. M. (Eds.). (2012). *Forgetting faith?: negotiating confessional conflict in early modern Europe* (Vol. 29). Walter de Gruyter.

Coleman, D. W., Hughes, A. A., & Perry, W. D. (2009, March). The role of data governance to relieve information sharing impairments in the federal government. In *2009 WRI World Congress on Computer Science and Information Engineering* (Vol. 4, pp. 267-271). IEEE.

Shih, C. C., Chen, C. Y., Yen-Hua, S., & Lin, S. H. (2010, April). Adaptive ecology m-learning for national park based on Scaffolding Theory. In *IEEE EDUCON 2010 Conference* (pp. 575-583). IEEE.

Thangavelu, M., Albarico, K., Enomoto, J., Doyle, M., Harmon, C., Mekonnen, E., ... & Hsu, B. (2009). Return to the Moon: Looking Glass 204. In *AIAA SPACE 2009 Conference & Exposition* (p. 6612).

Hoehle, H., Scornavacca, E., & Huff, S. (2012). Three decades of research on consumer adoption and utilization of electronic banking channels: A literature analysis. *Decision Support Systems*, *54*(1), 122-132.

McCallion, P., Swinburne, J., Burke, E., McGlinchey, E., & McCarron, M. (2013). Understanding the similarities and differences in aging with an intellectual disability: linking Irish general population and intellectual disability datasets. In *International Review of Research in Developmental Disabilities* (Vol. 45, pp. 225-256). Academic Press.

George, G., Haas, M. R., & Pentland, A. (2014). Academy of Management. *Briarcliff Manor, NY*.

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., & Traweek, S. (2014, September). The ups and downs of knowledge infrastructures in science: Implications for data management. In *IEEE/ACM Joint Conference on Digital Libraries* (pp. 257-266). IEEE.

Aydinoglu, A. U., Dogan, G., & Taskin, Z. (2017). Research data management in Turkey: Perceptions and practices. *Library Hi Tech*, *35*(2), 271-289.

Calvert, M., Thwaites, R., Kyte, D., & Devlin, N. (2015). Putting patient-reported outcomes on the 'Big Data Road Map'. *Journal of the Royal Society of Medicine*, *108*(8), 299-303.

Dorner, D. G., Gorman, G. E., & Calvert, P. J. (2014). *Information needs analysis: Principles and practice in information organizations*. Facet Publishing.

de Melo, B. A. G., Motta, F. L., & Santana, M. H. A. (2016). Humic acids: Structural properties and multiple functionalities for novel technological developments. *Materials Science and Engineering: C*, *62*, 967-974.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, *57*(7), 86-94.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Church, A. H., & DUTTA, S. The promise of big data for OD. *OD Practitioner*, *45*.

Park, S. (2020). *Marketing management* (Vol. 3). Seohee Academy.

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary vs. secondary. *Encyclopedia of social measurement*, *1*(1), 593-599.

Ticharwa, P. (2017). *Representation of environmental awareness communication campaigns: case of the Environmental Management Agency* (Doctoral dissertation, Midlands State University).

Dautovic, E., Gambacorta, L., & Reghezza, A. (2023). *Supervisory policy stimulus: evidence from the euro area dividend recommendation* (No. 2796). ECB Working Paper.

Obatolu, R. (2022). *A holistic framework for complex big data governance* (Doctoral dissertation, University of East London).

Igwe, P. A., Hack-Polay, D., Mendy, J., Fuller, T., & Lock, D. (2021). Improving higher education standards through reengineering in West African universities–A case study of Nigeria. *Studies in Higher Education*, *46*(8), 1635-1648.