HARNESSING DARK DATA FOR EFFECTIVE CORPORATE GOVERNANCE AT THE KENYA ELECTRICITY TRANSMISSION COMPANY LIMITED

BY

GLORIA REGINA MUKHONGO

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN RECORDS AND ARCHIVES MANAGEMENT, DEPARTMENT OF LIBRARY, RECORDS MANAGEMENT AND INFORMATION STUDIES, SCHOOL OF INFORMATION SCIENCES

MOI UNIVERSITY

ELDORET

2025

DECLARATION

DECLARATION BY THE CANDIDATE:

This thesis is my original work and has not been submitted for examination to any other university. No part of this thesis may be reproduced without prior written permission of the author and/or Moi University.

Sign……………………………………. Date…………………………………

Gloria Regina Mukhongo

IS/MSC/RAM/07/016

DECLARATION BY THE SUPERVISORS:

This thesis has been submitted for examination with our approval as university supervisors.

Sign……………………………………. Date…………………………………

Dr. Evans Munge Mwangi
Department of Library, Records Management and Information Studies
School of Information Sciences
Moi University

Sign……………………………………Date…………………………………

Dr. Elsebah Maseh
Department of Library, Records Management and Information Studies
School of Information Sciences
Moi University

ABSTRACT

Dark data is an emerging area in the field of records and information management. It entails the information assets organizations collect, process, store, access, use and share during regular business activities, however fail to use for any other purposes, yet it has the potential to be a very useful tool for informing business decisions. At the Kenya Electricity Transmission Company Limited (KETRACO), poor management of dark data continues to be a risk as evidenced through constant exposes in the media about delayed and alleged fraudulent compensation of land owners resulting to project delays loss of public trust and finance among others. This has been blamed on the lack of proper structures for data management. This necessitated this study which sought to investigate harnessing of dark data for effective corporate governance at KETRACO. The objectives of the study were to: examine the context within which dark data is generated at KETRACO; determine the existence of policies and standards that regulate harnessing of dark data at KETRACO; find out the implications of the risks and opportunities associated with dark data to corporate governance, and propose a framework for the effective management of dark data so that it can be harnessed to enhance corporate governance at KETRACO. The study was informed by the Data Curation Centre (DCC) Lifecycle Model which addresses the management of data throughout the lifecycle. The study was qualitative in nature and utilized an exploratory case study research design. Purposive sampling was used to select 91 respondents from a target population of 569 in all nine directorates at KETRACO. Personal interviews and document analysis were used to generate data, which was then analyzed using thematic analysis. The study confirmed that there were policies in KETRACO that could be used to harness dark data for corporate governance, but majority of the staff were either not aware of them or were not conversant with the policies' contents. Additionally, the findings revealed that dark data at KETRACO was accumulating in different locations and formats with no specific individuals responsible for its effective management. Findings also showed that dark data comes with risks and opportunities that have implications on corporate governance such as non-compliance and efficient service delivery respectively. The study concluded that failure to effectively manage dark data was a serious challenge to effective corporate governance at KETRACO, and thus the need for strategies to harness it. Key recommendations include: deployment of a centralized data management system for controlling the accumulation of dark data; maximizing the implications of the opportunities posed by dark data, while mitigating its risks to corporate governance; adequate sensitization of stakeholders on data management policies in the Company for effective corporate governance; and enhancement of strategies already in place for harnessing dark data in addition to existing policies and systems. This study is significant because it provides a framework for harnessing dark data for effective corporate governance, emphasizing adequate sensitization of data handlers on the importance of dark data management to effective corporate governance, promoting a data-sharing culture; and ensuring adherence to dark data management policies.

DEDICATION

This research project is dedicated to my parents for the foundation they laid for my education despite all the socio-economic challenges. My life would have unfolded in a different way had it not been for your sacrifices and encouragement.

I also dedicate this thesis to my siblings who kept encouraging and criticizing me as I pursued University education.

ACKNOWLEDGEMENTS

Special appreciation goes to my supervisors, Dr. Evans Munge and Dr. Elsebah Maseh of Moi University whose guidance, support and patience, has enabled this work to be successfully accomplished. May the almighty God richly reward you.

More gratitude to the late Professor Justus Wamukoya who guided me through the proposal and would have loved to see the conclusion of this Thesis. May he rest in the safe and loving arms of God Almighty.

I extend my sincere thanks to KETRACO for accepting my request to conduct my research there.

Above all, I thank the almighty God for giving me the strength and ability to prepare this Thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CCTV	Closed-Circuit Television
CIPU	Critical Infrastructure Protection Unit
GDC	Geothermal Development Company
GM	General Manager
GPS	Global Positioning System
ICT	Information Communication Technology
ILIS	Integrated Land Information System
IoT	Internet of Things
KETRACO	Kenya Electricity Transmission Company
KNEB	Kenya Nuclear Electricity Board
NEMA	National Environmental Management Authority
NuPEA	Nuclear Power and Energy Agency
REA	Rural Electrification Authority
REREC	Rural Electrification and Renewable Energy Corporation
SAP ERP	System Applications and Products - Enterprise Resource Planning System
SM	Senior Manager
GDPR	General Data Protection Rule

CHAPTER ONE

INTRODUCTION AND BACKGROUND TO THE STUDY

1.1 Introduction

Dark data is a subject which presents both potential problems and exciting opportunities for businesses because, while it holds hidden insights that could improve decision making, it may also lead to compliance, security and storage risks (Marumolwa and Marnewick, 2025). More so, the most important asset in any organization today is data. Organizations continue to swim in oceans of varied types of data; current, ancient, structured, unstructured, sensitive and trivial which they collect or create in the course of their various business transactions. Besides data, organizations accumulate and use other types of information and records for management, decision making and business intelligence among other reasons. However, not all this is used beyond the particular reason they were created or acquired for despite some having the ability to translate to records that can be used for documentary purposes. For example, visitors' details may be collected at the reception for the purposes of performance contract reporting only, yet if this data is analyzed it can be used to establish trends such as types of frequent visitors and their nature of the services they need. As a result, the organization will be able to get insights into customer journeys, identify customer dissatisfaction earlier, and solve customer problems faster (Feliu, 2020). In some instances, most of these data that is continuously collected by organizations is never used at all. Gartner (2025) and Martin (2018) describe such data as dark data, which entails information assets that organizations collect, process and store during regular business activities, but generally fail to use for other purposes other than the ones they were created for. Van Rees (2020) observes that dark data is generated within and outside of the organization. The sources of such data vary from organization to organization, from personal phones to personal

computers, CCTVs, biometric devices, log in/out devices, visitors' details form and registers, contract documents, social media, call records and free-text customer feedback among others.

Despite the huge rise in the generation, capture and accumulation of dark data, many organizations have been unable to discover or indeed take advantage of the opportunities that dark data can present to them, such as business intelligence, decision-making, and gaining competitive advantage. For an organization such as Kenya Electricity Transmission Company (KETRACO), leveraging dark data for corporate governance would be the ultimate benefit.

1.1.1 The Concept of Dark Data

Data is becoming the main asset for corporate governance today. As a result, the rate at which corporates are creating and acquiring and storing data both on paper and electronically is multiplying exponentially. IBM (2025) notes that unfortunately, a significant amount of these available and stored data is never analyzed nor operationalized. Rather, it remains in the dark, thus dark data. Van Rees (2020) explains that the term Dark Data has not been around too long, giving thanks to its older brother Big Data, which is the word everyone uses when they try to understand the complexity of just how much data they are producing and how they can start to make sense of it all. It is in these organizations' big data platforms that lies dark data forming the largest portion of over 80% (Van Rees, 2020). In fact, Trivedi and Gokulnath (2017) argue that as corporate data increases (forming big data platform), dark data also increases.

Raza (2025) and Shacklett (2025) say that electronic dark data is found in unmanaged and ungoverned repositories, as well as third-party vendors who house data on behalf of corporates in databases that may provide customer information, log files, previous

employee information, raw survey data, financial statements, email correspondences, IP addresses, account information, feedback forms, notes or presentations and old versions of relevant documents. Data Protection Education Ltd (2024), and Shacklett (2025) further points out that paper-based dark data is evidenced in organizations by unmanaged paper records located in abandoned practice group work rooms, storage areas, abandoned file cabinets, offsite storage, staff home offices, and staff secret file locations. This shows that corporate dark data is created everywhere and stored anywhere. Shacklett (2025) and Data Protection Ltd (2024) note that not all dark data is digital because sometimes people leave paper everywhere and the ultimate dark data may well be sitting in a third-party warehouse that businesses are billed monthly for but have almost forgotten about it. These statements dispel the myth that dark data is only electronic. Therefore, the irony of dark data is that at least one individual or stakeholder in a corporation is initially aware of it. This is because it is majorly a result of the data clutter that accumulates from each employee or stakeholder creating, sending and storing more unstructured content, as well as departments/divisions' different information needs among others.

On their part, Angelo (2024) and Martin (2018) observe further that dark data may be found in records an organization pays to store, protect and manage but are not effectively used to add value to the business. These include old files, deleted files, client phone numbers and completed contracts. Enterprise data capacity is growing at a high rate of annually, making the ability to manage the growth of such data even more challenging and potentially prohibitive (Shacklett, 2024). IBM (2015) estimated that 90 percent of this big data generated in organizations content systems like mobile devices, local computer drives, email, network file shares, legacy paper files, cloud file sharing services and even structured databases, such as the document management system is

never analyzed or acted on and as a result never used. Dimitrov (2018) adds to these by saying that dark data may also entail data generated naturally by the operating systems such as non-cleaned recycle bin in Windows, Linux and in UNIX, memory caches, disk caches, database engines cashes, proxy's cache and many more. Panorama (2018) likens dark data to a gate crusher at a party, who either turns out to be useful, or costly and risky.

Moreover, according to Raza (2025), dark data is predominately uncategorized, and has limited visibility to the organization if not completely obscured. Because of its obscurity, it serves no clear purpose. Domo (2021) goes on to posit that dark data entails unstructured data, such as real-time communications (like emails and instant messaging) and documents (like MS Word and excel spreadsheets), semi-structured data, like XML code, or structured data, like the one in a database, which organizations continue to incur the cost of storage when they might probably never use it at all. It may even contain duplicate copies of data.

Purushothaman (2019) goes on to place dark data into two major categories, structured and unstructured dark data. Structured dark data entails contracts and reports of an organization, which become dark over time while unstructured dark data entails bits of information like personally identifiable information like birth dates and billing details. Sarkar (2017) further classifies dark data into three categories. These are:

a) Traditional unstructured data – This is the information that an organization collects, processes, and stores during regular business activities, but generally remains unused for anything (iDashboards, 2019). Sarkar (2017) goes on to say that it encompasses untapped data which remains available with the organizations but is not explored and remains idle like emails, documents and messages among others.

b) Non-traditional unstructured data– This entails data attached or related to videos, images and audio files among others that cannot be mined and analyzed using traditional analytics techniques but requires more advanced technologies, such as computer vision, advanced pattern recognition, and video and sound analytics (iDashboards, 2019).

c) Data in the deep web – According to iDashboards (2019) and Sarkar (2017) this is the untapped information covering data from academics, government agencies, communities, and other third-party that is often hidden behind firewalls and requires specialized tools, vendors, or techniques to collect and analyze it. Data in the deep web is dark because it is not discoverable by means of standard search engines, including password-protected or dynamic pages and encrypted networks (Stevens, 2025).

The foregoing demonstrates not only the different dimensions of dark data and its complexity, but also the enormous potential it has in enhancing organizational growth.

1.1.2   Dark Data in Corporate Governance

Cilia (2025) describes governance as the way power is exercised over corporate entities in the management of economic and social resources for sustainable human development. Governance entails processes, systems, practices and procedures with an aim of acting to achieve the public good, rule of law, transparent processes, and a strong civil society that participates in public affairs (Canadian Audit and Accountability Foundation, 2024). The Governance Institute of Australia (2024) adds on that governance encompasses the system by which an organization is controlled and operates, and the mechanisms by which it, and its people, are held to account whereby

ethics, risk management, compliance and administration are all elements of governance.

On the other hand, corporate governance is the system that guides the control and management of corporations by specifying the various roles and duties of directors, senior executives, shareholders and other stakeholders in a corporation to ensure responsibility and accountability of the corporation's leaders in exercising their authority (CCG 2024 and Crawford, 2007). In addition, corporate governance is the framework of rules and practices by which a Board of Directors or Trustees or Coucil ensures accountability, fairness, and transparency in a company's relationship with all shareholders, community, financiers, customers, suppliers, employees and the Government with the aim of mitigating any conflict of interests (Kengen, 2024). Barac (2001) in Ngoepe and Ngulube (2013) contends that corporate governance comprises structures, processes, cultures and systems that stimulate the successful operation of organizations by putting in place components such as internal audit, audit committee, internal controls and a records-management program.

Day (2025) and Crawford (2007) explain that good corporate governance steers the company towards useful objectives and helps in monitoring and measuring the progress in achieving those objectives. It puts systems and processes in place to ensure accountability, transparency, fairness, uncompromised responsibility and proper probity in the conduct of an organization's business and when dealing with stakeholders (CCG, 2024 and Duffy, 2025).

Agwatu, Ajubare and Adumade (2025) posit that corporates create/acquire different types of data in the course of their transactions. However, the existing records management standards, practices and systems in these corporates are inadequate and

undermine the contribution this data could make in support of the governance functions. Moorthy (2018) also notes that dark data has existed ever since in organizations. This is as a result of organizations collecting and storing data with an aim of using it to achieve their corporate management objectives, yet they are not well equipped to deal with this constantly flooding data. Eventually, data is left unexploited, which translates to either an opportunity lost or a potential security risk.

However, based on an organization's intent and investment thirst, dark data can either be tapped to generate more opportunities or remain in the dark (Moorthy, 2018). In fact, Sarkar (2017) agrees that insights drawn from dark data may help in decision-making, creating new target paths for the future, giving new opportunities, reducing risk and increasing return on investment. Concurring with Mittal (2018), deep within the astonishing volumes of data generated by business transactions, search engines, social media, IoT and countless other sources, there lies valuable intelligence about customers, markets and organizations waiting to be discovered.

The accumulation of dark data impacts greatly on the role of records management in organizations. Shepherd and Yeo (2003) in Erima and Wamukoya (2012) assert that organizations create and manage records with an aim of providing evidence for some specific activity regardless of the medium or forms, which may include documentaries, databases, photographs and audiovisual.

Therefore, the data gathered in this study will help to understand the extent to which dark data can be analyzed to produce records that can be used to enhance corporate governance.

1.2 Background to the Kenya Electricity Transmission Company Limited

The Kenya Electricity Transmission Company Limited (KETRACO) is a public company which was incorporated on 2nd December 2008 and registered under the Companies Act, (Cap 486 of the Laws of Kenya) pursuant to Sessional Paper No. 4 of 2004 on Energy (KETRACO, 2020). Its core function is to plan, design, construct, operate and maintain high voltage electricity transmission grid and regional power interconnectors that form the backbone of the National Transmission Grid.

The vision of KETRACO is to be a world-class electricity transmission company and the leading interconnector in Africa. Its mission is to provide reliable, efficient and effective electricity transmission and promote power trade for sustainable socio-economic development (KETRACO, 2020). It comprises nine directorates, namely: Design and Construction; Project Development Services; System Operations and Power Management; Company Secretary and Legal Services; Human Resources and Administration; Supply Chain Management; Strategy, Research and Compliance; Internal Audit; and Finance. Records Management section headed by the Company Records Manager, falls under the Human Resource and Administration Directorate. To achieve the organization's vision and mission, the various directorates create/acquire a variety of records while carrying out their daily business transactions. Some of this data may provide great opportunities for the organization's attainment of its mandate if processed and managed properly. On the other hand, they may hold potential risks to KETRACO, risks that can only be exposed if it's analyzed.

From a corporate management perspective, KETRACO continues to create and maintain data, records and other information for legal, administrative and historical value.

The records management function is decentralized at KETRACO. Each directorate's data, information or record is managed by the directorate's secretary who are guided by the records personnel who are under the leadership of the records manager. Data is continuously being created or received in KETRACO from different sources, locations and devices. These include data created or received by officers in the field as well as those in the offices.

1.2.1   Dark data at KETRACO

As a result of decentralization of records management at KETRACO, data is generated or captured in different departments during the daily business transactions. This data comes in different formats and sizes as well as from different sources both externally and internally. With lack of staff mandated to handle this data, it ends up being locally within these different departments as well as other remote locations such as staff personal electronic devices and desks/workstations depending on who acquired it. With time the siloed is rarely shared among individuals within departments of or even across departments. Eventually this data is forgotten about, becoming dark data.

1.3 Problem Statement

Kenya Electricity Transmission Company (KETRACO) has faced persistent challenges related to irregular payments, including delayed compensation to landowners and contractors, overpayments, and disputed claims. Investigations by Nzomo (2025) in The Kenyan Wall Street reveal that these challenges are compounded by audit findings suggesting corruption and unauthorized payments, such as duplicate compensation to some landowners and losses in legal suits stemming from mismanagement. Further, Mwere (2025) highlights that unresolved compensation disputes are delaying project completion, with root causes including inadequate documentation, absence of clear compensation guidelines, unapproved valuation schedules, and unsupported payment

vouchers. These deficiencies undermine the accuracy, reliability, and legitimacy of compensation processes, despite KETRACO's existing procedural frameworks.

Central to this issue is the management of data—particularly "dark data," which refers to unstructured, unmanaged, or unanalyzed information that is generated or received but not effectively utilized. KETRACO's failure to adequately capture, curate, and leverage dark data has contributed to persistent irregularities in payments, operational disruptions, financial losses, and erosion of public trust. This data management gap hinders transparency and accountability, compromising corporate governance and risking compliance breaches that threaten continued donor funding, given that many of KETRACO's projects are externally financed (Streamline, 2025).

KETRACO's inability to provide comprehensive supporting documentation for compensation payments—including those made to unaffected landowners—and the lack of verification of pending payments exemplify systemic weaknesses in data governance. The company's contested responses to media reports further highlight the critical need for transparent, accurate, and accessible data management systems that can validate compliance and inform decision-making.

Addressing the failure to harness dark data presents significant opportunities for KETRACO. Effective management of all organizational data—from creation and appraisal to storage and sharing—can enhance the integrity of compensation processes, reduce disputes, accelerate project timelines, and improve legal and financial outcomes. Moreover, it can rebuild stakeholder trust and ensure adherence to funding conditions by fostering transparency, equity, and accountability. Harnessing dark data for corporate governance thus represents a strategic avenue for strengthening institutional performance and sustainable project delivery.

Accordingly, this study aims to investigate how dark data generated within KETRACO can be systematically managed and leveraged to enhance corporate governance. This inquiry is vital for developing data-driven policies and governance frameworks that promote compliance, operational efficiency, and public trust in the company's management of compensation and related processes.

1.4 Aim of the Study

The study sought to investigate the harnessing of dark data for effective corporate governance at KETRACO with a view to propose strategies that can be used to manage, promote access, use and sharing.

1.5 Objective of the Study

The objectives of the study were to:

a) Examine the context within which dark data is generated at Kenya Electricity Transmission Company (KETRACO).

b) Determine the existence of policies and standards that regulate management of dark data at KETRACO.

c) Find out the implications of the risks and opportunities associated with dark data on corporate governance at KETRACO.

d) Propose strategies for harnessing dark data to improve corporate governance at KETRACO.

1.6 Research Questions

The research questions were:

a) In what context is dark data created, captured and/or accumulated at KETRACO?

b) Are there policies and standards that exist to support harnessing of dark data for corporate governance at KETRACO?

c) What are the implications of the risks associated with dark data on corporate governance at KETRACO?

d) What are the implications of the opportunities posed by dark data on corporate governance at KETRACO?

e) What strategies can be adopted to harness dark data for improved corporate governance at KETRACO?

1.7 Assumptions

The study was based on the assumptions that:

- The management of dark data at KETRACO is inadequate and unproductive, and this has impacted negatively on the company's corporate governance.

- The absence of a structured method of managing dark data at KETRACO undermines its usage for effective corporate governance.

- KETRACO wished to develop new strategies or/and enhance the ones already in place to harness dark data for effective corporate governance.

1.8 Significance of the Study

The significance of this study is demonstrated by its theoretical implication, its practical importance and how it influences policy.

1.8.1 Theoretical Significance

Dark data is a unique topic in the world today in matters of information and records management. Hence the research will contribute to the broader discipline of information sciences, more specifically to the field of information and records management. This is by providing a body of knowledge for the interest of both

researchers and academicians who seek to explore or investigate how dark data can be harnessed for corporate governance.

1.8.2   Practical Significance

This research will aid KETRACO and other corporations in coming up with strategies for harnessing dark data for effective corporate governance. Other Organizations may also benefit from the outcomes and recommendations of this research.

1.8.3   Policy Significance

The study will provide guidance not only to KETRACO but other corporations during policy formulation and implementation on management of dark data as an information resource for effective corporate governance.

1.9 Scope of the Study

Bhosale (2025) explains that any research study being undertaken must have distinct boundaries indicating the start and end to make the study manageable by focusing the work of the study down to what is practically achievable within a given timeframe. Therefore, the scope of this study was KETRACO, which is a 100% Government owned entity being regulated by the *State Corporations Act, Cap 446 2010 (KE)*. It has regional offices, but the research will be confined to the head office. This is because, the staff at the regional offices are mostly doing fieldwork, thus hard to reach due to limitations such as poor network connection and tight schedules. In addition, the staff in the headquarters represents all directorates, managerial level and regional staff. Furthermore, policies and strategies are designed in headquarter offices and then implementation is cascaded to regional offices.

The study interconnected the concepts of dark data and corporate governance while bounding itself to dark dart as unused, unstructured, or semi-structured data held by

KETRACO in both electronic and physical formats. Additionally, the study also bounded itself to Corporate Governance in the lenses of risk management and strategic decision-making. This is whereby the research investigated how dark data, if harnessed, could identify operational risks, inform long-term infrastructure planning, and improve audit trails.

The research adopted a qualitative, case study design to achieve an in-depth understanding of dark data within KETRACO. The sample was purposively selected based on officers most likely to handle dark data in the top-level, middle-level and low-level management from various directorates in KETRACO. Data was collected using semi-structured interviews and document analysis of policies related to data management and governance in KETRACO.

1.10   Limitations of the Study

Rautenbach and Jansen (2024) discuss that limitations to the study are inevitable are the constraints and challenges that will invariably influence your ability to conduct your study and draw reliable conclusions. Good (2024) explains that these conditions or influences on the research cannot be controlled by the researcher but have an impact on the results of the research, especially the validity. They may include lack of previous research on the topic requiring the researcher to do more, inaccessibility to respondents, time constraints and a naturally restricting sample among other issues.

In the context of this study, the key limitations were as follows:

First, heads of departments granting access to the various directorates that hold sensitive information. This was because they were uncertain on how the researcher was going to store, analyze and use the information. However, the researcher through

the introductory letter assured them that the information was purely for academic purposes and was going to be treated with uttermost confidentiality.

Additionally, the researcher found it hard to tell whether the informants were responding to the questions truthfully and fully. However, with the analysis of the findings a certain trend was established, whereby almost similar responses to some questions were observed from different respondents. This was also supported by document analysis to confirm viability of the information the respondents gave about the policies.

Finally, there was sensitivity by KETRACO that some of the information from the findings could be divulged by the researcher. KETRACO, being a business company with competitors, was not readily open to allowing the researcher to look/discuss certain sensitive information, nor allow its staff to participate in a study that discusses company information. The researcher assured respondents of anonymity and even explained to them how the findings were going to be used. Respondents were informed that the findings were to be used for academic purposes and the same were to be shared with KETRACO as well.

1.11   Definition of Operational Terms

Big Data

Badman and Kosinsik (2024) describe massive, complex data sets that traditional data management systems cannot handle, however when managed and analyzed, it can help organizations discover new insights and make better business decisions. In the context of this study, big data is the collection of significantly large volumes of data in unstructured, semi-structured and structured that may contain hidden insight or

intelligence to an organization which can only be discovered upon proper management and analysis.

Corporate Governance

Duffy (2025) explains that corporate governance is a set of processes, customs, policies, laws and institutions affecting the way a corporate is directed, administered and controlled. Based on this research, corporate governance is a governance system at KETRACO that establishes clear structures regarding accountability, responsibility and transparency while defining roles and rights of different stakeholders in the organization. This is through providing structures by which KETRACO set objectives, attain these objectives and monitor their performance.

Data

Data is a collection of facts, numbers, words, observations or other useful information which when processed and analyzed, is transformed into valuable insights that improve decision making and drive better business outcomes (Badman and Kosinsik, 2024). However, in the context of this study data is viewed as unit of a record.

Dark Data

Acing to Techopedia (2018) dark data is a type of unstructured, untagged and untapped data that is found in data repositories and has not been analyzed or processed. However, in the context of this research, dark data represented data that is found in log files and data archives stored within and without the enterprise's data storage locations and includes all data objects and types that are yet to be analyzed for any business decision making.

Deep Web

According to Group-IB (2025), deep web refers to parts of the World Wide Web that are not discoverable by standard search engines such as content behind password-protected logins, dynamic pages generated in real time, and encrypted networks. On the other hand, the researcher uses deep web as large quantity of data created or received by KETRACO various stakeholders through different platforms that are insightful to the Company only if the right search tools/mechanisms are employed to reach this data.

Electronic Waste (E-waste)

National Environmental Management Authority (2021) describes e-waste as waste resulting from electrical and electronic equipment including components and sub-assemblies thereof after the end of their useful life either through redundancy, replacement, or breakage. The research used this term to represent the electronic devices that lie in offices and staff desks but are no longer actively in use at KETRACO due to obsoleteness, breakage and redundancy such as system units, old tablets and laptops.

Harness

The researcher uses the word harness mean how dark data can be collected, processed, transformed, and analyzed for effective corporate governance. It represents KETRACO exploiting dark data to make predictions about its customers, the market, and itself for its own advantage.

Record

Records are both evidence of business activity and information assets (ISO, 2016). The researcher used a record as a composition of related evidential data about a person,

processes, among other things created and maintained by KETRACO in the transaction of its core business.

1.12   Chapter Summary

This chapter gives the reader the context of the thesis whereby it discusses the background of dark data and corporate governance, particularly at KETRACO. It outlines the main aim of the study as investigating how dark data can be harnessed with a view of proposing a framework that can be used to manage it for effective corporate governance. The objectives of the study, research questions, and the significance of the study are also outlined in this chapter. It goes on to provide the assumptions of the study as well as the scope and limitation. Finally, it gives the definition of key terms in the thesis. The key theme that emerges in this chapter is dark data as records that needs appropriate analysis and processing for it to enhance corporate governance. However, because these records are dark, they are never used productively.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Literature review is a discussion and synthesis of published information in a particular subject area with an aim of giving new interpretation to old materials or combining new with old interpretations, tracing the progress of a particular subject area, and sometimes evaluate sources to advise the reader on the most pertinent or relevant ones (University of North Carolina, 2025). As posited by University of southern Carolina (2025), in academic writing, literature reviews are used to provide a foundation and support for new insight by giving a background of the research papers investigation. The sources of literature review include books, research articles, abstracts, reports, dissertations, policy documents and the internet among others. Print and electronic primary literature such as reports and previous studies, as well as secondary information sources such as books are reviewed in this chapter.

From the literature reviewed, the rate at which organizations are accumulating dark data is becoming a great issue of concern. To get a comprehensive overview of this literature, the current chapter starts with a theoretical framework for the study and reviewing of related literature on dark data. Most of it focuses on the types of dark data, how dark data is accumulated, and dark data's potential benefits and risks to corporate management.

Given the focus of this study on harnessing dark data for corporate governance, a model deemed to be core was selected to help understand how dark data can be harnessed for corporate governance.

2.2 Theoretical Framework

According to George (2022), a theoretical framework is a foundational review of existing theories that serves as a roadmap for developing the arguments you will use in your own work. It serves as a structure and support for the justification of the study, the approach to the problem, the purpose, the importance and the research questions while providing a basis, or an anchor, for literature review and, above all, for methods and analysis (Precision Consulting, LLC, 2024). Therefore, one model related to the research topic was reviewed. This is the Data Curation Centre (DCC) Lifecycle Model.

2.2.1   Data Curation Centre (DCC) Lifecycle Model

The Digital Curation Centre (2022) observes that there is need to ensure continuity of digital materials, while maintaining their authenticity, reliability, integrity and usability. This can only be achieved through data curation and preservation mechanisms by data creators, data curators and data users. Thus, the data curation center (DCC) lifecycle model created by the Digital Curation Centre with an aim to solve the extensive challenges of digital preservation and digital curation (University of California, 2019). This model identifies curation actions applicable across the whole digital lifecycle; those which need to be undertaken sequentially if curation is to be successful and those which are undertaken occasionally, as circumstances dictate. This model also outlines logical steps of receiving, appraising, selecting or disposing of data. This is then followed by ingesting and subsequent actions such as preservation, storage, access, and possibly transformations or reappraisals of the data (Digital Curation Centre, 2022).
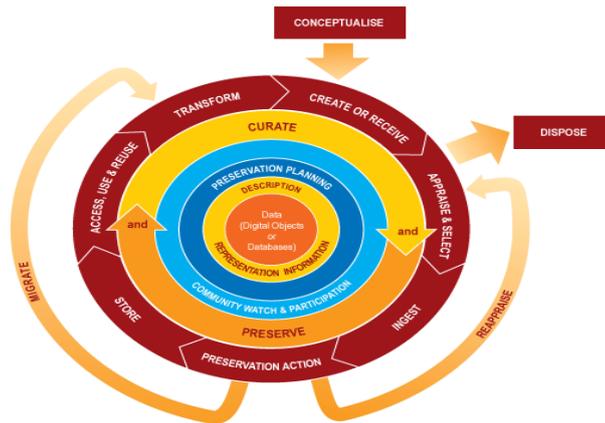
Figure 2.2: Data Curation Center (DCC) Lifecycle Model

Source: Huggin (2008) in Digital Curation Centre (2022)

According to Figure 2.2, Digital Curation Centre (2022) outlines the following as elements of the Data Curation Centre Lifecycle Model:

a) Simple digital objects such as text, image or sound files, with their related identifiers and metadata.

b) Complex digital objects. These are discrete digital objects that are made by combining a number of other digital objects, such as websites.

c) Databases. These entail structured collections of records or data stored in a computer system.

Full life cycle actions

These are activities that take place at any time during the digital curation lifecycle and are relevant to many different sequential actions. These are:

Description and Representation Information - This entails using appropriate standard to assign appropriate metadata (administrative, descriptive, technical and structural) to

ensure long-term preservation while reducing the risk of loss of information integrity (Digital Curation Centre, 2022 and Rhee, 2022).

Preservation Planning - Rhee (2022) and Digital Curation Centre (2022) observe that these actions include plans for the management and administration of all curation actions throughout the lifecycle of a digital material. For example, regular backups and data migration.

Community Watch and Participation - Digital Curation Centre (2022) and Rhee (2022) concur that these entail community participating in supervising data creation activities and assisting in the creation of the standards to be used, the tools to create data and appropriate software to create it.

Curate and Preserve - This is when one is aware of and undertakes management and administrative actions that have been planned to promote curation and preservation throughout the curation lifecycle of the digital material (Digital Curation Centre, 2022). These include encouraging best practices such as use of policies and standards (Rhee, 2022).

Sequential actions

These are steps taken repeatedly to ensure that data is curated according to best practice. They are:

Conceptualize – It entails conceiving and planning the creation of data, capture method and storage options (Digital Curation Centre, 2022). This is when strategies and policies are formulated and adopted to address how data will be captured and stored while maintaining its authenticity, integrity and accessibility in both short-term and long-term (Rhee, 2022).

Create and receive – Digital Curation Centre (2022) posits that these actions aim at creating well-formed data that is ready for curation by developing, documenting and applying policies that influence data creators to create curation-friendly data. Curation-friendly data is one that is created in standard formats and file types that can be processed with open-source, well documented programs; has collection and storage metadata (Digital Curation Centre, 2022).

Appraise and select – This comprises creating appraisal policies and evaluating and select data for long-term curation and preservation in adherence to these, policies, documented guidance or legal requirements (Digital Curation Centre, 2022 and Rhee, 2022).

Ingest (take in and transfer) – Digital Curation Centre (2022) asserts that this is the transferring of data to an archive, repository, data center or other custodian while adhering to documented guidance, policies or legal requirements.

Preservation action – It is when actions are taken to ensure long-term preservation and retention of authentic, reliable and usable data and its metadata (Digital Curation Centre, 2022). Preservation actions include data cleaning, validation, assigning preservation metadata and ensuring acceptable data structures or file formats (Digital Curation Centre, 2022).

Store – It entails storing data securely using relevant standards that will enable it to remain usable and easily retrievable (Digital Curation Centre, 2022). This entails ensuring that the integrity of data and its metadata as well as the system and physical security is maintained. Digital Curation Centre (2022) further discusses that storage also entails monitoring activities that might trigger other preservation actions.

Access, use and re-use – This action entails ensuring that data is accessible to both designated users and re-users, when needed (Digital Curation Centre, 2022).

Transform – This comprises creating new data from the original, for example, migration into a different format or creating a subset, by selection or query to create results; as well as technology emulation; information migration (Digital Curation Centre, 2022).

Occasional actions

These are actions that interrupt or reorder the sequential actions as a result of a decision, for example, transfer of data to another archives or data centre, or set for re-appraisal if it does not meet archival standards, or destruction of data (Digital Curation Centre, 2022). The actions include:

Dispose – Digital Curation Centre (2022) posits that this is occurs when the data is transferred to another archive, repository, data centre or other custodian and in some instances, data is destroyed. This should be guided by documented policies, guidance or legal requirements to reduce the risks of data breach.

Reappraise – This is when Data is returned for further appraisal and re-selection as a result of failing validation procedures (Digital Curation Centre, 2022).

Migrate – It entails the use of data policies to drift from one format to another to accord with the storage environment or to ensure the data's immunity from hardware or software obsolescence (Digital Curation Centre, 2022).

2.2.2.1 Application of the Data Curation Center (DCC) Lifecycle Model

This model is relevant to the study as it addresses the following research questions as

shown in table 2.2.

Table 2.1: Application of the Data Curation Center (DCC) Lifecycle Model

Research question	Variables in the Model	Application
How is dark data managed at KETRACO?	1. Create/receive 2. Appraisal/reappraisal 3. Selection 4. Disposal 5. Ingesting 6. Preservation 7. Storage 8. Access/use/reuse 9. Transformations	The DCC lifecycle model looks at issues such as conceiving and planning the creation of data, capture method and storage options. It goes on to emphasize on creation of curation-friendly data by use of policies and standards. It also looks at issues such as appraisal, reappraisal and selection. As a result, this model can assist KETRACO to minimize the accumulation of dark data by only creating or receiving data that is deemed to be relevant.
Are there policies and standards that exist to support harnessing of dark data for corporate governance at KETRACO?	1. Every stage of the life cycle is directed by standards and policies.	This model underscores the use of policies and standards as best practices in data management. For example, it encourages the use of policies to guide on appraisal and selection of data. It goes on to emphasize the use of appropriate standards to assign appropriate metadata to ensure long-term preservation while reducing the risk of loss of information integrity. Use of policies and standards can assist KETRACO to minimize the accumulation of dark data.
What opportunities does dark data present to corporate governance at KETRACO?	1. Preservation 2. Access, use and reuse 3. Storage 4. Transform 5. Migrate	This model looks at the need to provide users at KETRACO with high-quality data which can be proven to be authentic and have integrity for sound decision making. The models posit good practice in curation such as proper description of data in the creation and preservation stage. This metadata illuminates dark data so that it can be used to produce new

		and better results (Digital Curation Centre, 2022).
What strategies canbe adopted to harness dark data for improved corporate governance at KETRACO?	Full life cycle actions, sequential actions and occasional actions	It talks of strategies such as community participation in supervising data creation activities and assisting in the creation of the standards to be used, the tools to create data and appropriate software to create it. It also talks of use of appropriate metadata to ensure long-term data usability and ability to be retrieved. The model goes on to emphasize documentation of data management processes, application of standards and policy framework among others. Using this model at KETRACO will ensure sufficient documentation of data management processes, standards and policies as well as documentation of roles and responsibilities.

Source: Researcher (2025)

2.3 Review of Related Literature

The following themes as gleaned from the objectives of the study are reviewed in this section:

- Impetus for the creation and accumulation of dark data.

- Policies and standards for harnessing dark data.

- Risks and opportunities of dark data to corporate governance.

- Strategies for harnessing dark data as records for corporate governance.

2.3.1 Impetus for the Creation and Accumulation of Dark Data – Dark Data

Future Hosting Technologies LLC (2017) observes that most people hold on to a cliché that 'more data is good'. As a result, companies keep accumulating volumes of information streaming in from business operations, employee and customer activity,

email, social media and many other channels some of which become dark data. The following are some of the reasons as to why organizations are swimming in oceans of dark data.

2.3.1.1 Need to Comply with Business Regulations and Legal Requirements

Companies are trying to reduce the sanctions that come from incomplete or inaccurate disclosure during audit or eDiscovery (Spinner, 2018). Chant (2023) and Spinner (2018) contend that most organizations accumulate dark data in the course of trying to lower the risk of policy breach. This happens when organizations keep obsolete information beyond its scheduled retention period, or when they try to reduce the risk of fines by storing regulated information like personally identifiable information (PII) outside the appropriately secured systems. The main aim here is to try to comply with business and legal requirements such as having complete and available information whenever required, forgetting to analyze its posterity value.

2.3.1.2 Perceived Potential Value of Data

Future Hosting Technologies LLC (2017) discusses that because data is generally deemed to have potential value many organizations opt to store it despite having no idea of how to generate value from it. These organizations hope to figure out what to do with this data one day. For example, Kdnuggets (2019) discusses that organizations collect huge volumes of data such as data on usage of products or website visits, believing it will improve their products or services. However, a large portion of this collected data is never analyzed, resulting to the accumulation of dark data. This data which is kept 'just in case' but doesn't have proper usage in an organization is referred to as dark data. IBM (2025) agrees that the perceived potential and utility value of dark data to the organization explains why many organizations are not at ease to part with

their dark data, even if they have got no plans to put it to work, either in the near term or long-term planning horizon.

2.3.1.3 The "Bring Your Own Device" (BYOD) Phenomena

This is an organizational policy that gives employees authority to bring their personal devices like mobiles, laptops and tablets to get connected over the corporate network, so that they can access the corporate mails as well as the company's confidential data and applications using their personal devices (Comodo Group Inc, 2019 and Monnappa, 2025). With the big data phenomenon, Dimitrov (2018) notes that this practice generates information governance challenges that arise when information is generated by and stored on mobile devices, social networks, file sharing services, and unmanaged SharePoint sites. NetGovern (2022) adds to this that with employees disregarding corporate data policies, corporate storage is then used for dumping personal data, such as personal photos, IDs, legal documents, music, videos, and video games. As a result, dark data accumulates in organizations.

2.3.1.4 The Big Data Phenomenon

Negligence of big data is a major impetus of dark data accumulation. IBM (2025) and Karrmann (2025) opine that big data projects that support customer service, marketing or process improvement initiatives can open the door to a dramatic accumulation of data that quickly goes dormant in big data lakes. These unused or unidentified content that sits outside of the records retention schedules, classification schemes and retrieval systems that organizations rely upon to meet compliance obligations, ease the burden of electronic discovery and ensure decisions are derived from accurate and relevant information can also be described as dark data. Dimitrov (2018) observes that the uncontrolled growth in data volumes and formats plays a major role in the accumulation

of dark data. This is because it makes it increasingly more difficult to discover, retrieve, and reuse trusted information. Purushothaman (2019) goes on to say that these constitutes the largest volume of big data, and its ignored despite being available due to organizations not having bandwidth or technical capability or not believing the data will add value. In fact, SaaSholic (2021) also observe that big data results to the creation of dark data when employees drag and drop all kinds of work-related files easily, as well as personal files such as personal photos, MP3 music files, personal communications and many more in addition.

2.3.1.5 Lopsided Priorities

As discussed by Kdnuggets (2019) this is when organizations fail to assign priority to the data they collect such that it can be used for more purposes other than the main one they were collected for. Kdnuggets (2019) goes on to give an example of the online banking services where the credit card marketing team is focused exclusively on customer details and eligibility while no attention is given to the data on how customers arrive at the application page. If all these aspects are assigned priority, then the details can provide valuable insights on the usability of the bank website and the application page. This will also minimize the duplication of data that is collected by the two teams (i.e., credit card marketing time and website manager).

Answer Miner (2020) observes that a tricky situation for any company is when every interaction, transaction and engagement get captured and the companies need to prioritize which data to utilize and which data to push aside for safe keeping. Often, this results in vast amounts of unstructured or semi-structured data being stored in log files or data archives in case it is required in the future.

2.3.1.6 Disconnect Among Departments

Kdnuggets (2019) observes that in most organizations, departments have their own data collection and storage processes and systems which may not be known to other departments. These data silos result to collections of often overlapping-but-inconsistent data (Stitchdata, 2022). Mwiti et al (2016) further adds that advancement in data capture devices that have enhanced the capture of unstructured data such as voice, audio, video and graphics have also created a large pool of data within and outside organizational boundaries. Waida (2021) agrees that information silos created as a result of sing too many tools at once, so information is spread out and hard to look for. These hampers organizations from creating comprehensive data policies (Pal, 2021). The end result is that data, even if relevant to other departments, lies unused in the dark in another department which more often than not result to the accumulation of dark data.

2.3.1.7 Technology and Tool Constraints

This occurs when organizations use different tools and technologies to collect their data, yet there may be a probability that these technologies and tools don't interact with each other (Kdnuggets 2019). For example, companies with different IT systems and formats may find it hard to integrate audio file contents from call center with click data from websites. Bridge Head (2020) observes that in healthcare, a cardiology department might have all of its data in one system, with little or no access to other departments' files due to lack of system interoperability such that when a cardiologist needs the latest chest x-ray for one of their patients, they might not be able to easily pull the images and data from the radiology department's system. The result is that the data gathered can't be brought together to create a cohesive repository, rather massive amounts of dark data remain in the respective technologies, anonymous, irretrievable and therefore unusable. On the other hand, this lack of interoperability can disrupt or delay vital procedures,

and even put patients' health at risk if important information is not communicated properly (Bridge Head, 2020).

Answer Miner (2020) adds on that often, data stays in the dark because of a company's inability to process it efficiently. For example, it may not be easy to turn data into actionable information if it is stored in a format not supported by the company's analytics tools. This illustrates that dark data may contain crucial information which if analyzed and operationalized, can improve corporate efficiency, revenue and even enable discovery breakthroughs. However, many corporates are finding it hard to harness dark data due to, dark data's extreme volumes, dark data's dynamic schema, dark data's unknown value and many more factors.

Clark (2023) and Commvault (2014) further observe that dark data is mostly complex to analyze and found in unreachable places such as the deep web, as well as on connected devices such as personal spreadsheets, videos such as surveillance footage, text messages, social media, internet, inactive databases and unused customer information, previous employee files, server log files, account information and transaction history among others.

2.3.1.8 Backing up Data

Krause (2020) alludes that as much as the organization do the right thing to regularly back up their data (everything from email servers to file shares to individual user folders), this process accumulates tons of unnecessary unstructured data (dark data). Apart from the organization piling up dark data from its regular back up, Suciu (2023) also observe that employees create dark data when they back up workstation files yet afterwards, they never follow up on a majority of these files leaving them unmonitored and for a longer period than expected. As a result, they end up unmanaged and therefore

never looked at again by the employee or anyone else. As a result, clogs up storage arrays and hard drives with files that are obsolete or duplicates, creating big bills and unknown risks (Krause, 2020). This point is underscored by Munge (2019) who cites research findings from several studies (among them Veritas Technologies (2016) and IBM (2025) which suggest that information and data storage and disposal issues affect organisations world-wide, leading to accumulation of massive blocks of expensive data that clog up storage facilities.

2.3.1.9 Inexpensive Storage

With organizations embracing information technology, storage of electronic data has become inexpensive thus making it easy to store data (Spinner, 2018; Future Hosting Technologies LLC, 2017). Mwiti et al (2016) concur that organizations also accumulate dark data as a result of today's IT storage optimization and the global reduction in the cost of storage whereby organizations can now be able to store too much data and in diverse forms at a very low cost. NetGovern (2022) observes that with the mentality that cloud computing is free and with unlimited storage, most organizations are opting to push all their data (store-all mentality) there without looking at the data value. This simplifies the accumulation of dark data in the cloud. However, it is important to note that storing and securing the data usually entails greater expenses (or even risk) than the potential return profit in case of a breach (Gartner, 2025).

2.3.1.10 Lack f Motivation to Manage Data Properly

Zhao (2024) is of the opinion that today's researchers and scientist are professionally, economically, and socially unmotivated to correctly label data, adapt old and new file formats, and overall keep data in good health. As a result, there is a spontaneous growth

of unstructured data that that is hard to analyze and use. Thus, resulting in the accumulation of dark data.

UpKeep (2022) further observes that the users and other data handlers in organizations lack motivation due to poor leadership from those in charge of ensuring proper data management. UpKeep (2022) sites ambiguous instructions to employees by managers such as "Input quality data." This may seem like a simple request by the manager, yet "quality data" may be interpreted by employee differently such as meaning lots of data, or that they should input anything and everything they can find. This results in an influx of dark data that is not useful and doesn't point towards solving real problems.

2.3.2 Policies and Standards for Harnessing Dark Data

Need for regulatory compliance as well as risks posed by the sporadic accumulation of data in most organizations today has created great concern about ensuring that data is managed properly throughout the information lifecycle. Ho-Dac and Huang (2023) concurs that increasing data breaches have broad impacts on the customer as well as organization, thus the need for effective data governance. This may be achieved by putting in place policy frameworks and standard guidelines to facilitate data management in organizations.

2.3.2.1 Policies for Harnessing Dark Data

Smartsheet Inc (2019) describes a data governance policy as a set of rules for safeguarding an organization's data assets by establishing roles and responsibilities for data that include access, disposal, storage, backup, and protection. On the other hand, TechTarget (2025) describes a data governance policy as a documented set of guidelines for ensuring the proper management of an organization's digital information. It outlines how business activity monitoring should be carried out to ensure

organizational data is accurate, accessible, consistent and protected as well as establishing who is responsible for information under various circumstances and, specifies the procedures to be used to manage it. Anderson (2017) agrees that data governance policies address issues such as: assigning accountability to employees responsible for data assets; controlling access to data; maintaining data accurately and consistently; storing data securely; proper data back-up and protecting data from internal and external threats. TechTarget (2025) and Anderson (2017) concurs that effective implementation of a data management policy should be a cross-disciplinary and company-wide effort. Additionally, Sharma (2025) emphasizes continuous review of selected data governance frameworks such as policies in organizations to keep up with changing business needs and regulations.

However, Morris (2021) notes that data is critical for strategic operations in organizations. Hhstarling (2019) asserts that organizational data need to be of high quality and trustworthy. Morris (2021) suggests the use of formalized system of data governance such as use of policies, protocols, procedures and metrics that provides a methodical approach of dealing with a company's data to help companies avoid or reduce risks and liabilities as well as enable data mining for maximum use. Apart from assisting everyone in an organization to use data effectively, efficiently and responsibly these formalized tools also control how data is used, stored and operated (Morris, 2021).

However, Morris (2021) goes on to note that no specific data management policy can suit all organizations but rather has to be tailor-made to meet particular organization's data needs. A good data governance policy in an organization should focus on internal functionality requirements; data collection methods; quality and size of data; business

goals; data storage framework and extent of automation (Morris, 2021. Therefore, organizations need to come up with their own custom-made data governance policy.

The following are some of the policies and standards that may be used to harness dar::

a) *Privacy and Data protection policy 2018 – Kenya.* This policy aims to ensure privacy and data protection while complying with regulatory requirements; international best practices such as creation of data recovery strategies, classification and documentation (Office of the Data Protection Commissioner, 2021 and Government of Kenya, 2018). This policy may come in handy for harnessing dark data as it addresses issues like collecting data purposefully, storing data for periods not longer than it's necessary and accountability by data controllers/processors.

b) *The National Information and Communication Technology (ICT) Policy Guidelines 2020 – Kenya.* This policy is designed to facilitate universal access to ICT infrastructure and services all over Kenya with an aim of realizing the potential of a digital economy (Kenya Gazette, 2020). While providing public services digitally, this policy posits that data should always be kept in safe custody while ensuring that it is accessible to authorized individuals whenever needed. The policy also looks at Internet of Things (IoT), a factor which contributes to the accumulation of dark data as well. These policy guidelines aim to provide network, storage and processing infrastructure to enable Kenya to compete at a global scale. Theses entail investing in data centers, hardware and software among others. Therefore, the policy may provide a better platform for harnessing dark data.

2.3.2.2 Standards for Harnessing Dark Data

According to Seers (2022), data standards are rules or guidelines used to ensure that data is collected and formatted in a consistent, reliable manner. UConn (2022) and Seers (2022) further explain that data standards give rules for how data should be collected, formatted, and stored. They give technical specifications that describe how data should be stored or exchanged for the consistent collection and interoperability of that data across different systems, sources, and users (Federal Enterprise Data Resources, 2022).

Data standards are used to standardize data format as well as to enable sharing, exchanging, combining and understanding of data (Data Management, 2021). They allow for the reuse of data elements, thus reducing redundancy and enhancing reliability while also bringing down the cost (Satori, 2022). Clarivine (2022) further observes that standards play a major role in breaking down data silos and creating interoperability, or the ability to exchange data between disparate systems and teams.

A universe of standards has been developed that can be used for harnessing data. The following are some of these standards:

a) *ICTA-4.002:2019 – Electronic Records Management Standard* – This standard was developed by the ICT Authority of Kenya to provide guidance on management of electronic records such that they meet the same requirements as their regular paper record counterparts (ICTA, 2019). This standard covers digital objects created by email, word processing, spread sheet and imaging applications (such as text documents, and still and moving images) that are identified to be of business value. This standard also covers storage and formats of records managed by an electronic records management system. According to this standard, ICTA (2019) explains that records may be stored on a variety of

different media formats and may be managed in hybrid record aggregations that include both electronic and non-electronic elements.

b) *ISO 27001: The International Information Security Standard* – This is a best practice framework that sets out the specification for an information security management system (ISMS) by addressing the people, processes and technology (IT Governance, 2022). It comprises policies, procedures, processes and systems that manage information risks, such as cyber-attacks, hacks, data leaks or theft. Compliance with this standard ensures that organizational data, including dark data is secure. ISO 27001 certification assures customers, partners and other stakeholders that the company's information security infrastructure meets their expectations (isms.online, 2022).

c) *ISO/IEC 38505-1:2017: Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data* – ISO (2022) explains that this standard was developed to provides guiding principles for members of governing bodies of organizations on the effective, efficient, and acceptable use of data within their organizations. ISO (2017) observes that with the advent of cloud computing, the realization of the potential of the "internet of things" and the increasing use of "big data" analytics, data is becoming easier to generate, gather, store and mine for useful information. However, this flood of data brings with it an urgent requirement and responsibility for governing bodies to ensure that valuable opportunities are leveraged, and sensitive data is protected and secured. Therefore, if this standard is used, it may increase the value of dark data while decreasing the risk associated with it.

It looks at governance of the current and future use of data that is created, collected, stored or controlled by IT systems. In return, this standard aims to

inform organizations on the use and protection of data; ensure stakeholder confidence in governance of data by organizations; and establish vocabularies for data governance among others.

d) *Dublin Core Metadata Initiative* – According to DCMI Usage Board (2020), this is a descriptive metadata standard encompassing a set of fifteen core elements (properties) plus several dozen properties, classes, datatypes, and vocabulary encoding schemes used for describing networked resources. The core elements are title; subject; description; creator; publisher; contributor; date; type; format; identifier; source; language; relation; coverage; and rights. Ryte (2021) says that these standards mainly facilitate search engines in searching for documents by noting key content already in the metadata. This is because the standard is used for simple resource description; combining metadata vocabularies of different metadata standards; providing inter-operability for metadata vocabularies in the Linked data cloud and Semantic web implementations (Galaxy Consulting, 2014). Therefore, as suggested by Kdnuggets (2019) application of this standard can assist organizations using different tools and technologies for data management to integrate their data to create a cohesive repository minimizing creation of information silos that could result in the accumulation of dark.

e) *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF)* – First developed by the International Council of Archives (ICA), this standard provides guidance for preparing archival authority records by providing descriptions of entities (corporate bodies, persons and families) associated with the creation and maintenance of archives (Memorix Archives, 2022). It looks at both manual and automated

systems while focusing on full documentation and continuous maintenance of the context of records creation and use, most importantly the provenance of archives and records. Memorix Archives (2022) goes on to explain that this standard also documents relationships between different records creators and between those entities and the records created by them and/or other resources about or by them. As a result, this enables the linking of descriptions of records creators and contextual information to descriptions of records from the same creator(s) that may be held by more than one repository and to descriptions of other resources such as library and museum materials that relate to the entity in question. This means that if organizations use this standard, it will enhance sharing of archival authority records by promoting the preparation of consistent, appropriate and self-explanatory descriptions, a culture that can minimize the accumulation of dark data.

2.3.3 Implications of Risks and Opportunities of Dark Data on Corporate Governance

Rizzolo (2019) posits that organizations need data to enable them to establish baselines, benchmarks and goals to keep moving forward. However, Visix, Inc (2024) observes that data can also expose inefficiencies and disadvantages. Therefore, dark data being part of the big data is not left behind in these two scenarios as it may present both risks and opportunities on corporate governance as discussed in Sections 2.3.3.1 and 2.3.3.2.

2.3.3.1 Implication of Risks Posed by Dark Data to Organizations

IBM (2025) observe that dark data might be a relatively new term, but it's not a new problem for most medium to large organizations as its cost extends beyond the financial implication of its storage. Domo (2021) and Tittle (2014) observe that the degree of

risk from dark data is dependent on the ability of the investigators or hackers to draw from a collection of dark data made available to them. Champagne and Anderson (2021) note that 44% of security alerts go uninvestigated as a result of the overwhelming amount of information received by security officers. The following are some of the risks posed by dark data to organizations:

Exposure to legal or regulatory issues

Data covered by regulation that's kept but improperly stored can lead to costly sanctions to an organization (Dimitrov, 2018). Many organizations are not even aware that some of the dark data in their custody contain such vital information such as trade secrets, employees' personal identifiable information or customer payment information among others (Spinner, 2018 and Rao, 2024). As a result, organizations fail to provide the necessary security for such data. According to Spinner (2018), this may provide a good platform for attackers who most definitely zero in on such information, making data regulators to demand for explanations/answers. Sarkar (2017) goes on to say that data containing confidential information like credit card information as well as sensitive information which can compromise important business activities and relationships may appear anywhere in dark data collections. This may result in organizations facing legal and financial liability.

Matthews (2018) further discusses that with the implementation of the General Data Protection Regulation (GDPR) in the European union on 25th May 2018, consumers can exercise the so-called right to be forgotten and ask companies to delete their data under certain circumstances. For example, when the company doesn't need the information anymore. However, this might become challenging if companies continue ignoring their dark data as it even becomes hard to locate it. In addition, business

representatives who don't pay attention to dark data could also be at increased risks for GDPR fines if dark data gets exposed during a breach (Matthews, 2018).

Trajanov et al (2018) and Domo (2021) underscore that besides increased storage costs, having large amounts of unstructured or unorganized data can potentially lead to serious security risks. This is because despite being outdated and seemingly useless, dark data may also contain sensitive, proprietary information that hackers, lawyers and auditors are more willing to manage (Izenda Inc, 2019). Trajnov et al (2018) observes that dark data may conceal obsolete or inaccurate information that may be misinterpreted if discovered by auditors and hackers. This is exemplified by day-to-day news of data breaches like the one that rocked Sony Pictures in November 2014 (Alvarez, 2014).

Risk on organizational resources.

Organizations are often inclined to save everything which include data (Spinner, 2018). According to Kidd (2016) dark data consumes significant data center capacity on a global basis yet it is virtually invisible for use. This is because, accumulating dark data will require resources in terms of funds to purchase server space as well as physical space. As alluded by Gartner (2018) in Izenda Inc (2019) storing and securing data typically incurs more expense (and sometimes greater risk) than value. This is because more data storage translates to more overhead cost as well as increased security risks. Izenda Inc (2019) goes on to argue that as outdated and seemingly useless documents increase in an institution, this dark data is likely to also contain sensitive, proprietary information. All this happens despite of the fact that even the in-expensive on-premises and cloud storage is burying companies into excessive data management expenses while raising security red flags. Mathews (2018) gives an example of a company that had to spend $6 million to search for dark data needed for a court case. This illustrates

how costly it can be for an organization that does not prioritize figuring out ways to classify dark data in practical ways. Apart from financial resources, a lot of unnecessary time and energy may also be spent searching for data that is in the dark (Javanainen, 2015).

Data management complications

Dark data can also be described as data a company stores without any clear idea why they are storing it and the value it represents (Future Hosting Technologies LLC, 2017). Domo (2024) agrees that one of the main challenges presented by dark data is determining its real value, if any at all. This is because, a large chunk of data created or received in the organization is rarely thought about or seen. According to Future Hosting Technologies LLC (2017) one can only derive meaning and value from such data by harvesting using big data analytics or machine learning algorithms. Or else, it will continue occupying office space in the case of physical records or sitting on hard drives in data centers, and on the cloud racking up hosting bills without generating any value.

Intelligence risk

United States Data Science Institute (2025) discusses that dark data may entail proprietary or sensitive information that reflects the business operations, practices, competitive advantages, important partnerships and joint ventures among others. However, because it is not used people tend to forget about its content, purpose or even its existence. As a result, dark data goes unmanaged and become more vulnerable to security risks such as unauthorized access that may leak trade secrets and other proprietary knowledge. Opsitnick et al (2017) further point out that former employees may also pose a big threat to the company data whereby over 25% of employees have

reported stealing company data when leaving a job. This is because exiting employees may feel entitled to the data they create or receive in the course of their work which may include intellectual information such as customer lists, source code, secret formulas and strategy documents among other trade secrets which may be used against the organization when they go to work for a competitor or decides to start a new company.

In addition, Izenda Inc (2019) and Kdnuggets (2019) acknowledge that normally, dark data is usually neither encrypted nor subject to the same protection as data that is known to be sensitive. Therefore, its accumulation may result in deliberate or inadvertent disclosures, or even loss of proprietary or sensitive data on business operations, products, financial status and business plans. As a result, this may affect the baseline or even compromise important business activities and relationships. This poses an obvious risk to the business, bearing in mind the information contents that might be in the dark data (Future Hosting Technologies LLC, 2017). Future Hosting Technologies LLC (2017) goes on to discuss that dark data may contain information which if analyzed properly, it might generate actionable insights that can be used for product development, marketing and operational optimization among other positive applications. If such information land in the wrong hands, then the organization in context will face intelligence risk.

Loss of reputation

Any kind of data breach, dark data breach included, reflects badly on the organizations affected (IBM, 2025). Future Hosting Technologies LLC (2017) asserts that not subjecting dark data to encryption nor same protection as data that is known to be sensitive in an organization poses an obvious risk to the business. As discussed by Tittle

(2014), dark data's secrets may be too dark and damaging indeed, but one has no way of knowing for sure. Sarkar (2017) contends that dark data may appear to be potentially harmless yet carrying potential privacy risk. In case of a data breach to sensitive and confidential data, the company reputation mya be compromised leading to lose of stakeholder confidence and trust (Buckbee, 2022 and Sarkar, 2017).

Opportunity costs

Izenda Inc (2019) and Sarkar (2017) discuss that lack of analysis of dark data for use may result in organizations losing vital opportunities. This is because dark data may contain unknown and untapped sources of intelligence as well as serve as a source of exposure to loss or harm. Kdnuggets (2019) goes on to say that while some companies decided not to analyze and processes dark data their competitors do, making them a mile ahead because of the value and intelligence they extract from the dark data.

Cloud and hybrid storage concerns.

Spinner (2018) notes data as organization embrace both on premise cloud and hybrid data storage, they are not factoring in the security limitations and capabilities of both environments. Spinner (2018) goes on to say that with many organizations embracing cloud and hybrid storage, they are not into books that the two often lack security controls as those expected from the on-premises storage. Izenda Inc (2019) further asserts that storage of dark data that may contain vital data like customer information and organization operations on the cloud, out of their immediate control and management poses a great risk. Tittle (2014) concurs that cloud storage of data, outside the immediate control and management presents a risk to the continued business health and well-being. Therefore, it may become hard to secure sensitive information on need-to-know basis when storing data both on premises and on the cloud (Spinner, 2018).

Thus, the need to understand the capability and limitations of both environments, lock down your security and monitor them for threats.

In addition, Biscom Incorporated (2019) discusses that with the new technologies used to store data in organizations such as Dropbox, Box, Google Drive and email among others, taking files is nearly effortless with no feeling of wrongdoing.

Security risk

Dark data being data that is barely known and hard to locate, it becomes hard to attach value as well as security measures to it when one doesn't even know its content. Spinner (2018) alludes that as dark data accumulates in an organization and ages, it also becomes a low security priority. However, to competing companies or attackers looking for information to maximize for personal, political or monetary gain, this data may be extremely interesting and valuable (Spinner, 2018). Mathews (2018) asserts that cybercrime experts say that there has been a shift in hacking strategies, whereby unstructured data is getting pinpointed more than ever before by cybercriminals. Krause (2020) and IBM (2025) concur that lack of insight into dark data leads to permission challenges as it becomes impossible to determine who should have access to which data because when data is not inventoried and catalogued effectively, it is not classified properly and therefore also not protected adequately. T. With the low security priority given to dark data, then the attackers get a window of opportunity to place the organization into data breaches and intelligence risks.

2.3.3.2 Opportunities Derived from Dark Data to Organizations

Analyzing dark data could give a business insight they don't currently have to develop greater context and unveil trends, patterns and relationships that miss them during normal business intelligence and analytics activities (AnswerMiner, 2020 and Mittal,

2018). For example, server log files can give website visitor behavior, customer call detail records may reveal customer sentiments and feelings while mobile geo-location data can provide traffic patterns. Trajanov et al (2018) agrees that dark data may initially appear irrelevant, yet it may contain large opportunities that companies ignore.

Pal (2021) opines that if dark data is analyzed, it can be a real goldmine that can help to predict demand and solve issues. For example, by analyzing customer clickstream data and getting product telematics, companies can be able to accurately forecast demand and respond appropriately by optimizing the supply of goods. Purushothaman (2019) goes on to explain that with advances in machine learning, dark data can be extracted automatically and be connected with other data attributes to provide a more complete view. This in turn may help organizations to tap into great opportunities for example, leverage historical data to predict outcomes.

Efficiency in service delivery

More so, analysis and processing of dark data can help organizations to build a smarter supply chain (Pal, 2021). This is when dark data is used to provide granular information about the supply chain to enable organizations to accurately know the time and volume of demand so as to respond to the requirements appropriately. Thus, companies having a smart and robust supply chain.

More so, Gartner (2025) elucidates that some organizations accumulate dark data when they keep records on support interactions (like time customers contacted the business, communication channel, duration of engagement, reason for contacting, among others) hoping to use it later in case of a need to research on a customer issue. However, instead of keeping this data in the dark, organizations should always try to build them into their

analytic workflows to help get new insight like when customers are most likely to contact, their preferred methods of contact, issues they have and so on.

Customer satisfaction

Feliu (2020) asserts that using data-derived insights allows us to move from decision making based on intuition to making informed and sound decisions. Therefore, analysis and processing of dark data may help by relying on customer feedback to improve product/service quality (Pal, 2021). This is because customers are brand ambassadors of products/service by word of mouth, referral and social media. Therefore, dark data can help manufacturing companies by providing an all-rounded view of the product/service and as well as market view through leveraging customer feedback and improving product/service quality.

Gartner (2025) observes that life-science companies for pharmacovigilance, have put into practices the analysis of dark data embedded in doctors' notes, voice mails, or emails to determine any adverse experience associated with the use of a medical product in patients. This may enable companies to develop drugs safely through the use of digital technologies such as computer vision, computational linguistics, artificial intelligence, text classification, machine learning and predictive modeling that can enable track issues, predict and solve potential problems with less human error.

Competitive advantage

Cubeware GmbH (2021) asserts that companies that utilize all their data are ahead of the curve. This is because analyzing and leveraging dark data not only creates new revenue streams, streamline internal processes, and lower costs but may also provide new insights such as customer journeys, identify customer dissatisfaction earlier, and solve customer problems faster (Feliu, 2020 and Cubeware GmbH, 2021). Feliu (2020)

adds on that it can also uncover hidden correlations between pieces of information that were previously thought to be completely unrelated.

Elliot (2014) also notes that dark data may carry a great hidden potential if analyzed and used by organizations. Izenda Inc (2019) discusses the case of a project at Copenhagen, where useful information was drawn by analyzing the data in the log files of the Wi-Fi routers scattered in the airport. The analyzing team noted that the customer phones pinged routers while they walked through the terminals even if they did not connect to the Wi-Fi. This information could help track passenger movement and behavior precisely. For example, the data could help discover the most visited areas of duty free, typical passenger flow and choke points among others.

Creating new revenue opportunities and business growth

Data-derived insights have a significant economic impact and value when leveraged correctly and promptly (Feliu, 2020). Cubeware GmbH (2021) posits that for growth, companies can utilize dark data to identify crucial insights that would have otherwise remained hidden. As a result, this may be used to enhance employee performance.

Reducing cost

Diligent Corporations (2018) alludes that as users interact with multiple data sources there is a likelihood of duplication and inaccurate data that may accumulate in the delocalized storages without them knowing. This is bad data that may cost companies time, money and, potentially, reputation. This is because it can lead to inaccuracies and redundancies end result being incorrect customer contact information, inefficient billing and processing operations, or miscalculated manufacturing demands. Joshi (2019) advises that by analyzing and processing this dark data, organization may not

only fix the storage space issues but also foster financial savings. This is because they will be able to weed out redundant and unclean data.

Develop new partnerships.

Mackey (2017) posits that dark data incorporates a wealth of documents, proposals, account details, emails and a variety of other types of communiqués, which can be considered extremely relevant. Therefore, analyzing dark data may help to identify new target path, discovering new investment opportunities, reducing risks, or increasing the returns on investment (Reiley, 2022).

Reduced legal risks.

Dark data contains a wealth of documents, images, email, log files, ZIP files, videos, and images which may carry confidential information about the company and its customers (Mackey, 2017 and Joshi, 2019). Therefore, if this data is well managed, then there is reduced risk of exposure of confidential data to unauthorized parties thus reducing legal risk. Joshi (2019) emphasizes that if all the organizations data is put correctly into use, then organizations will strengthen their security procedures, safeguarding their data against theft.

2.3.4 Strategies for Harnessing Dark Data as Records for Good Governance

It is critical for organizations to set up information governance strategies to enable shining light and providing visibility to dark data as well as gain an upper hand on the risks posed by dark data (Krause,2020). According to Spinner (2018) and Krause (2020) the following are information governance strategies that organizations can set up to harness dark data:

Understanding data gathered in organization

The quality of the metadata appended to files and documents is key in illuminating dark data, through casting a spotlight on single documents amongst millions for quicker retrieval (Mackey, 2017). As a result, this successfully direct searches to the most relevant data for the request made.

In addition, metadata helps in understanding the data gathered by the organization and its value by identifying, linking, curating and cross-referencing information in a way that unlocks its relevance and usefulness (Javanainen, 2015). According to Hosting Technologies LLC (2017) understanding the nature of the data gathered in the organization and its value to the organization helps in determining the appropriate storage and protection. Through analysis and categorization, the personnel in charge will be able to determine whether to delete or encrypt data. Ghosh (2024) alludes that you are able to control dark data and realize its hidden value only when you understand it.

Regular database audit and pruning

Openshaw (2025) and Tittle (2014) observe that yesterday's dark data may become a bright source of insight, thanks to developing data analytic technologies that can thoughtfully harness it. Annalect (2017) and Sydell (2024) note that Artificial Intelligence (AI) has become a critical tool within dark data analysis since it can automatically mine data to add tags, structure, meta-information and more. Natural Language Processing (NLP) tool as a subset of AI can also be used to analyze free text or pick up on subtle social mentions. Shekhar (2021) add on Google's Video Analytics Application Programming Interface (API), can also be used to go through every scene in a video and identify specific elements in the scenes where a search engine can then

be implemented to look through the video to identify specific features and when they show up in the video. FindFace Multi Version 4.3 software of NtechLab has begun to collect and analyze dark data that is inaccessible to users of other systems using a built-in statistic that reflect visitor information on simple charts (Ntech Lab, 2020). The software is able to recognize gender, age, count the number of unique and repeated visits, as well as provide data for thermal maps of different cameras. These are just some of the many data analytic technologies that can be used.

Therefore, Kdnuggets (2019) and Tittle (2014) suggest that organizations' dark data holdings should be recognized and subject to periodic auditing and pruning. This should entail structuring and categorizing the old data so that you know the kind of dark data stored, where dark data resides, how it's stored, how it's protected and what kinds of access controls help to maintain its security. Tittle (2014) further suggest that organizations should drive ongoing research into new tools and technologies to help extract value from such data. In addition, there should be periodic security audits, risk evaluation, incident response and policy to manage risks of exposure and potential loss or harm (Tittle, 2014). According to Krause, (2020) it also entails taking a hard look at the regulated, out-of-policy data lurking within files and removing or archiving.

Maintain and enforce a strong security policy
This entails managing access control and data protection through application of strong encryption standards on data, use of biometric access to control access of storage rooms and many more (Kdnuggets, 2019). Tittle (2014) asserts that data should have strong encryption to make it extremely difficult to unlock its contents for those who do manage to access dark data. Additionally, there should be equally strong access controls and monitoring that makes it obvious as to who can (and has) access to such information

for any purposes whatsoever. These also entail monitoring and alerting whenever the information stored exhibits characteristics of unauthorized access (Krause, 2020).

Employment of retention and disposal policies

Tittle (2014) observes that it is always worth considering if and how dark data should be retained or properly disposed. This calls for retention and disposal policies to help guide and drive such decisions. Kdnuggets (2019) notes that carefully formulated disposal policies identify data for erasure or destruction while good retention policies help one retain valuable data for later use. Marumolwa and Marnewick (2025) further elucidate that the policies such as retention and disposal policies should be safe in that they should be able to determine whether only content, or both content and media should be disposed of. Therefore, these enable removing or archiving data that no longer provide value and retaining the one that has enduring value (Technologies LLC, 2017)

Goal oriented data collection

Technologies LLC (2017) discusses that businesses should only collect data with a goal in mind rather than indiscriminately. Data collected should always answer a 'why' question or fulfill a need so as to avoid gathering data that may pose a risk, increase costs, and offer dubious value to the organization (Technologies LLC, 2017). Ghosh (2024) argues that one should only collect data for a valid reason, not just because they can get data and store it. According to Laserfiche (2018), this entails businesses deciding on what to keep and get rid of the rest.

Assigning data priorities

The data created or received in organizations has different levels of usefulness to the organization, thus the need for evaluating and auditing the usefulness of these data.

Organizations need to classify dark data according to their various different priorities in practical ways to reduce unnecessary access to data (Mathews, 2018 and Technologies LLC, 2017). Mathews (2018) discusses that dark data may include sensitive details companies are legally required to submit for court subpoenas. However, it's not always easy to find that in-demand data just because it is classified as dark. Furthermore, Ghosh (2024) suggests that it is good to understand which regulations require what data to be kept and for how long so that data is not kept longer than necessary unless clear benefit can be derived from it.

Use of Software and artificial intelligence

Krause (2020) suggests that organizations can also use software/artificial intelligence to manage their dark data. Sydell (2024) and Viewpointe (2013) notes that use of software can enable automation of the data lifecycle, which entail automation of policies that classify, organize, retain and delete legacy data at reasonable limits. For example, the Veritas Data Insight tool from the Software House International can be used to reduce risk and improve accountability by analyzing and classifying data which goes to an extent of predicting information crisis before it happens (SHI International Corp, 2020). Once the software is installed and configured into the server, it scans through to produce and deliver actionable, all-encompassing reports and findings (SHI International Corp, 2020). Openshaw (2025) suggests that organizations can use an systems such as email management system that is centralized and not using local '.pst' files that automatically deletes messages as per the organization's policy.

Application of an information management policy

An information management policy helps organizations to manage all the data it produces (creates or acquires) even if it is for a short time value. Sharma (2025)

discusses that not having an information management policy is pretty bad in this modern age of data while having one that is not followed or which no-one takes responsibility for is even worse. This is because the policy helps understand the data that exists in an organization. TechTarget (2025) goes on to say that this is achieved through answering questions like who are the people in charge of all company data, which data should be kept, why it should be kept, where it should be kept, for how long and who should have access to it long before it disappears never to be seen again.

Sensitization

Decision makers and key data handlers in the organization need to be made aware of the existence of dark data in the organization (NetGovern, 2022). With almost all organization stakeholders creating or acquiring data, the personnel in charge of the organizational data are called upon to enlighten these data handlers on proper data management. According to NetGovern (2022) these include documenting the types of data you hold, its source and who you share it with. In addition, you should have the right procedures and tools in place to detect, report and investigate a personal data breach, as well as deal with data access requests.

More so, Marumolwa and Marnewick (2025) discusses that it's also critical to talk and communicate with people because not everything will be officially documented or known about. So, people should be made aware of the impact of not being able to manage their data effectively as well as how to manage it effectively. IBM (2025) goes on to say that with sensitization personnel in charge of data management in organization will be sure of full organizational support as they will be aware of dark data's existence and its value.

Well establish data management team

Holdsworth (2025) and Podraza (2014) suggest that a good dark data management program should have a data management team to develop and implement a strategic plan for the program as well as provide a governance structure. In addition, the team members should have their responsibilities clearly defined in matters legal data management to enhance accountability (Profisee Group, 2019). A well-established data management team has clearly defined roles and responsibilities of each member such that every action in regard to data can be attributed to a particular member or members.

Stakeholder involvement

Data management should be proactive in nature whereby data preservation and curation involve all stakeholders (Digital Curation Centre, 2022). Various data stakeholders in the organization are able to assist in the collaborative and managerial duties that ensure data is created appropriately and preserved. For example, stakeholders are involved in curation activities like standards and policy formulation. In return, there is enhanced data management in organizations as its more accepted and easier to practice by all stakeholders. Community participation also enhances support, wider access to expertise, tools and systems among other resources. This in-turn enhances data management planning, reducing wastage of resources and duplication of efforts. Thus, minimizing accumulation of dark data.

Budgetary allocation

DataONE (2021) emphasizes that organization should accept the fact that the process of data management incurs costs in terms of resources and money. Champagne and Anderson (2017) further observe that some small and midsize businesses (SMBs) lack adequate staff and/or budget to keep up with increasing cyber security demands.

Therefore, DataONE (2021) advises that organizations need to address these costs and the allocation of resources in the early planning phases of the data management project. According to DataONE (2021) budgetary allocation should be looked at in terms of: short-term costs, long-term costs, internal/external costs, equipment/services (i.e. computer cycles, storage, software, and hardware) costs, overhead costs, time costs and human resource costs.

2.4 Chapter Summary

This chapter brings out dark data as an issue that has become of concern in many organizations. This is because it poses both risks as well as potential/lost opportunities to an organization. Despite all this, dark data keeps accumulating in organizations due to lack of dark data analytic tools. Though the mechanisms and strategies needed to tap dark data's potential may be costly, it is worth the investment. The more organizations agree to sit on dark data and leave it unearthed, the more they expose themselves to several risks. Therefore, the only solution is to do something proper about dark data rather than treating it as mere thing.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

Research methodology is the specific procedures or techniques used to identify, select, process, and analyze information about a topic (University of Witwatersrand, 2022). The research methodology determines the overall validity and reliability of a study as it answers two main questions: How data was collected or generated and how data was analyzed. Jansen and Warren (2020) add on that research methodology assist in determining what data to collect and ignore as well as who to collect it from (sampling). This shows that a sound research methodology guarantees good finding.

The aim of this study was to find out how dark data can be harnessed for corporate governance at the Kenya Energy Transmission Company (KETRACO). Therefore, the purpose of this chapter was to identify methods that are suitable for collecting data, analyzing it and writing the final report. This chapter covers research design, target population and samples under consideration, research tools and instruments. It also talks about data collection and data analysis. This gave guidance on how the study was conducted and final reporting of the findings.

3.2 Research Paradigm

Ulz (2023) explains that a research paradigm is a worldview or philosophical framework, which includes ideas, beliefs, and biases, that guides the research process with an aim of helping to determine the way the research will be conducted. This is achieved by providing the framework into which the theories and practices of the study fit to create the research plan to guide all areas of research, including the aim of the study, research question, instruments or measurements used, and analysis

methods. Rautenbach and Jansen (2024) discuss that research paradigms are based on three models which are positivist research paradigm that lead to quantitative studies through assuming that reality is objective thus focusing on measurable results; interpretivist research paradigm that lead to qualitative studies by seeing reality as subjective thus exploring people's lived experience; and pragmatic research paradigm that leads to mixed studies by taking a flexible problem-solving approach. Nickerson (2024) explains that with a focus on understanding how people make sense of their social world, instead of searching for universal laws interpretivism research paradigm explores the meanings, experiences, and perspectives individuals hold by believing that reality is socially constructed, so they study human behaviour through empathy, dialogue, and interpretation rather than measurement or experimentation. Therefore, for this research the study adopted and interpretivism paradigm as the best fitting guiding framework.

Tanlak and Aryal (2025) explain regarding social context, interpretivists urgue that phenomena are best understood when their interpretation is situated in specific social, cultural, or historical contexts. Therefore, the study focused on examining how dark data can be harnessed for effective corporate governance at KETRACO. Additionally, Pervin and Mokhtar (2022) discuss that interpretive approach is subjective with an aim of understanding complex phenomena by deeply understanding the perspectives and meanings of respondents' experiences in unique context by use of qualitative methods like observation, focus groups, interviews and document analysis. Therefore, this study used interviews and document analysis with an aim of exploring the respondents' understanding of dark data and effective governance based on their experiences at KETRACO. This aimed to reveal the unique perceptions, challenges and meanings that staff at KETRACO ascribe to dark data and governance. Therefore, by placing the

research within the interpretivist paradigm, the study sought to develop a contextually rich, in-depth understanding of the potential, challenges, and perceived value of harnessing dark data to improve corporate governance within the unique setting of KETRACO rather than producing generalizable truths.

3.3 Research Approach

Research approaches can be defined as the collection of procedures and plans that decide the overall process of research as well as decide the methods for data collection, analysis, and interpretation in order to address the research problem (PubGenius Inc (2023). Extant literature on research methodology identifies three approaches to research: qualitative; quantitative and Mixed Methods Research approaches. Creswell (2015) explains that quantitative research approach, also called 'true science' heavily relies on mathematical and statistical research techniques to obtain the results of the research activity while using closed ended questions. On the other hand, qualitative research approach uses words while employing open-ended questions to generate various ideas and concepts that can be converted into logical and testable hypotheses in future (Creswell (2015). A mixed research approach incorporates elements of both qualitative and quantitative research approach.

The current study employed a qualitative research approach. QuestionPro (2022) and Teherani et al (2015) describe a qualitative research approach as a systematic inquiry into social phenomena in natural settings by focusing on obtaining data through open-ended and conversational communication. This method is used when one wants to get detailed responses as it allows in-depth and further probing and questioning of respondents based on their responses, while the researcher also tries to understand their motivation and feelings (QuestionPro, 2022). Vaughan (2021) and Wilson (2012) in

Rahman (2021) observe that a qualitative research approach can be more targeted, which means that it can be far more concentrated, sampling specific groups and key points in a company to gather meaningful data. This can both speed the process of data capture and keep the costs of data-gathering down while getting detailed information at the same time (Vaughan, 2021). Therefore, this made this approach suitable for the study which is hard to quantify to enable gathering of detailed information about harnessing dark data for corporate governance. This is in addition to giving respondents opportunity to voice their own accounts in matters of dark data. The qualitative approach entailed conducting interviews to enable recording of attitudes, feelings and behaviors while encouraging respondents to expand on their responses as well as seeking clarification.

3.4 Research Design

Bhandari (2022) describes a research design as a strategy for answering one's research question using empirical data. QuestionPro (2022) goes on to say that it is a framework of research methods and techniques chosen by a researcher to come up with research methods that are suitable for the subject matter and set up their studies for success. Qualitative research designs include ethnographic research design; historical research design; phenomenological research design; grounded theory research design and case study research design (Grand Canyon University, 2025).

Ethnographic research design is where the researcher actively participates in the group being studied in order for him to gain an insider perspective or have a similar experience as the group members (University of Virginia, 2020).

According to Deakin University (2023) and Voicedocs (2022) historical research design entails the careful study and analysis of data about past events with an intent to

gain a clearer understanding of the impact of the past on the present and future events related to the life process.

Phenomenological research design is used when the researcher wants to understand a phenomenon's universal nature by exploring the views of those who have experienced it, gain a deeper understanding of how human beings think, and expand a researcher's knowledge about a phenomenon (Delve, 2022).

Grounded theory research design comprises systematically obtaining data and analyzing the data using comparative analysis with an aim to discover or construct theory from this data (Tie, Birks and Francis, 2019).

Cherry (2024) explains that a case study research design is an in-depth study of one person, group, or event so as to explain, describe or explore events or phenomena in the everyday contexts in which they occur. It allows researchers to investigate things that are often difficult to impossible to replicate in a lab. According to Cherry (2024), the advantages of a case study research design include: allowing the researchers to collect a great deal of information; giving the researchers the chance to collect information on rare or unusual cases; enabling the study of things that the researchers could not replicate in experimental settings and permitting the researchers to develop hypotheses that can be explored in experimental research.

However, a case study research design cannot necessarily be generalized to the larger population; cannot demonstrate cause and effect; may not be scientifically rigorous; and can lead to bias (Cherry, 2024).

The suitability of a research design is influenced by many factors that include but are not limited to the research problem and the depth of knowledge required about the

phenomena in question. Cherry (2024) advises that a researcher may choose to perform a case study if they are interested in exploring a unique or recently discovered phenomenon so that the insights gained from such research can help the researchers develop additional ideas and study questions that might be explored in future studies.

Additionally, Angela (2025) and Cherry (2024) discuss that the main types of case studies include: illustrative that provides vivid examples to clarify a concept or a theory; exploratory that investigates a new or uninvestigated phenomena to generate new insight; cumulative that gathers insight by analyzing multiple cases over time; critical instance that investigates specific events or occurrence to understand the broader significance; intrinsic that focusses on understanding a particular case for its own unique characteristics; and descriptive that offers a detailed description of specific phenomena or situation. Therefore, the researcher adopted an exploratory case study research design to get in-depth knowledge by paving for deeper inquiry about dark data and how it can be harnessed for effective corporate governance.

3.5 Study Population

Akman (2025) and Hu (2014) describe a study population as a subset of the target population from which the sample is selected and perceived to derive results for a specific study. This is the group of individuals from which the respondents and participants are recruited by the researcher. In this case, the study population for the research was 569 staff from KETRACO stratified into three management levels. The management levels were:

- *Executive Management* - This comprised the Managing Director, General Managers (Head of directorates) and Senior Managers, all totaling to 15. This group oversees all company operations. They are also responsible for articulating

and implementing policies, presidential directives, circulars and executive orders in KETRACO.

- *Middle Level Management* – This entailed the Managers whose main responsibility is to execute plans and policies in KETRACO. Population for middle level management was 34.

- *Low level management* – This level encompassed officers whose responsibilities are to execute the actual tasks and deliverables in KETRACO using the set policies, procedures and standards to meet the stakeholders' expectations. Population for low level management was 520.

The respondents categorized at these three management levels were drawn from all KETRACO directorates. This was because all directorates are involved in receipt or creation of data as well as data management. These directorates are:

- *Design and Construction* – It is responsible for all processes towards design, construction and commissioning of electricity transmission infrastructure of KETRACO.

- *Project Development Services* – this directorate is in-charge of transmission system planning, project development, wayleave and land acquisition. It also monitors and evaluates completed projects.

- *System Operations and Power Management* – This directorate is in-charge of KETRACO's transmission system operations and maintenance; power management and market operations; control and data acquisition (SCADA) systems; substation automation systems (SAS) and telecommunications systems to ensure grid availability, reliability and stability.

- *Company Secretary and Legal Services* – It provides secretariat services and guidance to the Board; management of corporate affairs; management of contract administration; litigation, conveyancing; prosecution and compliance with legal and regulatory requirements.

- *Human Resources and Administration* – This directorate ensures optimal utilization of human resources in compliance with Human Resource Management Act 2012 and Labour Laws.

- *Supply Chain Management* – It oversees operations, logistics and projects functions of the Company's supply chain management as stipulated in the Public Procurement and Assets Disposal Act (2015).

- *Strategy, Research and Compliance* – This directorate oversees strategic planning, transmission master plan, performance contract and monitoring and evaluation in KETRACO. It also ensures compliance with economic regulation and tariffs formulation; economic appraisal of projects; business development; research & development; quality assurance, risk management; and ICT and innovation strategies and programs in KETRACO.

- *Internal Audit* – The Directorate is in-charge ensuring independent objective assurance on the effectiveness of the Company's risk management, internal controls and governance processes.

- *Finance* – This Directorate oversees the management of the Company's finances. It also ensures compliance with government regulations on finance matters.

Table 3.1 represents the study population stratified according to management levels.

Table 3.1: Total Study Population Based on Management Levels

Level of staff	Study population
Executive level	14
Middle management level	61
Low level management	494
Total	569

Source: Researcher (2023)

McCombes (2022) explains that sampling methods and techniques are the process of selecting the study sample. Byju's (2022) adds that sampling method or sampling technique is subdivided into two groups: probability sampling and non-probability sampling. For probability sampling, random selection is utilized thus giving all eligible candidates a chance to be selected. However, the method is time consuming and expensive (Byju's, 2022). On the other hand, non-probability sampling is a technique in which the researcher selects the sample based on subjective judgment rather than the random selection (Byju's, 2022). The study adopted a non-probability sampling method.

3.6 Non-Probability Sampling Method

StatCan (2021) describes non-probability sampling (also known as non-random sampling) method as a sampling wherein, it is not known that which individual from the population will be selected as a sample as compared to probability sampling where all members have a chance. McCombes (2022) goes on to say that it involves non-random selection based on convenience or other criteria, allowing the researcher to easily collect data. However, QuestionPro (2022) observes that the output of research conducted with a non-probable sample leads to skewed results, which may not represent

the desired target population. But it is imperative to note that there are situations such as the preliminary stages of research or cost constraints for conducting research, where non-probability sampling will be much more useful than probability sampling method. The following types of non-probability sampling explain this better:

- *Convenience sampling* – this sampling method is used when there are time and cost limitations in collecting feedback (QuestionPro, 2022). For example, startups will better conduct convenience sampling at a mall by standing at the mall entrance and giving out pamphlets or leaflets of upcoming events or promotion of a cause randomly.

- *Purposive sampling* - also referred to as judgmental sampling, QuestionPro (2022) explains that the researcher purely considers the purpose of the study, along with the understanding of the target audience when selecting the sample. However, Nikolopoulou (2022) points out that purpose sampling requires a thorough research plan to avoid researcher bias.

- *Snowball sampling* - this technique is applied when the subjects are difficult to trace such as illegal immigrants or when the topic is highly sensitive and not openly discussed such as research on rape victims (QuestionPro, 2022). Because not many victims will be ready to respond to the study, the researcher may need to contact people they might know, or volunteers associated with the cause to get in touch with the victims and collect information.

- *Quota sampling* – the selection of members in this sampling technique happens based on a pre-set standard thus forming the sample based on specific attributes, have the same qualities found in the total population (QuestionPro, 2022).

For this study the researcher adopted purposive sampling techniques as explained in the section below to get the target population. Target population is the subset of the entire study population with individuals that have similar characteristics or meet a certain criterion in whom the researcher is interested in conducting the research and generalize the findings (Willie, 2023). In this context, the target population were the staff who were involved in managing data in the Company.

3.6.1 Purposive Sampling Technique

This is also known as judgmental sampling. According to McCombes (2022) purposive sampling involves researchers using their expertise to select a sample that is most useful to the purposes of the research. Byju's (2022) and QuestionPro (2022) concur that purposive sampling provides high chances of obtaining highly accurate answers with a minimum marginal error within a short time due to the level of expertise of the respondents on the subject matter.

Therefore, purposive sampling technique was applied. Directorate secretaries were requested to provide list-based sampling frames of staff in their directorates, based on management levels. The records personnel assisted the researcher in identifying the most suitable respondents from the various directorate lists because they are the ones who are actively interacting with the staff involved in management of data in their respective directorates.

For the low-level management, information required was for short-term daily operations in KETRACO. These entail daily reports, summaries of activities completed and many others.

The researcher obtained short-range information from the middle-level management. This comprised information on resource distribution and market analysis among others. For example, project documentation, employee turnover, staff profiles and types of services offered among others.

Finally, the top-level management provided strategic information that can be used for long-term planning for the attainment of KETRACO objectives. These include policies and guidelines and information management of systems in place among others.

In addition, the records personnel provided the researcher with policies and standards they deemed were impacting on data management at KETRACO.

3.6.2 Sample Size

Elliott (2020) posits that a research sample size is those who partake in any given study and enables researchers to conduct studies of large populations without needing to reach every single person within a population. The sample size was arrived at by collecting all the lists of staff as provided by the directorate secretaries and assistance from records personnel. InterQResearch (2022) observes that a key component of qualitative research is smaller sample sizes that are homogenous in nature. This is achieved by segmenting the audience into similar personas to ensure that the research study is aimed at exploring themes or ideas from a specific subset of a population. For this study, the segment entailed staff who were deemed to be relevant to respond to the study questions as from the lists provided by the Records Section personnel and directorate secretaries. "The total number of purposively selected participants was eighty- four (84) from low-level management (Officers), four (4) from middle-level management, and three (3) were from top-level management. However, saturation was reached upon interviewing respondent number twenty-five (25) from lower-level management. This made the total

samples size to twenty-five (25) from low-level management, four (4) from middle-level management, while three (3) were from top-level management."

The researcher targeted to get responses from all officers until she hits saturation point. Quantilope (2022) and InterQResearch. (2022) explain that saturation point in qualitative research is that point during data collection when enough data has been collected to draw necessary conclusions, and any further data collection no longer produces value-added insights, instead the same themes start coming out repeatedly.

For top-level and middle-level management respondents the researcher reached out to all of them as advised by the records personnel. In addition, since they were small in number, Garcia (2017) advises that when the population is small enough, and a researcher has the resources to reach out to all of them then they should so as to ensure that everybody who matters to the research is represented accurately.

Table 3.2: Represents the Sample Size as Obtained by the Researcher.

Level of staff	Study population (List from directorate secretaries)	Purposively selected Sample
Top Level Management	14	3
Middle-level Management	61	4
Officers	494	84
Grand Total	**569**	**91**

Source: Researcher, (2023)

3.5.3 Coding of Respondents

Boughey (2016) asserts that maintaining anonymity is essential in qualitative research by using pseudonyms. To ensure anonymity of the respondents, the researcher coded them as illustrated in table 4.1:

Table 3.3: Respondents' Codes

Management Level	Respondent Codes
Top Management	T1, T2 and T3
Middle Level Management	M1, M2, M3, and M4
Low-level Management	L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, L17, L18, L19, L20, L21, L22, L23, L24 and L25

Source: Researcher (2023)

3.7 Data Collection Method/techniques

Egnyte (2021) posits that data collection method is a systematic approach to accurately collect information from various sources to provide insights and answers, such as testing a hypothesis or evaluating an outcome. Garcia (2017) goes on to say that data collection techniques are mechanisms that allows the researcher to systematically collect information about an object of study and about the settings in which they occur.

Data collection methods are broken down into two major categories. According to Egnyte (2021) these are:

- Primary data collection methods which gather information directly. Dudovuskiy (2022) adds on that primary data collection method gets data from first-hand experience thus, this is data that has not been published before. The techniques used include interviews, administration of questionnaires, observation and focus groups among others. Primary data results are highly accurate provided the researcher

collects the information direct from the source (Simplilearn, 2022). In addition, Burrey (2021) concurs that primary data collection methods ensure in-depth investigation and a greater level of problem understanding.

- Secondary data collection methods pull information from existing/published sources. It could be third-party source material or the output of the analysis such as books, magazines, reports and internet among others. Burrey, J. (2021

For this study, the researcher used interviews and document analysis as the data collection techniques.

3.7.1. Interview

An interview is a qualitative research method that involves two or more people (interviewer and interviewee), exchanging information through a series of questions and answers question (George, 2022 and DeCarlo et al, 2020). Interviews comprise collecting data by asking questions from a pre-prepared interview schedule and listening to individuals, recording, filming their responses, or a combination of methods (Abawi, 2014). According to George (2022) interviews allow the researcher to gather rich information and draw more detailed conclusions, taking into consideration nonverbal cues, off-the-cuff reactions, and emotional responses. However, George also notes that interviews can be time-consuming and deceptively challenging to conduct properly. For this study, four (4), three (3) and twenty-five (25) respondents from top-level, middle-level and low-level management respectively were interviewed.

3.7.2. Document Analysis

Document analysis is the process of reviewing or evaluating documents both printed and electronic in a methodical manner to uncover meaning, gain understanding, and come to a conclusion (Lumivero, 2024). Therefore, document analysis was used as a

complementary method of interviews to establish credibility of what respondents said about dark data management policies in KETRACO and as a basis of reducing bias. These documents are listed in *Appendix 5.* It was also key in guiding the research on subsequent questions to ask respondents about the policies.

3.8 Data Collection Instruments

Data collection instruments are the tools used by researchers to actually collect data in the research process (IGI Global, 2022). These may include questionnaires, interview schedules, survey plans and observation checklists among others. For this study, interview schedules were employed.

3.8.1. Interview Schedule

An interview schedule is a list containing a set of structured or semi-structured questions that have been prepared, to serve as a guide for interviewers, researchers and investigators in collecting information or data about a specific topic or issue (Luenendonk, 2019). The schedule may also be used by the interviewer to fill in the questions with the answers received from the interviewees. This study had a set of three interview schedules as follows:

- *Interview schedule for Executive/Top-level Management* was intended to give information about the implication of dark data to KETRACO, the policies and standards in use at KETRACO that can harness dark data, strategies that can be put in place to harness dark data at KETRACO.

- *Interview schedule for Middle-level Management* aimed to give information on the implications of the risks and opportunities of dark data to their Directorates and KETRACO at large; the availability and status of implementation of

policies and standards in KETRACO that harness dark data and strategies that can be used to harness dark data.

- *Interview schedule for Officers/Low-level Management* sought to provide the researcher with information about the awareness of dark data at KETRACO and the implications of its risks and opportunities to individuals and KETRACO at large; awareness of availability of policies/standards that can harness dark data and the extent of usage of the policies and standards; and their role in the accumulation of dark data.

Semi-structured interview schedules were used by the researcher while conducting the interview. Semi-structured questions allow open-ended responses from participants for more in-depth information (DeJonckheere and Vaughn, 2019). This is unlike structured interview schedules which pre-determine the answers by their questions which are closed-ended, featuring dichotomous (yes/no) or multiple-choice questions (George, 2022).

3.8.2. Document Analysis Guide

As deduced from Indeed (2023), a document analysis guide is a detailed systematic procedure that guides a researcher to analyze documentary evidence to answer specific research questions. In this context the researcher analyzed policies at KETRACO that impact on data with an aim to know their role in harnessing dark data for effective corporate governance and evaluate how they were being employed by staff. The roles of the policies in harnessing dark data were extracted.

3.9 Establishing Rigor

Queens University of Charlotte (2022) emphasizes the need to establish trust or confidence in the findings of a research study by assuring on the consistency in the

methods used over time while providing an accurate representation of the population studied. DeCarlo et al (2020) add on that to establish rigor in qualitative research, the researcher should ensure that the research takes place systematically and that consumers can be relatively confident that the findings are not fabricated and can be directly connected back to the primary sources of data that was gathered or the secondary data that was analyzed.

As demonstrated by Maher et al (2018), to establish trustworthiness and confidence in the finding the research had to ensure credibility, transferability, dependability, and confirmability of the research process. Therefore, the researcher conducted member check. Birt et al (2016) explains that member checking, also known as participant or respondent validation, is a technique for exploring the credibility of results. Respondent validation involves returning data or results to participants to check for accuracy and resonance with their experiences.

Mwita (2022), Saunders et al (2018) and Tay (2014) agree that making claims to saturation may retain a sense of methodological credibility in qualitative research. Therefore, credibility was also assured when the researcher engaged the participants until she hit point of saturation. This also involved asking the respondents for clarifications whenever they used ambiguous statement or terms. In addition, the researcher avoided leading questions.

The researcher also ensured that the findings are dependable by giving sufficient details of the research process such that any other researcher can repeat the work and get same results.

Conformity was achieved by the researcher minimizing biasness while conducting the study by remaining objective throughout the study. For example, the interview questions were according to research objectives and were open-ended while ensuring that the respondents understood them clearly before responding.

In addition, the researcher triangulated data source to eliminate biases. This was achieved by getting respondents from the various directorates at KETRACO as well as from different management levels and documents. Data triangulation, (also called participant or source triangulation) is where the researcher tries to examine data from different respondents but collected using the same method in this case, interviews with an aim of getting multiple perspectives to enhance e confidence in the accuracy and completeness of the findings (McLeod, 2024). This method considers that each participant has a unique and valid world view and the researcher's job is to try and look for a pattern or contradictions beyond the individual experience.

Moreso, the researcher also employed methodological triangulation when she combined stratified sampling to be able to reach all directorates/management levels and purposive sampling to be able to get respondents who were suitable to respond to the interviews. Bhandari (2023) and Indeed (2025) concur that methodological triangulation is the use of different methods to approach a research question with an aim of avoiding flaws and bias that come with using a single technique.

Finally, the researcher ensured that the findings can be applied/transferred to other institutions apart from KETRACO. The researcher clearly explained the research as well as the findings to ensure transferability.

3.10 Conducting the Pilot Study

A pilot study, also known as a feasibility study or feasibility test or pilot experiment or pilot trial, is a preliminary investigation designed to test the feasibility and design of a larger, more comprehensive study conducted on a smaller scale, aiming to identify any potential issues and make necessary adjustments before the main research is undertaken such as refining research questions, methods, and even the practicalities of data collection (CASP, 2025). Hu (2014) adds that this is when research instruments are tested on members of target population/study population, to evaluate the reliability, validity, practicability and sensitivity of the survey instruments prior to their final distribution. The pilot study helps to develop and test adequacy of the research instruments by identifying unclear issues in the instruments such as ambiguity, appropriateness of the interview questions and vastness among other inconsistencies. Therefore, to ensure validity and reliability of the interview schedule a pilot study was conducted on two (2) KETRACO staff. The researcher was able to identify ambiguous questions, repeated questions and spelling errors, which were then correct before they could be administered to the actual respondents.

3.11 Data Presentation, Analysis and Interpretation

Data analysis and interpretation is the process of assigning meaning to the collected information and determining the conclusions, significance, and implications of the findings (Whatagraph Team, 2021). Whatagraph Team (2021) emphasizes that the analysis of narrative (qualitative) data is conducted by organizing the data into common themes or categories whereby, the assessment/research purpose and questions can help direct the focus of the data organization. Perez (2022) goes on to say that data presentation is the process of organizing data into logical, sequential and meaningful categories and classifications to make them amenable to study and interpretation. Data

was collected then analyzed. Classification of findings was done based on the objectives of the study and research questions. The data obtained from the interview guide was analyzed using thematic analysis. Caulfield (2022) describes thematic analysis as a method of analyzing qualitative data where the researcher closely examines the data to identify common themes/topics, ideas and patterns of meaning that come up repeatedly. It entails searching across a data set to identify, analyze, and report repeated patterns (Kiger and Varpio 2020). The researcher also inspected keenly the data collected for errors.

The researcher organized the data she had collected. Then she categorized the findings into similar themes based on the research questions. This was followed by highlighting keywords/phrases to assist in categorizing the data. Then the researcher categorized the data based on recurring themes. This was aimed to build patterns that could give the researcher in-depth insight. Data was then presented based on the themes as understood from the opinions of the respondents. This included verbatim quotes from the respondents in some cases. Consequently, the findings were interpreted and recommendations made.

3.12 Ethical Consideration

To protect the participants, the researcher ensured that the research was conducted in a way that served interests of individuals, groups and/or society as a whole, while examining specific research activities for their ethical soundness, in terms of issues such as risk management, protection of confidentiality and informed consent. Using the introductory letter as attached in Appendix 1, consent was sought from the interviewees before conducting interview. The respondents were also informed of the reason behind

carrying out the research and the use of the findings of the research. In addition, respondents were also assured of anonymity.

In addition, to the best of my knowledge all sources cited in this work have been duly acknowledged, and further, the researcher obtained an anti-plagiarism certificate from Moi University.

Additionally, permission was sought by the Researcher and Moi University for the to conduct the study which by both NACOSTI and KETRACO.

3.13 Chapter Summary

This chapter focused on the research method used in the study. The areas covered include a qualitative study approach with purposive sampling as the non-probability sampling method used. Purposive sampling technique was imperative to this study as it enabled the researcher to get key respondents for the research. Interviews were employed with the use of open-ended interview schedules to enable get detailed information from the respondents. Rigor was established by doing member checks, triangulation of data sources and engaging the participants to saturation point. A pilot study was also conducted to pretest the viability of the interview schedules.

CHAPTER FOUR

DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.1 Introduction

To answer the study questions, this chapter presents, analyses and interprets data that was collected from the respondents. The study sought to establish how dark data can be harnessed for effective corporate governance at the Kenya Electricity Transmission Company (KETRACO). As explained in the previous chapter, data obtained from the interview guide was analyzed using thematic analysis.

The following sub-themes were covered in this study:

i.   The demographic information of respondents.

ii.   The context within which dark data is managed in organizations.

iii.   The existence of policies and standards that regulate harnessing of dark data.

iv.   Implication of risks and opportunities associated with dark data to corporate governance.

v.   Strategies of how dark data can be harnessed to enhance corporate governance.

4.2 Response Rates

The targeted respondents in this study were from the various directorates at KETRACO based on the management level. All KETRACO directorates and management levels were considered for the study to represent the entire Company. The respondents were reached from the lists provided by directorate secretaries and records personnel which were based on management level and suitability to respond to the study.

4.2.1 Directorates

The number of respondents from the various directorates were as follows: Human Resources and Administration seven (7); Company Secretary and Legal Services four

(4); Supply Chain Management two (2); Internal Audit two (2); Project Development Services two (2); Design and Construction two (2); Systems Operations and Power Management two (2); Finance five (5); and Strategy, Research and Compliance six (6).

4.2.2. Management Level

Management levels of respondents is very important as it brings out an all-inclusive organization. From the findings, twenty-five (25) interviewees were from lower-level management, four (4) from middle-level management, while three (3) were from top-level management. These made a total of thirty-two (32) respondents out of a total of ninety-one (91) targeted respondents from the lists provided by directorate secretaries and records personnel as in the following table.

Table 4.1: Response Rate

Level of staff	Sample size	Response rate
Top Level Management	3	3
Middle-level Management	4	4
Officers	84	25
Grand Total	**91**	**32**

Source: Researcher (2025)

4.3 The context within which Dark Data is Generated at KETRACO

The study sought to establish the context in which dark data is managed in KETARCO. To achieve this, the following sub-themes were covered.

i. Awareness and understanding the concept of dark data.

ii. Types of data created/received at KETRACO.

iii. Sources of data in KETRACO.

iv. Data storage, custody, access and usage at KETRACO.

4.3.1 Awareness and Understanding of the Concept of Dark Data

The understanding of the term dark data was the starting point of gauging the level of realization of the effects of dark data to corporate governance. This was because the level of awareness could (possibly) be an indicator of the perceived usefulness or value of dark data, and hence how it is managed for effective corporate governance. Therefore, the study posed a question to the respondents about their understanding of the term dark data. Four (4) respondents expressed their awareness of dark data. Majority described it as data that they are not aware of because it is hidden from their reach. Respondent L4 described dark data as data that the Company accumulates but is not much aware nor in control of. Respondent T3 opined that:

> *"Dark data is data that we have but we are not aware that we have. And it's hard to trace it when needed."*

Respondent L12 explained that dark data is data that exists somewhere in the Company, but not all the associated parties are aware of it. A different view was provided by respondent L2 who explained as follows:

> *"I would say that dark data is useful or harmful data that lies somewhere within our premises, yet we are not aware of it."*

The rest of the respondents could not describe dark data but acknowledged accumulating data that they were neither using nor in control of but did not give a specific term for it.

This revealed that majority staff at KETRACO could relate to the existence of dark data, but they could not clearly delineate it. The responses also revealed that respondents from Strategy, Research and Compliance Directorate with those from Human Resources and Administration Directorate were more aware of dark data as

compared to those from other directorates. This was because they were directly responsible for data management in the Company. Based on the management levels, respondents from top level management were the most aware of dark data followed by those from the middle level management. This was because top level management are in charge with the overall protection of the Company from data breaches. This acknowledgement by respondents of the existence of data within and beyond the Company that they were neither using nor in control of indicated that the researcher was engaging the right respondents. NUPPER (2020) agrees that good research is when the right queries are posed to the right individual because the right respondents have a noteworthy influence on the quality of research. Therefore, this meant that the responses and conclusions to the research question were most likely to be correct and agreeable.

This understanding of dark data by respondents therefore reflected a reduced dilemma of data governance for effective corporate governance at KETRACO. Grimm (2019) opines that under a spreading patchwork of digital privacy and data governance laws, the ability to appreciate dark data enhances the visibility of the risks embedded therein, thus reducing error and miscalculations. For example, an organization can only comply with laws governing the treatment of Personally Identifiable Information (PII) when it knows what PII possesses.

4.3.2 Types of Data Created/Received in KETRACO

Respondents were asked to explain the types of data that they create or receive during their daily transactions. According to the findings, the researcher categorized data created in KETRACO as follows:

.4.3.2.1 Staff Related Data.

According to Respondent M2, this comprises all data relating to staff affairs that KETRACO collects about employees for the duration of their employment. This data was mostly created by the Human Resources and Administration Directorate. However, the researcher also found out that in some instances other Directorates also created/received staff related data. Staff-related data at KETRACO includes staff names; employment and exit dates; staff numbers; staff targets; training needs; and appraisal scores. Respondent L3 also testified:

> *"Staff related data include hiring details such as employee requisition dates and reason; employment approval date; vacancy advertisements; application opening data; shortlisting data; interview and offer letters; and acceptances/decline. Appointment data included dates of appointment; contract nature; promotions; address; gender; tribe; marital status; photos (for staff and family, screenshots and general photos.)."*

Respondent L2 added that staff data include dates of birth for staff; education backgrounds; work experience; bank details; attendance details (various leave details); exit details and sometimes audios (voice recordings) among others.

4.3.2.2 Project Related Data

Respondent T2 explained that this entailed data that was related to the ongoing, completed and planned projects at KETRACO. This data was created or received by majority of the Directorates at KETRACO. The researcher placed it into three major categories as follows:

Cadastral GIS data

Dempsey (2022) describes cadastral GIS data as data that contains information about land ownership. This was in concurrence with respondent M5 who described cadastral data as data about land ownership particularly coordinates (alphanumeric numbers that

indicate the exact location of a piece of land on earth, land parcel number, landowner and land value among others.

PAPs' profile

Respondent L7 explained:

> *"These include data about the project affected person (PAP) that we use for compensation. For Example, PAP Names, bank details, ID number, pin number, land registration details, affected land size, number of crops and structures damaged, value of land/crops/structures and compensation amounts."*

Contractor profile

Respondent L6 said:

> *"Contractor profile is the data about the contractors who assist KETRACO in the various project constructions. These data entail contractor name, bank details, ID number, pin number, registration date, license number, duration of contract, line/project name, line capacity, ...yes among others."*

4.3.2.3 Finance and Procurement Data

According to Respondent M3, this was data that was related to procurement of services and assets in KETRACO as well as data related to KETRACO assets. It included supplier/service provider profile (statement of account and services being offered.); products purchase, requisition and reservation data; assets and liability details; procurement plan details (items to be procured, respective procurement dates and budgeted amounts). Respondent M4 added assets disposal register details (names of assets, quantity and conditions); asset register (names of assets, quantity and condition); tender process details (dates of tender opening, nominees, offers and acceptance of offers); utility bills; and photos.

4.3.2.4 General Data

This comprised data that was being created or received in course of daily activities by the respondents but could not be placed into a particular category. This included:

- Data on the current trends for various departments, for example engineering designs, new softwares in use and many more.

- Notes/drafts/sketches. These entailed electronically or manually created or received by staff while in course of their duties for particular reasons which either did or did not materialize or were overlooked.

- Guidelines and procedures from policies, standards and other statutory regulations used by KETRACO to ensure compliance.

- Audio-visuals. M9 said that audio-visual data entailed phone recording, CCTV footage, photo and videos captured by KETRACO staff on different devices KETRACO while performing their duties.

These findings show that to achieve the Company's mission and vision, KETRACO was creating or receiving different types of data in its various directorates. In addition, this data was being created or received in different formats through different devices or channels. The findings also indicated duplication of data created or received in KETRACO because the directorates needed either similar or closely related data for their daily activities. This showed that data in KETRACO was majorly unstructured. Maintaining the three categories of data discussed above is crucial for audit trial, which is an important part of corporate governance. However, its handling is a major course for concern because after their original function has been achieved, it is more often likely to be poorly managed and end up as dark data.

This finding substantiates the initial hypothesis of the big data phenomenon which results to the creation of dark data when employees drag and drop all kinds of work-related files easily, as well as personal files such as personal photos, MP3 music files, personal communications and many more according to (SaaSholic, 2021). IBM (2024)

also adds that email correspondences, PDFs, text documents, social media posts, call center recordings, chat logs and surveillance video footage are examples of dark data created from unstructured data sources. These varied types and formats of data make it hard for KETRACO to effectively integrate for it to be harnessed for effective governance. This then indicates that some of the decisions made at KETRACO are based on either inadequate or misconstrued information, a factor that hinders effective corporate governance. Finally, SaiPoint (2023) observes that the mix of structured, semi-structured, and unstructured data can be difficult to integrate and manage in a coordinated manner, which often results in inaccurate and inconsistent data sets across different data systems.

4.3.3 Sources of Data in KETRACO

Wanting to know how dark data is created in KETRACO, the researcher sought to establish the sources of the data used by respondents in their daily transactions. The researcher found two main sources of data in KETRACO. These were external sources and internal sources of data.

4.3.3.1 External Sources

According to respondent L17, this is data that comes from outside KETRACO. the respondent said:

> *"It comprises the bids/proposals, acceptance letters, pro-forma and invoice among other correspondences from the PAPs, contractors, suppliers and service providers; Financiers' and lawyers'/law firms' bids/proposals, demand/notice letters, fee notes and invoices.)"*

Respondent L11 also describing external data sources said that,

> *"External data is received through various incoming correspondences like letters, journals and invoices from professional bodies; Government documents such as Circulars, The Kenya Gazette, Constitution; visitors (both walk-ins and scheduled ones); external emails; Seminars, workshops, webinars among other trainer-based correspondences; internet; and incoming calls."*

Several respondents cited social platforms like Facebook, X and LinkedIn and the general web as some of the sources of data in KETRACO. For example, respondent L20 said that sometimes they receive information that affect the Company from either the news or social media just to confirm it later from the company communications. These entailed the Company's social media accounts (Facebook, X and TikTok) among other stakeholders' accounts.

Respondent M6 also said that the Company had recently hired an agency whose duties included media monitoring. This was aimed at ensuring that the Company gets daily reports on how they have been covered or mentioned internationally. It was up to the Company to be proactive with these daily reports received from the agency.

4.3.3.2 Internal Sources

This is data that comes from within KETRACO. Respondent L20 said:

> *"Internal sources of data at KETRACO entail staff themselves; work plans; transmission master plan; performance contracts; procurement plans, agreements (contracts, easements, MoUs); policies, standards and manuals; service charter; internal emails; ILIS; various subject files in the Company; calls and peer groups.*

Respondent L6 added:

> *"Internal data is also derived from employment master files, staff files; varied subject files (succession planning files, tender files, internal memo files among others.) in the organization; occurrence books; licenses; assets and liability registers and disposal registers."*

Internal data sources also comprised of reports. Respondent L25 said that they got data from statutory reports; Factory Acceptance Tests (FAT) reports monitoring and evaluation reports; valuation reports; environmental impact assessment (EIA) and environmental and social impact assessment (ESIA) reports; resettlement action plan (RAP) study reports; general field reports and training reports among others.

Another major source of data was the System Application and Products – Enterprise

Resource Planning (SAP-ERP) system which Respondent L3 described:

> *"SAP-ERP is a major source of internal data that provides a wide range of data in the whole Company because it is the main system in use. This is because, as an enterprise resource planning system, it supports the automation of finance, human resources, administration, supply chain, among other directorate processes."*

This respondent gave examples such as data about purchases of goods that can be

retrieved from purchase requisitions, purchase orders, service entry sheets, store

reservations and issuance of goods. In addition, SAP also gives data about staff such as

designation, terms of contract, employment data, salaries, and many more. Several

respondents from the finance department also pointed out balance sheets, income

statements and cash flow statements as their sources of data.

The above responses revealed that the sources of data at KETRACO were varied

because, as all staff had either a computer, or a mobile device and sometime both, so

they could easily create or receive data at their own convenience. Majority of the staff

also had access to internet services, thus making it easy to access data. In addition,

manual correspondences were being created and received from different sources by

different stakeholders. However, the study also showed that there were sources of data

specific to some directorates, and equally repeated sources of data by some directorates.

The findings of the study on sources of dark data at KETRACO revealed that there were

no clear mechanisms, frameworks and guidelines on how to manage sources of dark

data. The findings showed that data was being received or created in the Company

through a variety of electronic and manual platforms/channels. These findings align

with Dash's (2022)'s observation that one might wonder where dark data comes from,

yet it emanates from various sources such as log files and customer feedback. It was

also significant that there were seemingly no clear controls regarding receiving or creating data at KETRACO. It seemed that data, whether electronic or manual could be easily shared among staff and other stakeholders provided one had access to it and was willing to share. However, the lack of control of such data often meant that it would eventually reside in information silos across the Company, with no real prospect of access and re-use. Additionally, lack of control may lead to dark data landing into unauthorized hands.

Because of the bulk and complex nature of creating/receiving data in KETRACO (this entails varied formats and platforms/channels of creation/receipt), some data is frequently disregarded even though it might offer insightful information. Additionally, the findings underscore Davies (2024), who observes that such data is often set aside, deemed too complex, irrelevant, or simply forgotten or sometimes analyzed disparately which does not give a holistic view therefore making the insight derived much less valuable. This can be likened to Kdnuggets (2019) observation that in most organizations, departments have their own data collection and storage processes and systems which may not be known to other departments. The result is that data, even if relevant to other departments, lies unused in the dark in another department. This might negatively impact governance in various ways such as delayed justice or misconstrued judgement and lack of transparency. According to Chaks (2023) to maximize on multiple data sources, organizations need to establish a data governance framework that ensures that data is managed consistently across the organization; use integrated data platform to ease the movement of data between different warehouses and synchronize the data; and use data virtualization, a technique that allows organizations to access and query data from multiple data warehouses as if it were a single source. Therefore, this

means that KETRACO needs to come up with strategies on data governance that will ensure all this data is harnessed for effective corporate governance.

4.3.4 Data Storage, Custody and Usage At KETRACO

The researcher wanted to know how dark data accumulates in KETRACO. Therefore, the research asked how data was stored, who were the custodians of the data and the usage of the data at KETRACO. The following were the findings:

4.3.4.1 Storage of Data at KETRACO

The researcher found out the following ways of storing data in KETRACO:

a) *Electronic storage*

This storage contained data stored in digital format. Respondent M4 said electronic storage entailed the Company online repositories and shared storages such as System Applications and Products - Enterprise Resource Planning System (SAP-ERP), SharePoint, OneDrive ILIS, laptops and phones. Describing the data found on electronic storage, Respondent L22 said that data found in these storages included reservation lists, project data, staff details and service provider details among others. Majority of respondents from the Directorate of Research, Strategy and Compliance noted that the electronic storage, particularly the online repositories and shared storage contained some classified folders with controlled access to be used for official purposes. However, a few staff were aware of these folders, and the majority were not purely using the said folders for official purposes. Therefore, they were not being fully maximized, rather some staff were creating their own folders for official purposes. This was backed up by majority of respondents from the Lower-level management. Respondent L2 alluded to the fact that these storages were not only storing official data but even personal data. L2 noted:

> *"Since we discovered that the OneDrive option on SharePoint is good for back up and sharing big files, we have personal folders on our accounts where we store and share large file."*

Further, L3, L15 and L18 also agreed that they had backed up their sensitive documents on OneDrive.

Some respondents also cited personal and official mobile devices and laptops as storages of data they require for their daily transactions. Respondent L10 was particularly keen on these by saying:

> *"Despite my workstation being at the headquarters, I am always on the move because of the nature of my job. Therefore, a large percentage of the data I use is on my personal mobile phone and laptop. Accessing the online repositories can be tricky outside there."*

In addition, many respondents said that staff laptops and mobile phones contained a wide range of both official and non-official data. Respondent L12 justifying storing both official and non-official data on the staff laptop said, "I have two major separate folders for official and non-official things. And I try to keep them as such though sometimes they mix up."

Respondents M2 and L23 lamented that it was hard to control the kind of data that was created/received and stored on staff laptops and mobile phones. L23 went on to say that "Sometimes staff need to work offline, and this even makes it hard to control their devices."

b) *Manual storage*

Manual storage comprised filing cabinets and safes located in various locations, including the secretaries' or Head of Department's office; the Central Registry and Human Resource Registry, the Strong Room, and the archives/go-down which also housed archival boxes.

Many respondents also said that they were storing data manually in their desk drawers. This data was mostly in the form of notebooks and working files. This was confirmed by respondent L16 who said: "I have a lot of paperwork on my desk and drawer some which am not sure of the content or when they got there. But sometimes I use it in my daily duties". Respondent L3 alluding to the fact that sometimes Company data was being stored at Staff homes said:

> *"During the Corona work-from-home directive I created a mini-office in my house. It has a substantive amount of data that I can use. I have ensured it is updated just in case..."*

Respondent L9 and L12 alluded to some staff having a lot of non-documented data that had been used in some instances to solve daily transactions. The data is found on their desks, personal devices and sometimes at home.

These findings revealed the absence of a specific data storage or repository at KETRACO. The responses showed that data could be found almost anywhere, for example on staff desks, drawers, staff mobile devices, laptops/desktops and staff homes despite there being directorate secretaries' offices, the registry, archives as well as designated online repositories. When the researcher asked the respondents why this was the case, they argued that the mobile devices given to them by the Company had no controls on what they created, received nor stored. Respondents went on to claim that there were loose restrictions in terms of data storage in the online spaces allocated to each staff on OneDrive.

This does not only mean that the Company data is scattered in different electronic devices and manual locations, known and unknown, but there is also a high likelihood that it might be outside the official Company storage. This encourages the accumulation of dark data electronically (in the cloud and local devices) and manually shows

inadequate enforcement of data policy and guidelines or absence of a coordinating unit on data or advocacy of standardized programs concerning standardized storage and its relevance in data management. As a result, it can be very hard to tell who is accessing this data, and even when data is being lost. In addition, it is hard to use data for decision making when its location is unknown resulting in unfair decisions, lack of accountability and transparency. Indeed, the findings correspond to Spinner (2018) and Future Hosting Technologies LLC (2017) who argue that with the illusion of data storage increasingly becoming cheap it's becoming easier to store data, especially by using mobile devices. This accumulation of dark data electronically and manually puts KETRACO data at risks such as unauthorized access and loss of valuable data among others. Factors that can jeopardize effective governance by making the Company more susceptible to inequality and lack of transparency among other risks.

4.3.4.2 Custodians and Responsibility of Data Management in KETRACO

The researcher wanted to know who was responsible for the management of data in KETRACO. The following were the findings:

Several respondents said that the Records Section and ICT were responsible. "Records and ICT personnel are the skilled individuals on matters data management thus making them more responsible" said respondent L2. On the other hand, Respondents M1, L8 and L5 said that the Directorate secretaries were responsible for data management in their Directorates.

Respondent L24 was emphatic that they were in charge of their data files with guidance from the Records Section and not the Directorate Secretaries. Therefore, they got data either from their files in the departments/sections or from the Records Section. In consonance with respondent M2, respondent M4 said that they got data from the

Records Section or from the files they manage in their department and sometimes from colleagues in other departments and PAPS.

However, respondents L1, L9, L19 and L20 said that data management was a responsibility of everyone at KETRACO. When the researcher asked them what they meant by that, they said that all staff were in one way, or another directly linked to the management and storage of data that they need for their daily transactions. For example, Respondent L20 said:

> *"The payroll officer is somehow responsible for the completeness, accuracy and safety of the payroll files among other associated files."*

These findings reveal that the responsibility of data management is not well defined at KETRACO. In addition, it shows there are no specific custodians of data at KETRACO. This is seen when some respondents say they oversee their data, while others are certain that it's the Records Section and directorate secretaries that are responsible for data management.

This means that the responsibility of data management at KETRACO was not made clear to the staff. This is because majority of respondents felt there were some individuals who were more responsible for data management in the Company. These findings substantiate Profisee Group (2019) who posits that clearly defined roles and responsibilities for data management ensure that every action regarding data can be attributed to a particular member or members. Lack of clarity and awareness of the responsibility of data management could result to recklessness in data management, a factor that may contribute to the accumulation of dark data. This results in lack of accountability, which hampers effective governance.

4.3.4.3 Access and Usage of Data

The study sought to find out all the possible users of data found in various departments at KETRACO. The researcher received the following responses:

Many respondents emphasized on staff within the Directorates and from other Directorates as the major users of data. Respondent M4 said:

> *"In most instances, data created or received in our directorate is used by staff from this directorate. But sometimes we get requests for data from outside our directorate or even from outside the Company."*

Respondents L6 said:

> *"Our data is used by staff from our department, management, the Ministries of Energy (MoE) and Finance, and other departments. It depends."*

Respondent T1 and M1 in different interviews said that apart from their directorates being the main user of their data, contractors, service providers, financiers such as the World Bank and other government institutions such as the National Treasury, Ministry of Energy and parastatals such as Kenya Power and Lighting Company (KPLC) among others were also using their data.

Majority of respondents said that auditors (external and internal) were also users of their data. Respondents L22 and M7 also agreed that journalists and bloggers were using KETRACO data.

This revealed that KETRACO data was of interest to many parties, ranging from staff within departments using their own data, to getting requests from staff in other departments and even outside the Company. However, no respondent clearly addressed the issue of vetting the users to know whether they were actually supposed to use the data they requested or had access to. This therefore meant that there was a likelihood of KETRACO data being used by unauthorized people. This also meant that there was

a possibility that KETRACO data could be in unknown hands; people whose intentions are unknown, nor their contribution to corporate governance. This weakness in the Company data control systems predisposes it to inequality and unfair distribution of resources (data being one of these resources) where one group of stakeholders may benefit unfairly at the expense of other stakeholders resulting in loss of stakeholder confidence and trust. In addition, the Company is more exposed to legal, regulatory and reputational risk.

FormAssembly Inc (2024) postulates that if organization data is collected across several platforms for some time, it's tough to keep all this information accurate, updated, and reliable because it is hard to know where data is going or how it is being stored. There is also a high likelihood of creating data silos which are unintended thus resulting in the accumulation of dark data. This can then cause security vulnerabilities and compliance challenges. These findings uphold the Data Curation Center theory by encouraging access, use and reuse of data through ensuring that data is accessible to both designated users and re-users on a daily basis.

4.4 Impetus of the Accumulation of Dark Data

The researcher paused a question on the impetus of accumulating dark data in KETRACO. The various responses were grouped under different categories, as explained in the sections below:

4.4.1 Absence or Inadequate Sensitization on Standard Data Creation/Receipt and Management Procedures

The researcher found out that most respondents were not aware of standard data creation and management procedures in the Company. Therefore, data was just being created/received and managed haphazardly. Respondent L4 said:

"There are no standard data creation or receipt procedures in KETRACO to guide on the creation/receipt of data and its management."

Respondents L7, L9 and L7 concurred that they were just creating and receiving data from different sources and by different devices and that they were not aware of any guiding tools on this. Respondent L9 went on to say that almost everyone at KETRACO, including external stakeholders, was not only handling data, but also creating or receiving different types of data in a variety of formats. To add on, respondent L11 said that they knew that there is a *Records Management Policy* but were neither very conversant with the procedure of accessing it nor its content. Majority of respondents said they had not been adequately sensitized on proper data management in KETRACO, yet they were handling data daily.

The staff found to be conversant with some of the data management procedures in KETRACO were mainly from the Human Resource and Administration directorate, particularly the Records Management Section, and staff from Strategy, Research and Compliance Directorate. This was probably because data management was their direct responsibility. For example, respondent M4 said that the Computer Usage Policy discourages use of Company computers for personal rather than Company activities which could result to accumulation of dark data on personal device without one's knowledge.

This revealed that despite KETRACO having several policies on data management accessible to all staff, only a few staff were aware of them, their location and were using them. This indicated that the various data governance tools had not been well communicated to the staff at KETRACO. Therefore, this pointed to the fact that data creation and receipt was not standardized at KETRACO. This validates NetGovern (2022) who deduces that decision makers and data handlers in the organization need to

be made aware of the existence of dark data in the organization as well as proper data management procedures such as documenting the types of data one hold, its source and who to share it with. Additionally, the findings confirm Morris (2021)' explanation that the use of formalized system of data governance such as use of policies, protocols, procedures and metrics provides a methodical approach of dealing with a company's data to help companies avoid or reduce risks and liabilities as well as enable data mining for maximum use. Therefore, this means that due to lack of or inadequate staff sensitization on effective data management procedures and policies in place, shortcomings such as uncontrollable accumulation of data in different locations emerge, that result to dark data buildup.

4.4.2 Absence of Data Unit/Department

The respondents claimed that all staff at KETRACO were creating and receiving data but not all of it ended up with the Directorates' secretaries nor the Records staff. In fact, most respondents claimed to keep copies of the data they create or receive. Respondents L7 and L13 went on to say that they oversaw the data they created or received. Respondent L13 said that "In fact I manage majority of the Company data I create and receive on my mobile devices or desks." Some of the respondents went on to say that some of the data they hold has never reached the designated data personnel (Secretary) in the directorates.

This revealed that almost everyone at KETRACO was acting like a data manager who was most likely managing their data in varied styles. For accountability and transparency purposes, organizations need to have staff who are particularly responsible for the organizations' data. These may include staff such as record officers, data protection officers, data managers, among others who are under the data unit.

These are responsible for the organization's data throughout its life cycle. This is consistent with Podraza (2014) and Holdsworth (2025) who concede that a good dark data management program should have a data management team to develop and implement a strategic plan for the program as well as provide a governance structure where team members have their responsibilities clearly defined in matters legal data management to enhance accountability. Therefore, not focusing on designated teams to be responsible for dark data meant means that there is reduced accountability for data at KETRACO. It also means that there are unstandardized data management styles that are likely to give birth to lots of unstructured data in the company, which is also a recipe of accumulating dark data. A data management unit will oversee functions, roles, activities and services of data/dark data. This is in addition to spearheading formulation, implementation and updating of data policies; advocacy campaigns and creating awareness; and laying down strategies and mechanisms for dark data management among others.

4.4.3 Absence of Proper Policies and Guidelines on Fair use of Personal Devices

The researcher found out that KETRACO had greatly embraced technology whereby majority of staff used laptops, mobile phones, landlines/desk-phones and tablets, while some used electronic systems such as Biometric Systems, CCTVs, survey equipment like handheld GPS, ILIS equipment and many more.

Majority of respondents also said that they were creating and receive Company data through official and personal devices. Some went on to say that they were also storing a lot of Company data on their personal devices. Respondent L4 explained:

> *"Most of the time I am in the field, either receiving or creating data. My laptop and mobile phones are enough for these exercises. I capture PAPs' details, contractors' details and sometimes even staff details."*

Respondents L14 added that despite having PAPs' data on the personal laptop, the personal mobile phone carried the same. Majority of respondents agreed that they could access the work email on their mobile phones therefore creating an enabling platform to receive and store Company data on these devices. They also said that they could access SharePoint and OneDrive platforms on their personal and official devices.

Respondent L11 said:

> "To be sincere, this tablet that the Company bought for me has become so personal. It contains both official and non-official information. In my gallery, it has passport photos, IDs, bank details, pin certificate and more for the PAPs. It has my documents and photos too. There is no limiting factor that I know of."

In agreement, respondent L14 attested:

> "Since the Company bought me a tablet, I have never bothered to replace my lost phone. The tablet does not only serve me officially but personally too. I have my personal documents and photos here too. We are not restricted to do that."

This revealed that KETRACO had enabled staff to create or receive data on their personal devices. This is evidenced when respondents say that they could access their Company email, SharePoint and OneDrive on their personal smart devices. In addition, the findings showed that Company devices were being used to create and receive both official and personal data without any limitation unless maybe one ran out of storage.

The enablement of staff to use personal devices for official purposes as well as using official devices for personal reasons means that official data has a likelihood of lying in isolated personal storages. In addition, it also points to high chances of chunks of personal data occupying official storage. These results in hoarding a lot of data with time where majority translates to dark data. This is in line with the *"Bring Your Own Device"* (BYOD) phenomenon which, according to Dimitrov (2018), allows employees to source organizational data using various techniques, including through personal gadgets like phones, laptops and tablets. This then creates information governance

challenges as a result of information being generated by and stored on mobile devices, social networks, file sharing services, and unmanaged SharePoint sites. Indeed, data becomes more difficult to discover, retrieve, and reuse thus leaving some information in the dark, thus accumulation of dark data.

Additionally, this finding shows a risk of dark data being found in wrong or unauthorized hands. This is besides the risk of valuable dark databases becoming lost or discarded when staff or some stakeholders such as contractors exit or disengage with the Company which may result in dysfunctional succession in managing data.

4.4.4 Poor Data-Sharing Culture

The researcher found out that there was a poor data sharing culture in KETRACO. Several respondents in the Lower-Level Management said that they feared actively sharing data and knowledge lest they are mistaken for trying to outshine their supervisors. In concurrence with this, Respondent L5 said:

> "*Sometimes I may have certain data that will save the day needed in a meeting or share with colleagues in other sections, but I can't give it out, lest my supervisors look clueless and am accused of sabotage.*"

Respondent L2 suggested:

> "*To encourage information sharing, maybe something like a data hub can be created where frequently requested for and most shared data is put. This can be for the whole Company or in the directorates.*"

Respondent H6 explained that:

> "*Some staff don't like sharing information. Yet, this information may be useful in solving some daily problems by others. In addition, it is also hard to know who has the information you require sometime.*"

Respondent L22 emphasized these by saying that some staff don't share nor train others in their areas of expertise for the fear of being outshined, while on the other hand some

other staff are not willing to learn from experts for the fear of being overworked. So, when these experts leave, there is a big data gap.

These responses revealed that employees were holding on to their ideas due to fear of intimidation or being overworked, competitive advantage, lack of motivation and basically lack of sharing strategies among other reasons. In addition, despite respondents acknowledging that they could share some of the data they had with other departments, there was no defined channel for sharing where one could know who has what and who needs what. As a result, they held on to a lot of data in their personal storage. Data that could otherwise be shared to minimize duplication or losing opportunity. Consequently, this disconnect was likely to enhance the accumulation of dark data in the Company. Data that could rather give valuable insight that can be used for strategic decision making to enhance corporate governance.

This corresponds to Emerging India Analytics (2024)'s opinion that fostering a culture that values data exploration and innovation by encouraging collaboration between different niches of employees such as data scientists, analysts, and business stakeholders is crucial for successfully leveraging dark data in decision-making processes. As much as technology supports people in their jobs, staff input is essential. This is achieved by organizations promoting a culture of information sharing by encouraging their employees to make their knowledge available to others, which is not only good for the company, but also good for the employees who feel useful to the company's overall success. Additionally, the findings validate Kdnuggets (2019) and Pal (2021) when they explain that due to lack of comprehensive data sharing policy, departments in most organizations have their own data collection and storage processes and systems which may not be known to other departments. As a result, data, even if

relevant to other departments, lies unused in the dark in another department. Thus, the accumulation of dark data in the Company.

4.4.5. Poor Access Controls

The research found out that there were no proper data access control mechanisms at KETRACO. This was brought out by respondents who claimed it was hard for them to access some information they use daily. Respondent L17 explained that:

> *"Sometimes we are forwarded links to documents that have the data we need in OneDrive. But the procedure for getting access to view those documents is long and tedious. You must make a lot of requests, wasting a lot of time. The solution sometimes is constructing a new document and letting the one on OneDrive be."*

Some respondents said that there are instances when they could access storage locations that they are not authorized to see and were even able to do modifications on the content there if they wished to. Respondent M4 even disclosed that there was an instance when they were able to access their colleagues' laptop contents from their own laptop.

> *"One day when I clicked the network option on my file explorer I saw an icon with a colleague's name. On opening it, I realized that I could access everything on their machine. Someone with malicious intentions could have added or deleted files or documents without them knowing. I notified ICT who rectified that security hitch."*

This revealed that there were weaknesses in the system that could enable duplication of data and sometimes even the creation of unsolicited data in storage while exposing the Company data to unauthorized access. Therefore, this duplicated and unsolicited data is very prone to become dormant just eating up storage space and never to be used or re-used thus becoming dark data.

These findings also uncovered the likelihood of staff accessing colleagues' accounts and a possibility of hoarding data there without the account owners knowing. The evidence upholds preliminary findings that access control, is a key component of data security that regulates who can access a particular dataset, and specify whether they can

do modifications to that dataset by identifying an individual or entity, verifying that the person or application is who or what it claims to be, and authorizing the access level and set of actions associated with them.

Therefore, these findings mean the following: with poor access mechanisms there is a likelihood that staff can uncontrollably access and hoard unsolicited data in their colleagues' account without the account owners knowing; and staff may duplicate data when they could get it somewhere else due to difficulties in granting access to some data. These can thereafter fuel the accumulation of dark data in the Company. Issues such as hoarding unsolicited in colleagues' machines without their knowledge might result in lack of accountability, transparency, fairness and social responsibility. All these factors hinder corporate governance.

4.4.6 Budgetary Constraints to Purchase Data Processing/Management System

Respondents inferred that despite having data creation/receipt and storage tools, there was no proper data management system. Respondent L25 said that good data management systems were very expensive to purchase and maintain, therefore KETRACO relied on the minimal capability of the systems it had in place to manage data which was not capable of controlling growing dark data. For example, respondent M7 said that SharePoint did not provide mechanisms that would enable staff to easily know who has what data. Therefore, this resulted in a lot of duplications in the system. The respondent went on to say that their contact person for SharePoint had said that the features that the system came with were directly proportional to its cost.

Therefore, this revealed that there were some shortcomings in the system which were encouraging the accumulation of dark data, shortcomings that came up due to financial shortages. This corroborates Gartner (2025)' hypothesis that storing and securing the

data usually entails greater expenses (or even risk) than the potential return profit in case of a breach. This means that without enough budgetary allocation, organizations will be unable to purchase better data management systems with Artificial Intelligence (AI) and dark data machine learning software. Systems that can minimize the accumulation of dark data by digesting data from their source into an asset catalogue that has all the relevant metadata for its management. For example, such high-end software for managing dark data as advised by DFIN (2024) include AI-powered software that enable companies to scan millions of pages quickly and have pattern detection capabilities that allow the software to not only classify types of dark data but to pinpoint sensitive information, such as phone numbers, names, or Social Security numbers.

4.4.7 Absence of Involvement of all Data Stakeholders in Selection, Designing and Implementation of Data Management Systems

Most respondents, from the findings of the study expressed their concern about not being involved when KETRACO was designing or selecting its data management systems as well as not getting adequate training on the system. The systems mentioned most by respondents were SAP, SharePoint and OneDrive. Other respondents also talked about lack of involvement in designing and/or subsequent training/guidance on implementing the Company Master Records Classification System. Respondent L14 explained that:

> *"We were never consulted on features we wanted the system to have. It was just kind of delivered to us. In addition, we did not get adequate training on how to use the system. For example, I get a certain email at least twice daily saying that I recently submitted documents to Records Center and additional information is required to complete the submission, with a link which never leads me to where I can complete the submission. This repeated mail is eating up my mailbox storage and I don't know how to stop or complete it. And am sure am not the only one experiencing this."*

In agreement with this, respondent L23 said:

> *"Majority of staff have duplicate copies of documents in the storage, shared severally by individuals hoping that it could be worked on collaboratively just to fail repeatedly. This is not because of a faulty system but lack of proper training on the SharePoint features."*

Emphasizing the lack of involvement when selecting the system in place and subsequent training, respondent L6 said that some staff were not even aware of the classification system in the organization. L6 went on to say that that is why documents ended up in wrong files, never to be traced when needed due to wrong reference.

Respondents went on to note that lack of training of all staff or biased selection of staff to be trained on new data management system also added up to lack of involvement which made it hard to work with the rolled-out data management systems.

This revealed that despite there being a system that could possibly mitigate the duplication and accumulation of dark data, some employees were still not able to maximize on this. This was as a result of not involving all the staff/system users in designing the system to propose the system features. In addition, there was no adequate training or user support for using the system. This corresponds to recent literature on the subject that user involvement in information system development is usually considered as vital mechanism to enhance system quality and ensure successful system implementation as it helps system developers identify users' pain points, preferences, and behaviors, and incorporating these insights into the design process (Techved, 2023). This therefore means that dark data also was accumulating in KETRACO as a result of inadequate knowledge of using the data management systems in the company. This ranges from both electronic and manual as well as the Company's records classification scheme.

4.4.8 Perceived Potential Value f Data

The researcher found out that respondents had a lot of unnecessary data on their office desks and in their mobile devices. They even agreed that some of the data had been long outdated and was even redundant, just eating up storage space and even time when searching for specific information. When they were asked why they were hoarding the data, majority cited its possible potential which they could not explain clearly. Responded L23 said:

> *"I have files and documents on my computer for as far back as 2015. You never know when they might be needed. I can't delete them."*

Further, when asked how useful these documents or files for 2015 were, the respondent said:

> *"They don't have to be useful now. In fact, I have not opened some in years. But you never know when they might be needed. I have seen such cases before."*

These responses demonstrated that staff in KETRACO were holding on to data they created or received with anticipations to use in future; anticipations that they could not clearly explain or attach to each dataset they held on to. As result a lot of data accumulated in the company, both useful and irrelevant data.

This corroborates Future Hosting Technologies LLC (2017) who observe that because data is generally deemed to have potential value, many organizations opt to store it despite having no idea of how to generate value from it but rather hoping to figure out what to do with this data one day which is kept 'just in case' but doesn't have proper usage in an organization as dark data. Therefore, this means that as some staff keep hoarding data in KETRACO's prime storage facilities and locations with hopes of using it in future so do they give room for the accumulation of dark data in the company. This hinders effective governance due to delays in decision making when the needed data lies in the dark and one has to take a lot of time to retrieve it. This results in delayed

justice and unfairness thus hindering governance. Moreso, needed data may not be traced at all because it lies in the dark thus resulting in lack of accountability, transparency, fairness and responsibility. The Data Curation Center model emphasizes appraisal and reappraisal to allow preservation of valuable data.

4.4.9 Integrating Data from a Variety of Resource/Platforms

The research found out that there was no specific system/tool for integrating the data from the varieties of sources of data in KETRACO. Respondent L10 explained that:

> *"Whenever I go out for field work, I create and receive a lot of data in different formats. These include photos of structure, hardcopy documents like photocopies of bank details, sometimes I scan Identity Cards and Bank Cards, hard copies of correspondences and signed agreements among many. The problem sets in when I must harmonize this data to make it usable. It is hectic and confusing and many a times I have left out some details unknowingly, yet I have them in my cocktail of data collected. Therefore, I am forced to go to the field again to collect this data."*

In addition, respondent M2 said that the sources of data in KETRACO included social media pages, ERP applications, customer logs, CCTV footage, Biometric systems, e-mails, presentations, reports, and GPS systems among others.

This showed that in KETRACO, data was being created or received in different formats and using different devices. This resulted in the creation of data silos which need great efforts and attention to harmonize them for their usability. This challenge therefore resulted in some of the data lying untraced thus unused in some devices while also creating opportunities for duplication of data. This supports Sharma (2020) that data in an organization comes from a variety of sources, such as CCTV footage, ERP applications, biometric systems and emails among others which can be challenging to combine these data to create a report. In return, whatever is not used often ends up accumulating as dark data.

Therefore, this means that failure by KETRACO to integrate data from these different sources gave rise to data silos and eventual accumulation of dark data. These unexplored dark data silos could hold information which if analyzed and used could be used to make decisions that enhance corporate governance in the Company. The Data Curation Center model states clearly about description of data to enhance data management and guides on acquisition of data management systems with elaborate metadata to to allow accommodation of all formats.

4.5 Policies and Standards that Regulate Harnessing of Dark Data

The research also sought to find out the existence of policies and standards that can be used to harness dark data at KETRACO. From document analysis, it was established that KETRACO has several policies that can be used to harness dark data for effective corporate governance. The researcher also sought to know how accessible and familiar these policies were to the respondents. Respondent M6 said that these policies were available on SharePoint 365, the Company online data platform for storing, organizing, sharing, and accessing information for all employees in KETRACO, and this had been communicated on mail to all employees. Several other respondents also agreed with this. However, some were not aware of how to access the policies. For example, respondent L11 said, "I know there is a *Records Management Policy*, but I don't know how to access it." The policy documents that can be used to harness dark data at KETRACO are discussed below. However, the analysis and interpretation regarding the familiarity to these policies was done as a block (see section 4.5.1). This is because when the various respondents were interrogated, their responses were mostly the same, therefore the researcher wanted to minimize repetitions.

The findings were as follows:

a) Records Management Policy – April 2020

The researcher found out that KETRACO had a *Records Management Policy* that had been last reviewed in April 2020. A number of respondents were aware of this policy. Respondent M3 said that the *Records Management Policy* set out guidelines for the creation/receipt, use, maintenance, and disposal of records to ensure that records are properly created/received and maintained for accountability, administration and memory. In addition to this, respondent M6 said that this policy underscored records creation as a responsibility of all employees who should ensure completeness and accuracy of records with all the accompanying metadata.

After further interrogation, the researcher noted that despite respondents being aware of the existence of this policy, those from the Human Resource and Administration Directorate were the most conversant with the content of this Policy, followed by those from Research, Strategy and Compliance directorate. This was because the implementation of this Policy was largely the responsibility of the Human Resource and Administration Directorate but data management matters directly fall under the strategy Research and Compliance Directorate. Respondent L4 said:

> *"I am aware of the Records Management Policy's existence, but I have never read it. I am aware that it is somewhere on SharePoint. I just believe that I handle my information well, but I will get time and read it."*

This finding indicates that records management and by extension data management is the responsibility of everyone at KETRACO whereby data creation is to be done on need basis (this entails accountability, administration and memory) with complete metadata. In addition, the findings reveal that there is guidance on the use and disposal of data. This means that the accumulation of dark data can be mitigated by applying

this policy whereby only relevant/needed data is created or received with its complete metadata. In addition, it also means that with the policy, accumulation of inactive data that gives birth to dark data in the Company would be minimized as data would be disposed of appropriately. These findings substantiate Linnik (2023) who advises that such policies determine which types of data should be kept and which should be deleted, when it should be deleted, and the specific manners to do so securely. This is to avoid legal and financial implications, if data covered by a specific mandate or a regulation appears anywhere in dark data collection is exposed.

b) E-Waste Management Policy – October 2020

From majority of respondents from Strategy, Research and Compliance Directorate and Human Resource and Administration Directorate the researcher found out that there was an *E-Waste Management policy* whose purpose was to minimize generation of e-waste; mobilize and sensitize stakeholders on the proper management and handling of e-waste on a sustainable basis; and ensure proper handling and management of e-waste in KETRACO.

Respondents L25 said:

> *"This policy guides on disposal of obsolete electronic data storage devices. For example, when they should be destroyed and how they should be destroyed. Therefore, this policy is to be used to ensure that ephemeral electronic data or storages don't overstay lest their existence is forgotten since they are inactive."*

In addition to these, respondent L17 said that the policy had stipulations such as need for appropriate deletion or extraction of data to be done first in correspondence with classification whenever e-waste, is equipment containing electronic data. This is to ensure that no sensitive data lands in foreign hands during destruction.

This showed that the policy could be used to ensure that when data storage technologies become obsolete, they are appropriately disposed in a timely manner while ensuring that only relevant information is migrated to current storage. Therefore, this means that the policy can be used as a measure to mitigate the accumulation of dark data in obsolete storages and versions of documents. Linnik (2023) agrees that dark data is also found in outdated versions of documents and software components which are very easy for hackers to access. Therefore, with such a policy in place, then the chances of accumulating dark data in e-waste are minimized by timely migration of relevant data to current technologies, deleting data from outdated storages and appropriate disposal of electronic data storages.

However, this policy seemed not to be actively used by stakeholders as the majority of respondents from other directorates save for Strategy, Research and Compliance Directorate seemed not to be aware of this policy nor some of its content. This meant that there was a likelihood of critical data lying in inappropriately disposed obsolete technologies. This makes it easy for malicious groups such as hackers to access that data and use it against the Company. For example, they might leak personal identifiable information of stakeholders to the public resulting in legal suits or extortions.

c) **Acceptable use policy – October 2020**

From respondents in the Strategy, Research and Compliance Directorate, the researcher learnt that this policy is meant to guide KETRACO users (staff, partners, suppliers and contractors, auditors and consultants among others.) on the acceptable use of information assets, assets associated with information and information processing facilities. Respondent M7 said:

> *"In this context, an information asset or resource is an element of infrastructure that enables the transaction of certain selected significant and relevant data, prepared so as to provide content and information services that can be used directly by the user. For example, SAP invoicing interface for service providers and contractors."*

In addition, Respondent L10 said that the *Acceptable Use Policy* also gives employees steps of how to return all of Company assets (these include mobile devices, laptops and other electronic devices as well as Company information on personal devices) in their possession upon termination of their employment, contract, or agreement.

Respondent L3 went on to say:

> *"The Policy guides on issues such as incidental use and use of KETRACO information assets outside offices. It is strict on nominal use of information assets in terms of storages and network to avoid interference with an employee's performance."*

Nevertheless, from the previous responses majority of respondents from other directorate seemed not to be aware of this Policy. Respondent L11 said:

> *"To be sincere, this tablet that the Company bought for me has become so personal. It contains both official and non-official information. in my gallery, it has passport photos, IDs, bank details, pin certificate and more for the PAPs. It has my documents and photos too. There is no limiting factor that I know of."*

This indicates that there was a clear policy on the use of information assets in KETRACO that could assist in mitigating the accumulation of dark data. For example, only using the company information resources for official and authorized purposes rather than personal reasons. However, not all respondents were aware of it. This means that this policy was not being fully implemented. Therefore, dark data was creeping into the company through substantial use of official devices by staff for personal use or creating and receiving official information on personal devices. This could result in accumulation of ephemeral data in storage that could have rather been used to store important information. In return, the Company will have to unnecessarily incur more storage costs to accommodate the relevant data.

d) Quality Management System (QMS) Policy, 2020

This policy is based on ISO 9001:2015 Standard Clause 7.5 whose purpose is to ensure effectiveness and consistency in the control of Management System documented information. It provides guidelines on processing information and data that result in developed and approved documents. Respondent M6 who seemed to be the only one that was conversant with this policy said:

> *"On matters data, QMS emphasizes creating information only when there is need and after approval by the relevant authority. The Policy also puts emphasis on who should participate in creating or receiving particular documents. The standards further outline the management and use of the approved data such as storage, protection, back-up, retrieval, use, retention and disposal of records among others."*

This reveals that by using this policy one can ensure that KETRACO only creates or receives relevant data into its storage whether manual or electronic. In addition, this policy also enables accrediting creation of data in KETRACO to designated individuals. With this policy, it signifies that there will be minimal room for accumulating dark data due to creation of only approved data on need basis and by approved individuals. This agrees with Ghosh (2025) that to prevent dark data from clogging up storage, organizations should only accept and store data for a valid reason and not just because the storage is cheap. Additionally, the findings also support El Haj (2023) that to unlock the potential of dark data entails putting in place policies that will also define data ownership and access control among other issues. However, at KETRACO, the policy seemed not to be fully implemented by many respondents as they mostly agreed to creating and receiving Company data autonomously. This therefore meant that there was no guarantee that data storage contained only relevant data nor data could be attributed to their creators. Thus, there are high chances of dark data in the company eating up storages and loss of finances due to unwarranted acquisition of more storage space for relevant data.

However, its key to note that there are some ISO standards that would be very useful in dark data management, but which were not being used at KETRACO. These include ISO 8000-1:2022 that focuses on data quality. According to ISO (2022), the widespread adoption of digital computing and associated communication technologies are making organizations become dependent on digital data, a dependency that amplifies the negative consequences of lack of quality in these data. Therefore, unreliable decisions made based on low-quality data decrease organizational performance. Thus, the ISO 8000-1:2022 edition which aims to ensure data quality in organization to enhance performance. In turn, enhanced organizational performance is expected to go hand in hand with enhanced corporate governance.

4.5.1 Analysis and Interpretation of Familiarity with the Policies

The researcher went interrogated the respond to know their knowledge about the policies in terms of accessibility and content. From the above responses, there are policies in KETRACO that can be used to harness dark data. However, it seemed that majority were not familiar with these policies save for those from the directorate of Strategy, research and Compliance, and the directorate of Human Resource and Administration. This could probably be because the responsibility of implementing these policies lies directly under the two directorates. For example, for the *Records Management Policy,* Records Management Section is under the Human Resources and Administration directorate, while for the *Acceptable Use Policy,* the ICT department which is under Strategy, Research and Compliance Directorate oversees Computer usage in the Company. In addition, some respondents also attributed this to inadequate sensitization about the policies. These entailed awareness of the policies' existences, access, content and usage.

Therefore, this means that there was underutilization of these policies, particularly by expected users from other directorates. There could be also possibility that some staff do not use the policies at all. From the analysis, this forms the largest percentage. Bearing in mind the previous findings that showed that almost everyone was creating and receiving data at KETRACO, then dark data would thrive due to non-adherence to policy by most staff.

This substantiates El Haj (2023) underscore of the nexus between optimal use of dark data and implementation of data governance policies and procedures in organizations. This is because, data governance structures such as policies and procedures are key in encouraging cross-functional collaboration to maximize the utility of insights derived from dark data; implement continuous monitoring and auditing processes to ensure that dark data remains well-managed and compliant; disposing data that no longer services the purpose as per the defined data retention policies that align with legal and regulatory requirements and establishing clear guidelines for data management, security, and compliance, especially for sensitive dark data. Additionally, the finding validate Polymer (2024) argues that organizations should implement clear policies about data retention and destruction to minimize accumulation of dark data through necessitating regular data audits to ensure outdated or irrelevant data is properly discarded. The inadequate way in which KETRACO was utilizing existing policies that would be suitable for managing dark data had negative consequences for corporate governance as manifested by unauthorized access to company data.

4.6 Implications of Risks and Opportunities Associated with Dark Data to Corporate Governance

The study also sought to find out whether there were risks and opportunities associated with dark data to corporate governance at KETRACO. The researcher found out that

indeed, dark data had both risks and opportunities which could only be well manifested if the data is brought out to light as discussed in section 4.6.1 and 4.6.2.

4.6.1   Implications of Risks Associated with Dark Data to Corporate Governance

4.6.1.1 Loss of Time

Accumulation of dark data was resulting to a lot of wastage or loss of time when searching for needed data by respondents. Most respondents said that in most instances, it took them a lot of time to find information whenever it was requested for. Respondent L6 said:

> *"Sometimes I am sure that I have some of the information being requested for, but I can't trace it. It becomes even worse when I am not sure whether the information is in my machine or somewhere in my drawer or desk in a printed version. I end up wasting a lot of time to trace it. Sometimes I don't even get it completely."*

They attributed this to a lot of unrecognizable data that they had accumulated in their official storage (electronic and manual) where some were irrelevant to the Company. They said that they often took a lot of time to search through their machines and desks at time to get data which they felt that they possibly had. This finding agrees with Javanainen (2015) who posits that a lot of unnecessary time and energy is spent, searching for data that is in the dark which they never even get at times. In addition, the evidence confirmed International Data Group, Inc (2021)'s explaination that businesses suffer inefficiencies and reduced productivity when users must cobble together data from multiple siloed sources to analyze, make correlations and gain insights. This shows how time can be lost while trying to dig out needed or useful data from various dark data sources, resulting in reduced staff productivity which is a detrimental factor to corporate governance. (see also section 4.4).

4.6.1.2 Wastage of Storage Space

Majority of respondents acknowledged that they had a lot of data in their electronic devices as well as in physical locations such as their desks. Respondent L12 confirmed that occasionally, they had been forced to pause their daily activities to clean storage both on their personal devices (laptops, tablets, mobile phones and more.), mailboxes and even desks to be able to accommodate incoming files and documents. In addition, respondent M5 said that there was a time when they could no longer receive emails because their mailbox was filled up with transient and redundant email subscriptions mostly from social sites.

This shows how dark data eat up space that could be used to store relevant and productive data in organization storage. This is because of the clear indication that some of the data eating up storage at KETRACO is either of no use, or its use has not been discovered. This is in brings about Kidd (2023) who says that dark data consumes significant data center capacity on a global basis yet its invisible to be used. Moreover Domo (2024) also explains that dark data can take up valuable storage space in servers and data warehouses. This means that organizations, in this context KETRACO, risk missing relevant data in their storages, data that can be used for corporate decisions simply because, their databases are clogged with dark data.

4.6.1.3 Loss of Reputation

Majority of respondents also observed that access to dark data by unauthorized parties could result in loss of individual or Company reputation. They said that they had a lot of electronic and manual data in their possession but were not totally aware of its details and sensitivity level. Respondent L7 said:

> *"My computer has a lot both personal and official data. I cannot even account for everything that is on it. Especially the downloads are a lot. And drafts or versions of documents."*

This confirms previous literature by Tittle (2014) that dark data's secrets may be too dark and damaging indeed, but one has no way of knowing for sure, and Sarkar (2017) that dark data may appear to be potentially harmless yet carrying potential privacy risk. This shows that in case of data breach, sensitive and confidential data, will result in a compromise of reputation to the company losing stakeholder confidence and trust.

4.6.1.4 Loss of Opportunity

Some respondents said that there are instances where they realized they had let some opportunities go just because despite being sure they had the necessary requirements, either they could not retrieve this information in time, or they couldn't retrieve it at all. Respondent L19 said:

> *"During one of the ISO audits, we were flagged for non-compliance of a certain process because we couldn't provide the evidence. We were all sure that we were compliant but just because we couldn't trace the evidence in time, we failed. We did not get that certification. We later got the document with the data that was required a year later in one of the boxes in a certain office."*

This revealed that there were a lot of documents with useful data that lied undetected and not taken advantage of in KETRACO. This was most likely because the information had been clouded by a lot of unprocessed data that accumulated in the official storages. The findings corroborate Kdnuggets (2019) that while some companies decide not to analyze and process dark data, their competitors do, making them a mile ahead because of the value and intelligence they extract from the dark data. Also, this is in line with Izenda Inc (2019) and Sarkar (2017) that lack of analysis of dark data for use may result in organizations losing vital opportunities. This is because dark data may contain unknown and untapped sources of intelligence as well as sources of exposure to loss or harm. This therefore meant that despite KETRACO holding

useful data that it could use for its own advantage, it was not maximizing on this because the data lay hidden in the accumulated dark data. Thus, it leads to loss of serious strategic opportunities such as certifications.

4.6.1.5 Health Risk – Insects and Rodents

Respondents expressed their concerns about increasing dust and numbers of insects in the offices because of the accumulated papers and files in, on and around staff desks as well as other storage places. Respondent M4 said:

> *"It's about four years now since those files were put in that cabinet. We have never touched them, neither do we know their content. We just assume that whatever they hold is important."*

Such were the confessions by respondents to never having touched again some of the files on their desk since they created or filed them. Some did not even know the contents of the file other than the dust that had accumulated around them and the insects they saw occasionally around them. These could not only lead to common colds and coughs but also other health complications. This corresponds to Hollingshead (2024) that poor employee health means slacking on the job and eventual business losses. In addition, that total health related employee productivity loss accounts for 77 percent of all such losses, and costs employers two to three times more than annual healthcare expenses.

4.6.1.6 Legal Risk

Several respondents said that some of the data they randomly hold in their laptops, mobile phones, tablets, desks and drawers have legal value such as personal identification information, client/staff bank details, machine log-in details and many more. However, some of the data on these personal devices don't have appropriate security controls in place. Respondents observed that in case of any data breach there can be serious legal issues arising. Despite the awareness of the potential risks

associated with their data handling practices, most of the respondents said that they were not taking any specific action to remedy the situation; rather, they were either deleting or transferring this data to the designated storages only when they stumbled upon it. They said this was also dependent on availability of time, internet and sometimes psych.

This finding revealed that employees were not taking personal initiative to look for the information, and whenever they stumbled on it, the subsequent actions were dependent on convenience. This increased the risk of accumulating dark data of legal value in the Company. These findings resonate with Spinner (2018) that many organizations are not even aware that some of the dark data in their custody contain such vital information like employees' personal identifiable information or customer payment information among others. As a result, organizations fail to provide the necessary security for such data. This provides a good platform for attackers who most definitely zero in on such information, making data regulators to demand for explanations/answers. This means that these staff and KETRACO as a whole risk legal tussles in court in case there is a data breach. In addition, such poor data handling practices were in direct contravention of stipulations in existing policies, as noted in section 4.5.

4.6.2 Implications of Opportunities Associated with Dark Data to Corporate Governance

The research sought to find out whether there were any opportunities associated with dark data. Respondents conceded that indeed dark data had a lot of hidden potential which could only be achieved if strategies for data management from conceptualization, through acquisition, use to disposal are put in place at KETRACO. The following were the potential benefits found to be associated with dark data:

4..6.2.1 Efficient Service Delivery

Majority of respondents agreed that dark data could be used to enhance service delivery. They explained that dark data could provide details that will enable clients get value for their money. Verbatim response from Respondent L2:

> *"Well processed and analyzed data about PAPs will enable us to compensate them correctly. With this I mean being right about the parcels affected, the size of parcel, the structures and crops in the parcels that are affected and so on."*

This sheds light to Pal (2021) who says that analysis and processing of dark data can help organizations to build a smarter supply chain and Domo (2024) who discusses that leveraging dark data can help one to focus their efforts on the things that matter which in turn, can help them to save money and increase efficiency. This therefore reveals that if KETRACO makes it a habit to analyze dark data such as telephone conversations then they will respond to customer needs more efficiently.

4.6.2.2 Timely and Effective Decision Making

Majority of respondents observed that if all official data could be processed and made accessible to relevant users, it could result to timely and effective decision making. Respondent L10 underscored:

> *"Sometimes we delay some payments because of delays in provision of the data we require. This is because, some staff can't trace the required details from their machines and desks, yet they are sure they captured it. In some cases, it is worrying when someone assumes or thinks that they have all the data for a certain transmission line in their piles of data just to realize that some PAP's details were not collected at all."*

In agreement with this, respondent L13 explained:

> *"My machine and phone are full of both personal and project data. This is because while in the field as much as we are working, we are also creating personal memories. Unfortunately, these two mix up and if someone is not so keen on creating time to separate them, then the sequence continues over the years. That has been my case since I joined the Company in 2015. Therefore, I take time, even days to retrieve some information that could just take at most five minutes if I had been organizing it properly. Sometimes I am completely unable to retrieve it and am forced to go back to the field to collect it. Sadly,*

sometimes I stumble on it later when am back to the office. And this applies to many other colleagues too."

The findings agree with Feliu (2020) that using data-derived insights allows organizations to move from decision making based on intuition to making informed and sound decisions. KETRACO has a delivery service charter which dictates the turnaround time for specific services. If staff could only make the decision to process all the data they create or receive, then effective decision making could be on a higher level. When one needs data to make time-based decisions (for example make payments) they will not have to waste time searching for or recollecting data which they could already have.

4.6.2.3 Customer Satisfaction

Dark data can help meet customer needs. Respondent L24 noted as follows:

> *"We are continuously receiving calls both from our colleagues as well as from outside the Company. These calls carry a lot of data which ranges from complaints to complements, inquiries and advice. Most of the time, once the caller is disconnected, that is it with that call. I wish we could process this data to help KETRACO design its services to meet the clients' need."*

This reveals that there is a lot of input from both external and internal stakeholders that lies hidden in the dark. Input that can only be used when processed. This supports previous literature that dark data can help companies by providing an all-rounded view of the product/service and as well as market view through leveraging customer feedback and improving product/service quality (Pal, 2021). Therefore, this means that if KETRACO processes the dark data particularly from the customer feedback, they will be able to design their services in a manner that aligns with the customers' needs thus ensuring customer satisfaction.

4.6.2.4 Cost Reduction

The researcher found out that processing dark data at KETRACO could result in cutting

costs by relieving storage space occupied by redundant data. Respondent M21

explained:

> *"Analyzing dark data will enable removing the redundant data while retaining*
> *only useful one. As a result, this will create space for more useful data in the*
> *Company. Meaning, the Company will save money that could have otherwise*
> *purchased more storage."*

Respondent M4 added that processing dark data will reduce data duplication at the same

time remove ephemeral data from official storage thus creating more space for useful

data while saving on cost.

This confirms Joshi (2019)'s assertion that by analyzing and processing dark data,

organizations will not only fix the storage space issues but also foster financial savings.

Additionally, the findings also validate Bley (2024) that by identifying and analyzing

dark data, unused resources can be detected and removed leading to significant cost

savings, whether by discontinuing services that are no longer needed or by optimizing

processes. Therefore, this means that if KETRACO actively processes its dark data, it

will be able to not only create more storage spaces but also focus its efforts on things

that matter, thus cutting off unnecessary expenditures.

4.6.2.5 Compliance with Regulations

The researcher found out that processing dark data could not only result to assigning

appropriate security levels to data but also ensure that data access controls mechanisms

are in place. This will result in minimizing the legal risks posed to KETRACO due to

data breach as there will be minimal exposure to confidential data by unauthorized

parties. M3, M7 and L6 opined that if employees knew the content or were in a position

to process all the Company data in their possession, then the company will be more compliant with regulations. Indeed, Respondent M4 observed as follows:

> *"If we could only ensure that all the data in our machines and cabinets is well processed then it will also save us from a lot of legal tussles. We will be aware of the kind of information we have and how to protect it."*

This confirms previous literature that dark data may include information that is subject to various regulations and compliance requirements, such as personal data protected by data protection laws such as GDPR whereby failure to properly handle and protect such data can result in non-compliance and legal consequences (Mugala, 2023)). This underscores the fact that some contents of dark data have legal implications which if properly identified and managed, can assist KETRACO to be compliant thus evading data-related tussles. This implies that paying attention to dark data reduces chances of data breaches thus minimizing the legal risks associated with exposure to confidential information. For example, the company could be risking contravening the *Data Protection Act of 2019* particularly on its social media platforms. This can be exemplified by the Daystar University case where a woman sued the University in October 2023 for over three million shillings because her photo was used on the University's Instagram page without her consent on 7th August 2020 (Nation Media Group, 2024).

4.7 Strategies for Harnessing Dark Data to Enhance Corporate Governance

The study aimed at finding strategies that can be used to harness dark data for effective corporate governance. The following were the suggestions for harnessing dark data that the researcher found out:

4.7.1 Hiring more Qualified Data Management Personnel and Retraining/Retooling the Existing Records Staff

The researcher found out that there was need for hiring of more qualified data management personnel at KETRACO as well as retraining/retooling the existing records staff on matters of data management. This was after questions were posed to respondents on their roles at KETRACO as well as the person managing data in their departments. The study found out that despite all respondents not being data management personnel, they were managing data at one point or another. Podraza (2014) and Profisee Group (2019) advise that a dark data management program should have an established data management team to develop and implement a strategic plan for the program as well as provide a governance structure. In addition, the team members should have their responsibilities clearly defined in matters legal data management to enhance accountability. This, therefore, calls for the Company to invest in more qualified data personnel and retrain/retool the existing records staff on data management with a focus to ensure appropriate management of data in the various directorates. This will ensure that data and by extension dark data is appropriately managed to foster effective corporate governance.

4.7.2 Staff Involvement in Data Management System Processes

The researcher noted the need for involvement/participation of all staff in data management system design and selection as well as policy formulation. Respondents complained of not being involved in many key processes of data management. The following response from Respondent L9 was typical of their concerns:

> *"We get to hear about some of the data policies when they are already on SharePoint when we are being told to familiarize ourselves with them. We are never involved in the formulation processes."*

Consequently, respondents from the Records Section in the directorate of Human Resource and administration requested that the Company involve them more in designing/reviewing the documents and records identification manual and the classification systems, the data management policy formation and information management systems for them to better understand and appreciate these systems more. The study substantiates Techved (2023)'s call for user involvement in information system development to be considered as vital mechanism to enhance system quality and ensure successful system implementation. Therefore, this calls for KETRACO to ensure that all the anticipated users of data management policies and systems are always involved in the whole processes of system development. If possible, from conceptualization up to actualization and subsequent upgrades. This is because they will be in a better position to understand, appreciate and use the systems.

4.7.3 Top Management Support

The respondents underscored that support across the board in terms of equipment, staffing and morale were key to the success of harnessing dark data. Respondent M5 explained as follows:

> *"Management and the rest of staff need to support us in such a journey. Their commitment will ensure that we get enough funds for the right tools and software, adequate and qualified data management staff and the appropriate data management programs in place."*

The findings corroborate Data ONE (2021) that organizations should accept the fact that the process of data management incurs costs in terms of resources and money. Therefore, with support from top management in matters dark data management, KETRACO will be in a better place to harness dark data for effective corporate governance. The finding also confirms Shauna (2023) opinion that if you don't have senior leadership sponsorship and understanding of your Data Governance program,

your program is going to be at risk, where by some other shiny object, some other project will come along and steal the thunder from data governance. Additionally, the findings upholds previous conclusions that the success of data governance initiatives heavily relies on the support and endorsement from executive leadership which entails not just approving budgets but also championing data governance throughout the organization. This is because management is best placed to communicate the strategic importance of data governance in achieving business objectives, thereby fostering a data-centric culture and breaking down silos while promoting collaboration across different departments.

4.7.4 Raising Awareness of Importance of Dark Data Management

Majority of respondents pointed out the need for continuous sensitization, trainings, awareness campaigns and reminders on data management policies and standards in the Company. Respondents M19, L3 and L9 posited that continuous training and sensitizations will keep staff and the Company at large updated on data management which will respectively boost harnessing of dark data. Many respondents cited the *Records Management Policy,* which they seemed to be more aware of. A Policy that emphasized the importance of ensuring that data is only created/received on a need basis and ensuring its completeness with accompanying metadata.

This revealed that there was laxity in data management despite there being some structures in place that could harness dark data. The findings validate Gureje (2021) that for the success of any data governance program every stakeholder in the organization should understand why there is an investment in the health of the organization's data. This means that there is need for KETRACO stakeholders to be continuously reminded of; data management frameworks that impact on dark data in

the Company; their roles and responsibilities in dark data management for effective corporate governance; and generally, the importance of data management to effective corporate governance.

4.7.5 Promoting a Data Sharing Culture

The research found out that there was need to encourage a proper data sharing culture in KETRACO. Respondents complained of staff being mean with data due to misinformed fears such as losing their relevance to work and being seen as very important at the workplace. In addition, the researcher learnt that some respondents, majority from the lower-level management were just not willing to learn from experts or show their expertise claiming that it was a strategy to avoid "being misused."

This revealed that due to the poor data sharing culture dark data was not only accumulating in KETRACO's electronic and manual storage, but also experts' heads as a result of intentional withholding data or refusal to acquire data. This confirms previse literature that alludes that dark data is data that is supposed to be collected but currently not being collected as dark data too as Tacitdatalabs (2023) describes it as data embedded in experts' heads as tacit data, which represents information that is implied and cannot be written down or explained verbally, but is hidden in the experience of specialists and needs to be deduced through deliberate codification and visualization to be leveraged. This means that there is need to develop a culture that will act as a tool for extracting and combining dark data and other expertise to optimizing the business processes. These could be through mentorship, brainstorming, teamwork, succession planning and skills gap analysis among others.

4.7.6 Adherence to Data Management Policies

Strict adherence to data management policies and standards could also help harness dark data. Many respondents cited the *Records Management Policy* which they seemed to be more aware of. This Policy emphasizes on the importance of ensuring that data is only created/received on need basis and ensuring its completeness with accompanying metadata. Respondent H3 argued that strictly abiding with policies and standards of data management will control the accumulation of dark data. Respondent M13 went on to say:

> *"We have several policies in KETRACO that can assist to harness dark data. For example, Records Management Policy. However, I know that very few people use this policy to guide them in their data creation and receipt."*

They were mostly being used as a reference whenever someone went against them and was caught. In concurrence, Respondent M2 said:

> *"Staff rarely use the policies in the Company to guide them in their daily activities. They only realize their importance when they are in trouble as a result of non-compliance with the policies."*

This means that there was a great likelihood of data management mistakes going unnoticed which impended the harnessing of dark data for effective corporate governance in the Company. The findings align with Mizrahi (2024) that data governance policies are a set of rules and guidelines that help to ensure data accuracy, security and privacy while also helping businesses adhere to relevant compliance laws and regulations. As a result, they ensure an organization is closed to liability issues and fines while enhancing the quality of data. Therefore, the finding support Computer Transition Services, Inc (2023) that data management policies ensure that organizations can effectively capture, create, access, distribute, use, store, secure, retrieve, and dispose of data. Therefore, these are factors which when guided by the data

management policies in place can not only mitigate the accumulation of dark data in KETRACO but also harness dark data for effective governance.

4.7.7 Data Management Centre

Respondents from the Human Resource and Administration Directorate as well as those from Strategy, Research and Compliance requested for a centralized server or data management center where all the data entering into and going out of the Company could pass through. They said that this could aid in minimizing accumulation of unnecessary data. These respondents also argued that a central server or data management center would enhance the sharing of data to various users while minimizing duplication. Indeed, this validates Diligent Corporations (2018) that a centralized database enhances organizations' data governance capabilities because it eliminated data silos, it is easier to maintain and keep clean, providing an organization with data that is verified and accurate. The process of cleaning data removes outdated entries and eliminates duplicate copies of files, categorizing information into usefully organized indexes, thus improving corporate governance.

4.7.8 Enhancing Data Security

Respondents also suggested that more system restrictions be put in place in terms of data creation, receipt and exchange by devices connected to the Company network. They pointed out the ease of capturing different data formats and downloading materials by Company devices as some of the reasons behind the accumulation of unnecessary dark data in the Company. Therefore, respondents argued that this could be minimized if ICT and Innovation Division could be more vigilant on how devices are used to capture data into the Company storages. This corroborate the explanation by Dimitrov (2018) who notes that the "Bring Your Own Device" (BYOD) practices

generates information governance challenges that arise when information is generated by and stored on mobile devices, social networks, file sharing services, and unmanaged SharePoint sites. The findings go on to agree with NetGovern (2022) that if strict policies and controls are put into place there will be reduced misuse of corporate storage in terms of dumping personal data, such as personal photos, IDs, legal documents, music, videos, and video games thus minimizing the accumulation of dark data.

4.8 Chapter Summary

The findings of this chapter are significant since they contextualize dark data at KETRACO. This chapter reveals that KETRACO creates/receives a lot of different types and formats of data, from different sources, both external and internal. This data is stored in different places at KETRACO, some being official and others non-official storage ranging from SharePoint, OneDrive, local computers and mobile devices, filing cabinets, safes and desks. In addition, the users of KETRACO data are also shown to be both internal and external. The forces behind the incognito accumulation of dark data in KETRACO and its risks and opportunities to corporate governance have also been discussed. Suggestions on strategies that can be used to harness the dark data in order to improve corporate governance at KETRACO are also given.

CHAPTER FIVE

SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS

5.1. Introduction

The purpose of this study was to investigate how dark data can be harnessed with a view of proposing a framework that can be used to manage it for effective corporate governance. The case study was KETRACO. This chapter therefore discusses the summary of findings, conclusions and recommendations. It also suggests areas that require further research to harness dark data.

5.2. Summary of Major Findings

5.2.1. In what Context is Dark Data Created, Captured and/or Accumulated at KETRACO?

The findings revealed that:

Dark data was generated in the various Directorates at KETRACO during normal business transactions, in varied formats with expectation to use it in the future or for compliance purposes. These normal business transactions included onboarding on new staff that generated staff personal identifiable data; appraisal of staff that generated staff needs data such as training needs; engagement with external stakeholders such as financiers and contractors that generated projects/contractors/financiers/PAPs profile data; security surveillance that captured CCTV footages, login details, audit trails and biometrics; and general customer calls that gave customer feedbacks among others. Therefore, there was need to proactively manage these data to harness corporate governance. For example, appraisal reports could be analyzed to bring to light actual training needs for various staff to ensure that they are sent to the right training courses to enhance their productivity at work.

Dark data was generated by both internal and external stakeholders of the Company through different channels ranging from electronic mails to hardcopy correspondences. These channels included social media platforms; staff personal electronic devices; incoming correspondences both electronically and hardcopy such as journals; contracts; mainstream media houses; blogs and newspapers, among others. In addition, there were no individuals particularly responsible for the creation/receipt of data. Despite the Company having ICT and Records staff who were deemed to be skilled in matters data management everyone was handling data they created or received. This revealed that KETRACO data and by extension dark data was at a risk of being handled unprofessionally. This made the Company data more susceptible to breaches of privacy that could expose sensitive data. As a result, effective corporate governance may be hampered through financial losses and reputational damage among other factors.

There was no centralized storage of data at KETRACO, therefore dark data was in different locations and devices. These ranged from official storage locations such as SharePoint and OneDrive platform, official hard-drive spaces on staff laptops; physical filing and boxing equipment in the registry; archives and various directorate secretaries' offices, to personal storage devices such as staff personal mobile phones; flash-disks; staff desks/drawers and home offices, among others. Such diversity of locations and devices led to the growth of data silos that held dark data. Some of these data silos could hoard important data needed for decision making in the Company. As a result, delaying or even hampering effective decision making, thus negatively affecting effective corporate governance.

5.2.2. Are There Policies and Standards that Exist to Support Harnessing of Dark Data for Corporate Governance at KETRACO?

The research revealed that there were policies that support harnessing dark data at KETRACO. These policies included *Records Management Policy – October 2020; E-Waste Management Policy – October 2020; Computer Usage Policy – October 2020; Acceptable use policy – October 2020 and Quality Management System Policy – 2020.* It is key to note that the *Quality Management System Policy – 2020* is based on ISO 9001:2015 Standard. However, there are other data management standards that can be used to harness dark data for effective corporate governance at KETRACO, but which have not been adopted.

Majority of staff were not conversant with the policies that supported harnessing of dark data due to inadequate sensitization on their existence, location and application. This was except for staff from the Directorate of Strategy, Research and Compliance and those from the Directorate of Human Resources and Administration who were directly responsible for the implementation of these policies. This is despite the policies being available on SharePoint, an online platform on which all staff could access all the Company policies.

In addition, staff who were aware of the existence of these policies seemed not to be aware of the potential role of the policies in harnessing dark data, beyond the fact that the policies were intended to facilitate effective data management. Respondents merely associated these policies with handling "normal", as opposed to, dark data. This underutilization of these policies in management of dark data could result to non-compliance by the Company. In addition, it also poses a risk of widespread mismanagement of data in the Company that could breach privacy regulations, thus

exposing KETRACO to fines and sanctions thus hindering effective corporate governance.

5.2.3. What are the Implications of the Risks and Opportunities Associated with Dark Data on Corporate Governance at KETRACO?

The study findings revealed that dark data posed some risks and opportunities which had impact on effective corporate governance at KETRACO. The implications of risks entailed loss of time; wastage of space; loss of reputation; loss of opportunities; health risks; and legal risks. The implications of opportunities include efficient service delivery; timely and effective decision making; customer satisfaction; cost reduction; and compliance with regulations. Therefore, properly harnessing dark data will minimize corporate exposures to the above risks that could lead to penalties, while maximizing the opportunities to ensure effective corporate governance.

5.2.4. Research Question Four: What Strategies can be Adopted to Harness Dark Data for Improved Corporate Governance at KETRACO?

From the findings, it was revealed that there are several strategies that KETRACO can put in place to support harnessing of dark data for effective corporate governance. Some of the strategies suggested by the respondents include hiring of professional staff specifically for data management; involving staff in all data management system design and implementation process; having support from top-management; sensitization on the importance of dark data management; encouraging a data-sharing culture; adherence to data management policies; centralization of data management; and ensuring data security mechanisms are enhanced.

5.3 Conclusion

These study gives insight into the role of dark data for effective corporate governance at KETRACO. It is evident that dark data is indeed important at KETRACO due to its positive and negative impacts to corporate governance.

Dark data keeps accumulating in different locations and formats at KETRACO. This includes official and non-official storage locations. This is because all staff and other associated stakeholders are constantly creating/and receiving this data by both official and personal devices. The physical dark data location at KETRACO is also revealed to be ranging from official storages to non-official storages such as staff desks/drawers to their homes. This therefore puts dark data at risk of mismanagement thus ease of exposure whereby the content can have negative repercussions for the Company. For example, breach of stakeholder privacy can result in legal suits that cause financial losses, a factor that affects corporate governance.

In addition to these, there exists several policies at KETRACO that can be used to harness dark data for effective corporate governance. However, these policies are not being fully utilized in the Company due to inadequate sensitization of staff about them. Moreover, there are data management standards that the Company could also use to harness dark data but is not using save for ISO 9001:2015 standards which the Quality Management System Policy, 2020 of KETRACO is based on.

Dark data also poses risks and opportunities that have implications for corporate governance at KETRACO. The Company spends a lot of resources (manpower, finances and time among others) to deal with the implications of the risks with minimal impact of dark data potential. This negatively hinders effective corporate governance due to implications such as financial losses and inefficiencies among others. Conversely, the Company has not explored the opportunities that harnessing of dark data presents for effective corporate governance, including embedding an accountable and transparent work culture and ethics, achieving corporate objectives as well as performance monitoring.

There is, therefore, need for strategies to be put in place for harnessing dark data for effective corporate governance. This starts with creating awareness of dark data among stakeholders to ensure that the positive implications will surely outweigh the negative ones. In addition, the company needs to retool the existing records and ICT staff with dark data management skills while hiring more qualified staff. In addition, there is need to get top-management support, purchase better data management systems, ensure stakeholder involvement in data management process, promotion of a data sharing culture and adherence to the data management policies in the Company.

5.4 Recommendations

The study findings revealed gaps in the management of dark data at KETRACO, which need to be addressed so that dark data can be harnessed for more effective corporate governance. Consequently, the study makes the following recommendations to address those gaps:

5.4.1. Improvement in Dark Data Curation

With the varied formats of dark data flowing in at KETRACO, there is need to ensure that they are all maintained in manner that they will enhance corporate governance in the Company. Therefore, it is recommended that the ICT department in conjunction with the Records Section put in place a centralized data management system. According to Radar Healthcare (2020) by centralizing data, an organization can significantly reduce data redundancy, which minimizes errors and simplifies access to information while establishing a clear chain of command. Centralizing data management will streamline data/dark data capture, appraisal, description, storage/preservation, access, use/reuse and sharing. This will in turn minimize the accumulation of dark data.

In addition, KETRACO through ICT department and the Records Section should upgrade its data security measures. These comprise enhancing control/rights in terms of data creation, receipt and exchange by devices connected to the company network as well as in physical data storage. This will ensure that only necessary data is created or received in the company while ensuring efficiency in the sharing of data within the Company. This will ensure that the chances of creating/receiving and accumulating dark data through duplication are minimized.

5.4.2. Enforcement in Implementation of Policies and Standards

Despite there being several policies that could be used to harness dark data for effective corporate governance at KETRACO, the study revealed underutilization of these policies by staff. This is despite the Human Resource and Administration Directorate and the Strategy, Research and Compliance directorate being directly responsible for the implementation of these policies. Therefore, the study recommends that the two directorates should immediately endeavor to do continuous sensitization to all Company stakeholders about these policies in terms of their content, access and use.

In addition, the study revealed that KETRACO had a *Quality Management Systems Policy, 2020* based on ISO 9001:2015 that basically aimed at ensuring effectiveness and consistency in the control of Management System documented information. In addition to this, the study recommends that the management approves the adoption of other standards such as ISO 8000-1:2022 that aims to ensure data quality in organization to enhance performance as a medium-term recommendation.

5.4.3. Sensitization of Stakeholders on Dark Data and its Roles in Corporate Governance

The study revealed that there were implications of both risks and opportunities associated with dark data on corporate governance. Therefore, with the revelation that

dark data was potentially in custody of all staff and stakeholders, the study recommends that ICT and Records Section carry out sensitizations about dark data and the importance of its management to make them appreciate their role in harnessing dark data for effective corporate governance. This is whereby stakeholders will appreciate more by maximizing the implications of the opportunities of dark data to KETRACO.

5.4.4. What Strategies can be Adopted to Harness Dark Data for Improved Corporate Governance at KETRACO?

For the successful harnessing of dark data for effective corporate governance, the study recommends that the Human Resource and Administration Directorate, together with Strategy, Research and Compliance Directorate ensure the following:

i. Having top management support for this program that will not only spearhead securing of resources needed for the execution of various strategies for harnessing dark data but also work as pacesetters of various dark data management activities.

ii. Management to spearhead a data-sharing culture. This will promote a culture of collaboration and as a result break down information silos thus minimizing the accumulation of dark data. Human Resource and Administration Directorate should endeavor to hiring professional staff specifically for data management. These will be handy in implementing the recommendations of this study while giving a hand in harnessing dark data. This is because professionals not only bring their expertise and experience to an organization but also enhance efficiency.

iii. The Company management to ensure that staff are involved in designing/reviewing/revising dark data management systems/schemes. This is because, since the staff are the ones who are mostly in charge of the data they create/receive regardless of whether it is dark or not, they are in a better position

to suggest futures for data management system and classification schemes that will enhance harnessing dark data.

iv. Putting in place a centralized data management system. According to Radar Healthcare (2020) by centralizing data, an organization can significantly reduce data redundancy, which minimizes errors and simplifies access to information while establishing a clear chain of command. This will in turn minimize the accumulation of dark data.

v. Upgrading the Company's data security measures. These comprise enhancing control/rights in terms of data creation, receipt and exchange by devices connected to the company network as well as physical data storage. These may entail revising of data policies to make them stricter on the use of corporate storage, having stronger controls in the systems for creation/receipt of data, its storage and use.

vi. Retooling the existing records and ICT personnel with skills to manage dark data. In addition, the Company should hire more professional data management staff.

5.5 Suggestions for Further Study

The high rate at which dark data is accumulating in organization is becoming an issue of great concern. This is because of the associated issues that come with it such as security and storage space among others. Further studies in the following areas will enable successful harnessing of dark data:

- Findings revealed that despite there being guidelines on using metadata in data management at KETRACO, they were not being fully implemented. Thus, the suggestion to study the role of metadata creation and management in harnessing dark data for effective governance in the Kenyan Public Sector Organizations. Metadata management is critical to building a data-driven business as it comprises best-

practice processes and technologies that give organizations insights for more effective data management while allowing users of all kinds, business and technical to search for, understand and access the data they need to do their job (Informatica, 2024).

- Adopting ISO 8000-1:2022 for effective management of dark data in the Kenyan Public Sector Organizations. Adopting and implementing this comprehensive data management standard is likely to offer a wider base of solutions to the myriad of challenges in the management of dark data.

- Finding sustainable mechanisms to manage dark data in the wake of big data in the Kenyan Public Sector. With the sporadic accumulation of data in various formats where some are complex and siloed in Companies, there is need to have sustainable mechanisms of managing this big data. Dark data forms the biggest percentage of big data in any organization, however it is usually not analyzed or processed despite containing undiscovered insights and opportunities.

REFERENCES

Abawi, K. (2014). *Data Collection Instruments (Questionnaire & Interview).* Retrieved from https://www.gfmer.ch/SRH-Course-2013/Geneva-Workshop/pdf/Data-collection-instruments-Abawi-2014.pdf on 14/05/2022.

Agwatu, C.O., Ajubare, O.O., & Adumade, C.O. (2025). The Impact of Poor Records Management Practices on Service Delivery and Accountability in Local Government Councils (LGCs) in Oyo State, Nigeria. *Journal of Public Administration and Social Welfare Research*, 10(6), 103-119.

Akman, S. (2025). *Study population: Definition, techniques & more*. Retrieved from https://forms.app/en/blog/study-population on 22/11/2025

Alvarez, E. (2014). *Sony Pictures hack: the whole story*. Retrieved from https://www.engadget.com/2014/12/10/sony-pictures-hack-the-whole-story/ on 09/05/2019.

Anderson, K. (2017). *What Is Data Governance and Why Is It Important to Your Business?* [Web log post]. Retrieved from https://blog.capterra.com/what-is-data-governance/ retrieved on 13/10/2019.

Angela, T. (2024). *Types of Case Studies Explained in Simple Terms*. Retrieved from https://domyessay.com/blog/types-of-case-studies on 18/11/2025.

Angelo, L. (2024). Unlocking Hidden Value: A Framework for Transforming Dark Data in Organizational Decision-Making. Retrieved from https://mpra.ub.uni-muenchen.de/122776/1/MPRA_paper_122776.pdf 20/12/2024.

Annalect. (2017). *Digging into Dark Data Can Reap Benefits*. Retrieved from https://www.annalect.com/digging-into-dark-data-can-reap-benefits/ on 08/05/2022.

AnswerMiner. (2020, August 31). *Dark Data 101: Everything You Need to Know*. Retrieved from https://www.answerminer.com/blog/dark-data on 12/05/2022.

Badman, A. & Kosinsik, M. (2024). *What is big data?* Retrieved from https://www.ibm.com/think/topics/big-data on 19/11/2025.

Badman, A. & Kosinsik, M. (2024). *What is data?* Retrieved from https://www.ibm.com/think/topics/data on 19/11/2025.

Bhandari, P. (2022). *Research Design | A Step-by-Step Guide with Examples*. Retrieved from https://www.scribbr.com/methodology/research-design/ on 09/05/2022.

Bhandari, P. (2023). *Triangulation in Research | Guide, Types, Examples*. Retrieved from https://www.scribbr.com/methodology/triangulation/ on 20/11/2025.

Bhosale, U. (2025). *Setting Limits and Focusing Your Study: Exploring scope and délimitation.* Retrieved from https://www.enago.com/academy/scope-and-delimitations/ on 21/11/2025.

Birt, L. Et al. (2016). *Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation?* Retrieved from https://pubmed.ncbi.nlm.nih.gov/27340178/ on 13/10/2023.

Biscom Incorporated. (2019). *Employee Departure Creates Gaping Security Hole, Says New Data*. Retrieved from https://www.biscom.com/employee-departure-creates-gaping-security-hole-says-new-data/ on 12/10/2019.

Bley, C. (2024). *Dark data: Lost gold or cost trap?* Retrieved from https://www.sergroup.com/en/knowledge-center/blog/using-dark-data.html on 09/04/2024.

BridgeHead (2020). *Interoperability and the New Urgency for Data Sharing in Healthcare*. Retrieved from https://www.bridgeheadsoftware.com/2020/09/interoperability-data-sharing-effectiveness/ on 01/10/2022.

Buckbee, M. (2022). *Analyzing Company Reputation After a Data Breach*. Retrieved from https://www.varonis.com/blog/company-reputation-after-a-data-breach on 22/11/2025.

Burrey, J. (2021). *How to Write Your Dissertation Chapter 3?* Retrieved from https://grademiners.com/blog/how-to-write-chapter-3-of-the-dissertation on 14/05/2022.

Byju's. (2022). *Sampling Methods*. Retrieved from https://byjus.com/maths/sampling-methods/ on 10/10/2022.

Canadian Audit and Accountability Foundation. (2024). *What is governance?* Retrieved from https://www.caaf-fcar.ca/en/oversight-concepts-and-context/what-is-oversight-and-how-does-it-relate-to-governance/what-is-governance on 30/03/2025.

CASP. (2025). *What Is a Pilot Study?* Retrieved from https://casp-uk.net/news/what-is-a-pilot-study on 22/11/2025.

Caulfield, J. (2022). *How to Do Thematic Analysis | Step-by-Step Guide & Examples*. Retrieved from https://www.scribbr.com/methodology/thematic-analysis/ on 13/6/2022.

CCG. (2024). *Corporate Governance in Kenya: What It Is and Why It Matters*. Retrieved from https://ccg.or.ke/corporate-governance-in-kenya/ on 21/11/2025.

Chaks, S. (2023). *Strategies for Your Organization to Manage Multiple Data Warehouses and Data Lakes*. Retrieved from https://www.linkedin.com/pulse/strategies-your-organization-manage-multiple-data-lakes-srinivasan#:~:text=Use%20data%20virtualization%3A%20Data%20virtualization,data%20from%20multiple%20data%20warehouses on 02/03/2024.

Champagne, J & Anderson, K. (2021). *28 Truly Shocking Cyber Security Stats*. Retrieved from https://blog.capterra.com/12-shocking-cyber-security-stats/ on 22/10/2021.

Chant, G.G. (2023). *Dealing with Dark Data – Shining a Light*. Retrieved from https://www.researchgate.net/publication/371146680_Dealing_with_Dark_Data_-_Shining_a_Light on 30/01/2025

Cherry, K. (2024). *What Is a Case Study?* Retrieved from https://www.verywellmind.com/how-to-write-a-psychology-case-study-2795722 on 10/10/2022.

Cilia, C. (2025). *Corporate Governance.* Retrieved from https://corriericilia.com/publications/corporate-governance on 20/11/2025.

Clarivine. (2022). *Data Standards Defined and Explained: Learn How to Create Data Standards.* Retrieved from https://www.claravine.com/resources/what-are-data-standards/ on 06/10/2022.

Commvault. (2014). 5 Ways to Illuminate Your Dark Data. Retrieved from http://nth.com/wp-content/uploads/2015/03/5_Ways_to_Illuminate_Your_Dark_Data.pdf.

Comodo Group, Inc. (2019). *What is BYOD (Bring Your Own Device)?* Retrieved from https://dm.comodo.com/what-is-byod/ on 14/05/2019.

Computer Transition Services, Inc. (2023). *The Importance of Data Management Policies and Procedures in Your Business.* Retrieved from https://www.ctsinet.com/the-importance-of-data-management-policies-and-procedures-in-your-business/ on 19/04/2024.

Crawford, C. J. (2007). *Compliance & Conviction: The Evolution of Enlightened Cooperate Governance*. Santa Clara, Calif: XCEO.

Creswell, J.W. (2015). *Research Design – Qualitative, Quantitative, and Mixed Methods Approaches.* Retrieved from https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf on 14/10/2022.

Cubeware GmbH. (2021). *Dark Data: Definition, Types, Examples, Advantages, and Disadvantages*. Retrieved from https://www.linkedin.com/pulse/dark-data-definition-types-examples-advantages-disadvantages-/ on 07/05/2022.

Dash, U.K. (2022). *How to Leverage Dark Data for Customer Support*. Retrieved from https://fluentsupport.com/dark-data-for-customer-support/ on 29/2/2024.

Data Management. (2021). *Data Standards*. Retrieved from https://www.usgs.gov/data-management/data-standards on 06/10/2022.

Data Protection Education Ltd. (2024). What is dark data? (and why does it matter?). Retrieved from https://dataprotection.education/best-practice-update/what-is-dark-data-and-why-does-it-matter on 12/12/2024.

DataONE. (2021). *Best Practices of Data Management*. Retrieved from https://dataoneorg.github.io/Education/bestpractices/ on 6/10/2021.

DataONE. (2021). *Provide Budget Information for your Data Management Plan*. Retrieved from https://www.dataone.org/best-practices/provide-budget-information-your-data-management-plan on 14/10/2021.

Davies, N. (2024). *Unveiling the Power of Dark Data in Strategic Decision-Making*. Retrieved from https://www.dataversity.net/unveiling-the-power-of-dark-data-in-strategic-decision-making/ on 29/02/2024.

Day, M. (2025). What is good corporate governance? 9 characteristics (with examples). Retrieved from https://www.diligent.com/resources/blog/what-constitutes-good-governance on 21/11/2025.

DCMI Usage Board. (2020). *DCMI Metadata Terms*. Retrieved from https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ on 02/10/2022.

Deakin University. (2023). *Qualitative study design – Historical*. Retrieved from https://deakin.libguides.com/qualitative-study-designs/historical on 10/10/2023.

DeCarlo, M. Et al. (2020). *Graduate Research Methods in Social Work*. Retrieved from https://viva.pressbooks.pub/mswresearch/chapter/23-grounded-theory/ on 29/10/2022.

DeJonckheere M. & Vaughn L.M. (2019) *Semistructured Interviewing in Primary Care Research: A Balance of Relationship and Rigour*. Retrieved from https://fmch.bmj.com/content/7/2/e000057 on 15/05/2022.

Delve. (2022). *Phenomenological Research Design*. Retrieved from https://delvetool.com/blog/phenomenology on 15/4/2022.

DFIN. (2024). *What Is Dark Data? How to Manage and Protect it*. Retrieved from https://www.dfinsolutions.com/knowledge-hub/thought-leadership/knowledge-resources/dark-data on 30/3/2024.

Digital curation Centre. (2022). *DCC Curation Lifecycle Model*. Retrieved from https://www.dcc.ac.uk/faq/dcc-curation-lifecycle-model on 07/10/2022.

Diligent Corporation. (2024). *The Benefits of Centralized Data: A Single Source of Information for Your Business Entities*. Retrieved from https://www.diligent.com/resources/blog/the-benefits-of-centralized-data-a-single-source-of-information-for-your-business-entities on 6/9/2023.

Dimitrov, W. (2018). *Types of dark data and hidden cybersecurity risks*. Retrieved from https://www.researchgate.net/publication/329119026_Types_of_dark_data_and_hidden_cybersecurity_risks on 14/05/2019.

Domo. (2021).*The Risk of Collecting and Storing Dark Data.* Retrieved from https://www.domo.com/learn/article/the-risks-of-collecting-and-storing-dark-data on 23/3/2024.

Domo. (2024). *How to Properly Manage Dark Data*. Retrieved from https://www.domo.com/learn/article/how-to-properly-manage-dark-data on 6/4/2024.

Domo. (2024). *The benefits of Leveraging Dark Data*. Retrieved from https://www.domo.com/learn/article/the-benefits-of-leveraging-dark-data on 09/04/2024.

Dudovuskiy, J. (2022). *Data Collection Methods*. Retrieved from https://research-methodology.net/research-methods/data-collection/ on 14/5/2022.

Duffy, D.W. (2025). *What is Corporate Governance?* Retrieved from https://www.thecorporategovernanceinstitute.com/insights/lexicon/what-is-corporate-governance/?srsltid=AfmBOoq9FxtfzTLLEW8nFOrjDFbndoVyWKdGXPLRXH7ieGR2TTgzJl9j on 20/2025.

Egnyte. (2021). *Data Collection: Definition and Methods.* Retrieved from https://www.egnyte.com/guides/life-sciences/data-collection#:~:text=Data%20collection%20is%20a%20systematic,hypothesis%20or%20evaluating%20an%20outcome. on 14/05/2022.

El Haj, M. (2023). *Dark Data.* Retrieved from https://www.linkedin.com/pulse/dark-data-moatasem-el-haj on 5/4/2024.

Elliott, R. (2020). *Probability and Non-Probability Samples*. Retrieved from https://www.geopoll.com/blog/probability-and-non-probability-samples/ on 11/05/2022.

Elliott, T. (2014). *Are You Making the Most of Your Dark Data?* [Web log post]. Retrieved from https://timoelliott.com/blog/2014/03/are-you-making-the-most-of-your-dark-data.html on 12/05/2019.

Emerging India Analytics. (2024). *Unveiling the Power of Dark Data: How Leveraging Unstructured Information Revolutionizes Big Data Analytics.* Retrieved from https://www.linkedin.com/pulse/unveiling-power-dark-data-how-leveraging-kpebc/ on 28/3/2024.

Erima, J. A. & Wamukoya, J. (2012). Aligning records management and risk management with business processes: a case study of Moi University in Kenya. *Journal of the South African Society of Archivists*, 45, 24-38.

Federal Enterprise Data Resources. (2022). *Data Standards*. Retrieved from https://www.accesstomemory.org/en/docs/2.3/user-manual/data-templates/dacs-template/ on 06/10/2022.

Feliu, C. (2020). *The Hidden Opportunities of Dark Data*. Retrieved from https://blog.datumize.com/the-hidden-opportunities-of-dark-data#:~:text=Why%20is%20Dark%20Data%20a,insights%20originating%20in%20the%20past on 08/05/2022.

FormAssembly Inc. (2024). *Top 3 Risks of Multi-Platform Data Collection*. Retrieved from https://www.formassembly.com/blog/risks-multi-platform-data-collection/ on 04/05/2024.

Future Hosting Technologies LLC. (2017). *Dark Data Is a Risk and an Opportunity for Small Businesses.* Retrieved from https://www.futurehosting.com/blog/dark-data-is-a-risk-and-an-opportunity-for-small-businesses/ on 13/05/2019.

Galaxy Consulting. (2014). *Dublin Core Metadata*. Retrieved from https://www.galaxyconsulting.net/blog/dublin-core-metadata on 06/10/2022.

Garcia, C. (2017). *6 Sampling Techniques: How to Choose a Representative Subset of the Population*. Retrieved from https://humansofdata.atlan.com/2017/07/6-sampling-techniques-choose-representative-subset/ on 11/05/2022.

Gartner, Inc. (2025). *Dark Data.* Retrieved from https://www.gartner.com/en/information-technology/glossary/dark-data on 21/11/2025.

George, T. (2023). *What Is a Theoretical Framework? | Guide to Organizing.* Retrieved from https://www.scribbr.com/dissertation/theoretical-framework/ on 10/3/2023.

Ghosh, P. (2024). *Dark Data Explained*. Retrieved from https://www.dataversity.net/articles/dark-data-explained/ on 30/01/2025.

Good, H. (2024). *How to present limitations in research*. Retrieved from https://dovetail.com/research/how-to-write-limitations-in-research/ on 17/11/2025

Governance Institute of Australia. (2024). *What is governance?* Retrieved from https://www.governanceinstitute.com.au/resources/what-is-governance/ on 13/4/2024.

Government of Kenya. (2018). *Privacy and Data protection policy 2018.* Retrieved from https://www.ict.go.ke/wp-content/uploads/2018/08/Kenya-Data-Protection-Policy-2018-15-8-2018.pdf on 08/05/2022.

Grand Canyon University. (2025). *5 Qualitative Research Designs and Research Methods*. Retrieved from https://www.gcu.edu/blog/doctoral-journey/5-qualitative-research-designs-and-research-methods on 19/11/2025

Grimm, D.J. (2019). Dark Data Quandary. *American University Law Review, 68*(3), 3-11. Retrieved from https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=2086&context=aulr on 27/03/2024.

Group-IB. (2025). *What is the Deep Web?* Retrieved from https://www.group-ib.com/resources/knowledge-hub/deep-web/ on 19/11/2025.

Hhstarling. (2019). *Creating a Holistic Data Governance Framework*. Retrieved from https://datagovernancesociety.wordpress.com/2012/10/09/october-30th-creating-a-holistic-data-governance-framework/ on 12/10/2019.

Ho, F. N., Ho-Dac, N., & Huang, J. S. (2023). The Effects of Privacy and Data Breaches on Consumers' Online Self-Disclosure, Protection Behavior, and Message Valence. *Sage Open*, 13(3). https://doi.org/10.1177/21582440231181395.

Holdsworth, J. (2025). *What is data management?* Retrieved from https://www.ibm.com/think/topics/data-management on 21/11/2025.

Hollingshead, T. (2024). *Poor Employee Health Means Slacking on the Job, Business Losses*. Retrieved from https://ph.byu.edu/poor-employee-health-means-slacking-on-the-job-business-losses on 08/04/2024.

Hougen, A. (2025). *What is the Deep Web and How Do You Get There in 2019?* Retrieved from https://www.cloudwards.net/the-deep-web/ on 14/05/2019.

Hu S. (2014) Pretesting. In: Michalos A.C. (eds) *Encyclopedia of Quality of Life and Well-Being Research.* Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_2256.

IBM. (2025). *Hidden risk of shadow data and shadow AI leads to higher breach costs*. Retrieved from https://www.ibm.com/think/insights/hidden-risk-shadow-data-ai-higher-costs on 29/1/2025

IBM. (2015). *IBM Connects "Internet of Things" to the Enterprise*. Retrieved from https://www-03.ibm.com/press/us/en/pressrelease/46453.wss on 10/08/2018.

IBM. (2024). *What is Dark Data.* Retrieved from https://www.ibm.com/topics/dark-data on 28/02/2024.

ICTA. (2019). *Government ICT Standards - Electronic Records Management Standard*. Retrieved from https://cms.icta.go.ke/sites/default/files/2022-05/Electronic%20Records%20Management%20Standard.pdf on 14/4/2021.

iDashboards. (2019). *Dark Data – The Blind Spots in Your Analytics*. Retrieved from https://www.idashboards.com/blog/2019/01/30/dark-data-the-blind-spots-in-your-analytics/ on 16/05/2019.

IGI Global. (2022). *What is Data Collection Instruments?* Retrieved form https://www.igi-global.com/dictionary/data-collection-instruments/79103 on 14/05/2022.

Indeed. (2023). *Document Analysis Guide: Definition and How to Perform it*. Retrieved from https://www.indeed.com/career-advice/career-development/document-analysis on 20/03/2024.

Indeed. (2023). *Triangulation In Research: 7 Common Types (With Examples).* Retrieved from https://www.indeed.com/career-advice/career-development/triangulation-in-research on 20/11/2025

Informatica. (2024). *What is Metadata Management?* Retrieved from https://www.informatica.com/resources/articles/what-is-metadata-management.html on 25/05/2024.

International Data Group, Inc. (2021). *Into the light: Data integration brings value to dark data.* Retrieved from https://www.domo.com/blog/into-the-light-data-integration-brings-value-to-dark-data/ on 09/04/2024.

InterQResearch. (2022). *What is Data Saturation in Qualitative Research?* Retrieved from https://interq-research.com/what-is-data-saturation-in-qualitative-research/ on 29/10/2022.

isms.online. (2022). *Understanding ISO 27001*. Retrieved from https://www.isms.online/iso-27001/ on 08/05/2022.

ISO. (2016). *ISO 15489-1:2016(en) Information and documentation — Records management — Part 1: Concepts and principles.* Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:15489:-1:ed-2:v1:en on 23/09/2022.

ISO. (2017). *International Standard Iso/Iec 38505-1*. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjc9O7kn8H6AhVwhf0HHbP3DjQQFnoECAM

QAQ&url=https%3A%2F%2Fwebstore.iec.ch%2Fpreview%2Finfo_isoiec385
05-1%257Bed1.0%257Den.pdf&usg=AOvVaw3T5cdne5xKhUblv6NTPvjX
on 02/10/2022.

ISO. (2022). *ISO 8000-1:2022(en) Data quality — Part 1: Overview*. Retrieved from
https://www.iso.org/obp/ui/#iso:std:iso:8000:-1:ed-1:v1:en on 21/05/2024.

ISO. (2022). *ISO/IEC 38505-1:2017: Information technology — Governance of IT —
Governance of data — Part 1: Application of ISO/IEC 38500 to the
governance of data.* Retrieved from https://www.iso.org/standard/56639.html
on 02/10/2022.

IT Governance. (2022). *ISO 27001 definition: What is ISO 27001?* Retrieved from
https://www.itgovernance.co.uk/iso27001 on 08/05/2022.

Izenda Inc. (2019). *The Risks and Hidden Benefits of Dark Data.* Retrieved from
https://www.izenda.com/risks-benefits-dark-data/ on 12/05/2019.

Jansen, D. & Warren K. (2020). *What Is Research Methodology?* Retrieved from
https://gradcoach.com/what-is-research-methodology/ on 26/10/2022.

Javanainen, M. (2015). *Dark Data: What is it and Why Does it Matter?* Retrieved
from https://dataconomy.com/2015/09/dark-data-what-is-it-and-why-does-it-
matter/ on 19/05/2019.

Joshi, N. (2019). *How businesses can utilize dark data with dark analytics*. Retrieved
from https://www.allerin.com/blog/how-businesses-can-utilize-dark-data-
with-dark-analytics on 07/05/2022.

Karrmann, K. (2022). *Big Data Analytics (& Why Dark Data Matters).* Retrieved
from https://authenticx.com/resources/big-data-analytics-why-dark-data-
matters/ on 4/04/2025.

KDnuggets. (2019). *The Rise of Dark Data and How It Can Be Harnessed*. Retrieved
from https://www.kdnuggets.com/2016/03/rise-dark-data-how-harnessed.html
on 25/08/2019.

KDnuggets. (2019). *What is the importance of Dark Data in Big Data world?*
Retrieved from https://www.kdnuggets.com/2015/11/importance-dark-data-
big-data-world.html on 14/05/2019.

Kengen. (2024). *Corporate Governance.* Retrieved from
https://www.kengen.co.ke/index.php/about-us/corporate-governance.html on
05/02/2024.

Kenya Gazette. (2020). *The National Information and Communication Technology
(ICT) Policy Guidelines 2020.* Retrieved from https://www.ca.go.ke/wp-
content/uploads/2020/10/National-ICT-Policy-Guidelines-2020.pdf on
08/05/2022.

KETRACO. (2020). *Our Organization*. Retrieved from
https://www.KETRACO.co.ke/about-us/our-organization on10/3/2020.

Kidd, M. (2023). The elephant in the data center: Shedding light on dark data.
Retrieved from https://www.datacenterdynamics.com/en/opinions/the-
elephant-in-the-data-center-shedding-light-on-dark-data/ on 22/11/2025.

Kiger, M.E. & Varpio, L. (2020). *Thematic analysis of qualitative data: AMEE Guide No. 131*. Retrieved from https://doi/full/10.1080/0142159X.2020.1755030 on 26/6/2022.

Krause, T. (2020). *Here's why unstructured data is putting your organization at risk*. Retrieved from https://blog.shi.com/solutions/heres-why-unstructured-data-is-putting-your-organization-at-risk/ on 16/05/2020.

Laserfiche. (2018). 4 *Ways to Manage Unstructured Data with ECM*. Retrieved from https://www.laserfiche.com/resources/blog/4-ways-to-manage-unstructured-data-with-ecm/ on 14/10/2018.

Linnik, I. (2023). *What Is Dark Data and How Do You Manage It?* Retrieved from https://softteco.com/blog/dark-data-and-how-to-manage-it on 31/03/2024.

Luenendonk, M. (2019). *Interview Schedule: Definition, Types, Templates and Tips*. Retrieved from https://cleverism.com/interview-schedule-definition-types-templates-tips/ on 6/10/2022.

Lumivero. (2024). *The Basics of Document Analysis.* Retrieved from https://lumivero.com/resources/blog/the-basics-of-document-analysis/#:~:text=Document%20analysis%20is%20the%20process,and%20come%20to%20a%20conclusion. on 01/04/2024.

Mackey, S. (2017). *Why Storing Dark Data and Mining Its Secrets Benefits Businesses*. Retrieved from https://www.kefron.com/blog/storing-dark-data/ on 22/06/2019.

Maher, C. Et al. (2018). Ensuring Rigor in Qualitative Data Analysis: A Design Research Approach to Coding Combining NVivo With Traditional Material Methods. *International Journal of Qualitative Methods*, 17(1). Retrieved from https://doi.org/10.1177/1609406918786362 on 5/11/2022.

Martin, A.J. (2019). *What is access control? A key component of data security*. Retrieved from https://www.csoonline.com/article/3251714/what-is-access-control-a-key-component-of-data-security.html on 22/6/2021.

Martin, B. S. (2018). *Big data and machine learning*. Chicago, IL: Norwood House Press.

Marumolwa. L. & Marnewick, C. (2025). Unveiling Dark Data in Organisations. *International Journal of Service Science, Management,* 16(1). Retrieved from https://www.sciencedirect.com/org/science/article/pii/S1947959X25000075 on 21/11/2025.

Mathews, K. (2018). *5 Ways Dark Data Is Changing Data Analytics*. Retrieved from https://www.smartdatacollective.com/5-ways-dark-data-changing-data-analytics/ on 16/05/2019.

McCombes, S. (2022). *Sampling Methods | Types and Techniques Explained*. Retrieved from https://www.scribbr.com/methodology/sampling-methods/ on 11/05/2022.

McLeod, S. (2024). *What Is Triangulation In Qualitative Research?*. Retrieved from https://www.simplypsychology.org/what-is-triangulation-in-qualitative-research.html on 20/11/2025.

Memorix Archives. (2022). *ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*. Retrieved from https://archives.memorix.nl/standards/ISAAR(CPF) on 06/10/2022.

Mittal, N. (2018). Dark analytics: Shedding light on a new business asset. *Analytics Magazine*. Retrieved from http://analytics-magazine.org/dark-analytics-shedding-light-new-business-asset-2/ on 21/08/2018.

Mizrahi, E. (2024). *What is a data governance policy?* Retrieved from https://www.secoda.co/blog/what-is-a-data-governance-policy on 19/04/2024.

Monnappa, A. (2025). *What is BYOD (Bring Your Own Device) and Why Is It Important?* Retrieved from https://www.simplilearn.com/what-is-byod-and-why-it-is-important-article on 21/11/20 25.

Moorthy, G. (2018). Dark data: The two sides of the same coin. *Analytics Magazine*. Retrieved from http://analytics-magazine.org/dark-data-two-sides-coin/ on 04/06/2019.

Morris, E. (2021). *Top 5 Data Governance Framework Tools to Look Out for*. Retrieved from https://towardsdatascience.com/top-5-data-governance-framework-tools-to-look-out-for-8d753ab314de on 11/10/2021.

Mugala, C. (2023). *Dark Data: Implications and Best Practices.* Retrieved from https://chibili.medium.com/dark-data-implications-and-best-practices-f0bf5fa4d309 on 09/04/2024.

Munge, E. M. (2019). *State of Information Management in Selected Government Ministries in the Context of Kenya Vision 2030*. [Doctoral dissertation, Moi University]. Moi University Open Access Repository: Doctor of Philosophy Theses (mu.ac.ke).

Mwere, D. (2025, October 2029). Ketraco scandal: Auditor-general flags massive irregularities in wayleave compensation. *Daily Nation.* Retrieved from https://nation.africa/kenya/news/ketraco-scandal-auditor-general-flags-massive-irregularities-in-wayleave-compensation-5249738 on 17/11/2025.

Mwita, K. (2022). *Factors influencing data saturation in qualitative studies Factors influencing data saturation in qualitative studies*. Retrieved from https://www.researchgate.net/publication/361102862_Factors_influencing_data_saturation_in_qualitative_studies on 14/12/2022.

Mwiti, N.K. et al. (2016). Dark data: Business Analytical tools and Facilities for illuminating dark data. *Scientific Research Journal, IV*(IV), 1-2. Retrieved from http://www.scirj.org/papers-0416/scirj-P0416325.pdf.

Nation Media Group. (2024). *Woman sues Daystar University for Sh3m over Instagram photo*. Retrieved from https://nation.africa/kenya/counties/kilifi/woman-sues-daystar-university-for-sh3m-over-her-photo-4394728 on 18/05/2024.

National Environmental Management Authority. (2021). *E-waste guidelines.* Retrieved form https://www.nema.go.ke/index.php?option=com_content&view=article&id=3 7&Itemid=180 on 12/04/2021.

NetGovern. (2022). *Data Auditing with NetGovern*. Retrieved from https://docs.netmail.com/6_1/data_audit_guide_manager/help/netmail/get_start/data_audit_netgovern.htm on 12/08/2022.

Ngoepe, M. & Ngulube, P. (2013). An exploration of the role of records management in corporate governance in South Africa. *SA Journal of Information Management*. 15(2), 8. Retrieved from http://dx.doi.org/10.4102/sajim.v15i2.575 on 04/06/2019.

Nickerson, C. (2024). *Interpretivism Paradigm & Research Philosophy*. Retrieved from https://www.simplypsychology.org/interpretivism-paradigm.html on 24/11/2025.

Ntech Lab. (2020). *Dark Data: How to Use 100% Of Information*. Retrieved from https://ntechlab.com/blog/2020/11/03/dark-data-how-to-use-100-of-information/ on 08/07/2022.

NUPPER. (2020). *Why Selecting the Right Respondents is Significant*. Retrieved from https://www.linkedin.com/pulse/why-selecting-right-respondents-significant-nuuper- on 16/3/2024.

Nzomo, B. (2025, July 29). KETRACO Audit Uncovers Suspected Corruption, Double Payments as Insolvency Threat Looms. *The Kenyan Wall street*. Retrieved from https://kenyanwallstreet.com/ketraco-audit-uncovers-suspected-corruption-double-payment on 17/11/2025.

Office of the Data Protection Commissioner. (2021). *Guidance Notes on Data Protection Impact Assessment.* Retrieved from https://www.odpc.go.ke/wp-content/uploads/2022/01/ODPC-guidance-note-on-Data-Protection-Impact-assessment.pdf on 08/05/2022

Openshaw, H. (2025). *Dark data can save lives, strengthen economies, and transform governance. But first, it must be thoughtfully harnessed*. Retrieved from https://dial.global/thoughtfully-harnessing-dark-data/ on 17/11/2025

Opsitnick, T.M., Anguiland, J.M. & Tucke, T.B. (2017). Using Computer Forensics to Investigate Employee Data Theft. *Law Journal Newsletter.* Retrieved from http://www.lawjournalnewsletters.com/sites/lawjournalnewsletters/2017/04/01/using-computer-forensics-to-investigate-employee-data-theft/?slreturn=20190922131258 on 22/10/2019.

Pal, K. (2021). *How Dark Data Can Impact the Big Data World.* Retrieved from https://www.techopedia.com/2/31521/trends/big-data/how-dark-data-can-impact-the-big-data-world on 4/11/2021.

Panorama. (2018). *Dark Data what is all the fuss about?* Retrieved from https://www.panorama.com/wp-content/uploads/2016/06/Whats-Dark-Data-1.pdf on 10/08/2018.

Perez, M.E. (2022). *Presentation, analysis and interpretation of data*. Retrieved from https://www.slideshare.net/31mikaella/presentation-analysis-and-interpretation-of-data on 16/05/2022.

Pervin, N. & Mokhtar, M. (2022). The Interpretivist Research Paradigm: A Subjective Notion of a Social Context. *International Journal of Academic Research in*

Progressive Education and Development. 11. Retrieved from https://www.researchgate.net/publication/360180378_The_Interpretivist_Research_Paradigm_A_Subjective_Notion_of_a_Social_Context on 26/11/2025.

Podraza, A. (2014). *The 5 Steps to Starting a Records Management Program.* Retrieved from http://documentmedia.com/article-1469-The-5-Steps-to-Starting-a-Records-Management-Program.html.

Polymer. (2024). *Dark Data.* Retrieved from https://www.polymersearch.com/glossary/dark-data on 5/4/2024.

Precision Consulting, LLC. (2024). *Literature Review*. Retrieved from https://precisionconsultingcompany.com/literature-review.shtml on 14/4/2024.

Profisee Group. (2019). *Master Data Management - What, Why, How & Who*. Retrieved from https://profisee.com/master-data-management-what-why-how-who/ on 8/6/2019.

PubGenius Inc. (2023). *What is Research Approach?* Retrieved from https://typeset.io/questions/what-is-research-approach-2a7q513bzz on 29/12/2023.

Purushothaman, A. (2019). *Dark Data in Big Data & Analytics*. Retrieved from https://www.expresscomputer.in/news/dark-data-in-big-data-analytics/24568/ on 15/05/2019.

Quantilope. (2022). *Data Saturation in Qualitative Research*. Retrieved from https://www.quantilope.com/blog/glossary-data-saturation-in-qualitative-research on 29/10/2022.

Queens University of Charlotte. (2022). *A Guide to Qualitative Rigor in Research*. Retrieved from https://online.queens.edu/resources/article/guide-to-qualitative-rigor-in-research/ on 22/10/2022.

QuestionPro. (2022). *Qualitative Research: Definition, Types, Methods and Examples*. Retrieved from https://www.questionpro.com/blog/qualitative-research-methods/ on 09/05/2022.

Rahman, M. (2021). *Advantages and disadvantages of qualitative research*. Retrieved from https://howandwhat.net/advantages-disadvantages-qualitative-research/ on 09/05/2022.

Rao, V. (2024). *Extracting dark data*. Retrieved from https://developer.ibm.com/articles/ba-data-becomes-knowledge-3/ on 10/03/2024.

Rautenbach, E. & Jansen, D. (2024). *What are Research Limitations?* Retrieved from https://gradcoach.com/research-limitations/ on 21/11/2025.

Rautenbach, E. & Jansen, D. (2025). *What is Research Philosophy?* Retrieved from https://gradcoach.com/research-philosophy/ on 24/11/2025.

Raza, M. (2025). *Dark Data: An Introduction*. Retrieved from https://www.splunk.com/en_us/blog/learn/dark-data.html on 4/4/2025

Reiley, J. (2022). *A Guide to Everything You Need to Know About Dark Data*. Retrieved from https://www.influencive.com/a-guide-to-everything-you-need-to-know-about-dark-data/ on 16/05/2024.

Rhee, H. L. (2022). A New Lifecycle Model Enabling Optimal Digital Curation. *Journal of Librarianship and Information Science*, *56*(1), 241-266. Retrieved form https://doi.org/10.1177/09610006221125956 on 17/11/2025.

Rizzolo, M.K. (2019). *12 Reasons Why Data Is Important*. Retrieved from https://www.c-q-l.org/wp-content/uploads/2019/12/12-Reasons-Why-Data-Is-Important.pdf on 12/6/2023.

Ryte. (2021). *Dublin Core.* Retrieved from https://en.ryte.com/wiki/Dublin_Core on 02/10/2022.

SaaSholic. (2021). *Dark Data: Why You Need to Know About it*. Retrieved from https://blog.saasholic.com/p/dark-data-why-you-need-to-know-about-it on 25/3/2024.

Sarkar, S. (2017). *Understanding Dark Analytics: Potential, Advantages and Risks*. Retrieved from https://www.analyticsinsight.net/understanding-dark-analytics-potential-advantages-risks/ on 13/05/2019.

Satori. (2022). *Guide: Data Management What are Data Standards and Why Do You Need Them?* Retrieved from https://satoricyber.com/data-management/what-are-data-standards-and-why-do-you-need-them/ on 06/10/2022.

Saunders, B. et al. (2018). *Saturation in qualitative research: exploring its conceptualization and operationalization.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5993836/#:~:text=Saturation%20means%20that%20no%20additional,that%20a%20category%20is%20saturated on 14/12/2022.

Seers. (2022). *Why are Data Standards Important and What are their Benefits?* Retrieved from https://seersco.com/blogs/why-are-data-standards-important-and-what-are-their-benefits/ on 06/10/2022.

Shacklet, M. (2025). *What To Know About Dark Data in Higher Education*. Retrieved from https://edtechmagazine.com/higher/article/2025/07/what-know-about-dark-data-higher-education on 20/11/2025.

Sharma, K. (2025). *Understanding the Importance of Data Governance*. Retrieved from https://www.definian.com/articles/understanding-the-importance-of-data-governance on 22/11/2025.

Sharma, R. (2020). *Top 6 Major Challenges of Big Data & Simple Solutions to Solve Them*. Retrieved from https://www.upgrad.com/blog/major-challenges-of-big-data/ on 20/5/2021.

Shauna, F. (2023). *Gaining Leadership Support for Data Governance*. Retrieved from https://www.dataversity.net/gaining-leadership-support-for-data-governance/ on 21/05/2024.

Shekhar, P. (2021). *Dark Data Analytics*. Retrieved from https://articles.xebia.com/dark-data-

analytics#:~:text=Dark%20Analytics%20refers%20to%20the,they%20go%20 about%20their%20data on 07/05/2022.

SHI International Corp. (2020). *Dark Data Assessment, Bring your Deepest, Darkest Data to Light!* Retrieved from https://www.shi.com/CustomerServices/SHIMarketing.aspx?ContentID=9335 8 on 26/06/2020.

Shinydocs. (2024*). 7 Essential Data Governance Best Practices for Effective Management*. Retrieved from https://shinydocs.com/blog/7-essential-data-governance-best-practices-for-effective-management/ on 21/05/2024.

Simplilearn. (2022). *What Is Data Collection: Methods, Types, Tools, and Techniques*. Retrieved from https://www.simplilearn.com/what-is-data-collection-article on 14/05/2022.

Smartsheet Inc. (2019). *How to Create a Data Governance Plan to Gain Control of Your Data Assets*. Retrieved from https://www.smartsheet.com/data-governance on 11/10/2019.

Spinner, K. (2018). *Is Dark Data Putting Your Organization at Risk?* Retrieved from https://www.datacenterknowledge.com/industry-perspectives/dark-data-putting-your-organization-risk on 11/05/2019.

StatCan. (2021). *3.2.3 Non-probability sampling*. Retrieved from https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm on 9/04/2025.

Stitchdata. (2022). *The causes and costs of data silos.* Retrieved from https://www.stitchdata.com/resources/data-silos/ on 24/09/2022.

Streamline. (2025, September 26). Ketraco yet to pay landowners Sh4 billion in wayleave compensation. Retrieved from https://streamlinefeed.co.ke/news/ketraco-yet-to-pay-landowners-sh4-billion-in-wayleave-compensation on 17/11/2025.

Suciu, D. (2023). Dark Data or Data in Darkness. Retrieved from https://insights.pecb.com/dark-data-data-darkness/ on 12/12/2024.

Sydell, E. (2024). *What Is Generative AI's Role In Unlocking Dark Data?* Retrieved from https://www.forbes.com/councils/forbesbusinesscouncil/2024/12/18/what-is-generative-ais-role-in-unlocking-dark-data/ on 4/04/2025.

Tanlak, E.F. & Aryal, S. (2025). *Interpretivist Constructivism: A Valuable Approach for Qualitative Nursing Research*. Retrieved from https://www.scirp.org/journal/paperinformation?paperid=140486 on 25/11/2025.

Tay, I. (2014). *To what extent should data saturation be used as a quality criterion in qualitative research?* Retrieved from https://www.linkedin.com/pulse/20140824092647-82509310-to-what-extent-should-data-saturation-be-used-as-a-quality-criterion-in-qualitative-research on 14/12/2022.

Techopedia. (2018). *Dark Data*. Retrieved from
https://www.techopedia.com/definition/29373/dark-data on 07/08/2018

TechTarget. (2025). *Data Governance Policy*. Retrieved from
https://searchcompliance.techtarget.com/definition/Data-governance-policy on
11/01/2025.

Techved. 2023). *The Power of User Involvement: How It Can Drive Design Success*.
Retrieved from https://techved.medium.com/the-power-of-user-involvement-
how-it-can-drive-design-success-1544918225b9 on 31/07/2021.

Teherani, A., Martimianakis, T., Stenfors-Hayes, T., Wadhwa, A., & Varpio, L.
(2015). Choosing a Qualitative Research Approach. *Journal of graduate
medical education*, 7(4), 669–670. Retrieved form
https://doi.org/10.4300/JGME-D-15-00414.1 on 12/05/2022.

Tie, Y.C., Birks, M. & Francis, K. (2019). *Grounded theory research: A design
framework for novice researchers*. Retrieved from
https://pubmed.ncbi.nlm.nih.gov/30637106/ on 10/10/2022.

Tittle, E. (2014). *The Dangers of Dark Data and How to Minimize Your Exposure*.
Retrieved from https://www.cio.com/article/2686755/the-dangers-of-dark-
data-and-how-to-minimize-your-exposure.html on 21/05/2019.

Trajanov, D. Et al. (2018). Dark *Data in Internet of Things (IoT): Challenges and
Opportunities*. Retrieved from
https://www.researchgate.net/publication/323337110_Dark_Data_in_Internet_
of_Things_IoT_Challenges_and_Opportunities on 10/08/2018.

Trivedi, B. & Gokulnath, K. (2017). Research on Dark Data Analysis to Reduce Data
Complexity in Big Data. *International Education and Research Journal*. *3*(5).
Retrieved from http://ierj.in/journal/index.php/ierj/article/view/955 on
26/08/2018.

UConn. (2022). *What are data standards and why use them?* Retrieved from
https://guides.lib.uconn.edu/c.php?g=832372&p=8226287 on 06/10/2022.

Ulz, J. (2023). *What is a Research Paradigm? Types of Research Paradigms with
Examples.* Retrieved from https://researcher.life/blog/article/what-is-a-
research-paradigm-types-examples/ on 24/11/2025.

United States Data Science Institute. (2025). *Dark Data: Hitting Untapped
Opportunities.* Retrieved from https://www.usdsi.org/data-science-
insights/huge-potential-of-dark-data-revealing-untapped-opportunities on
20/11/2025

University of California. (2019). *Data Curation Centre (DCC) Model*. Retrieved from
https://guides.ucf.edu/metadata/dataCuration on 10/10/2019.

University of North Carolina. (2025). *Literature Reviews*. Retrieved from
https://writingcenter.unc.edu/tips-and-tools/literature-reviews/ on 18/11/2025.

University of southern Carolina. (2025). *Organizing Your Social Sciences Research
Paper.* Retrieved from http://libguides.usc.edu/writingguide/literaturereview
on 18/11/2025.

University of Virginia. (2020). *Ethnographic Research*. Retrieved from
https://research.virginia.edu/irb-sbs/ethnographic-research on 15/7/2020.

University of Witwatersrand. (2022). *Research Support: Research Methodology*.
Retrieved from https://libguides.wits.ac.za/c.php?g=693518&p=4914913 on
30/10/2022.

UpKeep. (2022). *How do we encourage useful data input from our employees?*
Retrieved from https://www.upkeep.com/learning/how-to-encourage-useful-
data-input on 01/10/2022.

Van Rees, I. (2020). *The hidden opportunities of Dark Data*. Retrieved from
https://blog.datumize.com/the-hidden-opportunities-of-dark-data on
12/05/2022.

Vaughan, T. (2021). *10 Advantages and Disadvantages of Qualitative Research*.
Retrieved from https://www.poppulo.com/blog/10-advantages-and-
disadvantages-of-qualitative-research on 09/05/2022.

ViewPointe. (2013). *Dark Data, Big Data, Your Data: Creating an Action Plan for
information governance*. Retrieved from
https://www.viewpointe.com/uploadedFiles/Viewpointe/PDFs/viewpointe-
dark-data-white-paper.pdf on 10/08/2018.

Visix, Inc. (2024). *The Value of Communications Analytics*. Retrieved from
https://www.visix.com/blog/the-value-of-communications-analytics/ on
14/4/2024.

Voicedocs. (2022). *Historical Research*. Retrieved from
https://voicedocs.com/en/blog/historical-research on 15/11/2020.

Waida, M. (2021). *What Is an Information Silo?* Retrieved from
https://www.wrike.com/blog/what-is-information-silo/ on 14/4/2024.

Warren, K. (2020). *Qualitative Data Analysis Methods 101: The "Big 6" Methods +
Examples*. Retrieved from https://gradcoach.com/qualitative-data-analysis-
methods/ on 02/03/2022.

Whatagraph Team. (2021). *Data Interpretation: Definition, Method, Benefits &
Examples*. Retrieved from https://whatagraph.com/blog/articles/data-
interpretation#:~:text=What%20is%20Data%20Interpretation%3F,collected%
20'cleaned'%20raw%20data. on 29/10/2022. Willie, M.M. (2023).
Distinguishing Between Population and Study Population: A Mini Review.
Surgery Research Journal.3. 1-2.

Willie, M.M. (2023). Distinguishing Between Population and Study Population: A
Mini Review. *Surgery Research Journal.3. 1-2.*

Zeenea Software. (2019). *What is the difference between Metadata and Data?* [Web
log post]. Retrieved from https://zeenea.com/the-difference-between-
metadata-data/ on 12/06/2020.

Zhao Y. (2024). *Development of big data assisted effective enterprise resource
planning framework for smart human resource management. PloS one*.
Retrieved form https://pmc.ncbi.nlm.nih.gov/articles/PMC11104621/ on
30/3/2025.

APPENDICES

APPENDIX I: INTRODUCTORY LETTER TO RESPONDENTS

Gloria Regina Mukhongo

P.O Box 444

Funyula.

October, 2019

Dear Respondents,

My name is Gloria Regina Mukhongo, a student of Masters of Science in Archives and Records Management. I am carrying out research for Master's degree at Moi University School of Information Sciences. The title of my study is 'Harnessing Dark Data for Effective Corporate Governance at The Kenya Electricity Transmission Company Limited'. I hereby request you to assist me in answering these question so as to achieve my research objective. The acquired information will be purely for academic purposes and be treated with maximum confidentiality.

Your assistance in facilitating this study is highly appreciated.

Thank you.

Yours faithfully

Gloria Regina Mukhongo

Master Student

APPENDIX II: INTERVIEW SCHEDULE FOR EXECUTIVE/TOP LEVEL MANAGEMENT

1. What are your duties and responsibilities at KETRACO?

2. How long have you been working with KETRACO?

3. a) What type of data do you deal with in KETRACO?

    b) What are the sources of data in KETRACO?

    c) How is data stored at KETRACO?

    d) Who is in-charge of data management at KETRACO?

    e) Who are the possible users of these data and for what purpose?

4. a) What is your understanding of "dark data"?

    b) What is the impetus of dark data accumulation at KETRACO?

    c) What are the implications of the risks and opportunities of dark data to corporate governance at KETRACO?

5. a) Which policies or standards are in place at KETRACO for harnessing dark data?

    b) How can they be used or how are they being used to harness dark data for effective corporate governance.

6. How can dark data be harnessed to enhance corporate governance at KETRACO?

THANK YOU VERY MUCH FOR YOUR SUPPORT AND CONTRIBUTION!

APPENDIX III: INTERVIEW SCHEDULE FOR MIDDLE LEVEL MANAGEMENT

1. Which directorate do you work in?

2. What are your duties and responsibilities at KETRACO?

3. a) What type of data do you deal with in your directorate?

    b) Where do you get the data that you use in your day-to-day work?

5. a) Who is in-charge of these data in your directorate?

    b) How is data stored in your department/Directorate?

    c) Who are the possible users of these data?

6. a) What is your understanding of dark data?

    b) Does KETRACO have policies and standards to guide the management of data/dark data?

    c) If yes, please name them and how they impact data management and dark data.

    d) What are the implications of risks of dark data to corporate governance in your directorate and KETRACO at large?

    e) What are the implications of opportunities of dark data to corporate governance in your directorate and KETRACO at large?

7. What strategies do you think can be put in place to mitigate the accumulation of dark data in your directorate?

THANK YOU VERY MUCH FOR YOUR SUPPORT AND CONTRIBUTION!

APPENDIX IV: INTERVIEW SCHEDULE FOR OFFICERS/LOW-LEVEL MANAGEMENT

1. Which directorate do you work in?

2. What are your duties in KETRACO?

3. How long have you been working with KETRACO?

4. a) What type of data do you deal with in your daily transactions?

    b) Where do you get the data that to use in your daily activities?

    c) Who manages data you use or create in your daily transactions?

    d) Where and how is data stored in your Department/Directorate?

    e) Who are the users of the data that you create or receive in your daily transactions?

5. a) Are there policies and standards in KETRACO that guide how you create/receive and manage data?

    b) If yes, please name them and their role in data creation/receipt and management in your directorate.

6. a) Have you ever heard of dark data?

    b) What do you think are the implications of the risks of dark data to your daily operations?

    c) What do you think are the implications of the opportunities of dark data to your daily operations?

    d) How do the policies and standards on the management of data in the Company affect the management of dark data?

7. What do you think can be done to mitigate the accumulation of dark data in your directorate/department?

THANK YOU VERY MUCH FOR YOUR SUPPORT AND CONTRIBUTION!

APPENDIX V: DOCUMENTS ANALYSED

1. Records Management Policy – October 2020
2. E-Waste Management Policy – October 2020
3. Computer Usage Policy – October 2020
4. Acceptable use policy – October 2020
5. Quality Management System Policy - 2020

APPENDIX VI: DOCUMENTS ANALYSIS GUIDE

1. Purpose of the document to KETRACO.

2. Impact of the document on data/records management and by extension dark data management.

3. Repercussions of the impact of the document to governance at KETRACO.

APPENDIX VII: NACOSTI RESEARCH PERMIT

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION

Ref No: **648943**

Date of Issue: **20/July/2020**

RESEARCH LICENSE

This is to Certify that Miss.. Regina Gloria Mukhongo of Moi University, has been licensed to conduct research in Nairobi on the topic: **HARNESSING DARK DATA FOR EFFECTIVE CORPORATE GOVERNANCE AT THE KENYA ELECTRICITY TRANSMISSION COMPANY LIMITED** for the period ending : **20/July/2021.**

License No: **NACOSTI/P/20/5798**

648943

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, 2013

The Grant of Research Licenses is Guided by the Science, Technology and Innovation (Research Licensing) Regulations, 2014

CONDITIONS

1. The License is valid for the proposed research, location and specified period
2. The License any rights thereunder are non-transferable
3. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research
4. Excavation, filming and collection of specimens are subject to further necessary clearence from relevant Government Agencies
5. The License does not give authority to tranfer research materials
6. NACOSTI may monitor and evaluate the licensed research project
7. The Licensee shall submit one hard copy and upload a soft copy of their final report (thesis) within one of completion of the research
8. NACOSTI reserves the right to modify the conditions of the License including cancellation without prior notice

National Commission for Science, Technology and Innovation
off Waiyaki Way, Upper Kabete,
P. O. Box 30623, 00100 Nairobi, KENYA
Land line: 020 4007000, 020 2241349, 020 3310571, 020 8001077
Mobile: 0713 788 787 / 0735 404 245
E-mail: dg@nacosti.go.ke / registry@nacosti.go.ke
Website: www.nacosti.go.ke

APPENDIX VIII: KETRACO PERMIT

DIRECTORATE OF HUMAN RESOURCE & ADMINISTRATION

KETRACO
Kenya Electricity Transmission Company Limited
"Building a World Class National Grid"

Kawi Complex, Block B,
Popo Lane, off Red Cross, South C,
P. O. Box 34942 - 00100, NAIROBI
Phone: 020 4956000, 0719018000, 0732128000
Web: www.ketraco.co.ke I email: info@ketraco.co.ke

Our Ref: KET/5/1C/42/JEA/gm

24 August 2020

Gloria Mukhongo
P.O Box 38 – 50406
Funyula

Dear Madam,

RE: AUTHORIZATION TO CONDUCT RESEARCH

We acknowledge receipt of your letter dated 15th July 2020 to conduct research in KETRACO titled: **HARNESSING DARK DATA FOR EFFECTIVE CORPORATE GOVERNANCE AT THE KENYA ELECTRICITY TRANSMISSION COMPANY LIMITED** as part of your Master of Science in Records and Archives Management at Moi University.

Permission is hereby granted for you to undertake the research in fulfilment of your program. Kindly observe confidentiality of the Company's data and share with Learning and Development a copy of the results of the research.

I wish you the very best as you undertake the research.

Yours faithfully,
For: KENYA ELECTRICITY TRANSMISSION COMPANY (KETRACO)

DR. JOSHUA AMWAYI
MANAGER, LEARNING & DEVELOPMENT

APPENDIX IX: INTRODUCTION LETTER FROM MOI

MOI UNIVERSITY
SCHOOL OF INFORMATION SCIENCES
DEPARTMENT OF LIBRARY, RECORDS MANAGEMENT AND INFORMATION STUDIES

Tel: (053) 43222
Fax No. (053) 43292
Telex NO. 35047 MOIVASITY
E-Mail: hodlis@mu.ac.ke OR deanis@mu.ac.ke

P. O. Box 3905
Eldoret
Kenya.

Our Ref: IS/MSC/RAM/07/16

25th February, 2020

The Executive Secretary,
Kenya National Council of Science and Technology,
P. O. Box 30623,
NAIROBI.

Dear Sir,

RE: MUKHONGO REGINA GLORIA – IS/MSC/RAM/07/16

The above named is a Master of Science student in the Department of Library, Records Management and Information Studies, School of Information Sciences, Moi University.

Ms. Mukhongo is intending to carry out research work entitled *"Examining the Rise of Dark Data and how it can be harnessed for Good Governance at the Kenya Electricity Transmission Company"*.

We are kindly requesting you to issue her with a research permit to enable her proceed with her research.

Thank you.

Yours sincerely,

DR. ELSEBAH MASEH
SENIOR LECTURER AND HEAD,
DEPARTMENT OF LIBRARY, RECORDS MANAGEMENT & INF. STUDIES

EM/ww

9001:2015 Certified Institution)