

**BIG DATA MANAGEMENT AT THE KENYA ELECTRICITY TRANSMISSION
COMPANY (KETRACO)**

**BY
MICHAEL W. WECHULI**

**A THESIS SUBMITTED IN PARTIAL FULFILMENT 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
STUDEIES, SCHOOL OF INFORMATION SCIENCES,**

**MOI UNIVERSITY,
ELDORET, KENYA.**

2025

DECLARATION

DECLARATION BY THE CANDIDATE:

I Michael W. Wechuli Registration Number IS/MSC/RAM/02/16, declare that this thesis is my own intellectual work. The document has not been submitted by any other student and has never been submitted to any other examination body. No part of this work should be reproduced without my consent, and I accept the responsibility of the information contained within this document.

Signature:

Date:

Michael W. Wechuli

IS/MSC/RAM/02/16

DECLARATION BY THE SUPERVISORS:

This Research thesis has been submitted for defence with our approval as the University Supervisors.

Signature:

Date:

Dr. Elsebah Maseh

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

Signature:

Date:

Dr. Emily Ng'eno

Department of Library, Records Management and Information Studies,
School of Information Sciences,
Moi University

ABSTRACT

Big data has revolutionized how organizations around the world perceive and harness information, offering unprecedented opportunities to derive valuable insights and drive innovation from varied and vast datasets. Big Data can be defined as the information or data asset characterized by a high volume, variety, velocity and veracity that require specific analytical methods and technologies for its transformation into meaningful value. However, the ability to manage and originate meaningful insights from these datasets to make data-driven decisions aimed at improving organizational performance and growth remains a key challenge for many organizations. The purpose of this study was to analyse Big Data Management at the Kenya Electricity Transmission Company (KETRACO) with a view of proposing strategies for improving its management and use. The objectives of the study were to; Analyse the current state of Big Data Management at KETRACO; Examine the existing infrastructure for Big Data Management at KETRACO; establish the influence of collaborative partnerships on Big Data Management and propose a strategy for the improvement of Big Data Management at KETRACO. The study was underpinned by the Data Centre Curation Lifecycle Model, and the Community Capability Model. The study employed concurrent mixed methods approach employing survey design within a singular exploratory case study. The target population of 563 was stratified into different strata and different sampling strategies were used on each stratum. Census sampling technique was applied to select 14 key informants comprising of 2 personnel from the Executive office of the CEO, 7 senior management team, and 5 system analysts with whom interviews were conducted to collect qualitative data. The MaCorr Research solution calculator was applied to the remaining target population of 549 to attain a sample size of 63 respondents from 8 departments. Thereafter, simple random sampling was used to select respondents from each of the 8 departments from which quantitative data was collected using questionnaires. The quantitative data was analysed using descriptive statistics and presented in tables, graphs and pie charts while the qualitative data was analysed thematically and presented in narrative descriptions. The findings revealed that there were no standardized and unified procedures for big data management at KETRACO and that there were no specific policies and guidelines with tight regulations that streamlined big data management and usage at every level of data life cycle; further, the findings showed that KETRACO had inadequate and equally under-utilized the existing ICT infrastructure to effectively support data management functions. A great concern noted from the survey was the existence of knowledge and skill gaps such as the metadata skills gaps as exhibited by 51 (87.93%) of the surveyed respondents. The survey also established that Ketraco had not established unified mechanisms for managing its data from creation or receipt throughout its life cycle as this was unanimously confirmed by all the 58 (100%) respondents. Further, the survey established that privacy, confidentiality, and complexity in handling cumulative data assets were the major hindrances to Big data management at Ketraco as this was recorded by 54 (93.10%) and 47 (81.03%) respondents respectively. Finally, the study revealed inadequate collaborative mechanisms for effective data management with other partners. In conclusion, KETRACO had not adopted effective policies and strategies for Big data management, had under-utilized its ICT infrastructure, and not put enough mechanisms for collaborative partnerships, hence hindering its capability of deriving maximum value from its data assets. The study recommended among other solutions, the establishment and operationalization of policies and procedural frameworks for managing data throughout its lifecycle, capacity building and data literacy training programme to bridge the skill gap; development of a sustainable ICT infrastructure and security approach in support of big data; and building collaborative partnership to develop big data capabilities.

DEDICATION

I would like to dedicate this proposal to all the communication, IT and Records Management at the Kenya Electricity Transmission Company Limited.

ACKNOWLEDGEMENT

First and foremost, I would like to express the deepest appreciation to my lecturers and my supervisors Dr. Elsebah Maseh and Dr. Emily Ng'eno for their valuable guidance, monitoring and constant encouragement throughout the course of this assignment. They have shown the attitude and the substance of a genius. Without their supervision and constant help, writing this proposal would not have been possible.

I also take this opportunity to express a deep sense of gratitude to my parents Mr. and Mrs. Wechuli for their moral support which helped me in completing this task.

I am obliged to my fellow classmates for their cooperation during the period of my assignment. Lastly, I thank the almighty God, brothers, sisters and friends for their constant encouragement without which this project would not have been possible.

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT.....	iii
DEDICATION.....	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS AND ACRONYMS	xiii
CHAPTER ONE	1
INTRODUCTION AND BACKGROUND TO THE STUDY	1
1.1 Introduction	1
1.1.1 Contextualization of Big Data Management	6
1.2 Kenya Electricity Transmission Company (KETRACO)	9
1.2.1 Core Objectives	11
1.2.2 Big Data Utilization at KETRACO	11
1.3 Problem Statement.....	13
1.4 Aim of the Study	18
1.5 Objectives of the Study	19
1.7 Assumption of the Study	20
1.8 Significance of the Study.....	21
1.9 Scope and Limitations of the Study.....	22
1.9.1 Scope of the Study.....	22
1.9.2 Limitations of the Study	23
1.10 Definition of terms and concepts.....	23
1.11 Structure of the Thesis.....	24
CHAPTER TWO	27
LITERATURE REVIEW	27
2.1 Introduction	27
2.2 Theoretical Framework.....	27
2.2.1 Data Curation Centre (DCC) Lifecycle Model.....	29
2.2.1.1 Relevancy of the DCC model to the study.....	33

2.2.1.2 Gaps in the use of the DCC model.....	34
2.2.2 Community Capability Model	34
2.2.2.1 Relevancy of the CCM model to the study.....	37
2.2.2.2 Gaps in the use of the CCM model.....	38
2.3 Overview of Literature	38
2.3.1 Big Data Management	38
2.3.2 Big Data in Organizations	41
2.3.2.1 Contributing Factors to the Growth of Big Data	43
2.3.2.2 Importance of Big Data Management on organization's Operations at Ketraco.....	45
2.3.3 Stages Involved in Big Data Management	48
2.3.4 Legal, policies and regulatory framework in Big Data Management.....	53
2.3.5 Knowledge, Skill, and Training for Big Data Management.....	56
2.3.6 Technical Infrastructure for Big Data Management.....	57
2.3.7 Collaborative Partnerships in Big Data Management	60
2.3.8 Support Services for Big Data Management	62
2.3.9 Challenges in Big Data Management	64
2.3.10 Strategies for Effective Big Data Management.....	68
2.3.11 Summary of the Linkages between Theories to Research Questions.....	70
2.4 Chapter Summary	72
CHAPTER THREE	74
RESEARCH METHODOLOGY	74
3.1 Introduction	74
3.2 Research Approach.....	74
3.2.1 Qualitative Research.....	75
3.2.2 Quantitative Research.....	75
3.2.3 Mixed Methods Research	76
3.3 Research Design	77
3.3.1 Survey Design.....	77
3.3.2 Case Study Design.....	78
3.4 Population of the Study	79
3.5 Sampling Methods.....	80

3.5.1 Sample Size	82
3. 6 Data Collection Methods	86
3.6.1 Interview Method	86
3.6.2 Self-administered Questionnaires	87
3.6.3 Document Analysis.....	88
3.7 Validity and Reliability of Data Collection Instruments	89
3.7.1 Validity of Data Collection Instruments.....	89
3.7.2 Reliability of Data Collection Instruments	90
3.7.2.1 Conducting the Pilot Study	91
3.8 Data Analysis and Presentation	92
3.9 Ethical Consideration	93
3.10 Chapter Summary	94
CHAPTER FOUR.....	96
DATA PRESENTATION, ANALYSIS, AND INTERPRETATION	96
4.1 Introduction	96
4.2. Response Rate	97
4.3 Demographic Information of the Respondents.....	99
4.3.1 Gender Information	100
4.3.2 Level of Education.....	101
4.3.3 Years of Experience.....	103
4.3.4 Departments of Respondents	104
4.3.5 Position of the Respondents in Level of Management	106
4.4 Current State of Big Data Management at Ketraco	108
4.4.1 Data Capture and Format.....	108
4.4.1.1 Data Capture	108
4.4.1.2 Data Format	112
4.4.2 Data Appraisal	114
4.4.3 Data Description	117
4.4.4 Storage of Big Data	120
4.4.5 Data Backup Format.....	124
4.4.6 Access to Big Data at Ketraco	127
4.4.7 Usage of Big Data.....	130

4.4.7.1 Use and Re-use of Big Data.....	132
4.4.8 Sharing of Big Data at Ketraco.....	134
4.5 Policies Guiding the Use of Big Data Ketraco	138
4.5.1 Strategies that can help in the Implementation of Big Data Management at Ketraco.....	142
4.5.2 Quality Assurance and Quality Control of Big Data	144
4.6 Skills, Knowledge, and Training for Big Data Management at Ketraco.....	147
4.6.1 Existing Skills for Big Data Management at Ketraco	147
4.6.2 Knowledge and Training for Big Data Management at Ketraco	149
4.6.3 Proficiency of Employees to Various Big Data Skills at Ketraco	153
4.7 ICT Infrastructure for Big Data Management	156
4.7.1 Security of Big Data Management	159
4.8 Collaborative Partnerships in Big Data Management at Ketraco.....	162
4.8.1 Challenges Facing Big Data Management at Ketraco.....	166
4.9 Chapter Summary	169
CHAPTER FIVE	171
SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS	171
5.1 Introduction	171
5.2 Summary of the Findings	171
5.2.1. Demographic Information of Respondents.....	171
5.2.2 Capture, Appraisal, Description, Preservation, Accessibility, Re-use, and Sharing of Big Data.	172
5.2.3 Policies and Regulation for Big Data Management at Ketraco.....	177
5.2.4 Skills, Knowledge, and Training for Big Data Management	178
5.2.5 Level of ICT Infrastructure Preparedness for Big Data Management.....	179
5.2.5.1 Security Measures for Big Data.....	180
5.2.6 Collaborative Partnerships for Big Data Management at Ketraco	181
5.2.7 Challenges Facing Big Data Management	181
5.3 Conclusion	182
5.4 Recommendations	184
5.4.1 Data curation for Big Data Management.....	184
5.4.2 Policies and Regulation for Big Data Management at Ketraco.....	185

5.4.3 Skills, Knowledge, and Training for Big Data Management	186
5.4.4 ICT Infrastructure for Big Data Management	187
5.4.5 Security Measures for Big Data Management.....	188
5.5 Contribution and Originality of the Study	190
5.6 Suggestion for Further Research	191
REFERENCES.....	193
APPENDICES.....	205
Appendix I: Survey Questionnaire	205
Appendix II: Interview Schedule.....	221
Appendix III: Moi University Introduction Letter	226
Appendix IV: Moi University Letter to Nacosti.....	227
APPENDIX V: Ketraco Approval Letter	228
Appendix VI: Research License	229

LIST OF TABLES

Table 2.2: Summary of the Linkages among research Gap, Research Question and Theories Underpinning the Study	70
Table 3 1: Target Population.....	80
Table 3.2: Sample Size for Respondents	85
Table 4. 1: Response Rate.....	98
Table 4. 2: Distribution of Respondents by Gender	100
Table 4. 3: Distribution of respondents by Level of Education	101
Table 4. 4: Distribution of respondents by Years of Experience	103
Table 4. 5: Data Capture	109
Table 4. 6: Big Data Usage at Ketraco	130
Table 4. 7: Uses and Re-use of Big Data at Ketraco	132
Table 4. 8: Employee skills for Data Management at Ketraco	148
Table 4. 9: Knowledge and Training for Big Data Management at Ketraco	150
Table 4. 10: Expertise and Proficiency in Various Big Data Skills	154
Table 4. 11: ICT infrastructure at Ketraco for Big Data	156
Table 4. 12: Efficacy of ICT Infrastructure at Ketraco	157
Table 4. 13: Challenges facing Big Data Management at Ketraco	167

LIST OF FIGURES

Figure 2.1: The Data Centre Curation (DCC) Lifecycle Model	30
Figure 2.2: Community Capability Model.....	35
Figure 3. 1: Macorr Research Calculator	84
Figure 4. 1: Surveyed Respondents per department	104
Figure 4. 2: Interviewed Respondents per department	105
Figure 4. 3: Distribution of respondents by management Level	106
Figure 4. 4: Data formats Exiting at Ketraco	112
Figure 4. 5: Data Appraisal	115
Figure 4. 6: Data Description	119
Figure 4. 7: Data Storage	121
Figure 4. 8: Data Back-up and Formats	124
Figure 4. 9: Data preservation and Disposal responsibility	126
Figure 4. 10: Access to Data at Ketraco	128
Figure 4. 11: Big data sharing channels at Ketraco	135
Figure 4. 12: Existence of Big Data Management Policy at Ketraco	139
Figure 4. 13: Big Data management procedures at Ketraco	140
Figure 4. 14: Strategies for implementation of Big Data Management at Ketraco	142
Figure 4. 15: Presence of Quality Assurance and Quality Controls at various stages of Big Data Management	144
Figure 4. 16: Security of Big Data Management	159
Figure 4. 17: Forms of Partnerships at Ketraco	163

LIST OF ABBREVIATIONS AND ACRONYMS

ERP	Enterprise Resource Planning
BDM	Big Data Management
ICT	Information and Communication Technology
IIMSPFM	Integrated Information Management Systems for Procurement and Financial Management
KENGEN	Kenya Electricity Generating Company Limited.
KETRACO	Kenya Electricity Transmission Company Limited.
KPLC	Kenya Power and Lighting Company Limited.
MD	Managing Director

CHAPTER ONE

INTRODUCTION AND BACKGROUND TO THE STUDY

1.1 Introduction

The Modern society is an interconnected world that generates a large volume of data every day, including the logging files of social network users, search engines, e-mail clients, and data generated by machines (Jeske and Calvard, 2020). Nowadays, data sets are identified in different environments and in large volumes never seen before hence referred as 'Big Data', and this data doubles within a short period due to different types of databases (Lynch, 2008).

A review of the existing literature on the definition of big data indicates lack of an appropriate interpretation of "Big Data" terminology. According to Katal, Wazid, and Goudar (2013), big data relates with the underlying technologies and new types of workloads required to address business-related challenges that could not be supported initially due to technology limitations and prohibitive costs. Therefore, Big Data is may not only refer to large data sets, but about analytical workloads associated to some combination of data velocity, data volume, and data variety that may involve complex data types and complex analytics. Labrinidis and Jagadish (2012), Kshetri (2014), and Lynch (2008) argue that, when analyzed alongside traditional organizational data, big data helps an organization to better - understand its business, and change it, and enable it to have new revenue generating sources, greater innovation and a more competitive position.

Lynch (2008) and Katal et al (2013) indicate that Big Data can be well defined using the three V's which include Volume meaning large data volumes; Variety meaning data comes in a variety of forms, such as traditional databases, complex records, images and documents; and Velocity meaning the data content is constantly changing through the introduction of legacy collections or previously archived data, through the absorption of complementary data collections, and from streamed data received from different sources. Sagioglu and Sinanc (2013) add that Veracity is another element that should be added to this definition meaning the purity of data is key for its value since the large volumes of data harnessed can result to statistical errors and misinterpretation of the collected information. From the above definitions, Big Data is a concept that refers to large data volumes both structured and unstructured that flow in an organization on a day-to-day basis. However, it is not the quantity of data that is important to an organization, but the management and analysis of this data for insights that can lead to better decisions and improved organizational performance (Erevelles, Fukawa and Swayne, 2016). Therefore, Big Data can be defined as the information or data asset characterized by a High Volume, Variety, Velocity, and Veracity to require specific Analytical Methods and Technologies for its transformation into meaningful Value.

According to Labrinidis and Jagadish (2012), Zhang et al. (2018), and Lynch (2008), Big Data Management (BDM) goes further to include different new business possibilities and analytical technologies other than just handling large data volumes. Labrinidis and Jagadish (2012) further claim that the emerging analytical processing technologies have made it possible what was traditionally not possible. For instance, there are new systems present that manage different data sets from sensor data, social media and web data,

enhanced operational business intelligence and analytical capabilities that improves business agility by facilitating automated real-time actions and decisions, effective hardware platforms and cloud computing such as on-demand software-as-a-service. Therefore, supporting big data requires application and utilization of these technologies to enable new solutions that in return generate great value to the business.

Further, Borgatti, Everett, and Johnson (2018) enlist different big data generating sources. They include social media and networks, where the authors state that there are more than 250 million, 700 million, and 156 million users of Twitter, Facebook and public blogs users respectively; and each Tweet, blog post, Facebook update or comment generates several new data points that could be unstructured, semi-structured or structured. Another important source of data is Mobile Devices where Borgatti, Everett, and Johnson (2018) noted that over 5 billion mobile devices are being used globally, and every text message or call is logged as data. The communication devices mostly the smartphones and tablets also make it easier to access and use social networks and other data-generating platforms. The other data source listed by Labrinidis and Jagadish (2012); Kshetri, 2014), and Lynch (2008) is e-commerce where there are billions of online transactions such as stock trades, purchases and other automated transactions happening daily over the internet. Each of these transactions generates different data points collected by credit agencies, banks and retailers.

Data explosion is currently being witnessed in companies globally, as scholars such as Dey et al. (2018) and Lynch (2008) have asserted that businesses and industry analysts are focusing on Big Data as the new strategy to access insights, opportunities, profits and

solutions. Both the government and private sector Agencies are now applying big data initiatives to manage the growing sets of data that exist in different formats (Gantz and Reinsel, 2012). Effective strategies for BDM enable business enterprises to identify valuable information from large structured and semi-structured data sets derived from various sources, including social media and system logs.

From internal sources to external sources such as social networking sites, Big Data continues to play a key role in enhancing businesses processes, innovation and growth. With the continued realization of big data value, businesses continue to generate quality information and utilize it for data analysis, decision making and other business functions (Zhang et al., 2018). The big data obtained from different resources come with numerous challenges such as unstructured and voluminous data, that hinders companies' capability in handling or storing them. Thus, effective data management techniques, practices, infrastructure and technology can help address these challenges.

Many public and private organizations have recognized the value of big data management in driving business operations and growth and in this regard, government depends on big data for accountability, governance and administration for its citizens (Yin and Kaynak, 2015). Consequently, as a way of enhancing good governance and service delivery, many governments have embraced ICT as a way of improving the management of their big data that in turn improve service delivery and operational efficiency. According to Kwanya (2014), government agencies continually generated a lot of data that form as evidence of transactions.

In Kenya, the National government has embraced the use of ICT through the e-citizen platform to enhance citizens' access to government services as well as attain data on the use of the services. Similarly, many government agencies such as the Directorate of Criminal Investigation, the Judiciary, Ketraco, and Kenya Power, among others are using social media sites and other online platforms to engage more citizens and customers on matters to do with service delivery. Kimani (2015) stated that many organizations in Kenya such as commercial Banks, Insurance firms, Telecommunication companies, Government Agencies, and Institutions of higher learning have already shown goodwill in leveraging on Big Data. For instance, institutions such as the University of Nairobi have established online repositories to support E-research and data sharing; companies such as Safaricom have embraced cloud computing for storing their big data; Government bodies are currently using the Enterprise Resource Planning (ERP) platforms and Integrated Information Management Systems for Procurement and Financial Management (IIMSPFM) (Ndemo, 2017).

However, the unavailability of proper guidelines and mechanisms makes it difficult for organizations to capture, analyze, and manage data that comes in large volumes hence hindering their ability to derive many benefits from their data assets (Zhang et al., 2018). With the initiatives established by organizations in response to big data, organizations can derive lots of benefits if there was a Data agency established by law to protect and manage all data generated by all institutions in Kenya. In addition, a state Data management agency can enhance institutional collaboration and partnerships through data sharing, thus improving learning, skills development, and performance of organizations.

1.1.1 Contextualization of Big Data Management

Data management is a function that has been embraced by many countries across the world. From the UK perspective, initiatives to manage data have been in existence for many decades. For instance, Lewis (2010) argues that, in the UK, it is over 40 years since the Data Archive was established at the University of Essex, and many institutions of higher learning have long embraced data management, especially in managing their research collections. However, the growing interest in the long-term curation and storage of data has been driven by the growth of digital research data, and many institutions are looking into ways of preserving their data for long-term use and reuse.

In the United Kingdom, the government sponsored the e-Science Core Programme in 2001 administered by the Engineering and Physical Sciences Research Council on behalf of the UK's Research Councils. The purpose of this program was to establish the middleware (software applications and databases), infrastructure, and documentation that enhance e-research (Lewis, 2010). In this regard, the next generation would be characterized by large data volumes, and it would be vital for multiple institutions to collaborate in exchanging these data sets. Although data management was not directly addressed in the first phase of the Core Programme, the implications of large-scale e-research projects for data management were soon apparent (Lewis, 2010). In the meantime, working groups tasked to explore different aspects such as Information and Data creation, curation and preservation, computer and data storage, search, navigation, and retrieval were set up. It was these working groups that led to the development of a national infrastructure for managing digital research data (Van den Eynden et al., 2011).

On the same note on big data, The United States National Science Foundation has established a clear vision for investment in cyberinfrastructure by moving swiftly to invest funds in digital research data curation by establishing a DataNet program in 2007. The program is a research initiative of the National Science Foundation Office of Cyberinfrastructure (Lewis, 2010; NSF, 2007). This led the National Science and Technology Council's Committee on Science to establish an interagency working group to oversee a roadmap for coordinated national activities that outlines how various data management practices such as data creation, documentation, dissemination, and sharing are managed.

Australia has also embraced an e-research roadmap by establishing the Australian National Data Service (Lewis, 2010). They have invested in major components of the e-infrastructure such as ICT systems and frameworks for managing big data. They also published a roadmap for future development and management of big data (Australian National Data Services Technical Working Group, 2007).

From the African perspective, countries have started to embrace Big Data management. In South Africa, the Data Intensive Research Initiative of South Africa (DIRISA) enables researchers to plan, store, manage, and share their data using various technologies and cloud-based tools (Vahed, 2014). In addition, the Agency organizes local and international workshops around Big Data Management circles such as the procedures, policies, and services for managing research data. For instance, DIRISA organized a workshop in July 2019 and the key areas that were addressed include the initiatives and infrastructures for managing Big Data at the local, national, and international levels.

DIRISA supports and promotes data-intensive research in various disciplines hence contributing to Big Data management in South Africa.

From a Kenyan perspective, many organizations, both public and private sectors have developed mechanisms that enhance the management of their content, especially data storage on online platforms (Ndemo, 2017). A good example is the National Council for Law Reporting, a state corporation in the Judiciary. The Agency has the sole responsibility of reporting jurisprudences in the country, and this is done through the publication of Kenya Law reports (Kenya Law 2019). On their website, one can search a case, access the laws of Kenya, view the tribunals, search the archive, and search the available case list, access the Kenya Gazette, Bills, and other publications. Making this information accessible to the general public has been vital, especially in helping the public understand the past, present, and future legal environment.

Kenya Electricity Transmission Company Limited (KETRACO) has not been left behind in terms of data management. The organization has demonstrated efforts to utilize Big Data in facilitating the organizational process, decision-making, and problem-solving across the departments. The data has been critical in facilitating the organization's transactions and hence forming the foundation of KETRACO's functions, programs, services, policies, and in making sound decisions. There is a lot of collaborative partnership within and outside the organization in exchanging big data within which includes data from the ministries, other government agencies, and the public. For instance, the company requires data relating to energy consumption and distribution in various parts of the country; in addition, its mandate relies on geographical and cadastral

data, survey, and mapping. Therefore, appropriate Big Data Management practices should be taken seriously by KETRACO to fulfil its mandate.

Big Data Management and Records Management

The emergence of Big data has transformed how companies create, maintain, access, use and preserve information. Records management focuses on the systematic control of records throughout their life cycle. Big data management and records management are related disciplines but emerge from different needs and operate with different scopes. While records management concerns the systematic control of records from creation to disposition, big data management focuses on handling large and diverse datasets (Mariani and Baggio, 2022). Both disciplines focus on improving the utilization the value of information while minimizing operational and legal risks. This implies that certain datasets qualify as records if they document transactions, decisions or business activities. Records management applies a systematic and structured approach to manage data sets. For instance, it provides guidance on managing information assets at distinct stages, metadata standards, audit trails, legal admissibility as well as retention and disposal frameworks. It is therefore paramount for Big data management to incorporate governance principles from records management setup. Equally, Asnawi et al (2022) are of the view that records management should redefine what records is composed of, and modernize to handle complex, large scale and technologies of big data. Applying this complimentary approach will form a comprehensive information governance structure.

1.2 Kenya Electricity Transmission Company (KETRACO)

The Government of Kenya in Sessional Paper No. 4 of 2004 on Energy indicated the necessity to un-bundle Kenya power's function of transmitting and distributing electricity.

On further analysis, it was concluded that a separate parastatal, fully owned by the Kenyan government and financed by the government be established to construct new additional high voltage transmission lines (KETRACO, 2020). It would have been difficult to unbundle Kenya Power since it is a publicly quoted company. As a result, in November 2008, the government registered a separate company, The Kenya Electricity Transmission Company Limited (KETRACO). KETRACO's mandate is to design and construct high voltage transmission lines and associated sub stations for purposes of transmitting power at a high voltage from generating plants such as KENGEN and other independent power producers to the National Grid. The Kenya Power and Lighting Company (KPLC) then taps this power and distributes it to consumers in various parts of the country (KETRACO, 2020).

KETRACO is 100% owned by the government, and being a state agency, is regulated under the State Corporations Act, Cap 446 (Ketraco Annual Report 2017-2018). KETRACO was established to develop the new electricity infrastructure in line with Kenya Vision 2030. Its core business is to plan, design, build, and maintain electricity transmission lines and associated substations. The creation of KETRACO was also compelled by the need to transform power transmission system into an open-access platform that allows large power consumers to purchase electricity from generating sources. With the ongoing interconnections of electricity grid across Kenya, and with Tanzania, Ethiopia, Uganda, and other Southern Africa Power Pool, the Kenyan government views open access to electricity as having the capability to enhance market and supply options for both power generation and large consumers.

KETRACO'S vision is to be a world-class electricity transmission company and the leading interconnector in Africa and its mission is to provide reliable, efficient, and effective electricity transmission and promote power trade for sustainable socio-economic development. Consequently, the core functions are to plan, design and construct, own, operate, and maintain high-voltage electricity transmission lines, substations, and fibre optic cables across the country.

1.2.1 Core Objectives

1. Improve the reliability and quality of electricity supply in different parts of the country.
2. Evacuate power from generating plants.
3. Transmit power to regions that are not yet supplied by the national grid.
4. Establish a link with the neighbouring countries to promote power exchange and promote electricity trade.
5. Reduce power transmission losses encountered every year.
6. Reduce electricity costs to consumers by absorbing the capital cost of transmission infrastructure.

1.2.2 Big Data Utilization at KETRACO

In carrying out its mandate, KETRACO generates different types of data, especially from the technical department that performs the core mandate of the company. The Data generated in this department include:

- i. The Environmental Assessment reports that detail the impact of the company's projects on the environment;

- ii. Land valuation reports that detail the market value of land that Ketraco intends to use to construct power lines;
- iii. Wayleave records that document the land wayleave corridors the company intends to use to construct power lines;
- iv. Land Negotiation agreements that the company generates while negotiating with landowners on the use of their land;
- v. Mapping and survey records are also generated during cadastral surveys of different geographical locations, including physical locations where the company intends to construct its power lines;
- vi. Technical drawings that illustrate the plans and designs of power lines are also generated and;
- vii. Project Affected Persons (PAPs) data is also critical information captured by Project clerks in the technical division (It refers to information of persons who have been affected by Ketraco's projects especially when the company uses their land to construct power lines and substations).

Other support departments that also generate data consist of the following:

- Legal data that includes easements and contracts. Easements are agreement records between the PAP and the organization that give Ketraco the right to cross or otherwise use their land for power transmissions. Contracts on the other hand are Executed agreements between Ketraco and external clients such as suppliers and contractors to undertake a particular task for the company.
- The Human Resources department is another support department that generates data such as recruitment data, Staff data, and student Internship data, among others.

- The finance department generates records such as payment vouchers that show payments made to contractors, project-affected persons (PAPS), and other parties for providing a service to the company. The procurement department also generates data such as tendering process data, and tender evaluation reports, among others.
- The corporate department has embraced social media sites and the corporate website to engage and collect feedback from members of the public.
- Externally, the organization receives data such as approvals, guidelines, and other correspondences from the Ministries of Treasury, Public Service, and Energy. The organization also exchanges data with partners such as the Kenya Power, Energy and Petroleum Regulatory Authority, and KENGEN.

Consequently, every department has shared folders on the intranet where employees access and share data relating to their day-to-day activities. The workflows have also been enhanced especially with the introduction of the Enterprise Resource Planning (ERP) system that allows data to easily flow from one user or department to the other. In conclusion, KETRACO interacts with data assets on a day-to-day basis. The business operations generate and rely on both internal and external data, and this data helps the company achieve its mandate. Therefore, embracing proper big data management strategies will ensure data availability and utilization in its decisions, and this ensures the company achieves its objectives.

1.3 Problem Statement

In the recent past, the source of power has drifted from finance, capital, land to information, knowledge and data. Currently, organizations have embraced the use of data

to understand their customers, competitors, governments, and industries. Organizations are also using data to collect background information, such as data about technology and society, and to monitor the changing business environment and keep updated with emerging trends (Sagiroglu and Sinanc, 2013). Traditionally, organizations have benefited from their internal data assets in understanding their business processes and in improving decision-making at a tactical and on an operational level (Kwanya, 2014). In recent years, companies have embraced analysis of data from external sources originating from multiple dispersed information sources such as the government, competitors, and other industries to align with legal requirements especially for multinationals operating in different jurisdictions.

Big Data and data mining research are now of tremendous importance to organizations because better results are obtained from larger volumes of data (Kwanya, 2014). The interest in Big Data and its value to organizations has increased because of the significant amount of data generated every day for decision-making, enhancing customer experience, fostering innovation, and improving operational efficiency (Dey et al., 2018; Kwanya, 2014; Zhang et al., 2018). However, McAfee et al. (2012) argued that along with the steady enlargement of data volume, the percentage of data that are processed and analyzed by organizations is in constant decline because many organizations lack proper strategies for the effective utilization of their data assets. In the same breadth, KETRACO has continually increased its data generation; this can be witnessed by new systems being launched to support data collection and workflows (Ketraco Annual Report 2017-2018). However, the procedures of capturing, appraising, describing storing, preserving, accessing, or sharing these data that exist in systems are not communicated in KETRACO's

publications. This implies that there could be no clear mechanisms, policies, regulations, and guidelines that outline how KETRACO collects, describes, preserves, accesses, uses, or shares its data resources.

In the recent past, Ketraco has been in the limelight with both the media and the Office of the Auditor General for failure to justify spending, and this has had a negative reputation with members of the public (Office of the Auditor General, Kenya, 2022; The National Treasury, 2023). The Audit report for the year that ended June 2021, for instance, pointed out some financial spending that could not be documented or justified (Auditor-General Report, 2021; Office of the Auditor General, Kenya, 2022; The National Treasury, 2023). The report noted the inability of the company to provide supporting data and records evidencing how the company utilized funds. The irregularities pointed out by the auditor included missing invoices, expenditures not aligned with the approved budgets, and payments made to projects with no clear activities, unaccounted grants, unaccountable subsistence allowance, inaccurate data on amounts owed to Kengen for transferred assets, inability to confirm the accuracy in receivable balances amounting to 9,726,373,000, inability to prove that VAT returns amounting to 1,163,953,516 were filed and acknowledged by the Kenya revenue authority (Auditor General Report, 2021, P. 2). All these issues can be attributed to poor data management at KETRACO, leading to the indictment by the Auditor General.

Similarly, financiers of KETRACO's projects through monitoring and evaluation have also highlighted the need for proper management of project data to enhance effectiveness and efficiency in project delivery (World Bank Report, 2024). For instance, in the project

assessment report by the World Bank on the Eastern electricity highway project, the financier of a high-voltage power transmission project from Ethiopia to Kenya pointed out that KETRACO, as the implementing agency, faced numerous challenges that included delays in assembling data and key documents such as data related to donor funded projects (World Bank Report, 2024). The delays affected timely compensations to the affected local people, leading to project resistance by the local communities hence project completion delays. The report further outlined that KETRACO performed poorly in providing the requested information to the World Bank project team to facilitate the approvals for addendums relating to contract variations in terms of scope, time extensions, and cost adjustments (World Bank Report, 2024, P. 39). In turn, this resulted in cost overruns and delays in implementing the construction of the high voltage transmission line from Ethiopia to Kenya.

The World Bank also rated KETRACO's finance management as moderately unsatisfactory (World Bank Report, 2024). This was after the inability of the internal auditors from the World Bank to satisfy themselves with the completeness and accuracy of the accounting records they reviewed while auditing the Ethiopia Kenya High Voltage Transmission line project which they were the main Financiers (World Bank Report, 2024, P. 43-44). The issues highlighted in audit pointed towards variable and irreconcilable data on physical/fixed asset register with the ledger balance in the system, inaccurate reporting on data on goods on transit, and payable balances. Furthermore, difference between confirmation of balances obtained from financiers and ledger balances in KETRACO's books, and the inability for KETRACO to provide evidence that the fully complied with procurement regulations while undertaking the project were part of the

inadequacy to manage big data. The issues raised by the World Bank are critical to KETRACO's image since it heavily relies on financial organizations for funding to undertake its massive power transmission projects. Therefore, a study on Big Data management at KETRACO is vital to help KETRACO unlock the value of effective Big Data management for organizational operations and efficiency.

Further, Mwere (2020) in the Daily Nation newspaper dated Monday, January 02, 2023, exposed the issues that contributed to delays in KETRACO's completion of its high-voltage electricity transmission line projects. One of the key issues highlighted in the newspaper was the inability of the company to collect key data and documents from members of the community affected by the project. This data is necessary for internal workflow of processing financial compensation to community members affected by the projects (Mwere, 2020). Some of these documents include survey data and land ownership documents verified by the local administration that must be presented to Ketraco legal department for verification before forwarding to the finance department for compensation processing. Thus, the inability of the company to collect and repackage the listed data hindered the compensation process, resulting in resistance from local communities as well as legal suits that ultimately delayed the company in completing its projects.

From the foregoing, KETRACO is faced with a myriad of challenges related to data management. The scenario conforms to observations made by Kshetri (2014) and Sagioglu and Sinanc (2013) that, in many organizations, the gap between data quantity and percentage of their usage is becoming larger and larger; and that, data is now available

in different platforms and formats. For KETRACO, its data could exist on different platforms such as personal computers, shared online folders, website, databases, and computer systems. However, this technical infrastructure might not be coordinated in a way that harmonizes how data is created, described, preserved, accessed, used, and shared within the organization.

Equally, the data protection act, 2019 of Kenya majorly governs the acquisition, processing, storage, use and sharing of personal data to ensure security and privacy of individuals' personal information (Wanekeya, 2023). However, the act does not extend to provide a framework on how Companies can manage their big data throughout its life cycle or how companies can collaborate and share their data assets for mutual benefits such as innovation and product development, enhanced decision making, risk reduction and operational efficiency. It is therefore imperative to explore the technical infrastructure for managing Big Data in terms of ICT, staff skills in using data, security, and coordinated computer systems to harness all these forms of data and come up with regulatory frameworks that guide the management of big Data throughout its life cycle, as well as big data collaborations amongst organizations. It was against this backdrop that the study was undertaken to examine the strategies and interventions for improving the management, sharing, and reuse of Big Data in KETRACO.

1.4 Aim of the Study

The purpose of this study was to examine Big Data Management at the Kenya Electricity Transmission Company (KETRACO), Kenya, with the view of proposing interventions

to improve the management, sharing, and reuse of big data to facilitate proper planning, improved organizational processes and better decision making.

1.5 Objectives of the Study

The objectives of the study were to;

1. Analyze the current state of Big Data Management at KETRACO;
2. Examine the existing infrastructure for Big Data Management at KETRACO;
3. Establish the influence of collaborative partnerships on Big Data Management
4. Propose a strategy for the improvement of Big Data Management.

1.6 Research questions

The specific research questions that will guide this study were;

1. How does KETRACO capture, appraise, describe, preserve, and make accessible for Re-use and sharing of its Big Data?
2. What policies and regulations guide the management of data received or created by KETRACO?
3. What skills, knowledge, and training are needed to improve Big Data Management at KETRACO?
4. What is the level of ICT Infrastructure preparedness in supporting Big Data Management at KETRACO?
5. How do collaborative partnerships influence Big Data management at KETRACO?

1.7 Assumption of the Study

1. The study assumed that KETRACO collects, processes, preserves, makes use and disposes of data in its custody. As such, the absence of proper Big Data management mechanisms hinders the effective utilization of large volumes of KETRACO's data.
2. The study was also based on the assumption that KETRACO has some level of ICT infrastructure such as networks, servers and other hardware capable of supporting big data activities. It was assumed that respondents would accurately report on the functionality, adequacy and challenges in using KETRACO's infrastructure.
3. The study assumed that the respondents possessed varying but identifiable levels of knowledge, skills and training relevant to Big data management, and that participants can describe or self-assess their competencies objectively and that KETRACO had made at least minimal investment towards capacity building.
4. The study assumed that there are emerging policies and legal frameworks that govern and guide Big data management practices at KETRACO, and that the inability of KETRACO to establish proper systems, processes, updated policy and regulations on Big Data Management affects its utilization.
5. Further, the study assumed that KETRACO engages in certain forms of collaboration internally, external or cross-institutional partnerships with private companies, research institutions, or government bodies relating to data sharing or access. It was equally assumed that respondents would provide accurate information relating to the nature and effectiveness of these partnerships.

1.8 Significance of the Study

Big data management as part of information management is essential for any organization to effectively derive maximum benefits from the information it creates and receives. Therefore, the findings of this study will assist in policymaking for Big Data management at KETRACO. In addition, the study also benefits KETRACO by recommending appropriate policies and regulations that if adopted can guide Big Data Management practices. As such, various BDM aspects such as Data creation, description, preservation, and sharing will be improved hence contributing significantly to decision-making and other organizational processes. Consequently, any other institutions/organization dealing with Big Data Management will profit from policy making as well as referencing KETRACO BDM as the base line for their BDM practices.

Academicians will also gain from the findings of this study as it contributes to the existing body of knowledge or literature in the field of Big Data Management. As such, the study provides a basis for other investigations by future researchers and academicians into the area of big data management in Kenyan energy sector and other government Agencies.

The data volumes that KETRACO accumulates over the years will offer new opportunities for the Agency to use Big Data technologies to extract meaningful insights and keep track of the needs of their clients and partners. With the ability to derive meaning from its data, this study will therefore transform the work of KETRACO by helping it operate more efficiently and make more informed and transparent decisions.

1.9 Scope and Limitations of the Study

This section outlines the coverage and focus of the study as well as the limitations that the researcher could have encountered while carrying out the research.

1.9.1 Scope of the Study

The study focused on Big Data Management at KETRACO with the aim of proposing interventions that can improve the management, sharing, and reuse of big data. The study was limited to five variables namely: data curation, policies and regulations, skill, knowledge and training, ICT infrastructure, and collaborative partnerships for big data management, with specific reference to this study.

Kenya Electricity Transmission Company (KETRACO) Headquarters was taken as the case study and the focus of the study. The researcher considered KETRACO as the case study since the company had received national recognition in the utilization of ICT platforms such as SAP Ariba systems for e-procurement, among others. As a result, the incorporation of these new technologies into business processes generates a lot of data, hence ideal for this study.

The study targeted only the employees of KETRACO stationed at the headquarters. These are employees who interact with data regularly while performing their tasks. The departments or categories considered include senior management, systems analysts, finance, technical, supply chain, corporate planning and development, human resource and admin, corporate communications, legal and company secretary, ICT officers, and executive/MD's office.

1.9.2 Limitations of the Study

Given that the study was limited to KETRACO and its employees, obtaining adequate literature was a challenge, especially in cases where respondents handled confidential information hence unwilling to share information that was vital in answering the research questions. Thus, the researcher signed consent with them and affirmed that the data shared was for the purpose of this research and it would remain confidential and not be referenced in any other matter; this improved respondents' participation in the study. Secondly, employees at KETRACO were involved in diverse activities moving from one place to another. As such, the study anticipated that some respondents would not be available during the time of the study, and this could affect the way the study was carried out and the amount of data collected from the expected number of respondents. To address this challenge, the researcher came up with a sample that targeted respondents who were likely to be available during the study.

1.10 Definition of terms and concepts

This study concentrates on major key areas and issues of big data management namely;

Data: - Information in unorganized or raw form such as numbers, symbols, or alphabets, which represent or refer to an idea, condition, or object

Big Data: Big Data refers to large volume of both structured and unstructured information that it is complex to process using traditional methods (Dey et al, 2018).

Big Data Management: Refers to the administration and governance of large sets of unstructured and structured data (Katal, Wazid, and Goudar, 2013). The main objective of big data management is to ensure efficiency and a high level of data quality and its application for business intelligence.

Information: Refers to data that has been processed and placed in a certain context in such a way that it has meaning to the person who receives it.

Information and Communication Technology (ICT): Refers to technologies that facilitate access to information using telecommunications channels and platforms that include computer devices, networks, cell phones, and other communication mediums.

Records: Information created, received, and maintained as evidence and information by an entity, in pursuance of legal obligations and in the transaction of its main business (Sheng et al., 2017).

Records Management: A field of management that ensures the systematic process of the creation, use, and disposition of organizational records, including the process of managing evidence of a business transaction that is in the form of a record (Sheng et al., 2017).

An organization: An entity comprising different people aimed at performing a business-related function by providing products or services to meet the expectations of customers.

1.11 Structure of the Thesis

This research study is organized into five chapters that collectively form the research problem, research methodology, presents, analyses and discusses findings and their implications. The structure of the thesis guides the reader progressively from a broader perspective of the research to the specific contributions of the study.

Chapter One: Introduction and Background to the Study

The chapter covers the background and motivation of the study. It also defines the research problem, aim of the study, objectives as well as research questions. The chapter equally outlines the significance, scope and limitations of the study.

Chapter Two: Literature Review

The chapter reviews the existing literature relevant to the research topic. It synthesizes key concepts, theoretical frameworks, empirical studies as well as identifying limitations and gaps in the existing knowledge base. Reviews of literature and theories relevant to the study. The Literature review forms the basis upon which the research questions and hypotheses were developed.

Chapter Three: Research Methodology

The chapter outlines the research approach, research design, population of the study, sampling procedures, and data collection instruments. The chapter aims to demonstrate the reliability and validity of the research instruments.

Chapter Four: Data Presentation, Analysis and Discussion

The chapter presents and analyses the findings of the study and further interprets and discusses the said findings by exploring the implications based on the research questions. The presentation and analysis was done in a clear and systematic manner. The objective is to present the findings objectively.

Chapter Five: Summary of Findings, Conclusion and Recommendation

The chapter gives a summary of findings, conclusion and recommendations for the study.

1.12 Chapter Summary

The chapter provided a detailed overview of the study by introducing the concept of Big data, its background, context and application in various business sectors. The chapter begun with the general introduction and background of big data, highlighting trends and vital issues that motivated the study.

The chapter also pointed out the gap in the existing literature that the study intended to address. The aim of the study, problem statement and research questions were equally outlined to guide the study.

The significance of the study was examined, with the key focus on its contribution to literature, Big data practice and future research. Further, the chapter presented the objectives of the study with a brief outline of the scope and limitations of the study that frame the study boundaries. Big data in the context of Ketraco was also explored as well as inclusion of definition of key terms to ensure conceptual clarity.

Generally, the chapter established the foundation for the thesis by outlining the rationale, purpose, aim, and structure of the study. The subsequent chapter will explore the theoretical perspectives and relevant literature that inform the research framework.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

A literature review refers to an account of what has been published on a given topic by accredited researchers and scholars (Osanloo and Grant, 2016). The literature review is not just a set of summaries or a description of the available materials. Instead, Xiaofeng and Xiang (2013) argue that it is a review guided by a defined guiding concept such as the research problem, aim, or the research objective the researcher intends to achieve. It is vital to review and provide a description, summary, and critical evaluation of these works in relation to the research problem being investigated (Osanloo and Grant, 2016). Literature reviews are designed to provide an overview of sources the researcher has explored while researching a particular topic and to demonstrate to the readers how the research fits within a larger field of study. The preliminary literature presented in this section is from primary sources including theses, conference proceedings, government publications, and secondary sources such as journals, books, and newspaper articles. The literature review covers the following aspects of Big Data Management (BDM): a theoretical framework that underpins BDM, policies, and regulations for BDM, knowledge, skills, and training in BDM, ICT infrastructure for BDM, and collaborative partnership in BDM.

2.2 Theoretical Framework

A theoretical framework is a conceptual model of how one theorizes or makes logical sense of the relationships among several factors that have been identified as important to a problem (Osanloo and Grant, 2016). As such, theories are found early in the research

plan of the research and basically act as an orienting lens shaping the questions to be asked, who participates in the study, how data is collected and the overall plan of the research (Creswell, 2014). According to Osanloo and Grant (2016), the purpose of a theoretical framework is to make research findings meaningful and generalizable, thus stimulating research and the extension of knowledge by providing both direction and impetus. There are various theories and models that address different aspects of BDM. They include the information theory, technology acceptance theory, Data Curation Centre (DCC) Lifecycle Model (Higgins, 2008) and the Community Capability Model (CCM) (Lyon et al., 2012). The information theory developed by Claude Shannon stresses on quantifying information and gaining an understanding on how to efficiently store, transmit or compress data. Ash (2012) stated that information theory focuses on syntactic properties of data such as quantity and redundancy. However, it ignores assigning value to data which is key to BDM. Therefore, the theory would be inadequate to study Big data.

On the other hand, the technology acceptance theory developed by Davis (1986, 1989) elaborates how users reject or adopt new information systems. The theory suggests that system acceptance depends largely on users 'perception and perceived ease of use. This implies that the model focuses on individual behaviour not data life cycle hence not capable of evaluating how data flows across systems (Momani, 2020). In addition, technology acceptance model does not address the ICT infrastructure required for BDM implying that it does not address key system features such as scalability and capacity which are key for effective BDM. Therefore, this theory does not guide decisions relating to ICT infrastructure hence not adequate to study the research problem.

Therefore, to effectively address the research questions, the study adopted the Data Curation Centre (DCC) Lifecycle Model (Higgins, 2008) and the Community Capability Model (CCM) (Lyon et al., 2012) that were deemed relevant for investigating and underpinning BDM.

2.2.1 Data Curation Centre (DCC) Lifecycle Model

The Data Curation Centre (DCC) Lifecycle Model was developed at the UK Digital Curation Centre. The model focuses on the necessary steps for successful curation and preservation of data. The model presents a graphical, high-level overview of the stages required for effective curation and preservation of data from the first phase of conceptualization through the iterative curation cycle (Haggins, 2008, p. 136). The model can be used to plan activities within a specific research project, organization, or consortium to ensure all necessary stages are undertaken, each in the correct sequence (Pouchard, 2016). It is important to note that the description, preservation planning, community watch, and curate and preservation elements of the model should be considered at all stages of activity. The Data curation lifecycle of electronic materials is appropriate to ensure their continuous management throughout the curation lifecycle. The DCC lifecycle model is used to underpin the first and second research questions as illustrated in Figure 2.1

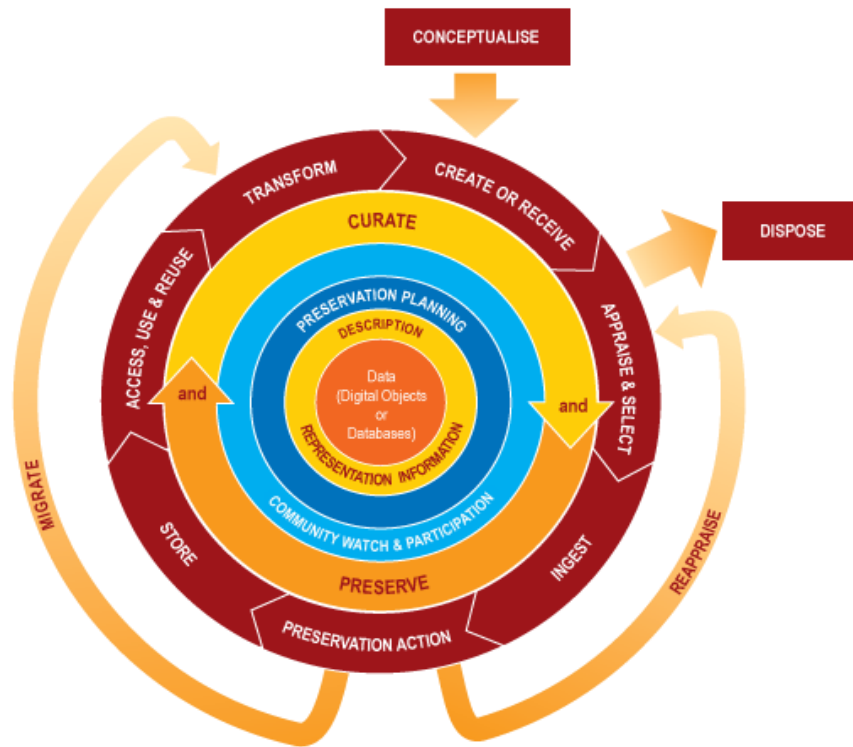


Figure 2.1: The Data Centre Curation (DCC) Lifecycle Model

Source: Haggins 2008, p. 136) (

The Actions in the Lifecycle include:

Capture/Create

It involves creating data and assigning descriptive, administrative, technical, and structural metadata. Receiving data from data creators, data centres, other archives, or repositories can be guided by existing policies. Thereafter, appropriate metadata is assigned. Pouchard (2016) indicates that it is vital to have a collecting policy in place in order to receive data from data centers, data creators, and other archives within a particular framework.

Appraisal and selection

This phase involves evaluating digital material and selecting those for long-term preservation. While performing this step, Phiri and Tough (2018) recommend adherence to documented policies, guidelines, and legal requirements. This implies that it is necessary to develop an appraisal and selection policy with curators and data creators. Therefore, data evaluation and selection for long-term curation and preservation can be guided by the policy. Data not selected for long-term preservation can be disposed of in accordance with the existing guidelines and policies.

Description (metadata)

Data description involves the use of appropriate standards to describe metadata, and this helps in managing and controlling data over the long term. It is necessary to ensure that all metadata and associated digital material can be represented and understood in appropriate formats. Administrative metadata describes how, where, and when the digital asset was created. It also involves the description of the usage or administrative rights. On the other hand, descriptive metadata describes the data asset for purposes such as identification and discovery. It includes elements such as the abstract, title, keywords, and author (Pouchard, 2016).

Preservation/Storage

Preservation is necessary for long-term storage and retention of data. Preservation actions include assigning preservation metadata, data cleaning, and validation (Cox and Tam, 2018). Preservation ensures data remains reliable, authentic, and usable. Data should be stored securely in accordance with the relevant standards and procedures. In the same

vein, data migration involves moving data to a different format, and this is done depending on the existing storage environment or to avoid incidences of software or hardware obsolescence. In equal measure, storage infrastructure can include cloud, servers, flash disks, and repositories, among others.

Access, Use, and Reuse

Data is much more valuable if it is accessible to authorized users to perform their day-to-day business. When choosing a particular mode of access, it is vital to consider the needs of users across the departments, their rights of access, and levels of expertise (Esteva et al., 2016). For instance, some users may only require a graphical interface that has no querying capability while other users might have technical needs to build their only reports and queries.

Big data usage has been embraced in different sectors across the world. Organizations have been using and re-using big data to enhance their competitiveness (Kimble and Milolidakis, 2015) thus, making it a vital asset for any organization. Re-using big data is vital in organization because similar data can be used more than once to undertake a similar job (Toshniwal, Dastidar, and Nath, 2015). For effective re-use of data, suitable metadata schemas can be used in the description of datasets to ensure proper preservation and reuse in the future.

Sharing

Many researchers and organizations opt to share data internally using their intranet or self-host on their websites or in other publicly accessible locations. Data sharing can lead to more collaboration between data creators and users or between one organization and

the other (within and outside the organization). However, Cox and Tam (2018) state that there must be clear policies and guidelines that stipulate how data should be shared.

Transform

Data transformation involves transforming existing data by creating a new set of data from the original one. This can be achieved by creating data subsets and re-analyzing the existing data.

2.2.1.1 Relevancy of the DCC Model to the Study

By applying the model, one can visualize the activities, relationships, and processes necessary for the effective management of big data and its long-term sustainability. The model allows granular functionality to be mapped against it to outline roles, responsibilities and establish a framework of technologies and standards to be implement in big data management (Haggins, 2008). In addition, the model allows the identification of whether additional steps are necessary, if there are missing links, or whether certain steps need to be eliminated to enhance data management processes at different stages. Furthermore, the sequential stage in this model supports the documentation of policies and processes between and by different stakeholders and also encourages the establishment of frameworks of technologies and standards; also identifies needs for specific services and tools to support data management at every level (Haggins, 2008; Pouchard, 2016). In relation to KETRACO, the model can illustrate how data is created/received, appraised, described, preserved, accessed, used/reused, and shared. It can also demonstrate whether the organization has put in place mechanisms and policies for managing data at different stages.

2.2.1.2 Gaps in the use of the DCC Model

The limitation of this model is that it does not fully address the research questions on aspects of collaboration and partnerships, the key enablers of Big Data Management such as human skills, knowledge, training as well as ICT Infrastructure. Similarly, the model emphasizes only policies and regulations at the individual stages of big data curation but leaves out the overall policies and regulations of big data management in its entirety. Furthermore, the key enablers of BDM such as human skills, knowledge, training, and ICT Infrastructure are not factored in the DCC model. As a result, the researcher proposes the Community Capability Model for Big Data Management to fill the gap.

2.2.2 Community Capability Model

The Community Capability Model was developed by Microsoft Research and the University of Bath to help research funders, researchers, and institutions enhance the capability of their communities to undertake data-intensive research (Lyon et al., 2012). The model profiles the current capability or readiness of the community, outlines the priority areas for investment or change, and then establishes the roadmaps for achieving the set objective. The community represents people who share collective or common activities and develop shared vocabularies, strategies, or values. Therefore, a group can be a group of researchers, individuals, or an institution. On the other hand, a capability model represents the readiness to accomplish a particular task. Therefore, the community capability model represents the readiness of a particular group to accomplish a given task. The model has eight capability factors represented by environmental, technical, and human issues. Every factor has several characteristics that influence the readiness or capability of a community to conduct data-intensive research. The CCM Framework is

used to underpin research questions two, three, four, and five as illustrated in Figure 2.2 below.

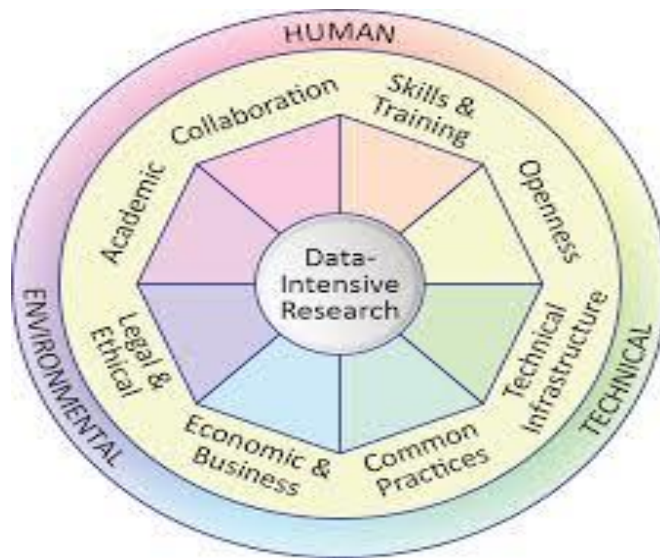


Figure 2.2: Community Capability Model

(Source: Lyon et al., 2012, p. 3).

The elements outlined in the model as outlined by Lyon et al (2012) include;

The Technical Infrastructure (ICT)

It involves the software for data capture and processing, the network platform for data sharing, cloud storage capabilities, data description, and access to large volumes of data (Lyon et al., 2012). Additionally, the technical infrastructure has the components for data integration and collaboration. Security, technical infrastructure for capturing, appraising, describing, preserving, accessing, using, and sharing big data.

Legal, Policies, and Ethical Issues

The CCM addresses the set standards for best practices and requirements for compliance with existing laws and regulations governing BDM. A community must establish policies, regulations, and ethics aligned to the legal framework to support the usage of data. The

Kenya's Data Protection Law Act of 2019' builds a framework that governs the acquisition, processing, use and sharing of personal data to ensure privacy and security of personal data. In relation to this, a community for instance a Company should integrate data protection mechanisms to safeguard personal data. With this understanding, Ketraco's big data should be safeguarded with policies and regulations aligned to The Kenya's Data Protection Law Act of 2019'. The key issues addressed in this section are the policies and regulations set in place to guide how data should be managed at every stage of the cycle. On ethical consideration, data ethics is concerned with the principles of privacy, ownership, openness, transparency, and consent (Lyon et al., 2012). Handling of data should be guided by the mentioned principles to ensure effective big data management.

Skills, Knowledge, and Training

It is appropriate for anyone working with Big Data to learn the skills related to big data management. Knowledge and training equip data creators, curators, technology professionals, and other data users with knowledge and skills needed to work with Big Data.

Collaborative partnerships

The community capability model outlines the areas of collaboration that can enhance elaborative data research. These areas include collaboration with the public, collaboration, and interaction with other sectors and disciplines (Lyon et al., 2012). This aspect is relevant to the study as it underpins the Big Data management between KETRACO and members of the public as well as collaboration with other partners and

stakeholders in the study. Having a collaborative partnership is key for the organization to enhance data sharing, knowledge, and skills exchange that subsequently result in better decision-making, innovation, and improved organizational performance.

Openness

The principle of openness applies at different levels of data-intensive research. By following the established guidelines, the community should ensure openness and access to published data, openness to data, and workflows. Openness improves reproducibility, validation, and re-usability of data. This principle is vital in Big Data Management to ensure data reaches a wider community and the value of data can be realized by a wider audience.

2.2.2.1 Relevancy of the CCM model to the study

The Community Capability Model is appropriate in BDM as it encompasses the key elements necessary for Big Data management. From Figure 2.2 above, it can be observed that different factors are strategic for an organization to enhance Big Data Management. The common practices are necessary to replicate or make improvements in the management of Big Data. In relation to KETRACO, the legal and policy considerations are necessary to provide a framework and guidelines for Big Data management, skills knowledge, and training are necessary to ensure staff perform big data management tasks effectively. Similarly, collaborative partnerships in the case of KETRACO are necessary to enhance data sharing and innovations. Finally, the technical infrastructure is an enabling factor for assembling the computational tools that support Data capture, processing, storage, curation, and access.

2.2.2.2 Gaps in the use of the CCM model

The CCM model is ideal for making coherence of various aspects needed for successful implementation of big data. In essence, it brings together elements of skills, training, policies, and partnerships for big data management. However, the limitation of this model is that it does not focus in detail on data curation which is a significant part of BDM. This implies that the distinct stages of BDM such as creation, appraisal, description, preservation access, use and reuse, and sharing are not well articulated in this model. However, the DCC model is used instead to fill the gap since it demonstrates how data can be curated at different stages.

2.3 Overview of Literature

This section reviews the themes concerning Big Data Management

2.3.1 Big Data Management

In the recent past, big data has advanced into a significant topic that has gained great attention from various industries and society. The concept of big data and its applications have developed from the growing volumes of internal and external data in organizations. Based on previous literature reviews, big data can be defined as enormous volumes of complex, structured, and unstructured data generated by different computer applications (Wang, Kung, and Byrd, 2018). Phrases such as enormous growth of data, massive amounts of data, and large data sets are basically found within the literature that attempts to define big data. According to Kitchin and McArdle (2016), velocity, variety, and volume (3Vs) are the characteristics that distinguish big data from standard data hence, big data is a high velocity, high variety, and high-volume which requires innovative and

cost-effective approach of big data processing. Further, Dezi et al. (2018) characterized big data with additional 2Vs that include veracity and value. Effective handling of big data demands one to create value against the (3Vs) qualities of data while ensuring this data is still in motion (velocity), and not just after data is at rest.

As it has been witnessed, billions of people across the world are connected via the internet, generating large volumes of data. Big Data emanates in two major forms that include structured and unstructured Data. According to Wang Kung and Byrd (2018), structured data has a certain meaning attached to it, while unstructured data does not encompass any latent meaning. The data that has recently increased in volume and received much attention in many companies is mostly unstructured data and this form of data includes data from social media platforms, online blogs and websites, and transactional activities, among others. All these data sets constitute Big Data, and they are generated by different devices and applications.

Given the role that big data has in business processes and its growing interest in different areas, relevant research works have attempted to explore big data, and those relevant to this study should be examined. A study done by Kshetri (2016) examined the contribution of big data in enhancing access to financial services for economically active low-income micro-enterprises and families in China. The author argues that in emerging economies such as China, finance companies had little success in providing services to micro-enterprises and economically active low-income households suitably, and in a cost-effective manner. However, recent developments and advances in telecommunications and computing technologies have significantly transformed this landscape by

successfully changing how financial institutions function. The key strategy behind this transformation has been the utilization of big data in evaluating, assessing, and refining the creditworthiness of their borrowers, and this has helped in enhancing their financial transactions and performance.

A study done by Höchtl et al. (2016) titled: 'Big Data in the policy cycle: Policy decision making in the digital era' explored the roles of big data in the decision-making processes. In this regard, the study established that decision-making process is based on two sets of elements namely technical and organizational. The organizational elements are those associated with organizational day-to-day operations where decisions made are aligned with the organizational strategy. On the other hand, the technical elements comprise the toolset used in supporting the decision-making process such as the data repositories, information systems, and analysis of decisions. Furthermore, the study by Höchtl et al. (2016) revealed that different elements when combined make an integrated model that supports decision making. These elements include business intelligence, decision support systems, big data, and organizational culture. If these elements work together, they enable decision-makers to have a reliable visualization of the opportunities relating to decision-making. Therefore, from this study, it can be concluded that big data is one of the key enablers in effective decision-making.

Another study by Horta et al. (2015) explored the challenges associated with handling large volumes of data sets. The study used active learning on extreme learning machines as an approach to solving big data problems. In their demonstration, the authors argued that big data comes in large volumes, and this in itself is a challenge to induct users on

using the large data sets. As a result, the authors suggested that dealing with big data requires some techniques and their proposal was to use a sample of data or data sets instead of the entire data set to induct learners, and this makes the handling of big data more attractive.

In another separate study on big data management, Song et al. (2015) discussed the development of big data technology, especially in the areas of generation, management, analysis, and use. In essence, the study found out that there has been a speedy growth in information generation and there are potential benefits hidden in these large data volumes. If effectively managed, big data can enhance productivity in business processes and a breakthrough across scientific disciplines, enabling people to make great progress in their fields. However, Chen and Zhang (2014) outlined that despite the many opportunities, big data arises with many challenges including data capture, storage, analysis, and visualization and these are the areas that still need further analysis. Data becomes more complex as it increases in volume exponentially, but few organizations realize that highly valuable information is hidden in these volumes of data, and business organizations can derive many benefits if this data is effectively managed (Chen and Zhang, 2014; Song et al., 2015).

2.3.2 Big Data in Organizations

The internet has made information readily available, and this information took many years for companies to acquire in the recent past. Currently, data is generated from different platforms including social media which generates huge amounts of data. Kitchen and McArdle (2016) observe that the data generated relates to various domains. The fields

that currently apply big data are quite large, and they include Telecom, Retailing, Media and Entertainment, Finance, Energy, Health, and Government, among others.

Big data is used by businesses and organizations to provide fully customized services to their customers, understand customer behaviours, needs, and habits, and help in managing stocks and understanding and predicting customer needs for certain products. More importantly, much of organizational data is helping organizations in making informed decisions based on facts, and in improving business processes. Generally, by managing big data Kimble and Milolidakis (2015) stated that organizations can improve their decision-making, competitive edge, organizational capital, and innovation. More effective big data management can lead to better and more confident decision-making resulting in greater operational efficiencies in organizations.

According to Dezi et al. (2018), organizations have recently started to explore and analyze information from external sources that come from many dispersed information sources including information from governments, industries, and competitors. However, the practice has demonstrated questionable success from the use of big data since organizations do not attain an appropriate value from big data applications. The reasons for this failure have not been clearly investigated hence more investigation is needed to explore how big data can effectively be implemented. As opposed to firms established in the digital age, traditional companies face a significant challenge in their efforts to integrate big data into their existing information technology systems (Kitchin and McArdle, 2016). Implementing big data initiatives in traditional companies requires the management to consider the socio-technical elements of an organization such as IT architecture,

analytics platforms, IT infrastructure, required expertise, security measures, and organizational structure.

2.3.2.1 Contributing Factors to the Growth of Big Data

Many organizations have traditionally been capturing and maintaining transactional data. However, there has been a significant shift in the modern world where companies are now generating or acquiring additional data from their business environment at a growing rate. Examples of this data captured include text data such as emails, documents, web data, and social network data (Dezi et al, 2018). The process of interacting with large data sets in an organization offers various benefits as it allows data users to give more attention in their decisions that rely on data. However, Labrinidis and Jagadish (2012) argue that increase in data volume needs to have a corresponding improvement in its quality. Different challenges arise in several dimensions. From a technological perspective, it is vital for emerging technologies to handle large data sets in a reliable and most effective way. On the business point of view, it becomes appropriate for organizations to be able to filter data that is valuable and separate from the invaluable ones (Labrinidis and Jagadish, 2012). Finally, from the social perspective, it is vital to ensure confidentiality and security in the handling of data. Creating the balance of all the business, technological and social dimension has significantly contributed to the growth of big data as companies try to leverage on technologies to derive much value from their data assets in a more efficient and secure approach.

There are several key trends that have considerably resulted in data generation growth. To begin with, an increase in traditional transactional databases is one of the trends and

is majorly linked with the noticeable fact that companies are increasingly collecting data with greater granularity and frequency. This growth is associated with various factors such as the increasing competition, the growing turbulence in the competitive business environment, varying customer needs and increasing customer expectations (Zhang et al., 2018). All these pressing issues demand more from organizations to swiftly react and adapt with maximum flexibility to the emerging trends.

Another contributing trend to the growth of data is the surge of multimedia content which is associated to the growth in the application of multimedia content in different industries of the contemporary environment (Information Today Inc., 2020; McAfee et al., 2012). Currently, data in multimedia platforms accounts for over half of the existing internet traffic, and it is projected that by the end of 2022, this share will rise to over 80% (Information Today Inc., 2020).

Another trend that has resulted in an increase in the volume of data being created is the growth of the term commonly known as the Internet of Things where there has been an increase in the number of physical objects or devices that communicate to one another without the interference of human beings. The devices link with each other through a wireless or wired connection, mostly using IP protocols. Since these devices are equipped with actuators and sensors, they collect and send huge amounts of data (Katal, Wazid, and Goudar, 2013). With the increasing use of electronic devices and accessibility of the Internet in remote areas, the volume of data generated from the 'Internet of Things' has grown exponentially, and this number is expected to increase further as the number of

connected devices existing globally is anticipated to rise at an annual rate of over 30% (Kumar, Tiwari and Zymbler, 2019).

Social media is another key source of the growing volume of data generated globally. For instance, Facebook users generate large volumes of data. The statistics by Borgatti, Everett, and Johnson (2018) indicated that there are about 2.27 billion users of Facebook worldwide, about 1.45 billion active ones who spent at least 20 mins on these social platforms, and with every user generating an average of 90 pieces of content (notes, photos, blog posts, news stories, and links. Research conducted by Booker, Barnett & Cribbin (2016); Villanti et al. (2017); Trifiro & Gerson (2019); Bengtsson & Johansson (2022); and Zachlad et al. (2022) indicate that if only messages are considered, Facebook users receive an average of about twelve messages a month. In the case of YouTube, a video is uploaded every month, and in the same timeframe, Twitter users send 98000 tweets. According to McAfee et al. (2012), smartphones perform a vital role in social network communication. Although the growth and use of social networks has increased for both PCs and smartphones, the access of social media networks is higher with smartphones. If frequent users are considered in the case of PCs, it is 11% p.a. while in the case of smartphones, it is 28% p.a. This has resulted in a rapid growth in mobile data traffic, and it is projected that mobile data traffic will increase higher by 2020 (Booker, Barnett & Cribbin, 2016; Chen, Cheng & Chen, 2018, and Trifiro & Gerson, 2019).

2.3.2.2 Importance of Big Data Management on organization's Operations at Ketraco

Big data plays a big role in the functions of any organization in the current economy. Data is useful in supporting the organization's operations and performance. Gantz and Reinsel

(2012) stated that while many IT-based organizations established recently have been able to collect and utilize data for their business processes and growth, many other organizations have also tried to embed big data practices in their business procedures.

To begin with, KETRACO successfully relies on data to enhance its operations across the country. Data collected from the field is crucial in supporting the functions of the organization (Ketraco Annual Report 2017-2018). For example, Cadastral data, surveys and mapping data, and personal data of persons affected by Ketraco projects are all crucial to Ketraco. These types of data form the foundation of any Ketraco project. For instance, to construct an electricity transmission line from one region to another, one needs data on feasibility studies, surveys, and land parcels, among other types of data to be able to compensate the affected persons and thereafter construct a power line. The effective collection, description, and sharing of this data can support the organizational mandate by reducing time wastage and ensuring timely decision-making.

Correspondingly, Ketraco has branches, assets, and other electricity infrastructure distributed across the country, and it invested in a Geographical Intelligence system and Integrated Location Information systems to integrate the company infrastructure in different locations and into one complex framework (Ketraco Annual Report, 2017-2018). The systems optimize organizational efficiency by pooling together data, and this enables the company to identify and respond quickly to any emergencies such as a broken power line (Barth and Bean, 2012). Ketraco has been able to collect and preserve data, and this has seen the company make instant decisions that are crucial to its operations.

Another practice being witnessed in organizations is heavy investment in systems and software that support big data and create a collaborative work environment (Barth and Bean, 2012). For instance, Ketraco has invested heavily in business applications such as the Systems Applications Program (SAP), SharePoint, computer network facilities, and storage platforms. Employees at different departments key in data that flows in the system to another person for execution. These new applications have improved workflows, efficiency, and better decision-making.

According to the findings of a survey carried out by Davenport, Barth, and Bean (2012) on different functional heads and C-level executives in different leading organizations in America revealed that there were benefits associated with shared data which included improved fact-based decisions and enhanced customer experience. Another important benefit was anchored on the expectations to make better organizational decisions. Other benefits include new product innovations, reduced risk, improved quality of products and services, increased sales, and efficiency (Davenport, Barth, and Bean, 2012). As a result, organizations are using technological tools to create a network of communities where data and information are easily shared amongst staff. Organizations are also using software to collect and manage a pool of data, and this data is instantly used to solve a particular need. Thus, the significant value of Big Data is being embraced to speed up the time taken to solve a problem hence increasing the pace of making decisions both at the tactical and operational levels. Generally, Big Data creates much value for Ketraco and other organizations as well. Incorporating data in business procedures ensures the availability of this data to support business functions. It also ensures the accessibility of this data to all relevant stakeholders.

2.3.3 Stages Involved in Big Data Management

The following are the stages involved in big data management from an information management perspective (Higgins, 2008; Erevelles, Fukawa, and Swayne, 2016; Penn and Pennix, 2017; Kitchin and McArdle, 2016; Dezi et al., 2018; Toshniwal, Dastidar and Nath, 2015; and Wang, Kung and Byrd, 2018).

a) Data Capture/Creation

Data creation is the first stage in the data management process, and it involves capturing or creating data itself. Data in organizations is collected for various purposes such as planning, operations, maintenance, or solving a particular objective or goal (Erevelles, Fakawa, and Swayne, 2016). Information users are key stakeholders in the process of creating and analysing data. The types of methodology used in data collection vary at different geographical and jurisdictional levels depending on the needs of different users. Therefore, the systems and methods used in creating data should be structured in a way that meets the internal and external needs of different information users as well as the needs of the agency (Penn and Pennix, 2017). The design and planning of data capture methodology involve the identification of data needs and objectives, the establishment of data providers, designing and planning methods of collecting data, and documenting data collection approaches. Factors such as funding availability, data quantity, target populations and research questions, and length of collection period influence the methods and strategies applied in the data acquisition process and hence should be considered when planning for data acquisition methods.

b) Data Appraisal

Data appraisal is the process of examining data to determine those with long-term value from those that no longer have value. In appraisal, those with continuing value are selected for long-term preservation and curation, while those that have no value are eliminated (Wamba & Akter, 2023; Zellner, Tilahun & Thakuriah, 2016). The appraisal of data for long-term curation and storage should be guided by well-documented policies, guidelines, and legal requirements.

c) Data Description

Data description involves the use of appropriate standards to provide descriptive, administrative, technical, preservation, and structural metadata required to describe data adequately over a long period. Description is necessary to manage and control a digital object in the long term and ensures that data is curated and stored in appropriate formats. Once data has been aggregated, it is necessary to package and store it for other uses in a well-structured format. The structuring process involves organizing data in a certain schema, and this is critical as it allows users to make queries on the data (Dezi et al., 2018; Zellner, Tilahun & Thakuriah, 2016). Different technology platforms such as NoSQL and Hadoop applications (data processing software) allow one to query even on unstructured data and these platforms are increasingly being embraced in big data analysis (Kitchin and McArdle, 2016). A major concern with big data is providing instant results after the query; therefore, data structuring needs to be conducted at a rapid speed. After data structuring, queries are conducted on the data. Thereafter, data presentation is done in a visual format (Dezi et al., 2018). The process of analyzing data encompasses targeting key areas of interest and outputting results depending on the structured data.

d) Data Preservation/ Storage

Preservation is important to ensure data remains authentic and reliable for the entire period it is retained. Preservation measures include assigning preservation metadata, data cleaning, and validation. Data should be stored securely in accordance with the relevant standards and procedures. Data identified to have long-term value can be preserved in flash disks, servers, and personal computers, among other external storage platforms. Penn and Pennix (2017) define data archiving as the identification and transferring inactive but vital data from active systems into archival storage. The purpose of archiving is to transfer inactive data from active systems to optimize performance and also store inactive data in cost-effective archival systems that allow future retrieval.

Data can be archived in an Institutional repository. An Institutional Repository is an online platform for collecting, preserving, and providing access to digital content of a community. Institutional Repository provides mechanisms and tools that help an institution to share its Institutional Repository data with the audience both within and outside the institution. Betz and Hall (2015) stated that the objectives of an institution repository are to collect and store digital content in a single collection, create global visibility of stored data, and also provide access to institutional data.

Cloud storage is another technique for storing data online in the cloud. The stored data can be accessed virtually from multiple connected and distributed resources. Toshniwal, Dastidar, and Nath (2015) stated that companies are opting for outsourcing storage services or cloud storage since the cloud option enables users to view storage medium as a service hence, they will not be required to maintain the platforms themselves. By

applying this approach, it becomes the responsibility of the service providers themselves to ensure the scalability and reliability of the computing platform while users can focus on data analysis and its use (Alharthi, Krotov, and Bowman, 2017).

e) Data Open Access

Ensure accessibility of data by both re-users and designated users on a day-to-day basis, and this can be in the form of published information for public access (Penn and Pennix, 2017). To ensure access by only authorized users, authentication procedures and access controls should be well established. Various international agencies, media, legal and human rights entities have advocated for open access to data. For instance, during the Berlin Declaration (2003), the conference advocated for institutions and researchers to publish their work and encouraged open access to help in decision-making and development (Klump et al., 2006). In addition, the proceedings from the 2016 European Commission Conference on Big Data in Space deliberated that the value of data generated depends on an organization's capacity to extract and derive meaning from its information sources. From the European Commission Conference and the Berlin Declaration, it can be concluded that the value of data depends on the capacity of an organization to facilitate access to it, so that users can derive meaning that informs their decisions.

f) Data Use and Reuse

Big data usage has been embraced in different sectors across the world. Business firms have been utilizing big data to improve their competitive advantages. Kimble and Milolidakis (2015) state that business organizations that fully apply or embrace big data get competitive advantages over their competitors, and these varying applications of big

data make it a vital asset for any organization. Toshniwal, Dastidar, and Nath (2015) define data reuse as using similar data more than once to undertake a similar job while repurpose is using similar data to undertake a different task from the original one. Once data has reached the end of its life cycle, it will eventually be reused, repurposed, or destroyed when its value has been exhausted. With repurposing and reusing, the cycle is not considered linear; instead, it is circular. Using suitable metadata schemas can enhance the description of datasets so that they can be properly preserved and reused over time. In addition, countries have set up measures to ensure that their data sources from various sectors such as business, agriculture, medical, tertiary institutions, and other government information are captured, documented, and made available for use and reuse. For instance, in Australia, the government funds the Australian National Data Service, a program that provides visibility to research data in search engines such as Bing and Google (Kim, 2013). Such portals enable users to access, use, and reuse data for research-related purposes by governments, research organizations, and cultural institutions.

g) Data Sharing

Sharing data with authorized users is a key to saving time as data can easily be accessed to perform vital tasks. It is important to grant users access to company data that is key to their respective duties or functions. Access to well-analyzed and packaged data can result in efficiency and effectiveness in decision-making processes as well as ensuring instant responses to information inquiries. However, while ensuring access, Wang, Kung, and Byrd (2018) recommend that the process should be balanced against the necessity of controlling the sharing of sensitive and classified data. Information is shared has social, economic, policy, and legal implications mostly in cross-cutting aspects such as data

ownership, security, privacy, and liability.

h) Data Transformation

Data transformation refers to the process of converting data from one structure into another structure or format. The process involves transforming existing data by creating data from the original one, and this can be achieved by creating data subsets or migrating into a different format (Kimble and Milolidakis, 2015). Data transformation is an essential process for activities such as data management and Data integration. The activities that can be included in data transformation are converting data types, cleansing data by removing duplicate or null data or enriching the data. Data transformation is necessary especially when one wants to make it compatible with other data, join it with other data, and move it to another system or aggregate information in the data (Kimble and Milolidakis, 2015; Toshniwal, Dastidar and Nath, 2015).

2.3.4 Legal, Policies and Regulatory Framework in Big Data Management

In many countries, legal issues have focused on matters of privacy, disclosure, improper collection, and improper use of data (Xue and Xin, 2016). Also, most mobile Apps, online services, and websites have privacy policy statements and terms of service agreements that users agree with before using the site (Gantz and Reinsel, 2012). In the US, there is no single national law that regulates how personal information is collected, used, or shared (Vilita and Chore 2024). However, different states have passed laws that regulate how certain information is collected or used. The state regulations have addressed certain types of personal information such as financial or health information.

In the UK, Big Data is affected by the European Union's General Data Protection Regulation that came into effect in May 2018. The law has provisions limiting the type of data collected by organizations, and it gives individuals the right to have their information deleted from databases even if the organization had permission to include that type of information (Tabesh,et al, 2019). In South Africa, the law on protection of personal data was introduced in 2013 and it provides citizens with the right to privacy. The law has provisions on the minimum condition of collecting and processing personal data. In Kenya, Article 31 of the Kenya Constitution recognizes the right to privacy and the protection of personal data. The Data Protection Act of 2019 was established to govern data in Kenya. This Act formed the basis of establishing the Office of the Data Protection Commissioner whose mandate is to implement and address issues relating to privacy and general data management (Kenyalaw.org, 2021; Office of the Data Protection Commissioner (ODPC), 2024). In addition, there are established policies on confidentiality, storage, use, and dissemination of personal data.

On the issue of regulations, these are the core of the data laws as established in various jurisdictions. There are about seven principles guiding the data regulations, and these apply across the board, and they include lawfulness, fairness, and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability (Vilita and Chore, 2024). These core elements inform and permeate the various provisions of the law on data management. They are critical in ensuring compliance with the established laws on data protection, as well as ensuring that individual rights are protected (Data Protection Commission, 2019).

Equally, the enhancement of policy framework and regulations vary from one nation to the another, and this is influenced by regional legal systems, public expectations, data governance priorities and technological developments (Mao et al, 2022). For instance, the European union has promoted data management policy by enhancing right based approach by enacting data protection regulations that established guidelines around access, consent, transparency in data governance (Tabesh, Mousavidin and Hasani, 2019). The same approach has been adopted in many countries such as United States, China, India Canada, Nigeria and Kenya by setting out strong principles for lawful user rights, data processing, storage and transfers. Generally, many countries are adopting citizen centric policies for data governance. The policy framework help define responsibilities, roles and standards by guiding the establishment of data catalogues and metadata standards for effective data management (Mao et al, 2022).

From the above review of legal issues in Big Data management, it is evident that many countries have established regulations addressing the collection, storage, use, and dissemination of personal data. Therefore, it is paramount for any organization operating in any country to understand and comply with the legal and regulatory requirements regarding Big Data Management through company policies. According to Pouchard (2016), big data is extensively used in different businesses, and its application can result in improved decision-making and better organizational performance. However, the successful implementation of big data in an organization is affected by the organization's compliance with various legal issues of privacy, international access, government regulation, and the growing criticisms of the digital gathering of information.

2.3.5 Knowledge, Skill, and Training for Big Data Management

People in an organization are crucial in ensuring the implementation of the big data management agenda. Big data is a task that requires people with skills and technical knowledge about how to process data into meaningful information. However, the requirement of having critical knowledge is not limited to people managing data, or to those that process, or interact with data systems and technology, but also to the decision makers and information users so that they can effectively interact with large volumes of data to filter the most relevant information appropriate for making key decisions (Poletto et al., 2015).

Every person in an organization requires training on cultural and organizational changes when implementing big to enhance its acceptance and application. In addition, workshops can be organized within the organization advocating for best practices in enhancing the use of big data among the employees. Awareness and data literacy programs can also be enhanced to equip employees with data skills that are vital in handling complex corporate data. Dezi et al. (2018) argued that since data is crucial to the success of any business, employees can be trained on which data is appropriate to use and also understand how to access or communicate data that supports business processes.

The personnel identified in an organization to manage large data sets as the information and data management team, IT team, data warehouse group, and also the department or business units need to possess the relevant skills and also focus on training needs that enhance the appropriate use of data to achieve greater results from big data management. Dezi et al. (2018) propose appropriate measures that an organization can use to enhance

data management, and one of these measures is training staff on how to work with large data sets and data infrastructure platforms. Appropriate measures should also include hiring data scientists and analysts who have the capability of developing and configuring systems for data discovery, analytics, and monitoring of these data sets. Wang Kung and Byrd (2018) claimed that people who are conversant with big data platforms need to closely work with the right individuals who largely interact with big data in their departments so that they can attain greater outcomes from these large volumes of data.

2.3.6 Technical Infrastructure for Big Data Management

Big data environment requires different techniques and technologies for gathering, processing, and storing data. Most of these technologies have evolved recently alongside big data; however, others have been available for a long time only that they have been improved to enhance their efficiency in managing big data (Katal, Wazid, and Goudar, 2013). At Ketraco, there is a need for several technologies for managing big data to be designed and adapted to improve the processes and procedures that must be performed in big data management. The key technical infrastructure for BDM includes:

i. Network Infrastructure

The large volumes of data acquired, processed, managed, and shared need robust network infrastructure for it to be used effectively for its intended purpose. Zhang et al. (2018) stated that big data projects and initiatives require network solutions in various phases of big data management. Many enterprises are already running their systems with networking hardware that enables large and faster connections and may only require

minimal adjustments like the installation of internet ports to facilitate the implementation of big data.

ii Storage Platforms

Most organizations have adequate in-house storage facilities that support Big Data projects. However, the relevant authorities within a company may make decisions to invest in storage solutions optimized for Big Data. Flash storage is attractive because of its high availability and performance advantages although it is not necessary for every Big Data deployment. Large Big Data users (i.e., organizations like Facebook and Google) make use of hyperscale computing environments that comprise commodity servers with direct-attached storage, use PCIe-based flash storage for reducing latency, and run such frameworks as Cassandra or Hadoop (Toshniwal, Dastidar and Nath, 2015). However, smaller organizations use clustered network-attached storage or object storage. For purposes of disaster recovery and backup of on-premises Big Data, cloud storage is a better option. The cloud may be available as a primary storage source, but most companies find it to be less cost-effective compared to on-premises storage because of the expense of transporting data into the cloud (Cui, Yu, and Yan, 2016).

ii. Big data Management Software

The software is a key component in the adoption of big data initiatives in an organization. Software is a collection or set of programs, data, or instructions used to command computers to execute a specific function or task data (Cui, Yu, and Yan, 2016). Ketraco should use big data management software to catalog, distribute, query, and manage. Some of the software available to process data include:

a) Hadoop: is an open-source software for processing huge amounts of data in distributed systems (Yan & Ma, 2019). The Hadoop software has the capability of processing extremely large volumes of data by largely allocating partitioned data sets to different servers (nodes), which independently handle different parts of the larger problem and then integrate them back for the final result. The software can serve two roles as an analytics tool or a data organizer (Katal, Wazid, and Goudar, 2013).

b) NOSQL: is a database good at processing dynamic and semi-structured data whose latency is low and this makes them suit a Big Data environment (Katal, Wazid, and Goudar, 2013; and Sullivan, 2015)). NoSQL is most suitable for operational tasks, such as interactive workloads that are based on selective criteria where data may be processed in almost real-time (Shen et al., 2013).

c) Cloud: cloud computing is a model that allows convenient access to the shared pool of computing resources such as servers, storage, networks, applications, and services that can be released easily with minimal service provider interactions (Poletto et al., 2015). In this technology, there are minimal up-front costs since one has to pay for the services used and then scale out from there. Shen et al. (2013) stated that apart from cost, cloud computing is also advantageous in terms of faster delivery of insights.

iii. Data Security

Data security is a key aspect of big data management that focuses on the safe handling of data. Big data security refers to the tools and measures put in place to guard data from theft, attacks, alterations, or other malicious activities that could negatively affect it. Users often have much trust and confidence in using data that is well-regulated and

secured (Wang, Kung, and Byrd, 2018). Organizational data needs to be reliable and trustworthy, and this can be achieved if it is safeguarded against unauthorized access. The possible threats to data include theft and unauthorized access to data stored online, viruses, and threat attacks that can attach to the servers and storage platforms, attacks on big data storage can lead to serious financial implications such as litigation costs, losses, and sanctions or fines.

Shen et al. (2013) recommend measures for data protection, and they include encryption of data by coding it in such a way that only users with access to the secret key can have access. Encrypted data appears unreadable to an entity or a person accessing without the key. Another security measure is building a strong firewall. Firewalls filter traffic that enters or leaves the servers and hence can be used to prevent attacks by restricting any third parties from accessing big data storage platforms (Wang, Kung and Byrd, 2018). Another approach to enhancing data security is by setting user access controls to regulate access privileges and protect data against unauthorized access and misuse.

2.3.7 Collaborative Partnerships in Big Data Management

According to Dezi et al. (2018), effective Big Data management is community-based. A company that deals with national or global issues should establish groups that can work together through collaborations to address a particular challenge. In the case of Ketraco, its projects traverse peoples' land, and the company collaborates with the National Land Commission to get land data of the persons whose land parcels have been affected. This collaboration ensures that only genuine landowners are compensated, and this measure

has saved Ketraco a lot of money that would have otherwise been paid to people with fake title deeds.

An organization can partner with another agency to create common platforms for collecting and sharing data and use that data to innovate and make informed decisions (Borgatti, Everett, and Johnson, 2018). Ketraco has also partnered with local administration authorities in order to act on behalf of Ketraco to receive a lot of data from people affected by Ketraco Projects. By using the local administration, who understand their locations and occupants better, Ketraco receives data that is authentic and reliable, and it is this data that informs Ketraco's decisions when undertaking a project in a certain area.

According to Cricelli and Grimaldi (2010), teamwork and collaboration across borders are key institutions complementing each other while executing their mandate. Ketraco works closely with other Energy sector players to drive the Kenyan government's agenda of building a robust energy infrastructure by 2030. In this regard, Ketraco's function is facilitated by the continuous exchange of data with other players in the energy sector. For example, Ketraco usually engages KPLC and Kengen, who are critical players in the Kenyan energy sector, regularly on matters relating to the transmission of power from one region to another. Partnerships between Ketraco and other players within the energy sector ensure that data is easily exchanged and the key players can also share their skills and expertise that are vital for better performance of any organization.

In their study, Yildirim et al. (2016) examined the opportunities that come with private-public partnerships and the application of Big Data in the manufacturing of pharmaceuticals. The study established that private-public partnerships and the exchange of data contribute to efficiency in the development of drugs by shortening timelines. Therefore, by sharing data, organizations like Ketraco can learn to address similar challenges experienced by other organizations in the past and address the challenges encountered by others while performing a task in the past in advance hence reducing inefficiencies and saving on costs.

According to Cricelli and Grimaldi (2010), there is a great opportunity for institutions that work together to amplify their success. Non-profit firms can exchange data and expertise on key areas that can help them repurpose or replicate each other's data solutions and inspire one another on the best approaches to harnessing the power of data to drive their mandates. By working together, firms can learn from one another, build and improve on their previous projects, and ultimately improve their processes and performance (McAfee et al., 2012). Therefore, collaborations can help an organization to recreate and grow stronger together.

2.3.8 Support Services for Big Data Management

The recent growth and adoption of Big Data management has been enhanced by information professionals and Big Data's closely related fields of Information technology, Records and Archives management, and library science. Information technology plays a key role in Big Data management by providing platforms for collecting data, analyzing, processing, and preserving data (Sheng et al., 2017). ICT professionals are vital in

providing technical support in the Big Data management systems. They help in offering support such as data modelling, data mining and extraction, analysis, version control, security, and access control, among others (Song et al., 2015). Their input ensures that the big data management function of an institution succeeds.

Records management supports Big Data management in various aspects. With Big Data management, data exists in large volumes and minimal value can be realized from these data unless proper guidelines are put in place to guide data appraisal, classification, storage, and access (Kimble and Milolidakis, 2015). Records management professionals play a key role, especially in advocating for an organization to comply with the procedures of collecting, describing, and preserving data. The growth in data production results in the growth in storage media; therefore, the guidelines applied in records management on functions such as appraisal and archiving can still be applied in big data management to ensure that the organization stores data that is needed and this saves on storage costs (Sheng et al., 2017). Records management techniques such as archiving can be applied in Big Data Management hence improving how organizational data is managed for a long period.

Librarians play a key role in the era of big data because they have the skills and the knowledge to help an organization of any kind utilize its massive digital resources. Librarians have a role to play in big data management in their core function to organize and describe data (Koltay, 2017). In addition, libraries are supposed to provide access to internal and external data sets and assist the processes of reusing data. On the same note, librarians are better placed to advocate for training on data literacy education, offer data

mining support, and provide data mining support (Kimble and Milolidakis, 2015). According to Kimble and Milolidakis (2015) and Koltay (2017), libraries take up the role of sharing vast volumes of data to benefit a larger audience hence highlighting the roles of information professionals in a big data environment. For instance, curators are vital in determining how and where to obtain data; data cleansers help to remove duplicated and erroneous data, while data archivists build, maintain, and operate data warehouses.

2.3.9 Challenges in Big Data Management

Digital information and large sets of data are being generated in large quantities in many organizations across the world. However, Xue and Xin (2016) state that challenges emerge on how to access these data in the best formats and channels that enhance the usability of data. Some of these challenges include:

a) Privacy Issues

Alharthi, Krotov, and Bowman (2017) define privacy as the individuals' rights to influence or control the type of information connected with them that can be disclosed or utilized. Privacy is a contemporary issue, and it has various implications for any organization intending to explore big data initiatives through data acquisition, storage, use, retention, and presentation. Therefore, when handling big data, measures should be in place to ensure personal data is not accessed by unauthorized persons. Similarly, an organization should ensure its data is well-secured and not accessed by competitors (George et al., 2016). Some of the measures to protect data sets include the use of codes and passwords on different computer gadgets, the use of controlled access mechanisms,

and installing security firewalls to safeguard systems and storage medium against online hackers.

b) Insufficient Acceptance and Understanding of Big Data in Organizations

In many cases, organizations fail to understand the big data concept including its basics, its benefits, and the required infrastructure. Without clear procedures and understanding, the adoption of big data projects risks failing. As a result, organizations may end up wasting many resources and time on a project which they cannot effectively implement. To address this challenge, Penn and Pennix (2017) recommend that big data initiatives should be first embraced by top-level management and then down the ladder to enhance its acceptance and understanding at organizational levels. In addition, training and workshops can help improve big data recognition.

c) Confusing Variety of Big Data Technologies

Currently, there are varieties of technologies that offer big data solutions. As a result, it becomes difficult for an organization to choose the ideal solutions that best satisfy their needs. Similarly, some of the solutions differ in functionalities, storage capabilities, speed, and other specifications hence easier to choose poorly the technologies that serve the business needs (Kimble and Milolidakis, 2015). It is recommended for organizations to clearly understand and define their needs before opting for these varieties of big data solutions.

d) High Costs in Big Data Implementation and Management

Dezi et al. (2018) argue that big data management comes with costs and expenses that an organization is likely to incur. There are costs in the management, infrastructure,

operation, and integration of Big Data with other organizational systems. As Big Data environments rise, more hardware is needed to provide the additional storage capacity. In addition, the organization needs to establish security mechanisms for its data, and disaster recovery as well as need to have skilled resources. For effective implementation and management, all these Big Data aspects are associated with costs to the organization.

e) Complexity of Managing Data Quality

The large organizational data sets come from various sources and in a variety of formats. Organizations need to analyze data from social media, website logs, competitors' websites, and other online platforms, hence implying that big data is never 100% accurate. Therefore, it is necessary to control this data and improve its reliability since not only can it contain inaccurate information, but also duplicated and contradicting information. Poleto et al. (2015) recommended the need to increase the need for data quality by considering several data characteristics that ensure data managed by an organization is clean. These characteristics include consistency, accuracy, and completeness.

f) ICT Infrastructure Challenges

Gantz and Reinsel (2012) conducted a study on how to effectively build a big data infrastructure for dynamic workflows from which the infrastructure challenges were explored. Storage capacity in many big data projects must increase to accommodate new sets of data increase. However, Alharthi, Krotov, and Bowman (2017) argue that simply adding raw capacity can lead to challenges in the efficient utilization of resources unless other infrastructure issues are taken into consideration. One of the key considerations is the interdependence of infrastructure elements. Many firms are increasingly embracing

virtualization to cut operating costs. However, this results in more architecture becoming widely used, hence the need for organizations to address other infrastructure challenges that are likely to arise (Chen and Zhang, 2014). It also becomes more complex if a server is running many virtualized applications for different clients. The main challenge is how to integrate storage, network, and server elements effectively. This has been a great challenge since it places new demands on the storage solution more specifically, on the fact that big data analytics need storage devices that are flexible and can dynamically expand to meet the growing needs, performance requirements, and varying capacity.

g) Management Related Challenges

McAfee et al. (2012) outlined management challenges that hinder organizations from deriving the maximum benefits of big data; they include decision-making, talent management, leadership, technology and company culture. Organizational leaders have a vision of the organization's growth hence they are required to set organizational goals so that they develop big data practices aligned with organizational goals.

Talent management relates to the need for an organization's commitment to recruit, develop, manage, and retain a talented workforce in areas of big data management who can work with huge sets of data (McAfee et al., 2012). Many organizations have not developed appropriate strategies for retaining potential workforce who can effectively work with large sets of data. The other management challenge is associated with initiating big data management practices so that they form part of the organizational culture. An organization can become data-driven by opening data access, becoming data literate, and

aligning big data with business objectives. These practices can create a data-centric culture in an organization.

h) Legal and Regulatory Requirements

Another key challenge in data management is connected to legal rights in data use and the key issues include copyright, trademarks, confidentiality, competition laws, database rights, and transparency in data collection procedures. Therefore, it is paramount for organizations to become data compliant by ensuring that all big data management practices adhere to established legal frameworks and existing principles (Gantz and Reinsel, 2012). The practices that can enhance this compliance include anonymizing data, controlling access to databases, and embracing best practices that adhere to data-related laws in a given authority.

2.3.10 Strategies for Effective Big Data Management

Organizations have already embraced big data to generate value, and others are also required to explore different ways to implement big data if they intend to compete in the current economy. According to Penn and Pennix (2017), big data contributes a significant amount of money to a country's economy, and its continuous adoption in different sectors is projected to increase income further. Furthermore, the study by Braganza et al. (2017) anticipates that the sectors most likely to derive many benefits from Big Data will be retail and manufacturing, financial institutions, and public bodies. However, the benefits can only be realized when effective strategies have been established to ensure effective big data management. Some of the strategies outlined in literature reviewed are anchored on the following:

Establishing a big data management culture: According to Labrinidis and Jagadish (2012), developing a big data culture in an organization requires adherence to data governance principles. Organizations that implement Data Governance by applying processes and tools that do not fit the organizational culture often fail. Proper data governance should include issues of accountability, responsibility, policies, and procedures, data standards as well as a collaborative environment (Penn and Pennix, 2017).

Adapting to new changes: Keeping up with trends in data management is among the best practices in the effective management of big data (Woerner and Wixom, 2015). It should be noted that software and data change from time to time due to dynamic technology and therefore, there are very many new products and tools that are introduced into the market on a daily basis.

Interlinking the company's data: According to Labrinidis and Jagadish (2012), there are very many channels that can be used to access a database. All business entities want all the data and applications in their possession to synchronize seamlessly. Cloud storage and a remote database administrator are some tools that can be used by organizations to interlink their data hence improving how data can be analyzed, secured, or accessed to support organizational functions.

Outlining the organizational goals: An organization should define its goals to know the data that it needs in order to thrive. Failure to do so may result in the gathering of large pools of data that are irrelevant to the organizational needs. Penn and Pennix (2017) argue

that no company or business entity intends to collect irrelevant data that will not have any impact on business performance but the vice versa is true.

Lastly, implementing data protection measures: According to Woerner and Wixom (2015), data management strategies for an organization should come along with appropriate measures to ensure the security of this data. Data can be secured by implementing legal framework, firewall security, malware scanning, and spam filtering. Security measures of organizational data should not be taken for granted at any given time because this can either propel a venture or bring it down.

2.3.11 Summary of the Linkages between Theories to Research Questions

Table 2.1: Summary of the Linkages among research Gap, Research Question and Theories Underpinning the Study

Research Gap	Research Question	How the Gap is Addressed	Underpinning Theory
Cui, Yu and Yan, 2016), (Kitchin and McArdle, 2016), Penn and Pennix (2017), Betz and Hall (2015), Toshniwal, Dastidar and Nath (2015) reviewed various processes and stages in big data management from creation to final disposal, and points out some of the stages that are often overlooked by data creators and users hence affecting the management of this data. The researchers recommended a detailed study on examining	How does Ketraco capture, appraise, describe, preserve, and make accessible for Re-use and sharing of its Big Data?	To analyse the current state of Big Data Management at Ketraco;	The Data Centre Curation (DCC) Lifecycle Model

various overlooked stages in managing data across its lifecycle			
(Gantz and Reinsel, 2012) explains how many regulations in different countries focuses on data privacy other than managing data across its life cycle. Same case was noted with Wanekeya (2023) who had similar sentiments on the data protection act of Kenya 2019. Equally, Woerner and Wixomn (2015) recommended that further investigation is needed on policies that are applicable to distinct stages of managing data from creation to ultimate disposal. and procedures for the management of big data within public institutions was vital to improve its overall management across its life cycle.	What policies and regulations guide the management of data received or created by Ketraco?	Examine the existing infrastructure for Big Data Management at Ketraco	The Data Centre Curation (DCC) Lifecycle Model
Poleto et al., (2015); Dezi et al (2018) and Wang Kung and Byrd (2018) suggested further research should be done to identify skill gaps and training for data users, data personnel and decision makers; for effective data management;	What skills, knowledge and training are needed to improve Big Data Management at Ketraco?		Community Capability Model

Zhang et al (2018), and (Dastidar and Nath, 2015) and Cui, Yu and Yan, 2016) stated the technical challenges associated with ICT infrastructure that hinder their application in managing data. Bowman (2017) recommended the need for further studies in understanding the emerging trends in ICT application for managing data.	What is the level of ICT Infrastructure preparedness in supporting Big Data Management at Ketraco?	Community Capability Model
Cricelli and Grimaldi (2010), and (McAfee et al., 2012). et al (2012) recommended that further investigation is needed to an art the unexploited opportunities and challenges that comes when companies work together to amply their success BDM	How do collaborative partnerships influence Big Data Management at Ketraco?	Establish the influence of collaborative partnerships on Big Data Management Community capability model

Table 2.2 above provides a summary of linkages among research gaps, research questions and models underpinning the study. The literature reviewed outlined the various gaps that existed in big data management and recommended effective strategies for enhancing the management of Big Data in organizations. The gaps were addressed in this study through the theoretical framework as well as the research questions posed and addressed by the respondents.

2.4 Chapter Summary

This chapter has reviewed the previous literature that has addressed big data management. Various authors describe big data as information presented in large volumes, the speed or

velocity at which it is created, and the scope or scope of data points being covered. Government agencies, corporations, and other agencies use big data management strategies to enable them to contain the growing pool of data so that they can derive valuable information. Big data solutions can help organizations make informed decisions and optimize business processes. As opposed to the companies that were formed recently in the digital environment, the literature has indicated that traditional companies face challenges in their efforts to integrate big data management into their existing organizational processes and challenges. For instance, issues of information security, too much data making analysis difficult, and lack of skilled employees to manage data assets, among other challenges. To address this challenge, the paper has outlined various measures that can improve big data management in an organization. In an organizational setting, processes, people, technologies, and organizational support, are key components that can facilitate big data management.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

Research methodology refers to the process by which a researcher carries out an empirical study by applying specific procedures and data collection instruments (Creswell, 2014). The methodology section is a key segment of any research assignment. It provides the methods the researcher utilizes in collecting data that adequately answer the research questions and meet the objectives formulated in Chapter One. The purpose of this study was to assess Big Data Management at the Kenya Electricity Transmission Company (KETRACO), with the view of proposing interventions to improve the management, sharing, and reuse of big data to facilitate proper planning, improved organizational processes, and better decision making. The key areas that were explored in this study were; the status of Big Data management at Ketraco, policies and legal framework for Big Data Management, Knowledge, skill, and training of data managers, the technical infrastructure for Big Data Management, collaborative partnerships in Big Data Management, and the contribution of Big Data in organizational functions and processes. The researcher chose a methodology with the following thematic sections that ensured all the key elements of the research problem were addressed.

3.2 Research Approach

Cresswell (2014) asserted the importance of outlining the value of the research approach as an effective strategy to improve the validity of a research study. Further, Bernard (2017) defined the research approach as a procedure and plan that involves steps of general assumptions to detailed data collection methods, analysis, and interpretation.

Cresswell (2014) points out that it is usually possible to characterize a research approach as quantitative, qualitative, or involving both quantitative and qualitative methods in which case it is commonly referred to as mixed methods research approach.

3.2.1 Qualitative Research

A qualitative approach is a research technique that focuses on understanding and discovering the perspectives, experiences, and thoughts of respondents, implying that qualitative research explores purpose, meaning, or reality (Brannen, 2017). This approach is used to gain an understanding of underlying opinions, reasons, and motivations (Creswell, 2014; Brannen, 2017). The approach is used to provide answers to questions about meaning, perspective, and experience, usually from the viewpoint of the respondent. These data are often not amenable to measuring or counting (Creswell, 2014). Qualitative research approaches include small discussion groups for studying attitudes, concepts, and beliefs of normative behaviour; semi-structured interviews to find out views on a particular topic; in-depth interviews to understand a condition, experience, or event from a personal perspective; and 'analysis of texts and documents', such as government reports, media articles, websites or diaries, to learn about distributed or private knowledge (Brannen, 2017). Qualitative methods are usually described as allowing a detailed exploration of a topic of interest in which information is collected by a researcher through interviews, and case studies, among others.

3.2.2 Quantitative Research

A quantitative approach is used to quantify opinions, behaviours, attitudes, and other defined variables, and generalize results from a bigger population sample

(Stockemer, 2018). It is suitable when factual data is required to answer various research questions (Creswell, 2014). Quantitative techniques can reveal, for instance, what percentage of the study population supports a certain concept, their distribution by age, residential area, and marital status, among others. Therefore, the quantitative approach uses measurable data to uncover patterns and formulate facts in a study. Integral to the quantitative method is the prospect that the researcher will put aside his or her perceptions, biases, and experiences to ensure objectivity in the process of conducting the study (Stockemer, 2018). Key features of quantitative methods are the use of instruments, such as surveys and questionnaires, which have been structured to produce computable data.

3.2.3 Mixed Methods Research

A mixed approach is an emergent methodology that advances the systematic mixing or integration of qualitative and quantitative data within a sustained program of inquiry (Quinlan et al., 2019). The fundamental premise of this approach is that such a mix allows a more synergistic and complete utilization of data compared to doing separate qualitative and quantitative data collection and analysis. The study was underpinned by mixed methods approach combining both qualitative and quantitative research approaches in a single study. The justification for its adoption is to legitimize the use of different techniques in answering research questions, rather than constraining or restricting researchers (Creswell, 2014). It is a creative and expansive form of research, not a limiting method of research.

Based on the mixed method approach adopted, the study employed a concurrent mixed methods to be able to collect both the quantifiable and non-quantifiable data. In this regard, certain procedures required data to be quantified, while in other incidences, the research collected the views and opinions of respondents making the mixed approach more appropriate. In a concurrent mixed methods approach, both quantitative and qualitative data were collected and analyzed simultaneously to enhance reliability and validity. Therefore, a concurrent mixed methods approach was used in this study to ensure the collection of comprehensive and accurate data that addressed the research problem.

3.3 Research Design

Research design refers to the researcher's overall approach selected to integrate different elements of the study logically and coherently to effectively help in answering the research questions (Allen, 2017). Brannen (2017) argued that when the purpose of a study ends up being an accurate analysis of a phenomenon or an association of different variables, then the preferred design should capitalize on the reliability and minimize the incidences of biases in data to be collected and analysed. This study aimed to assess the Big Data Management at the Kenya Electricity Transmission Company with the key focus on proposing interventions to enhance the management, access, use, and sharing of this data to enhance proper planning, improved organizational processes, and better decisions. The study employed survey design within a singular exploratory case study to get an in-depth understanding of Big Data Management at Ketraco.

3.3.1 Survey Design

A survey design refers to a structured methodology used in collecting data from a specified group employing a survey instrument (Allen, 2017). It is the common method

applied in descriptive research to effectively collect standardized information. This method facilitates observation and capturing information naturally as it occurs, hence ideal for recording real world opinions, behaviours, perceptions and demographic information. Marsh (2025) argues that provided the sampled population is adequately identified and well represented, the survey design can yield high generalizable results. The survey was administered to the sampled respondents selected from all the departments at Ketraco. This method was preferred as Brannen (2017) stated that it allows data collection in a systematic way, relatively inexpensive and could be distributed to the respondents quickly. Equally, the ability to come up with same set of structured questions on a survey and then distributed to different respondents helps minimize bias in how questions are answered or asked, thus the method was preferred to ensure reliability and consistency in data collection.

3.3.2 Case Study Design

The study also applied the case study approach, which refers to an in-depth study of a singular unit with the objective of making generalization across multiple sets of units (Stockemer, 2018). The case study allows the researcher to examine the data within a specific context diligently. According to Hancock, Algozzine and Lim (2021), a case study approach examines a contemporary phenomenon within its real-life context; and when the boundaries between phenomenon and context are not well-defined, and in which multiple sources of evidence are used. Case study enabled concentration on a single subject, which is Big data Management at Ketraco. By using Ketraco as a singular case, the researcher was able to obtain in-depth information from the respondents.

3.4 Population of the Study

Population from a broad term of view refers to the collection of various elements about which inferences and conclusions are made (Bernard, 2017). In a research study, the target population refers to the complete set of objects, cases, or individuals with some similar characteristics from which the researcher intends to generalize the findings of the study (Quinlan et al., 2019). In this regard, the population of this study comprises Ketraco employees only. Since it is not possible to include all the available population in the study, it is important to define an accessible population. The target population is a component or subset of the population targeted in the study and this subset reflects specific characteristics in reference to gender or age and those who are accessible for the research (Allen, 2017). Defining a population for any study is very important since an effective statistical exercise is based on established population definitions. Brannen (2017) argues that a study should concentrate on a specific population that can provide data that can adequately address the research problem. In this study, the target population was 563 members of staff drawn from every department who in one way or the other interacted with data while performing their duties. It included the senior management team, system analysts, and information users. The departments selected interacted with data regularly while performing their functions. Therefore, their inclusion in the study enabled the researcher to uncover the various aspects of big data at Ketraco. Table 3.1 outlines the target population.

Table 3 1: Target Population

Department/Category	Target Population
Senior Management	7
Systems Analysts	5
Finance	36
Technical	283
Supply Chain	15
Corporate Planning and Development	31
Human Resource and Admin	129
Corporate Communications	14
Legal & Company Secretary	18
ICT Offices	23
Executive/MD's Office	2
Total	563

3.5 Sampling Methods

Malterud, Siersma, and Guassora (2016) define the sampling method as the strategy of singling out a particular segment of the population which is a representation of the entire population from which the researcher intends to conduct the study. The sample is the identified elements selected for participation in the study, and sampling is the process of identifying and selecting a group of subjects with which to carry out the research. Bernard (2017) stated that sampling methods can broadly be classified into two categories: nonprobability sampling and probability sampling methods. With the non-probability method, units are selected at the discretion of the researcher, and the units derive their control from the judgment of the researcher (Allen, 2017). In probability sampling, the units are not identified at the discretion of the researcher but by using certain means of

procedures. This ensures that every unit has a fixed probability of ensuring it is included in the study (Stockemer, 2018). Probability sampling methods was used in the study as explained in the section that follows.

Non-probability sampling methods include such techniques as purposive sampling, convenient sampling, and snowball sampling to name but a few (Allen, 2017). On the other hand, probability sampling includes techniques such as random sampling, stratified random sampling, cluster sampling, and systematic sampling (Stockemer, 2018). In random sampling, a study sample is selected in a way that every individual has an equal chance of being included in the sample. Stratified sampling is a technique where the target population is portioned into different strata or subpopulations (Kumar, 2019).

In this case, the population was stratified into 11 strata (Senior Management, Systems Analysts, Finance, Technical, Supply Chain, Corporate Planning and Development, Human Resources and Administration, Corporate Communications, Legal & Company Secretary, ICT Offices, and Executive/MD's Office). The researcher applied stratified sampling since he intended to ensure that every person in the target population had an equal opportunity to participate in the study.

The census method was applied to select respondents from the Executive Office of the Managing Director, senior managers, and system analysts. Etikan, Musa, and Alkassim, (2016) posits that in census sampling method, the researcher completely examines each unit and collects important data for intensive study. The census method is used when the research intends to collect data from the entire population. Every unit/respondent in the population is studied before the researcher can draw any conclusions. The census method

is used in a situation where the categories of respondents that the researcher wants to engage will provide vital data that addresses the research problem (Kumar, 2019). For instance, system analysts oversee the information management systems, and they are largely involved in handling data that flows through these systems. The MD and senior managers were targeted in this study to address policy and regulatory issues that the study intended to investigate. They were also in better positions to address the question of collaborative partnership between Ketraco and other players within the energy sector, and this is one of the key themes the study aimed to investigate.

A simple random sampling method was used to select samples from the other remaining strata of ICT, Corporate communications, HR, finance, supply chain, and technical departments. The researcher first divided the target population into different strata and then drew simple random samples in each stratum (department) giving a number based on the proportionality of the departmental representation in the sample size. The study also intended to have a representation of all the departments in the study hence making simple random sampling a suitable technique to ensure this is achieved. This approach guarantees representativeness as claimed by Malterud, Siersma, and Guassora (2016), and it was also used to provide adequate data for analyzing several sub-populations hence increasing the statistical efficiency of the samples.

3.5.1 Sample Size

The sample of any given study can have a profound effect on the findings of a study. According to Kumar (2019), a representative sample of the population should be chosen carefully to ascertain accurate data. A sample was selected appropriately from the study

population outlined above. The study drew a sample from all the departments of the organization, and this aimed to ensure the study had an equal representation.

A census sampling technique was applied to select 14 key informants comprising 2 personnel from the Executive Office of the CEO, 7 senior management team, and 5 system analysts. In addition, the MaCorr Research solution calculator (Malhotra, Nunan, & Birks, 2020) was applied to the remaining target population of 549 to attain a sample size of 63 respondents from the 8 departments. The Macorr Research solution calculator is an online tool that helps identify an appropriate number of respondents required for reliable study results (Malhotra, Nunan and Birks, 2020). The sample size outputted by the calculator is determined by the confidence level of 95% and a confidence interval of 11.6% and the population size.

According to Malhotra, Nunan and Birks (2020), the confidence level is expressed in percentage and represents how often the same population will pick the same answer if the study was conducted repeatedly. For instance, 95% confidence level implies that if the same survey is conducted 100 times, 95 of the 100 surveys will produce the same results. Therefore, the study picked 95% as the confidence level since The Macorr Research solution calculator caps the confidence level as either 90%, 95% or 100%.

The confidence interval refers to a statistical measure applied to indicate a range or an array of estimates from which a particular unknown statistical value is likely to fall. The confidence interval is written as $\pm$ (a range) or (some value), and it can be a percentage or the actual value or range of values (Malhotra, Nunan and Birks, 2020). The confidence level of 95% indicated above always gives a corresponding z-score of 1.96, which when

calculated against the $\pm$ (error range and deviation measured against the total population) outputs the confidence interval of 11.6%. An extract of the tabulation of the online calculator was captured as shown in fig 3.1 below;

The image shows a web-based calculator titled "Determine Sample Size". It has three input fields on the left: "Confidence Level:" with a dropdown menu set to "95%", "Confidence Interval:" with a text box containing "11.6" followed by a "(%)" label, and "Population:" with a text box containing "549". To the right of these are three buttons, each with a question mark icon. Below the input fields are two buttons: "Calculate" and "Clear". At the bottom, there is a label "Sample Size:" followed by a text box containing "63" and a question mark button.

Figure 3. 1: Macorr Research Calculator

Source: Macorr Research website Calculator (2019)

Figure 3.1 above indicates that a sample of 63 respondents was achieved using the Macorr Research solution calculator of October 2019 calculated at a confidence level of 95% and a confidence interval of 11.6%. The researcher preferred this formula since it is more recent and provides accurate data (Malhotra, Nunan, & Birks, 2020).

Thereafter, the 63 respondents were distributed to eight departments using the proportionate technique. In this technique, the sample size of each stratum was proportionated to the population size of the stratum and then viewed against the entire population. Therefore, to get the sample size in every stratum, the researcher applied the formula below;

$$\text{Sample size in every stratum/dept.} = \frac{\text{population of every strata}}{\text{Total population size}} \times \text{the total sample size.}$$

For example, to get a sample size from the finance dept.

$$\text{Size} = \frac{36}{549} \times 63$$

$$= 4$$

After attaining the sample size in every department, simple random sampling was used to select respondents from each of the 8 departments. Table 3.2 below represents the sample size of the study.

Table 3.2: Sample Size for Respondents

Department/Category	Target Population	Sample Population	Percentage (%)
Interview			
Senior Management	7	7	9.09
Systems Analysts	5	5	6.49
Executive/MD's Office	2	2	2.6
Subtotal	14	14	
Questionnaires			
Finance	36	4	5.2
Technical	283	31	40
Supply Chain	15	2	2.6
Corporate Planning and Development	31	4	5.2
Human Resource and Admin	129	14	18
Communications	14	2	2.6
Legal & Company Secretary	18	3	3.9
ICT	23	3	3.9
Executive/MD's Office	2	2	2.6
Total	549	63	100

Table 3.2 above depicts the target population and the sample size of the respondents. The researcher arrived at the sample size of 14 and 63 respondents respectively (Kumar, 2019; Malhotra, Nunan, & Birks, 2020).

In conclusion, the researcher used the census method to reach out to 14 interviewed respondents who engage in big data formulation and implementation strategies, and they included the MD's office, the system analysts, and the senior management team. The sample size of the remaining 63 who were administered with questionnaires were from the eight (8) departments/stratums. The 63 number was achieved using The MaCorr Research solution calculator and then selected using simple random sampling, and this helped the researcher have a representation of all departments in the study.

3. 6 Data Collection Methods

The proposed study relied largely on primary data which, according to Roberts (2015), necessitates the process of data collection from the real world; it also involves the researcher going to the field to collect data. For the purpose of this study, questionnaire and interview methods were used as data collection methods.

3.6.1 Interview Method

Hussein (2022) defines a research interview "as a purposeful conversation between two or more people requiring the interviewer to establish rapport, to ask concise and unambiguous questions to which the interviewee is willing to respond and to listen attentively". Essentially there are three types of research interviews: structured, unstructured, and semi-structured interviews. Structured interviews are questionnaires administered verbally, and they contain already predetermined questions, with little or no

variations and scope to answer the follow-up questions. Unstructured interviews on the other hand are interviews that do not necessarily preconceive theories or ideas and are administered with little organization or structure. However, Roberts (2015) defined semi-structured interviews as containing various key questions that enable the researcher to define the subject being investigated and allow the respondents to diverge so that they explore the idea in detail.

The researcher utilized semi-structured questions to define the different aspects of the problem being investigated. The researcher conducted face-to-face interviews to staff in the Executive Office of the MD, the systems analyst, and the senior management team of Ketraco. This type of interview enabled the researcher to collect opinions, facts, and information on respondents' general understanding of big data regarding the research problem. In addition, it was in the form of in-depth interviews, which Cyr (2016) defined as a qualitative technique used in carrying out intensive one-on-one engagement with the respondents. This interview provided freedom for both the respondent and the interviewer to explore and discover additional points and where possible change direction. Since some of the respondents may not have come across some of the key terms such as big data, this form of interview was more of a guided approach between the interviewer and the interviewee enabling the respondents to flush out their views and understandings of the research questions.

3.6.2 Self-Administered Questionnaires

The researcher utilized questionnaires that had open and close ended questions to collect data from respondents who were information users drawn from every department. A

questionnaire is a research tool comprising a set of questions to collect data from respondents (Roberts, 2015). They are preferred and considered more suitable because despite the resources and time constraints, they can still be administered to all respondents in a short time, and Krosnick (2018) stated that this saves time. In addition, questionnaires allow the collection of a wide range of data hence improving the accuracy of the results. This method allowed each respondent to get the same set of questions in a similar manner. Questionnaires also yield more comparable data than information collected in verbal interviews (Bolarinwa, 2015). The researcher administered questionnaires to respondents who were employees of Ketraco drawn from systems analysts, finance department, technical department, supply chain department, corporate planning and development department, human resource and administration department, corporate communications department, legal and company secretary department, ICT department, senior management, and the executive office.

In this study, the interview schedules and questionnaires were the main tools for collecting data. For these research instruments to provide accurate and reliable data, they were checked for reliability and validity, and this has been indicated in section 3.8.

3.6.3 Document Analysis

Document analysis is a type of qualitative research in which the researcher reviews and interprets documents to give meaning to an assessment topic (Roberts, 2015). Before the actual date of analysis takes place, the researcher needs to establish a detailed planning process to ensure reliable results. The issues to consider include advanced identification

of sources to explore, developing appropriate research skills, and addressing ethical issues such as confidential documents.

The advantage of the document analysis method is that documents are commonplace and exist in various formats making them very accessible and reliable data sources (Roberts, 2015). In addition, obtaining and analyzing documents is often time-efficient and more cost-efficient and time saving than when conducting their own research. The study used document analysis methods to complement interviews and self-administered questionnaires. The documents that were reviewed included organizational policies, departmental procedures manuals, 2017-2019 Annual reports, Strategic plan 2018-2013, and project completed reports.

3.7 Validity and Reliability of Data Collection Instruments

The data collection instruments are essential in determining the kind of data collected to address a problem under investigation. The main objective of using the most relevant instruments in a research study is to collect relevant data in the most valid and reliable manner. Therefore, the accuracy and consistency of instruments form an important aspect of a research methodology, and this accuracy/consistency is known as validity and reliability.

3.7.1 Validity of Data Collection Instruments

Validity is the degree to which an idea, concept, measurement, or conclusion is well-reached and reflected correctly in the real world (Leung, 2015). From a research perspective, Flick (2014) defined validity as the accuracy of an item being assessed. It represents the extent to which the scale measures what it is intended to measure. Testing

the validity of the research instruments is key to the success of any study. The questions on a questionnaire can be said to be valid if they measure what they are intended to. The researcher first did a face validity which is a test that examines if a subject or test can measure what it purports to measure. The researcher made prior discussions with a small sample from Geothermal Development company, a sister Agency in the Energy sector. The selected sample had some general understanding of the topic being investigated. As a result, they expressed their views on the first draft of the questionnaire. This provided the researcher with opportunities to make modifications and adjustments to the questionnaire elements so that the items could adequately measure the issues they were supposed to measure. Peer debriefing was also undertaken to enhance the validity of the research instruments. This involved preparing and submitting the research instruments to the supervisors and other specialists in the field to check their validity and examine the relevance and suitability of the content and wording proposed to be used in the actual questionnaires.

3.7.2 Reliability of Data Collection Instruments

Bolarinwa (2015) defined reliability as the degree to which data collection instruments can provide similar results consistently. Reliability is the degree to which the inferences are made based on scores from meaningful, useful, and appropriate instruments for conducting the study. It is the judgment of the suitability of measure for decision or certain inferences on which the result from the scores is generated (Roberts, 2015). It simply means the measure of the extent to which a research instrument provides consistent data outcomes after repeated trials where the test-retest approach is used in a study. It implies that the research instruments yield similar results in repeated trials. If the

results of research can be regenerated under the same methodology, then the instruments can be described as reliable. Francis et al. (2017) refer to a more reliable instrument as a stable instrument implying that it is an instrument that generates results that are similar or closely related. On the other hand, the less reliable instruments are often referred to as unstable instruments, meaning that they provide varying results in different trials.

The reliability of the research instruments was enhanced through a pilot study, a process that was undertaken to protect the tools and methods of data collection. The main tool in this research was the research questionnaire. The researcher examined the research questions on the questionnaire to confirm their accuracy so that they can facilitate the collection of adequate and relevant data that addresses the research problem. The content validity of the instrument was enhanced by the researcher discussing with the supervisors and specialists and getting their input on enhancing the suitability of content used in the research questionnaires.

3.7.2.1 Conducting the Pilot Study

The pilot test is a procedure through which the researcher sends structured questions to a small number of the sample population to ascertain the accuracy of the information and make necessary modifications accordingly (Bolarinwa, 2015). As indicated above, the researcher conducted a pre-test on the questionnaires before administering them to respondents. According to Leung (2015), a pilot study is necessary since it can provide an advance warning on areas and elements that can cause the research project to fail especially in incidences where the proposed instruments for data collection may be too complicated. Since the researcher utilized questionnaires, the pilot procedure was vital to

enhance the internal validity of the questionnaire. The pilot study was also done to pre-test the tools and methods of data collection. A total of thirteen (13) respondents from Geothermal Development Company, a sister Agency were randomly selected to participate in the pilot process. The following are the activities that were undertaken during the piloting process.

- The researcher administered questionnaires to pilot the ten (10) selected respondents and interviewed three (3) senior managers exactly in the same way they would be administered during the actual study.
- The researcher requested the subjects to provide their feedback and views to identify ambiguities and irrelevant or difficult questions in the questionnaire.
- The researcher recorded the duration subjects took to complete the questionnaire and determine whether it was realistic.
- The researcher discarded any difficult, unnecessary, and ambiguous questions identified during the piloting process.
- The researcher confirmed that all questions in the questionnaire were answered as anticipated.
- Leung (2015) stated that the preparation of the research instruments prior to the actual study is vital in enhancing the outcome of the study. Therefore, the researcher re-scaled and re-worded any questions that would not have been answered as expected.

3.8 Data Analysis and Presentation

Flick (2014) defined data analysis as the process of evaluating, sorting, cleansing, modelling, and transforming data with the objective of obtaining useful information to

help support decisions and inform conclusions. Therefore, data collected in this study was evaluated to obtain useful information that could help address the research problem.

The proposed study collected both qualitative and quantitative data from interviews and questionnaires respectively. The qualitative data was analyzed thematically and presented in narrative descriptions. This means that the researcher sought to identify themes and relationships that emanated from the data collected, and thereafter developed concepts and arrived at some generalized statements that led to drawing conclusions. On the other hand, quantitative data collected through questionnaires was analyzed using descriptive statistics which was presented in tables, graphs, and pie-charts.

3.9 Ethical Consideration

According to Roberts (2015), the researcher must observe codes of ethics during the entire process of the study. The main ethical concern was the requirement for informed consent. The researcher informed the respondents about the nature and purpose of the study before conducting the study. This allowed the participants and the researcher to agree to the terms governing the study and to confirm their willingness to participate in the research. Also, the researcher informed the respondents about their rights of data privacy and confidentiality as stipulated in the data protection act 2019. Thus, Assurance was made to the respondents that the process of exchanging information would not cause any harm. These practices ensured that the researcher observed the codes of ethics and privacy during the whole process of data collection.

Before the actual study, the researcher obtained a research permit from the National Commission for Science, Technology and Innovation (NACOSTI); this was done online through NACOSTI's corporate website. Before the commencement of the study, the researcher also sought permission from Ketraco. Subsequently, the researcher discussed in detail the nature of the study with senior managers of Ketraco to make them aware of the intended study and seek their permission to allow the collection of relevant data from their respective departments. The researcher also booked the interview dates in advance, and the same was communicated to respondents before the actual day of the study. The researcher also established a good rapport with the targeted audience and explained to them the nature and purpose of the research before the actual day of the study. In order to get responses to questions freely and honestly, the researcher made respondents aware that there is no wrong or right answer to research questions, and the study would not in any way affect their job; in addition, their responses were to be kept strictly confidential.

3.10 Chapter Summary

The focus of this chapter is on the methodology that was applied to conduct the study. The study utilized both the qualitative and quantitative methods because as the researcher was collecting facts from the respondents, some information such as the period in which the respondent had worked at Ketraco could not be quantified. The study targeted employees of Ketraco and the researcher used stratified sampling techniques to obtain a sample that represented all the departments. Similarly, purposive sampling techniques were used to hand-pick the system analysts and all seven senior management staff since they played unique roles in administrative matters that are relevant to this study. The study used questionnaires and interviews to collect data from respondents. Quantitative

data was analysed using descriptive statistics and presented in tables, graphs, and pie charts while the qualitative data was analysed thematically and presented in narrative descriptions.

CHAPTER FOUR

DATA PRESENTATION, ANALYSIS, AND INTERPRETATION

4.1 Introduction

This chapter provides a detailed analysis and interpretation of data collected from the selected respondents who are employees of Ketraco. The chapter is key for the researcher to articulate research findings, make comparisons of the current and previous studies relating to the research theme, and make conclusions emanating from the findings. Data collection was through the administration of questionnaires and interview schedules to collect both quantitative and qualitative data. Questionnaires were administered to respondents drawn from the following directorates: Finance, Technical Services, Supply Chain, Corporate Planning, Human Resources and Administration, Corporate Communications, Legal & Company Secretary, and ICT. The study interviewed the top management team drawn from the Directorates of Technical Services, Supply Chain Management, Corporate Planning, Human Resources & Administration, and Legal Services & Company Secretary. Further, the study interviewed the Senior Executive Principal Office Administrator from the Executive Office of the MD as well as system analysts from the IT department.

Data presented in this study is descriptive in nature and the analysis and interpretation undertaken were based on research questions. Similarly, data presentations have been done sequentially according to the way questions were structured on the questionnaire and the interview schedule. Data collected was presented in tables, charts, and graphical forms. On the other hand, Data analysis focused on the respondents' opinion on big data at Ketraco, and the extent to which the organization was leveraging Big Data functionalities to achieve

operational efficiency and competitiveness. The detailed analysis helps to uncover deeper insights that align with the questions and objectives of the study.

The chapter is divided into themes that include: Demographic information; Capture, appraisal, description, preservation, accessibility, re-use, and sharing of Big data; Legal, policies and guidelines that guide the management of Big Data at Ketraco; skills, knowledge and training required for capturing, describing, and using Big data; Level of ICT Infrastructure preparedness for Big Data management, Influence of collaborative partnerships on Big data management at Ketraco; Challenges, and finally Strategies for effective Big data management at Ketraco.

4.2. Response Rate

According to Esser and Vliegthart (2017), the response rate is the number of respondents who answered a survey divided by the sample size. Therefore, the response rates were calculated on the number of Ketraco employees who completed the questionnaire divided by those who were eligible to participate in the study. Table 4.1 illustrates the response rate from a sample size of 63 respondents administered with questionnaires and 14 interviewees.

Table 4. 1: Response Rate

Respondents (Strata)	Sample Size	Response Rate	Response Rate (%)
Interviewed			
Executive	2	1	50.00
Senior Management	7	5	71.40
System Analyst	5	4	80.00
Total	14	10	71.43%
Questionnaires			
Finance	4	4	100.00
Technical Services	31	29	93.55
Supply Chain	2	2	100.00
Corporate Planning	4	2	50.00
Human Resource & Admin	14	13	92.86
Corporate Communications	2	2	100.00
Legal & Company Sec	3	3	100.00
ICT	3	3	100.00
Total	63	58	92.06 %
Total	77	68	88.31 %

Questionnaires

Source: Field data (2023)

Table 4.1 illustrates both the interviews and questionnaires administered to different categories of respondents. On the interviews, the researcher conducted ten (10) interviews from the sample size of fourteen (14) and this comprised four (4) System analysts, one (1)

Senior Executive Principal Office Administrator from the Executive Office of the MD, and five (5) Senior management team.

On the other hand, the researcher administered questionnaires to four (4) respondents from the Finance directorate, Twenty-nine (29) respondents from the Technical Services, two (2) from Supply chain, two (2) from corporate planning, thirteen (13) from Human Resource & Administration, two (2) from corporate communication, three (3) from Legal Services and Company Secretary, and three (3) from ICT.

Cumulatively, questionnaires were received from 58 respondents out of the sample of 63 respondents, representing a response rate of 92.06%. Equally, the ten (10) face-to-face interviews conducted out of a sample size of 14 represented a response rate of 71.43%. According to Hathcoat, and Meixner (2017), a response rate higher than 70% is considered adequate for presentation and analysis in social studies because it presents a minimal chance of bias in responses. This is consistent with Leung (2015) who argued that a sample size of 50% is considered sufficient to permit generalization of the results to the entire population. Mugenda and Mugenda (2006) also affirmed that a response rate of 50% is adequate for analysis, 60% is acceptable while a response rate that is over 70% is considered excellent. Therefore, the response rate of 88.31% for this study was considered excellent to make a generalization on examining the status of Big Data management at Ketraco.

4.3 Demographic Information of the Respondents

This section examined the respondents' general information including their gender, level of education, years of working experience, the department, and level of management at

Ketraco. The objective was to evaluate and determine whether the respondents possessed the requisite knowledge to provide reliable and adequate information in line with the objectives of the study.

4.3.1 Gender Information

Lauring and Villesèche (2019) define gender as the socially constructed behaviours, roles, and identities of male, female, and gender-diverse individuals. To echo these sentiments, Hathcoat and Meixner (2017) stated that gender influences how individuals perceive and identify with one another and how they interact and behave; this ultimately affects the distribution of resources and power in society. To ensure this, the respondents were asked to indicate their gender. Table 4.2 below shows the responses.

Table 4. 2: Distribution of Respondents by Gender (n=68)

Respondents	Frequency	Percentage	Frequency	Percentage
(Strata)				
Male	32	55.17%	5	50.00%
Female	26	48.83%	5	50.00%
Total	58	100.00%	10	100.00%

Source: Field data (2023)

Table 4.2 shows the distribution by gender for the two categories of respondents (surveys and interviews). Out of the fifty-eight (58) respondents surveyed, thirty-two 32(55.17%) were male while twenty-six 26(48.83%) were female. On the other hand, the ten respondents interviewed, both male and female corresponded by each having five 5(50.00%) representations. Thus, out of the cumulative sixty-eight (68) respondents, thirty-seven (37) were male while thirty-one (31) were female, representing 54.41% and 45.59%

respectively. An increasing number of studies indicate that diversity, including gender perspectives and gender balance, helps to enhance the social relevance and the scientific quality of research (Kumar, 2019). Therefore, it was more likely that having both males and females in this study would create more inclusive study results.

4.3.2 Level of Education

Big data management in an organization demands data creators, users, and managers to possess a wide range of skills ranging from analytical skills, and ICT skills to Big Data frameworks to effectively curate and manage large sets of unstructured data (Dezi et al., 2018). Therefore, a certain minimum level of education is necessary for one to manage data across its life cycle effectively. The findings on the distribution of respondents by level of education are presented in Table 4.3.

Table 4. 3: Distribution of Respondents by Level of Education (n=68)

	Frequency	Percentage
Certificate	5	7.35%
Diploma	13	19.12%
Bachelor's Degree	34	50.00%
Master's	14	20.59%
PhD	2	2.94%
TOTAL	68	100%

Source: Field data (2023)

On the presentation and analysis of data on the level of education shown in Table 4.3, 5(7.35%) of the participants possessed a certificate, 13(19.12%) had a diploma, 34(50%) had a bachelor's degree, 14(20.59%) had a masters, and finally 2(2.94%) had a PhD. The findings reveal that all the respondents had the requisite knowledge required to adequately

participate in the study and were therefore considered to respond to the survey questions based on the minimum educational qualifications. Consequently, by all respondents aligning with different levels of education, it can be interpreted that employees at Ketraco possess a wide range of knowledge; therefore, they can effectively implement big data strategies across the Company. This is in line with Dezi et al. (2018) who argued that the successful application of Big Data in an organization is not the sole responsibility of certain groups of professionals. Instead, it requires a community-based approach that involves collaborative responsibility for all members of the community to engage at different levels of Big Data management actively and this is in line with the Community Capability Model (2012) and Data Curation Centre lifecycle model (2008) on skills and training and aspects of collaborative responsibility.

The relevancy of growing the capability of the Ketraco community in big data was also illustrated in chapter two through the Community capability model. The model emphasizes the interoperability of various elements within a community to achieve a common goal. Therefore, it can be generalized that Ketraco employees possess varying skills ranging from basic to advanced levels, and this is key for the successful implementation of the Big Data management program. Furthermore, it can be ascertained that all the respondents possess soft skills such as document sharing, database search, document creation in MS Word, and the use of passwords in controlling access to data. Thus, Ketraco can tap on the already existing vast knowledge amongst its employees to drive the Big Data management strategy and ensure its effective implementation.

4.3.3 Years of Experience

The study sought to establish the duration of employment of the respondents at Ketraco. This was intended to establish whether the respondents had adequate working experience and acquainted themselves with the already existing Big Data activities at the Company. The findings were presented in Table 4.4.

Table 4. 4: Distribution of Respondents by Years of Experience (n=68)

	Frequency	Percentage
0-4 Years	12	18.75 %
5-8 Years	38	59.38 %
8-12 Years	10	15.63 %
Over 12 years	4	6.25 %
TOTAL	68	100%

Source: Field data (2023)

Table 4.4 summarizes the work experience as measured by the number of years that respondents have worked at Ketraco. The frequency represents the number of respondents in each category and the percentage shows the proportional representation of the total sample. The findings revealed that the majority 38(59.38%) of the participants had worked at the Company between five and eight years, while 12(18.75%) of the participants had worked for less than four years. In addition, 10(15.63%) of the participants worked between 8 and 12 years, and finally, 4(6.25%) participants worked for over 12 years. From the findings, the majority of employees had served in the organization for over five years

and thus had the requisite understanding and experience of the company's data management status. According to Wikström, et al (2018), employees' long-term experience at a workplace exposes them to different processes, challenges, and systems and this makes them highly knowledgeable and informative, especially in the execution of their functions. Therefore, experienced respondents had accumulated skills while interacting with Big Data and were in a position to provide credible information and valuable insights into gaps and improvement areas in the organization's processes and management of big data.

4.3.4 Departments of Respondents

The respondents were asked to indicate their department to establish whether the study was inclusive in capturing views from all big data operational areas in the Company. Figure 4.1 gives the breakdown of the distribution of surveyed respondents per department.

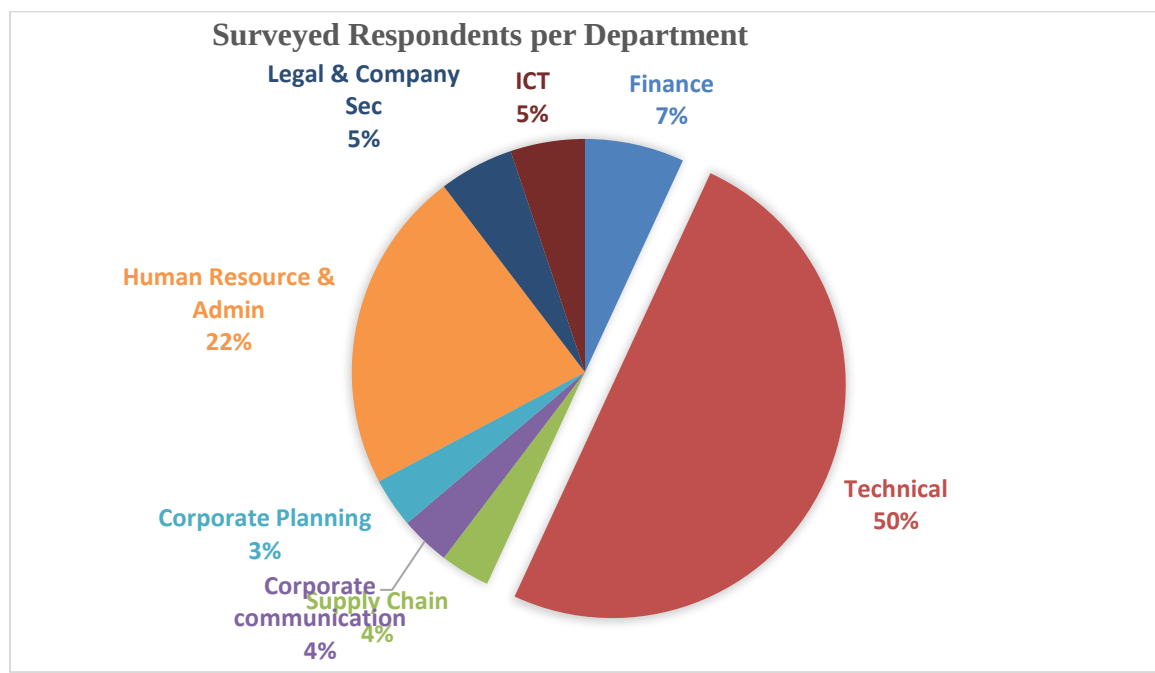


Figure 4. 1: Surveyed Respondents per Department (n=58).

Figure 4.1 revealed participation representation from each department with Finance having 4(6.90%), Technical Services 29(50.00%), and Supply Chain, Corporate Planning, and Corporate Communications each having a representation of 2(3.45%). Further, the Department of Human Resource & Administration had a representation of 13(22.41%) respondents, while legal and ICT departments were each represented by 3(5.17%) respondents in this study. From these statistics, it is more likely to get a general understanding of Big data across the company, with a possibility of collecting a wide range of views from the respondents and coming up with appropriate strategies for effective implementation of big data across the company. Consequently, the interviewees were asked to mention their departments; and the findings were tabulated in Figure 4.2

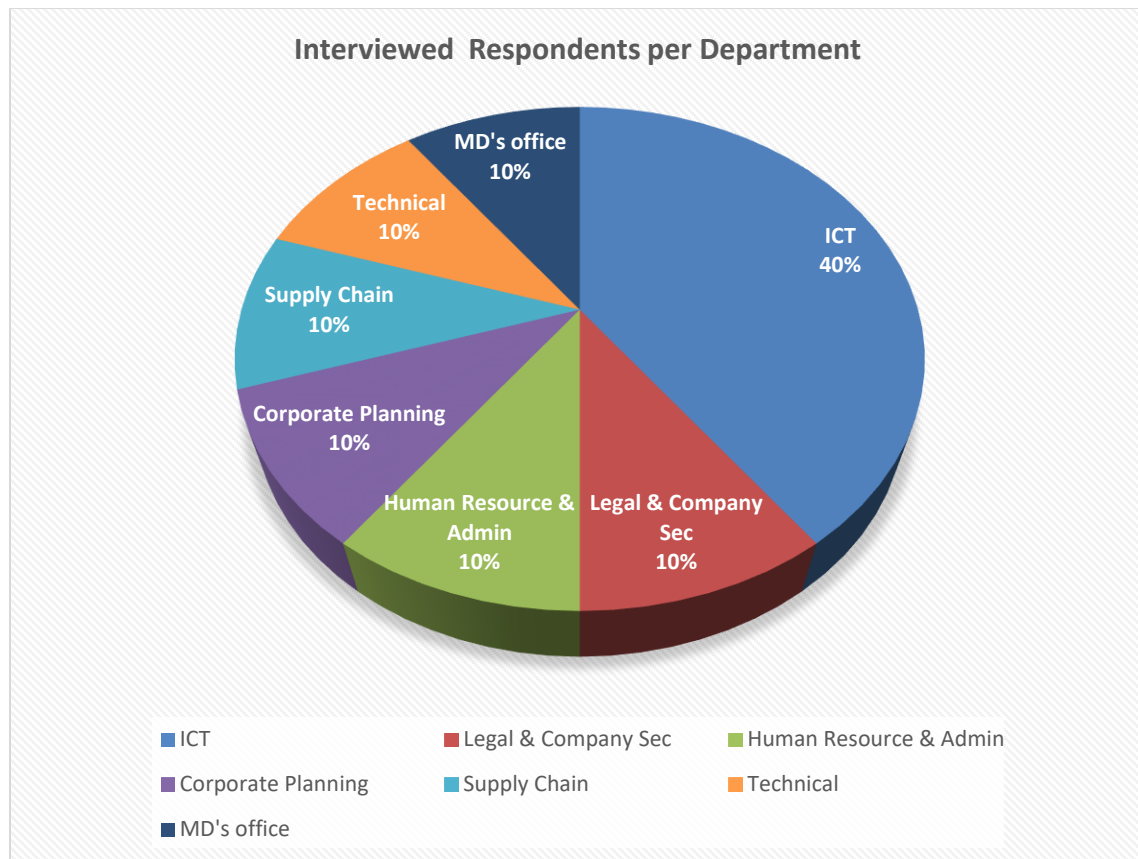


Figure 4. 2: Interviewed Respondents per Department (n=10).

On the assessment of the distribution of the Interviewees per department, the results from Figure 4.2 revealed that the Executive Office of the MD, the departments of Technical Services, Supply chain, Corporate Planning, Human Resources & Administration, Legal & Company secretary each had 1(10%) respondent in the interview. In this regard, there were high chances of collecting a wide range of Big Data activities as per the respondent's department. In addition, 4(40%) of the interviewees were from the ICT department (system analyst), and these were the system Analysts who understood Ketraco's ICT infrastructure and workflows. In essence, incorporating more System Analysts was critical for the researcher in getting a wider understanding of themes around big data and ICT processes.

4.3.5 Position of the Respondents in Level of Management

The management level in this study is measured against the contribution of each respondent's level in management, and how each level contributes to the management of data across its life cycle at Ketraco. Therefore, respondents were required to indicate their management level, and the findings were presented in Figure 4.3.

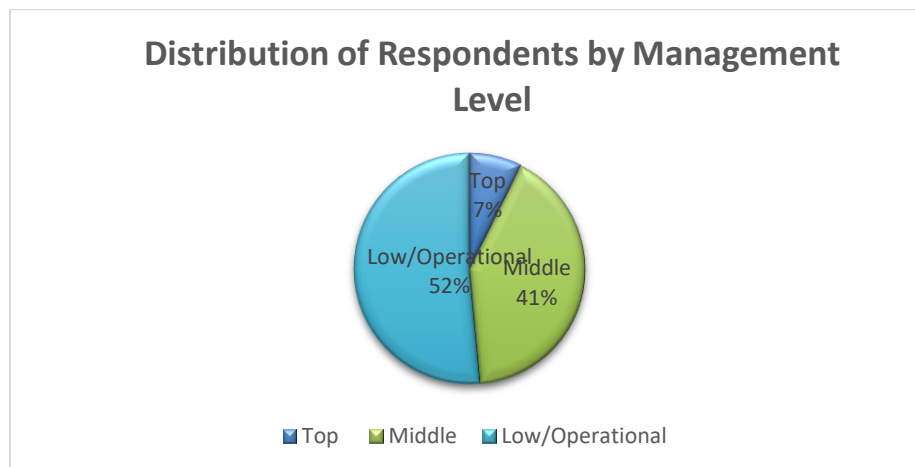


Figure 4. 3: Distribution of Respondents by Management Level (n=68).

The results in Figure 4.3 revealed that the representation of respondents in top management, middle-level management, and low-level management were 5(7.35%), 28(41.18%), and 35(51.47%) respectively. The top management team is responsible for providing strategic leadership and direction to the company (Griffin, 2021). Therefore, having 5(7.35%) respondents in this study was ideal for collecting vital information around the themes of Big Data training, budgeting, management support, partnerships, and policy formulation.

Further, 28(41.18%) respondents indicated that they were in middle-level management; according to Fiaz et al. (2017), middle-level management officers train, guide, inspire, delegate, coordinate, and communicate daily assignments to employees, targets, and any emerging tasks. Therefore, they were critical to this study in providing meaningful insights, especially on embedding the Big Data strategy into the day-to-day workflow of the Company.

Finally, 35(51.47%) respondents from the lower management level were also vital since this category according to Griffin (2021) plays a vital role in reporting, and ensuring the effective implementation of company plans, policies, and procedures. They are the hands-on staff who perform day-to-day operational tasks in an organization. Thus, lower-level management was key to providing an overview of Big Data management in the Company, the status of Big Data, and how data is managed across its lifecycle at Ketraco. Therefore, this study was represented by employees across the management levels, and this inclusion is essential for the effective development and implementation of a Big Data management

program that relies on teamwork, and equal distribution of big data-related activities across the employees in an organization.

4.4 Current State of Big Data Management at Ketraco

One of the underpinning models for this study was the Data Curation Centre (DCC) lifecycle model as illustrated in Chapter Two. The DCC model outlines coherent stages of managing data from creation through the curation cycle (Pouchard, 2016). The DCC Model highlights the stages required for an organization to create and preserve data for its use. These phases include creation, appraisal and selection, transfer, storage and preservation, and access and re-use (Higgins, 2008). The DCC model helps in data management workflows as well as in assigning various data responsibilities to the management of big data within the organization (Higgins, 2008). Thus, at Ketraco, it is important to understand how big data management activities are undertaken and mapped to support various operational units within the Company. To have a general understanding of Big Data management at Ketraco, the study, through the first objective, sought to assess the current state of Big Data Management at Ketraco. The researcher examined how Ketraco captures, appraises, describes, preserves, accesses, re-uses, and shares its data. Question 1 on Part B of the questionnaire (see Appendix 1) and question 1 on Part B of the Interview schedule (see Appendix 2) required the respondents to share the current state of big data management at Ketraco.

4.4.1 Data Capture and Format

4.4.1.1 Data Capture

Data capture is the extraction of information from various sources including electronic and paper formats and converting it into data for use by a system (Lyon et al., 2012). This is

the initial process of data management in many organizations in their digital transformation course. To understand how data is created at Ketraco, respondents were asked to state how the company creates/receives data by allowing them to select more than one option from the choices provided.

Table 4. 5: Data Capture (n=58)

Data Capture	Frequency	Percentage
Manual	45	77.00%
Scanning	37	63.79%
Legacy data imports	34	58.62%
External/incoming mail	7	12.07%

Source: Field data (2023)

The findings from the survey revealed that manual data entry, scanning, and Legacy data imports were the common techniques for data capture by 45(77.00%), 37(63.79%), and 34(58.62%) respondents respectively. External/incoming mail was another method used by employees to receive/capture data 7(12.07%). Voice recognition and barcoding were the least chosen methods for capturing data.

Further, senior managers interviewed had a general view that there existed several approaches through which the Company receives/generates data. For instance, the acting Manager of Human Resource Services [MHRS] stated:

Generally, we create/receive data in different formats and this depends on the setting. For example, when in meetings, the minutes and resolutions are manually captured, and these documents are accessed by the Records management staff for scanning and all that. But internally, we initiate our documents through memos (manual data keying)

Another senior manager, Corporate Planning [MCP] had this to say:

While in the office, lots of documents are scanned and forwarded to us for action, while others are brought to us directly (manually) through other members of staff.

From the surveys and the interviews, the study established that there were several methods of how data is created/received by Ketraco employees. Ketraco captures data using various methods such as manual entry, scanning, and legacy data imports. Much of the data capture process is also manually handled by employees. This confirms the findings by Penn and Pennix (2017) who in their publication on records management digitization observed that internal company staff are the major stakeholders (creators) of data. Similarly, Hamzah et al. (2020) pointed out that manual and digital formats are the two common methods of capturing data. Digital capture involves the use of advanced technologies such as bar codes, optical character recognition, intelligent document processing, and digital signatures (Penn and Pennix, 2017). On the other hand, manual data capture is the human-centric method of extracting data, and manually inputting information using mediums such as mice, touch displays, and keyboards. Hag, Li and Hassan (2019) in another study examined the text mining techniques used in cloud computing and big data processing. Their study observes an increase in the use of text mining for data capture and big data processing. By establishing that both manual keyboard keying and scanning were the common methods of capturing data, it was evident that Ketraco has to an extent embraced digital mechanisms in creating data, and that the Company staff were key in the process of creating data. This was attributed to the existing internal guidelines that were developed to guide the creation of internal documents such as forms and templates for creating memos and for report writing.

In addition, scanning was the commonly cited method of creating/receiving data across the company, and this was necessitated by the existing hardware such as the scanners, and large format printers that support document management solutions such as scanning in multiple formats. In their study, Roh, Heo, and Whang (2019) analyzed data collection for machine learning from the perspective of integration of big data and artificial intelligence. According to their study, data collection is a major challenge in machine learning. Therefore, they recommend the need for integration of AI and data management for purposes of data collection to enable seamless data collection and labelling of data for collection.

Further, a study done by Al-Najran, and Dahanayake (2015) on analyzing the key requirements for the development of a framework for the collection and capture of big data revealed that in the age of massive data, the ad hoc processes, or frameworks of capturing big data used in many organizations are inadequate and thus not feasible. Many organizations still use analogue data management systems such as mobile webs and weblogs to collect a bunch of general data that takes a long time to analyze and generate insights. An organization needs to create a structured approach for collecting data across various domains. The study revealed that at Ketraco, the existence of multiple approaches to data collection as well as users creating or receiving data from various sources in unstructured methods implies that there are no well-founded guidelines that support the data collection process. Therefore, there is a need to harmonize data creation at Ketraco by ensuring the adoption and implementation of a sound framework that guides the data collection or capturing process.

4.4.1.2 Data Format

A data format refers to the structure of data within a database or a file system and serves to give meaning to the information (Ghiringhelli et al., 2017). Different data formats are used for the preservation of various data types, including PDFs, audio, images, videos, and databases among others (Erevelles, Fukawa, and Swayne, 2016). The study thus sought to assess different formats in which data is captured at Ketraco, and the findings are presented in Figure 4.4.

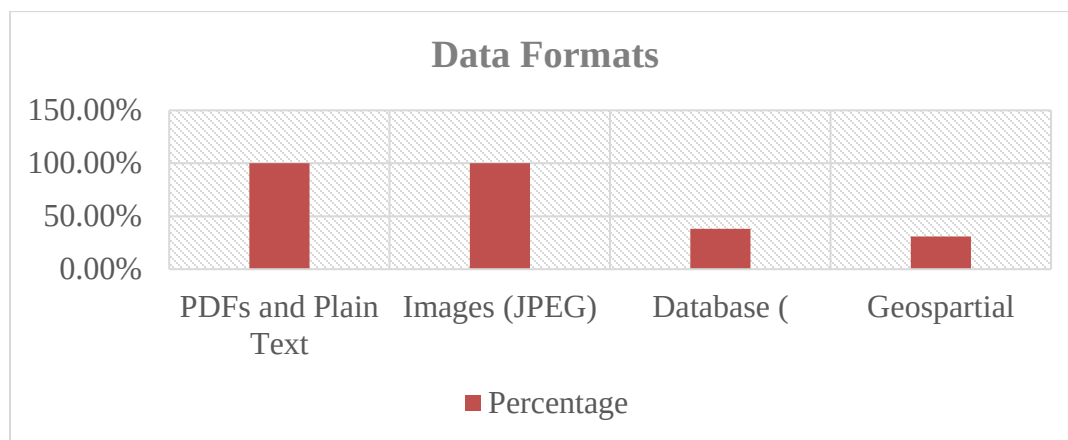


Figure 4. 4: Data Formats Exiting at Ketraco (n=58)

Figure 4.4 summarizes that all the respondents surveyed (100%) indicated the use of text documentation (PDF and plain text) and still images (JPEG) formats to capture data. Since the respondents could select more than one format, Database (XML, CSV) and Geospatial formats were chosen by 26 (38.24%) and 21(30.88% %) respondents respectively as other options for data formats. In this regard, Borgatti, Everett, and Johnson (2018) opined that different formats are designed to store specific types of information, such as CSV and XLSX. Additionally, all the interviewees indicated managing their data files in PDF, Word, Excel, and JPEG formats. However, the senior manager from the technical departments

indicated to have other formats for creating and storing data. In an interview with the acting General Manager of Technical Services [Ag. GM, TS], the interviewee stated:

For me and my team, we have a lot of Geospatial data, and we even have some software for generating this type of data.

This response indicates that at Ketraco, certain data formats largely depend on the activities of each department. For instance, the Department of Design and Construction indicated having large quantities of data generated and managed in ArchCard software, while that of survey is likely to generate a lot of Geospatial data and survey maps using ArcGIS Pro platform. Therefore, the departmental functions play a key role in understanding data and file formats existing at Ketraco. These findings indicate that Ketraco's data is captured in different formats based on the needs of every department and the available software tools to support the data capture process. The multiple formats existing at Ketraco can also be attributed to the existing enterprise resource planning software (SAP) that allows input of data in specific file formats. Therefore, for efficient workflows and continuity of business processes within the system, users at Ketraco are required to scan and upload their documents into the system using specified file formats.

A study by Belov, Tatarintsev, and Nikulchev (2021) examining the factors that influence the choice of data storage formats in the Apache Hadoop Systems established that data format is an important function of any big data processing platform. Further, the study findings indicated that various systems have different storage requirements for Big Data which makes it difficult to select an optimal big data storage format. Consequently, the existing multiple formats in data creation at Ketraco according to Koltay (2017) are appropriate for data sharing, preservation, and long-term access. However, Ketraco can

only realize these benefits if the format types are driven by the specific requirements and needs, as well as existing ICT applications that support those formats. Further, to ensure standardized data formats, policies that guide data formats are necessary to ensure uniformity and consistency within the dataset attributes to ensure effective analysis and management.

4.4.2 Data Appraisal

Data appraisal refers to the process of reviewing and classifying data before capturing and storing it (Phiri and Tough, 2018). Appraisal is essential since organizations are overwhelmed by large quantities of data. Additionally, a lot of data is created or received by organizations without having a specific purpose for that data. As a result, an appraisal is essential to help organizations assign meaning to their data sets; this helps to determine the value of data (Phiri and Tough, 2018). In appraisal, data with continuing value is identified for the long term, while data with no value may be destroyed or stored temporarily based on the existing company policy. Furthermore, Kimble and Milolidakis (2015) opined that the appraisal of data for long-term curation should be guided by well-established policies, guidelines, and legal frameworks.

The researcher sought to evaluate the status of Ketraco's Big Data appraisal. Both the surveyed and interviewed respondents were asked to state whether they appraised their data and to indicate whether Ketraco had any mechanisms for guiding these appraisals.

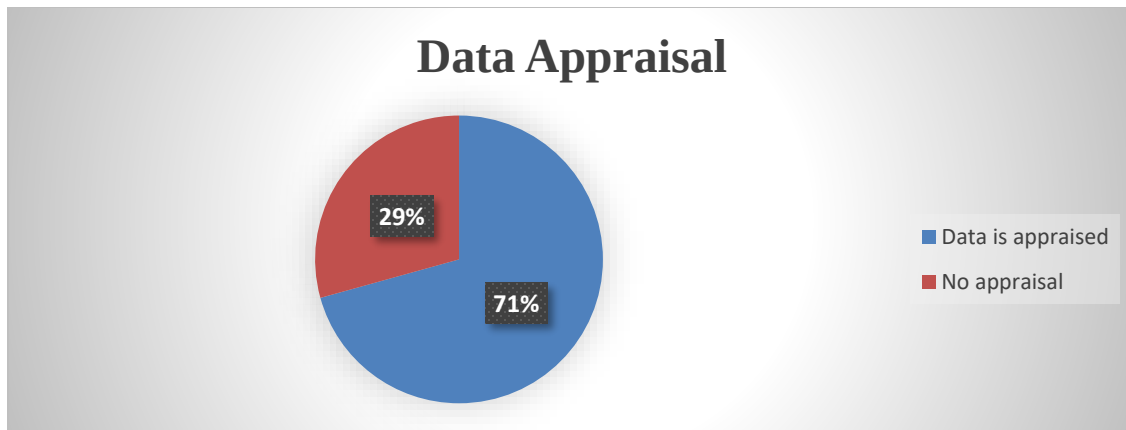


Figure 4. 5: Data Appraisal (n=58)

From Figure 4.5 most of the surveyed respondents 41(70.69%) indicated that they don't subject data to any form of appraisal, and the explanation for this was that there are no clear guidelines that support data evaluation. However, of the remaining surveyed respondents, 17(29.31%) indicated that they do appraise data received before taking subsequent actions of preserving or destroying it. Further, they stated that they are guided by their own judgment by examining data, its subject, value, and source before deciding subsequent actions of storage and preservation.

The question of appraisal was also posed to the interviews, and the responses were captured as follows.

Interviewee 1: Senior Manager Supply Chain [SMSCM]

I have never heard of data appraisal. The only appraisal I know normally comes from Human resources when we are being evaluated on our performance targets.

Interviewee 2: Manager Corporate Planning [MCP]

Appraisal? (Appears surprised), I don't think I do it since I always action or mark the document to the relevant officer for necessary action. I have never received any communication that guides the same.

Interviewee 3: Ag General Manager, Technical Services [Ag. GM, TS],

Only when free or when my laptop lags, is when this idea of freeing space crosses my mind. In these cases, I normally open the stored files, make my own judgment, and depending on the need of that document, I end up either deleting or moving the document and storing it somewhere for future reference.

Interviewee 4: Manager Human Resource Services [MHRS]

How is it supposed to be done? (The interviewee asked back with a smile on her face.)

Both the surveys and interviews seem to suggest that Ketraco has not considered appraising its existing data, as well as established any mechanisms, procedures, or guidelines that guide or direct its staff on appraising big data. The responses captured demonstrate that Ketraco does not have an established institutional methodology and practice for appraising digital data, and this also confirms findings from the study by Katuu (1999) on the status of electronic appraisal in Africa who established that many companies across Africa face challenges in Archiving digital data because of the unavailability of well-established policy frameworks for guiding appraisal and digital archiving of data.

In a study on cross-examination of the appraisal of electronic records, Katal, Wazid, and Goudar (2013) stated that the appraisal of digital texts should follow similar principles as paper-based records. Katal, Wazid, and Goudar further pointed out that the data that require appraisal and continued preservation is one documenting the essential functions of an organization and contains key information about organizational transactions. Similarly,

appraisal should involve content analysis that requires examining information and the evidential value of data to ensure that the data collected is viable and cannot be analysed to generate useful insights into the organization's operations. In addition, it requires a technical analysis that entails examination of the usability of data. The current study indicates deficiencies in the company's systems to support content appraisal and text mining for big data capture and processing. Instead, the company relies primarily on manual processes such as scanning and physical data entry. The unavailability of appraisal mechanisms at Ketraco can be attributed to the lack of proper structures that guide the appraisal process, and this leaves the company with unstructured and unusable data that consumes a lot of storage space in the Company's storage platforms. The findings suggest that the factors contributing to the current state of data appraisal Ketraco include the unavailability of a Big Data appraisal policy, a comprehensive electronic data inventory worksheet, lack of retention schedule and forms, and the absence of appraisal checklists. In addition, lack of awareness, Standard Operating Procedures (SOPs) of data management and usage, and training on appraisal procedures are additional reasons that lead to poor application of data appraisal. All the listed factors therein constitute the need for Ketraco to develop a comprehensive data appraisal policy that guides the company in data appraisal and selection of data sets deemed for long-term storage.

4.4.3 Data Description

An organization that has embraced a big data management framework should have data description guidelines that support the description of its data. Dezi et al. (2018) argued that it is easier for data users to locate, understand, share, and use data that is simply consistent and clearly defined. As a result, organizations can establish mechanisms that

ensure consistency and effectiveness in describing their data. Kitchin and McArdle (2016) in a study on the ontological attributes that organizations can have data models that outline a graphical methodology of describing data that comes in various formats revealed that organizations can establish guidelines on data naming, data definition, data dictionary, and data catalog to aid in defining and naming data sets.

Data Consistency Constraint (DCC) models are statistical models used in various data analysis tasks. These models incorporate the construct of data description, which refers to the process of summarizing and representing data in a meaningful way. DCC models utilize the construct of data description to enhance their performance and accuracy (Pouchard, 2016). By incorporating the knowledge about the data's characteristics, these models can capture the underlying patterns, relationships, and dependencies in the data more effectively. In turn, this improves the quality of analysis and predictions generated by the models. Thus, the researcher sought to understand how data description is done at Ketraco. Respondents both surveyed and interviewed were required to identify any elements such as subject, date, title, author, and Reference number they use in describing data. However, the interviews required respondents to indicate whether there were any established procedures guiding data description at Ketraco. The findings were presented in Figure 4.6.

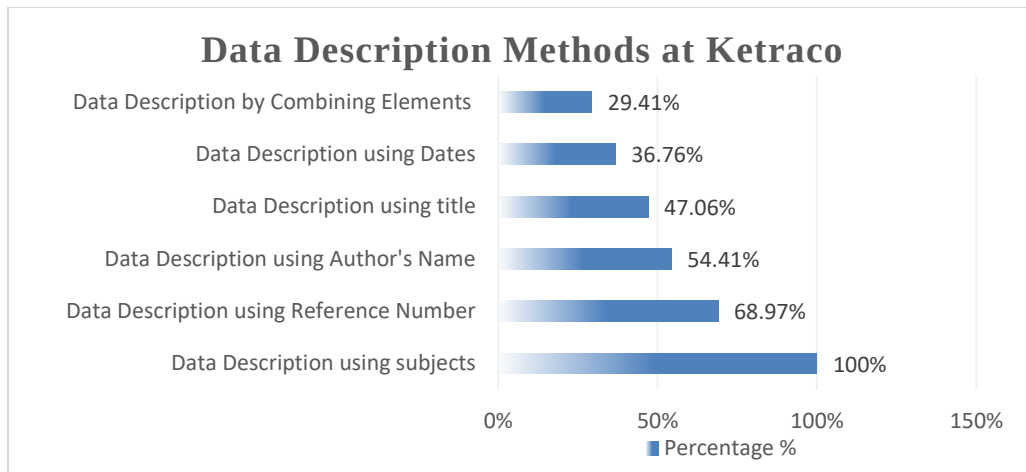


Figure 4. 6: Data Description (n=58)

All the respondents (100%) anonymously agreed that they describe data based on the subjects. Besides, 40(68.97%), 37(54.41%), 32(47.06%), and 25(36.76%) of the surveyed respondents pointed out that they describe data using document reference number, Author's name, title, and dates respectively. However, 20(29.41%) respondents indicated that they use a combination of various metadata elements such as data, title, date, and author in data description. Consequently, the responses captured from the interviews indicated that there are no standard formats and procedures for data description. For instance, the Manager of Human Resource Services {MHRS} in an interview stated:

I think I save most of my documents using the subject, and I only use a few keywords to save so long as it will help me to remember.

The findings suggest that there was no specific standardized description procedure, description policy, or standard operating procedures (SOPs) to guide the description of big data at Ketraco hence leading to variance in the description of big data. Katal, Wazid, and Goudar (2013) in a study examined the Big Data issues, challenges, and best practices and observed that a link between standardization of metadata and access to reusable data is with the use of standardized data description, the employees and management at the company

can enjoy access to reusable data using the established standard methods. The divergent approach to data description at Ketraco as presented in Figure 4.6 demonstrates that the Company does not have a standard procedure for describing its data, and this leads to inconsistency in data capture and preservation. Furthermore, inconsistencies in data description hinder other big data management practices such as data sharing, use, and re-use since it can be tiresome to locate or retrieve a data set. Therefore, for Ketraco to effectively utilize its existing data there is a need to establish tools and models that systematically describe metadata of data sets, and this will ease the proceeding stages such as data storage, preservation, access and use, and re-use.

4.4.4 Storage of Big Data

After data receipt, capture, and description, data should be safely stored to guarantee its accessibility, integrity, and reliability. Big data storage refers to the storage infrastructure designed to store, manage, and facilitate access to massive amounts of data (Dezi et al., 2018). Storage enables preservation and sorting of data in a manner that enables data to be accessed, processed, and used. According to the DCC Model, storage is necessary for long-term preservation and retention of data (Pouchard, 2016). Storage actions include assigning preservation metadata, data cleaning, and validation. Thus, the study sought to understand whether Ketraco has a storage platform, how data is stored, and establish the person responsible for day-to-day storage of Big data. The respondents were allowed to tick more than one option from the options given in the questionnaire. Concerning the existence of storage platforms at Ketraco, all the respondents (surveyed and interviewed) agreed that Ketraco has a mechanism for storing data as illustrated in Figure 4.7.

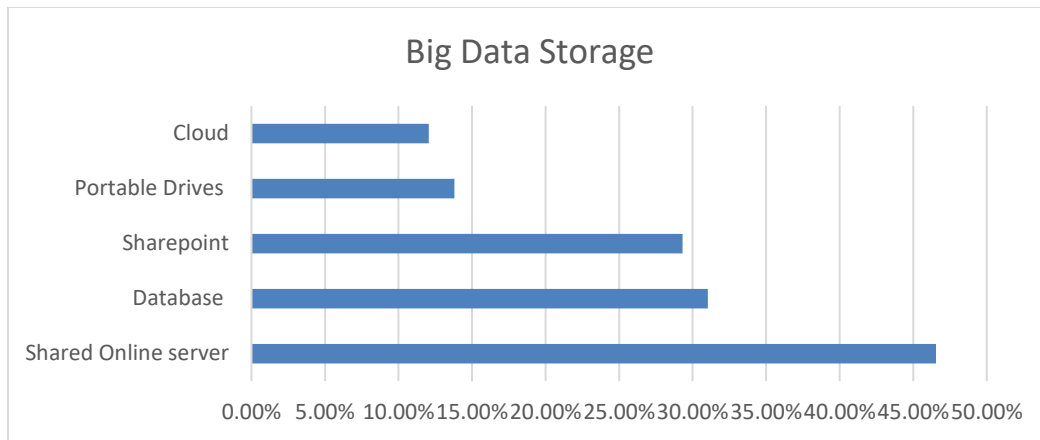


Figure 4. 7: Data Storage (n=58)

Based on findings in Figure 4.7 above, shared online servers and databases were the common methods used for big data storage 27(46.55%) and 18(31.58%) respectively. SharePoint system and external portable drivers/hard disks were also mentioned by respondents 17(29.31%) and 9 (15.52%) respectively as part of the storage point of big data. In addition, 7(12.07%) respondents also mentioned cloud computing solutions for providing cloud-based data storage. Equally, from the general answers given by the interviews, the ICT department provides storage platforms for the company data. The findings from both the surveys and interviews indicate that Ketraco has leaned significantly toward the use of shared online servers, databases, and SharePoint. However, cloud storage remains the least used method for the storage of big data in the organization. The findings thus suggest that there is the absence of a standard storage platform across the board in the organization leading to different storage frameworks.

In relation to this, Rubinfeld and Gal (2017) did a study examining the challenges in storage and access to big data in organizations today. The study highlighted technological barriers, legal and lock-in, and switching costs as part of the challenges in data storage. The slow

technological adaptation, legal barriers, and the lock-in and high switching costs may explain Ketraco's overreliance on traditional storage solutions such as Databases and shared online servers while disregarding novel solutions such as Cloud storage.

Similarly, Siddiqa, Karim and Gani (2017) examined the various big data storage solutions and established that the last decade witnessed a massive explosion of data leading to the shift from scale-up to scale-out hardware. Scale-up is the traditional data storage approach that involves adding more hard drives to manage data while scaling-out is a software-driven storage that includes clusters of software nodes that operate together, the nodes can be removed or added, allowing elements such as throughput, and bandwidth to decrease or increase as needed. Equally, Kumar et al. (2014) in a similar study on examining data management through NewSQL reverberated Siddiqa, Karim and Ganin (2017) sentiments by pointing out that the new trends in managing data have led to the move from traditional databases to relational databases. In a relational database, data is managed within tables composed of columns and rows, which are then linked to data in other tables by using unique identifying keys. Thus, the databases are designed for storing and retrieving structured data. On the other hand, a non-relational or traditional database doesn't store data in tables but in any suitable format depending on the data being stored. These databases are designed to contain unstructured data, such as videos, images, business documents, and email messages that are not easily standardized.

In this regard, Kaur and Sachdeva (2017) posited that the new storage technologies include Distributed File Systems, NoSQL Databases, NewSQL Databases, and Big Data Querying Platforms that need to be embraced by organizations generating Big Data. The new state-

of-the-art solutions to data storage as observed in the studies done by Kaur and Sachdeva (2017); Kumar et al (2014); and Siddiqa, Karim and Ganin (2017) are capable of handling volumes of data and able to search, retrieve, share, and archive data. The observations from these studies are indications that Ketraco has primarily stuck to the traditional relational databases which are incapable of addressing the five qualities (5Vs) of Big Data, namely value, velocity, variety, volume, and veracity. The findings of the study on data storage seem to imply that cloud computing and other software-based data management solutions have not been fully embraced at Ketraco, an observation that is contrary to Fang et al. (2010) publication on cloud storage as the infrastructure for cloud computing whose argument was that cloud and warehouse storage are the common types of storing big data in organizations.

Consequently, on examining the various factors that influence the choice of organizations to adopt cloud computing, Stieninger et al. (2018) surveyed cloud workers to examine their feedback on the state of adoption of cloud computing as well as the barriers and enablers of the adoption. The findings of the Stieninger et al. (2018) study revealed that while cloud computing offers benefits such as affordability, agility, and scalability, many organizations were reluctant to adopt these technologies due to issues relating to compatibility, relative cost advantage, information security, trust and the level of compatibility. This may explain the slow acceptance of cloud computing technology at Ketraco. In essence, warehouse storage involves on-site data storage, while cloud storage involves data storage in an offsite secure location. Various data storage technologies have been embraced across the world such as Snowflake, Apache Hbase, and Apache Hadoop (Siddiqa, Karim, and Gani, 2017).

Hadoop is a distributed processing platform that is based on open-source software that allows the processing and storage of large datasets across clusters of computers (Raj and D’Souza, 2019). On the other hand, Hbase is designed to handle large tables with millions of columns and billions of rows (Hussain, Sanga, and Mongia, 2019). Apart from the absence of a standard media or platform, Ketraco had also not developed the procedures for preservation and storage of Big Data as indicated by all categories of respondents. Watson (2014) stated that inconsistencies in data storage/preservation do not provide an optimum platform to support the management of both structured and unstructured data in an organization. Evidently, Ketraco leverages various storage mechanisms for its data which hinders the capability of providing input/output operations that rely on data in terms of access, use and reuse, sharing, and transformation.

4.4.5 Data Backup Format

Data backup entails the physical devices and locations used for storing data for recovery in cases of data loss or failure (Ramesh, Logeshwaran, and Aravindarajan, 2022). On the other hand, data format refers to the structure of data within a file system or database that gives information its meaning. Therefore, the researcher sought to investigate the data formats and methods used for managing data. The findings were presented in Figure 4.8.

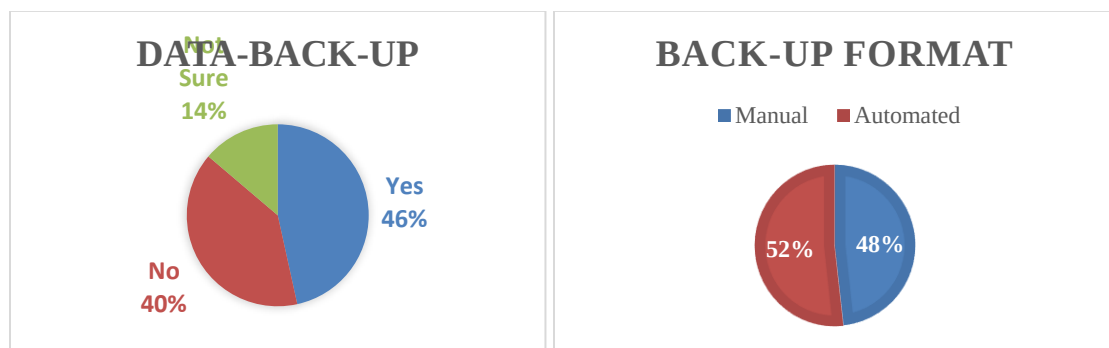


Figure 4. 8: Data Back-up and Formats (n=58)

In terms of data backup as illustrated in Figure 4.8, 27(46.00%) of respondents surveyed were aware of the backup for data at Ketraco. On the other hand, 23(40%) respondents were not aware of data backup while 8 (14.00%) were not sure of the existence of any data backup mechanism.

In relation to the backup formats, respondents indicated that Ketraco uses both manual and automated backup with a representation of 32(48.00%) and 36(52.00%) respectively. The manual backup is organized and performed regularly while the automated one is backed up using automated software to minimize human intervention. Equally, all the interviewees responded that issues to do with data backup is purely an ICT function. For instance, the Head of Supply Chain [SMSCM] had this to say in the interview.

On this one, I don't know how it is done, and whether it is done, maybe you ask the ICT team.

The findings on Data Backup and formats at Ketraco indicate that the Company to an extent has initiated mechanisms for data back up and since the function is domiciled in the ICT department, the majority of other data users besides the ICT team are not aware of whether Data backup is undertaken, and the processes involved therein.

On exploring disaster recovery as a cloud service; economic benefits and deployment challenges, Wood, et al. (2010) found out that it has become a robust IT Strategy for any organization to establish data backup and disaster recovery mechanisms to safeguard their data loss against malicious attacks, natural disasters or system failures, natural disasters. Consequently, Suguna and Suhasini (2014) outlined how many companies have transitioned from the traditional data backup strategies that involved creating copies of vital

data and preserving them on external devices or in on-premises storage systems. Currently, organizations are leveraging the scalability and power of the cloud for data backups, recovery, and protection capabilities. Perhaps this explains why many data users at Ketraco were unaware of how data backup was undertaken since it was mostly handled by the ICT department.

Further, the study examined how the decision on data preservation and Archiving is undertaken. The findings were presented in Figure 4.9.

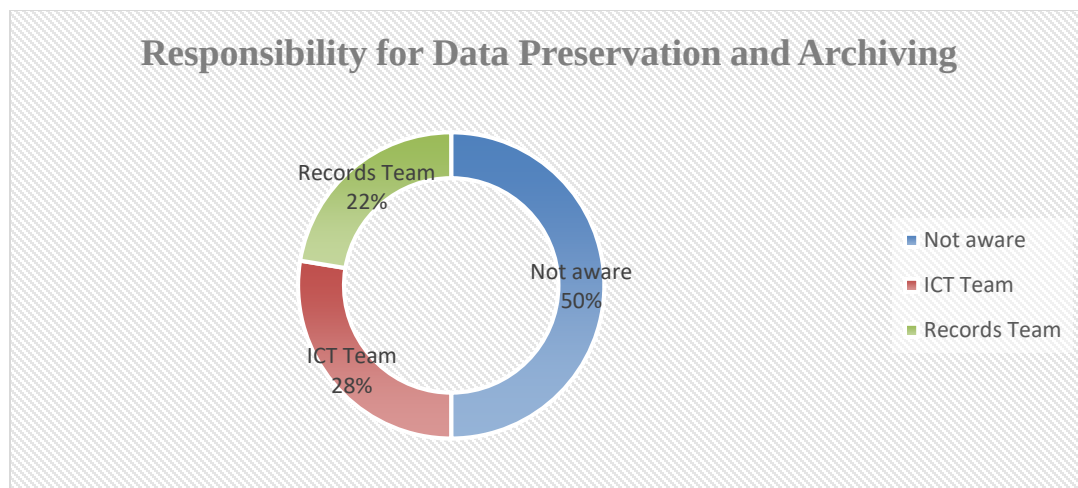


Figure 4. 9: Data Preservation and Disposal Responsibility n=58

From the responses as captured in Fig. 4.9, half of the respondents 29(50.00%) indicated that they are not aware of who decides on decisions relating to data preservation and archiving, while 16(28.00%) stated it was ICT, and finally, 13(22.00%) stated that it is the responsibility of the Records management staff to ensure data preservation and archiving. On the other hand, the interviewees concurred by stating that data preservation and archiving is a shared responsibility between the ICT and the Records Management staff. The Manager of Human Resource Services [MHRS] stated:

I think the Records officers work alongside the ICT to get storage for lots of electronic data they handle.

The findings recorded at Ketraco imply that both the Records Management team and the ICT work jointly to undertake data preservation and Archiving. The archival scholar Yeo (2018) argued that Records management professionals have a key role to play in managing information assets at various stages including preservation and archiving in the digital era. According to Yeo (2018), this ensures consistency and integrity of information materials are upheld. In a study exploring the role of digitization in the preservation of corporate archives at Kenya Power, Ronoh (2019) delineated that information and telecommunications technologies have positively impacted the preservation of digital information materials, and this in turn enhanced access and use of these digital assets. The study by Ronoh (2019) adopted a descriptive research design and a mixed-method approach to investigate how ICT plays a key role in coming up with tools for data preservation. Therefore, besides the Records management team, the ICT team is equally critical in data preservation for choosing the best solutions for data preservation, assigning access controls to electronically preserved data, and setting up systems for automatic data backups and archiving.

4.4.6 Access to Big Data at Ketraco

This section examined the extent to which big data at Ketraco is accessible to users. Esteva et al. (2016) defined data access as the on-request to retrieve, copy, modify, or move from IT systems. Therefore, access is the means and guidelines users can use to access stored data. Respondents were required to indicate the existing access to data at Ketraco. The findings were presented in Figure 4.10.

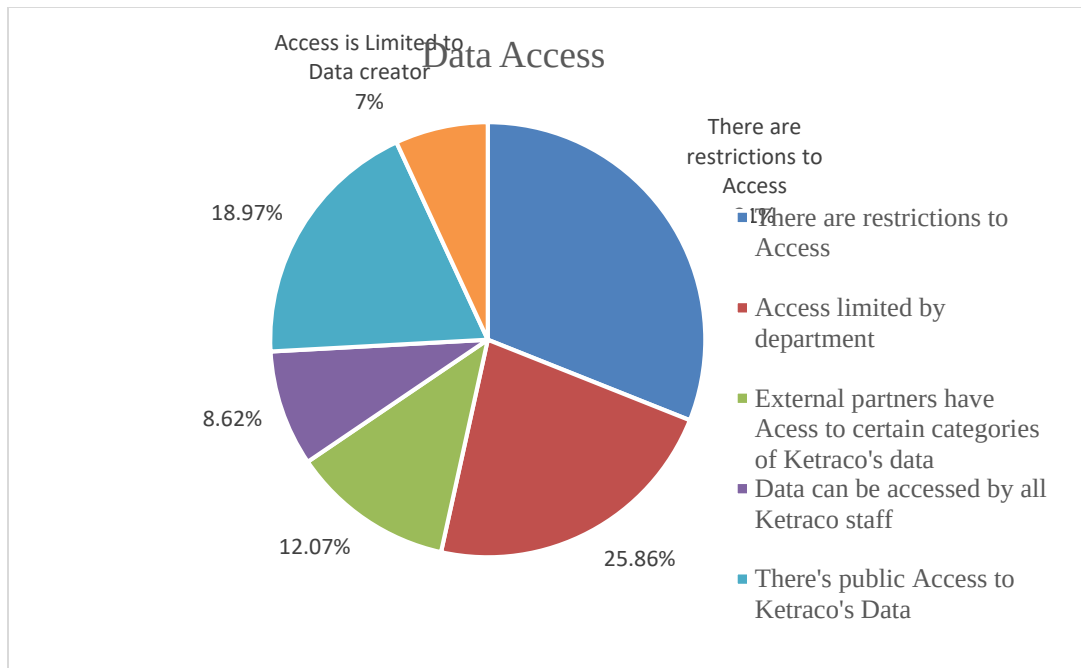


Figure 4. 10: Access to Data at Ketraco (n=58)

Figure 4.10 indicates the extent to which data is accessed at Ketraco. From the findings, 18(31.03%) respondents indicated that data restrictions exist within the organization, and this controls unauthorized access to data. Similarly, 13(22.41%) respondents indicated that access to data is limited to one's department, and thus staff from one department have limited authorization to access other departmental information unless through the internal procedures that involve information request and access approvals. On the other hand, 7(12.07%) respondents indicated that data access is limited to data creators while 5 (8.62%) stated that data can be accessed by any Ketraco staff. Finally, 11(18.97%) respondents stated that access extends to the public, while only 4(6.90%) indicated access to be limited to data creators. Equally, all respondents interviewed agreed on the existence of restrictive measures to control access and use of data at Ketraco.

The findings above indicate that big data at Ketraco is classified, and that access is role-based. In a study on the importance of access controls in big data management, Ali, Sabir and Ullah (2019) observed that access controls are critical in the management of big data as they help to authenticate and authorize individuals to access and use the information they need. Therefore, a data access policy should be in place to determine who accesses the data, how to ensure that only authorized individuals access the data, and the circumstances under which access to the data may be denied (Ali, Sabir and Ullah, 2019). Access to data and rights to data protection in Kenya are key issues contained in the Kenyan constitution and the Acts of Parliament (Access to Information Act 2016 and Data Protection Act of 2019). Colombo and Ferrari (2019) conducted a literature review to examine the key trends and technologies used in access control in big data. The authors argue that data security and privacy have become too complex with the increased volume, velocity, and variety attributes of big data. With this, companies are coming up with various technological solutions to aid in restricting and controlling access to valuable data to guarantee data privacy, integrity, and security. The legislations (Access to Information Act 2016 and Data Protection Act of 2019) provide the framework for a Kenyan citizen or government Agency to hold, access or share data or information, as well as exceptions or limitations to the rights of access to data or information. The publication by Gathu and Kahindi (2015) in the article titled "Access to information in Kenya" pointed out that Kenyans have for a long time been denied access to public data or information and hence excluded from making decisions and contributions in public affairs. This was confirmed by the few respondents 11(18.97%) who had a general view that only a limited fraction of company information can be accessed by members of the public.

4.4.7 Usage of Big Data

This section examined the key usage of big data at Ketraco. The findings indicated that all the departments at Ketraco including Human Resources, Corporate Communication, ICT, Corporate Planning, Audit, Legal, Admin, and Finance Departments are beneficiaries of the big data management at Ketraco. An interview with the top executives revealed that all departments are key users of data, and the same sentiments were echoed by the Acting General Manager of Legal Services who reiterated that all departments within and across Ketraco are key users of data. According to the Ketraco management, the usage of Big Data is critical for the operations of Ketraco such as evaluating contractors and awarding contracts all is based on data from the records on previous project reports, milestones, and achievements or failures.

In addition, respondents in the survey were prompted to rate their organizations' ability to leverage the potential of big data on a scale of 1-5 where 1 denotes Poor, 2= Fair, 3=Good 4=Very Good, and 5=Excellent. The findings were computed and presented in Table 4.6 below:

Table 4. 6: Big Data Usage at Ketraco (n=58)

	Poor	Fair	Good	Very Good	Excellent
Rating of Big Data Usage at Ketraco	0 (0%)	9 (15.52%)	15 (25.86%)	28 (48.28%)	6 (10.34%)

The findings presented in Table 4.6 revealed that the majority 28(48.28%) of the respondents rated the usability of Big data as very good followed by Good 15(25.86%), fair 9(15.52%), and 6(14/10.43%) as excellent. However, no respondent described

Ketraco's big data usability as poor. Equally, the interviewees also described the status of data usability as good, and this was majorly attributed to the efforts the ICT department has made towards the operationalization of data management and storage platforms that support the workflows and storage of data. For instance, the System Analyst [SAP Finance and Controlling] stated:

The integration of SharePoint and Office 365 into our business modules has greatly enhanced our workflows, especially on seamless storage, accessibility, and storage of data.

The findings above indicate that Ketraco is on a positive trajectory towards the adoption of big data although it is yet to realize the full scale of implementation. This was attributed to the respondents' revelation that the Company has invested largely in Microsoft 365, SharePoint, and Integrated location and intelligence systems for data collection and management. For instance, the SharePoint platform allows departments to create their own folders for departmental documents which are accessed by staff across the departments. These findings correlate with the publication by Business Daily dated 11th Feb 2020 titled “*Is Kenya Tapping its Vast Big Data Opportunities?*” which pointed out that the Kenyan government is increasingly making data-driven decisions across its various arms of the government including the enactment of the data protection Act 2019 (Ngila, 2020). However, the same publication (Business Daily) stated that the adoption of Big data in Kenya is still slow, and there are only a few existing firms that offer data science solutions in Kenya. While Ketraco has taken strides to implement big data, the company remains on course, the divergent views demonstrate that Ketraco is yet to achieve the full-scale adoption of big data across its operations.

Meru and Kinoti (2022) argued that the slow big data adoption in Kenyan private and public sectors is attributed to the low big data understanding and acceptance, as well as incorporation into company business strategy. Converting big data into meaningful insights has been a great challenge for many companies in Kenya and across the world (Kshetri, 2014). Therefore, Ketraco, having already invested much in ICT infrastructure such as document management systems, could tap into its existing platform to advocate and enhance its Big Data management strategy.

4.4.7.1 Use and Re-use of Big Data

Data is an important resource in many organizations due to its role in decision-making and daily operations (Kwanya, 2014). In the digital age, companies collect huge volumes of data used to support organizational operations. The questions in this section were designed to examine respondents on how Ketraco utilizes Big Data. Therefore, the respondents indicated the extent to which big data was used to support Ketraco's operations on a five-point Likert Scale where 1 denotes strongly disagree, 2= Disagree, 3= Not sure, 4= Moderate, 5=Agree; 6=Strongly Agree. The findings were presented in Table 4.7 with a computed mean that detailed how the respondents agreed or disagreed with the statements on the uses of Big Data at Ketraco.

Table 4. 7: Uses and Re-use of Big Data at Ketraco (n=58)

Uses and Re-use of big data at Ketraco.	1 Strongly Disagree	2 Disagree	3 Not sure	4 Moderate	5 Agree	6 Strongly Agree	N
Ketraco uses Big Data to engage customers	11 (18.97 %)	29 (50%)	11 (18.97 %)	4 (6.90%)	1 (1.72%)	2 (3.45%)	58

Ketraco uses Big Data to make strategic decisions	0 (0%)	0 (0%)	3 (5.17%)	3 (5.17%)	38 (65.52%)	14 (24.14%)	58
Ketraco uses Big Data to make operational decisions and transactions	0 (0%)	1 (1.72%)	1 (1.72%)	1 (1.72%)	2 (3.45%)	53 (91.38%)	58
Ketraco relies on its data to improve and innovate its services	6 (10.34%)	11 (18.97%)	18 (31.03%)	12 (20.69%)	9 (15.52%)	2 (3.45%)	58
Ketraco uses Data to reach out to new markets and access new business opportunities	2 (3.45%)	11 (18.97%)	14 (24.14%)	15 (25.86%)	13 (22.41%)	3 (5.17%)	58
Ketraco uses data to assess its risks	0 (0%)	1 (1.72%)	1 (1.72%)	2 (3.45%)	35 (60.34%)	19 (32.76%)	58
Ketraco relies on data to measure its performance and evaluate its operations	0 (0%)	0 (0%)	3 (5.17%)	4 (6.90%)	17 (29.31%)	34 (58.62%)	58

The survey sought to understand how Big Data streamlines Ketraco's operations, as well as provide tactical and strategic support in decision-making. The findings in Table 4.7 indicate that the majority of the respondents, 53 (91.38%) strongly agreed that Big data at Ketraco is mostly used to support operational decisions and transactions. In addition, it was noted by majority agreement 51 (87.93%) and 54 (93.10%) that Ketraco relies on big data to measure performance and evaluate its operations and to assess and reduce operational risks respectively. Further, the findings depict that Big data is an important resource that supports Ketraco's strategic and investment decisions as captured by the majority of the respondents 52 (89.66%) who agreed with this statement. However, there was an indication of deficiencies in the company's big data strategy as a sizeable number of respondents 29 (50%) were not sure whether Ketraco utilizes big data to support its need to access new

markets and new business opportunities and innovation for products and services respectively. Equally, the interviewees appreciated that every person in the Company relies on data to inform the decision-making process. For instance, the Manager, Human Resource Services testified:

For us to make any decision on any matter and make objective conclusions, our key reference point is the existing information. You will always find us referring somewhere when responding or transacting on any processes within our directorate.

It is evident from the findings that big data plays a critical role in the company's operations and support for decision-making processes. Equally, it is vital for improvements, especially in using big data to engage customers and improve service delivery. Similarly, there still exist opportunities to leverage the data generated by the company to inform continuous improvements and innovation in its grid and access new market opportunities. A study by Zhan et al. (2017) on evaluating the use of big data as a framework for promoting innovation established that the use of big data could offer a paradigm shift and enable firms to develop new products faster and at a lower cost. This demonstrates the power of big data as an ideal solution that Ketraco could leverage to spur innovation in its business process, products, goods, and services.

4.4.8 Sharing of Big Data at Ketraco

This section examines how big data at Ketraco is shared among various stakeholders including employees, management, the media, government agencies, and the public. To explore this, respondents were asked to indicate the existing channels that support the flow and sharing of data. The responses were summarized in Figure 4.11.

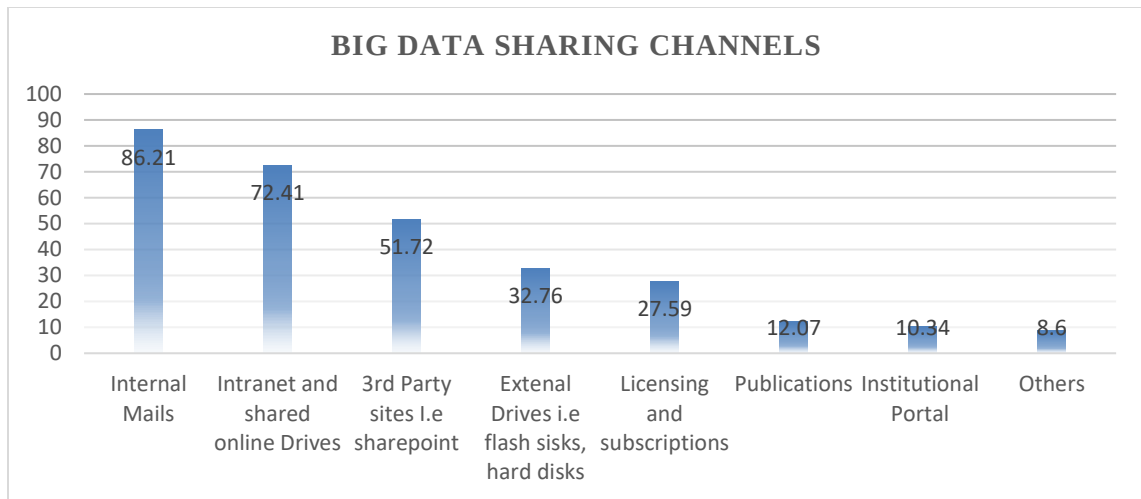


Figure 4. 11: Big Data Sharing Channels at Ketraco (n=58)

Findings from fig 4.11 indicate that at Ketraco, the majority of data is made available for use internally via the internal email communication 50(86.21%), 42(72.41%) intranet and shared online drives, 30(51.72%) third-party sites such as SharePoint, 19(32.76%) external drives such as flash and hard discs, 16(27.59%) licensing and publications, 7(12.07%) publications, 6(10.34%) institutional portal and 5(8.6%) chose other options that had not been listed on the questionnaire. In a similar way, interviewees 10(100%) unanimously agreed to the existence of multiple channels for data sharing by indicating that emails (Outlook), OneDrive, and SharePoint as the common channels for data sharing. The findings do not demonstrate the existence of a consistent and variety of channels for availing data at Ketraco, and this can be attributed to the lack of a well-established framework that guides users on data sharing and channels to use as well as the incorporation of big data sharing in departmental work plans. Whereas there are various channels of availing data for internal usage, Bertino, Ghinita and Kamra (2011) argue that organizations should consolidate information in a single platform and thereafter make its data available to users. In other words, data should be availed in a structured approach with

a guiding framework as a mechanism for controlling access and usage, as well as improving its security.

Furthermore, Gandomi and Haider (2015) encourage companies to establish formal channels of sharing data such as an institutional repository and company databases, and to develop effective and efficient analytical techniques to exploit large volumes of heterogeneous data in various formats such as unstructured text, audio, and video formats. Gandomi and Haider (2015) went ahead to state that the big data channels used by an organization or individual vary greatly depending on the audience, message, goals, and the Company's capability to invest in multiple varieties that have been necessitated by technological advancements. Therefore, the existence of multiple platforms for accessing data at Ketraco can be attributed to its ability to invest in multiple technological infrastructure that supports multiple storage, access, and sharing. However, there is a need for companies to put in place a structured mechanism to control how the multiple data generated and used in the organization is used to guarantee the security of the data.

In a separate question, respondents were asked to choose the key partners with whom they exchange/share data with and outline challenges encountered in the sharing process. The findings revealed that 58(100%) respondents share data internally with colleagues within the department, 43(74.14%) share with Colleagues outside their departments, 42(72.41%) share with stakeholders such as contractors and suppliers, 28(48.28%) share with regulators and Auditor General, and 25(43.10%) share with other partners in the energy sectors. Only 12 (20.69%) 5(8.62%) and 4(6.89%) indicated sharing data with other public agencies, the public, and the Media respectively. Internal data exchange at Ketraco is enhanced by the

existing medium and ICT systems such as SharePoint that support data storage, access, and sharing. This demonstrates that Ketraco has put in place a sound framework that guides how internal 5s in the organization can manage and access data within the organization. This supports this study's theme of improving management, sharing, and the use of big data to facilitate effective planning, drive efficiency in its processes, and improvements in decision-making. The findings of this study disagreed with those by Magina (2019) who in a study observed that Kenya still encounters challenges in providing access to data and information despite having an elaborate institutional and legislative framework for data access. Instead, the current study findings align with the study by Kabata and Garaba (2019) on leadership and political will for implementation of the Access to information in Kenya which revealed that the Kenyan leadership and political class have demonstrated goodwill to enhance access to government information to the public. The current study shows that Ketraco has demonstrated goodwill and a proactive approach to controlling how the organizational data is accessed, stored, and shared with different stakeholders.

Further, respondents highlighted the aspect of fear of breaching the confidentiality of information as one of the key challenges that hinder data sharing. The respondents pointed out the issue of confidentiality and secrecy as the major hindrances to data sharing. Other challenges listed include the limited document size that can be shared using the company software platform, and the reporting structure that requires employees to seek authorization before sharing data. In essence, many employees in Kenya often see data as a source or commodity for power, hence hold on it.

4.5 Policies Guiding the Use of Big Data Ketraco

McNeely and Hahm (2014) define a data management policy as a written document that guides an organization on its data usage. The policy establishes how data is created, appraised, described, preserved/stored, accessed, used, shared, and secured in the organizational IT structure. A functional policy reduces risks associated with data breaches and helps the company to comply with relevant laws. According to the Data Curation Centre (DCC) Lifecycle Model, every stage in managing the data lifecycle should be guided by a policy framework to ensure successful curation and preservation of data. The sequential stages in the model encourage the documentation of policies and processes by different stakeholders and also support the establishment of frameworks of technologies and standards that guide the management of data. On the same note, the Community Capability Model (CCM) recognizes the capability of every group with shared goals and capabilities to undertake a common goal. To achieve this, the model highlights the need to establish legal and ethical frameworks for the community to leverage its data asset capabilities. Thus, Big Data policies are necessary for any community or organization to guide data curation, human resources, ICT preparedness, data security, data literacy, and collaborative partnerships among others. Thus, the study sought to find out if there were any policies and guidelines for Big Data management at Ketraco. The researcher requested the respondents to Indicate the existence of data policy at Ketraco. The findings were presented in Fig 4.12

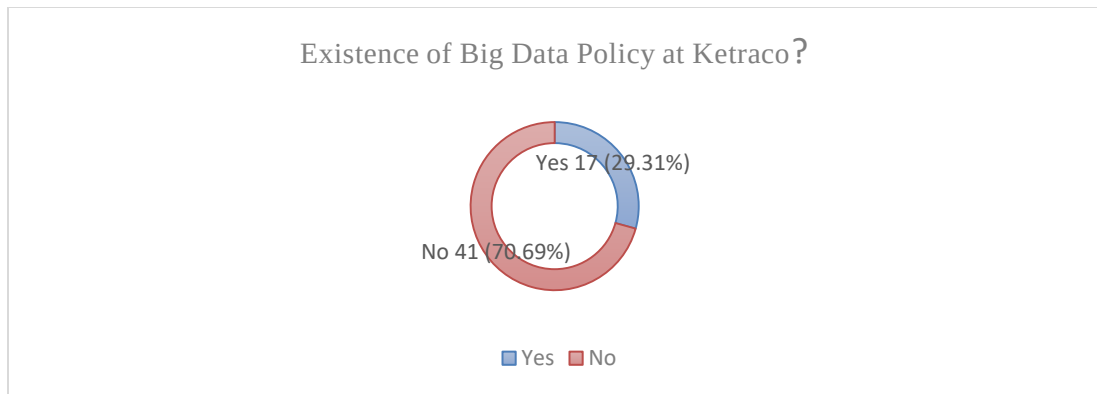


Figure 4. 12: Existence of Big Data Management Policy at Ketraco (n=58)

From the presentation in Fig 4.12, many of the respondents 41(70.69%) were not aware of the existence of the data policy while 17(29.31%) of the responses gave a positive response. As a follow-up question, those who said there were no policies indicated that they were not aware of the prevalence of such a policy. On the other hand, those who said yes and were aware of the existence of the policy emphasized the presence of a policy related to the management of data and information but not a specific Big Data policy. Thus, the response implies that the Company relies on an in-house framework that guides data management therein. Similar observations were made from the interviewers, who all indicated not having interacted with a data management policy.

McNeely and Hahm (2014) argue that organizations that have developed their policies to guide Big Data usage are in the right direction to adopting and implementing the universally established procedures. Therefore, despite Ketraco not having a specific policy for data management, and instead, having policies that can be associated with big data management implies that the company is in a positive direction toward the implementation of more comprehensive guidelines that are based on government laws and regulations.

Further, the absence of the policy was demonstrated when the respondents were asked to rate various procedures related to managing data across its life cycle. The procedures include how the company captures, appraises, describes, provides access, uses, re-uses, and preserves data. The respondents rated these procedures in terms of assuring accessibility and security of the data as outlined in Figure 4.13.

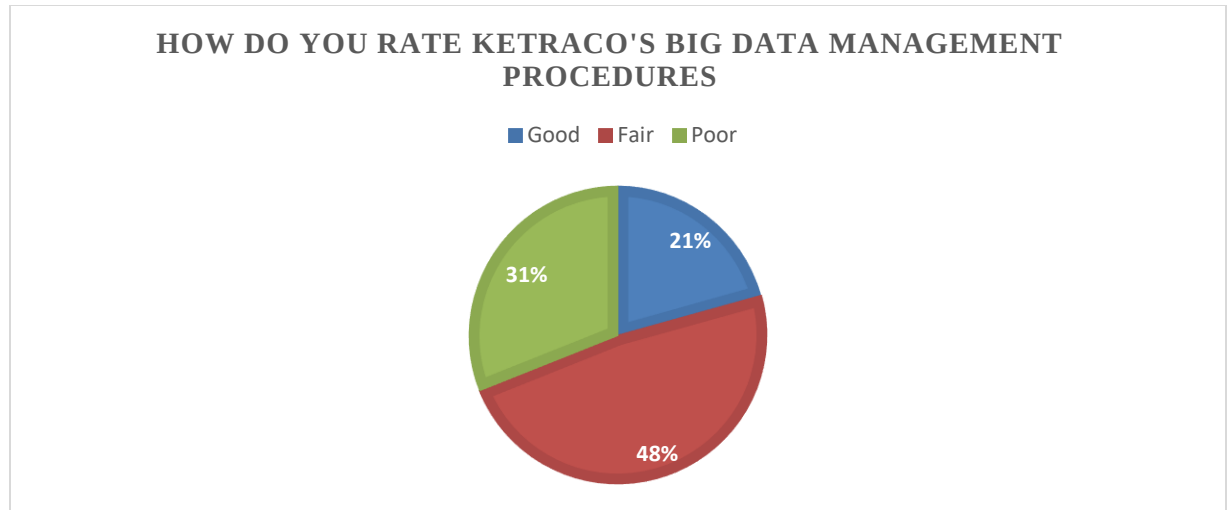


Figure 4. 13: Big Data Management Procedures at Ketraco (n=58)

Figure 4.13 revealed that 12(20.69%) respondents rated the procedures as good, while 28(48.27%) and 18(31.03%) rated them as fair and poor respectively. Thus, the results in Figure 4.13 suggest that the in-house mechanisms and procedures related to big data have to an extent allowed Ketraco to manage big data activities. These findings coincide with Bertot and Choi's (2013) sentiments who argued that irrespective of the organizational model (internal or external third-party cloud-based solutions), every organization should establish clear mechanisms for managing every phase of big data management such as the data creation, description, storage, access, sharing, disposal, quality assurance, knowledge and skill, ICT preparedness for big data, collaborative premiership among others. These systematic big data management activities require well-established guidelines as outlined

in the Data Centre Curation (DCC) lifecycle model (Higgins, 2008), as well as the community capability model (CCM) (Lyon et al., 2012). According to the DCC lifecycle model, every process or stage in data curation should be embedded with appropriate policies and guidelines that ensure effective data management across the lifecycle (Higgins, 2008). More so, community capabilities like knowledge, skill, and training; ICT preparedness, and collaborative partnership should be engulfed with policies for proper guidance and administration (Lyon et al., 2012). Therefore, organizations should have specific policies across big data management to ensure effective management of their data.

Furthermore, all the interviewees had a general view that Ketraco had indirect measures captured in the Company policy documents that hold them accountable for the use, sharing, and management of data. For instance, the Ag. General Manager Legal Services in the interview stated:

We have a communication policy that establishes procedures for sharing information or communicating with an external entity like the Media on behalf of the company.

As a public corporation, Ketraco has established a clear framework that clarifies who should communicate and the extent of the information to communicate to the public or other external stakeholders. However, the response fails to demonstrate the existence of a clear policy to guide the sharing of data with external parties. In their article, Kim and Cho (2018) call for a big data governance policy framework focused on timely, reliable, meaningful, and adequate data services. This is important to guarantee data quality. Broeders et al. (2017) also called for tight regulations at every level of big data from the level of use, user profile, and semi-automated decision-making processes. They also recommend the need for the responsibility of big data management and decision-making

to be strongly anchored in policy or legislation to strengthen the general and security-specific functions. Based on this, Ketraco has exhibited gaps in its big data policy, which compromises the quality of data as well as the general security.

4.5.1 Strategies that can help in the Implementation of Big Data Management at Ketraco.

Respondents were asked to identify strategies that can help in the implementation of Big Data management at Ketraco. They were to tick from the recommendations provided in the questionnaire and provide their feedback on a scale of 1-5 where 1 indicates disagreement while 5 indicates they strongly agree with the assertions as presented in figure 4.14.

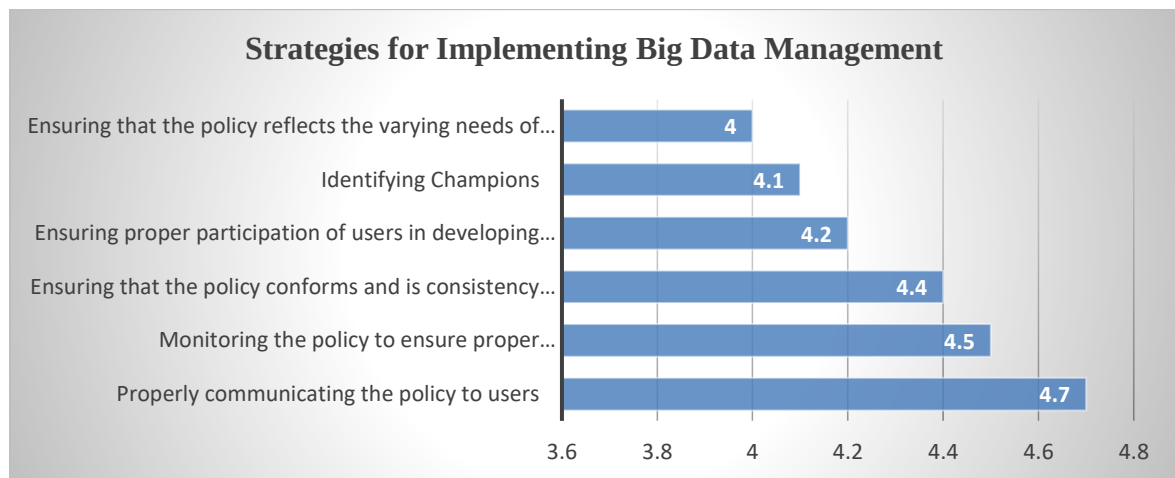


Figure 4. 14: Strategies for Implementation of Big Data Management at Ketraco (n=58)

Figure 4.14 presents the views of participants on how Ketraco should implement and manage its big data to ensure it satisfies the needs of all stakeholders and promotes maximum implementation and compliance by the users. Several statements were made by participants who provided their recommendations based on their knowledge and experience working at Ketraco. For the company to effectively implement a Big Data strategy, respondents concurred with the recommendations posed in the questionnaire. For

instance, the respondents pointed out the need to establish and communicate a Big data policy (mean= 4.7), monitoring the policy (mean= 4.5), conforming the policy with other related policies (mean= 4.4), ensuring participation of various stakeholders (mean= 4.2), identifying champions for the policy (mean= 4.1), and finally ensuring the policy reflects the diverse needs of various departments (mean=4.0). Similarly, 10(100%) of the respondents interviewed agreed that there was a need to establish a data management policy with various strategies and accompaniments that will ensure its successful implementation. These recommendations by respondents demonstrate that there exists good will especially from staff to help the Company develop a successful big data implementation program.

The findings suggest that employees at Ketraco support the need to implement a robust big data management policy provided that top management establishes measures to monitor the policies for effective implementation, and ensure it is consistent with other policies. Successful adoption of the policy also requires those formulating it to consider the needs of departments in its design and implementation. Tabesh, Mousavidin, and Hasani (2019) observed that many organizations have failed to implement big data due to managerial commitment and support. With this and the observations recorded from the respondents, the study recommends that management should prioritize effective communication of the policy and provide managerial support and coordination. Consequently, the Office of the Deputy Prime Minister in London on the Principles of Good Data Management (2005) highlighted that data policy formulation requires best practices to be adopted across the company with the guidance and leadership of the Executive Board (Paskaleva et al 2017). Therefore, the implementation of data governance best practices requires collaborative efforts to ensure the outlined strategies are adaptable and manageable. In addition, the

process demands effective communication that trickles from top leadership, through the selected champions, to the rest of the staff. This extensive approach will enhance its acceptance across the company.

4.5.2 Quality Assurance and Quality Control of Big Data

Quality assurance and control refers to the criteria and processes put in place to ascertain and verify that data meets the specific data quality objectives throughout its lifecycle. It was in this light that the study sought to get insights from respondents on their knowledge of the company's quality assurance and control mechanisms in the management of its data. Respondents were allowed to record more than one answer on the stages of big data management they felt had quality controls. The findings were recorded and presented in Figure 4.15.

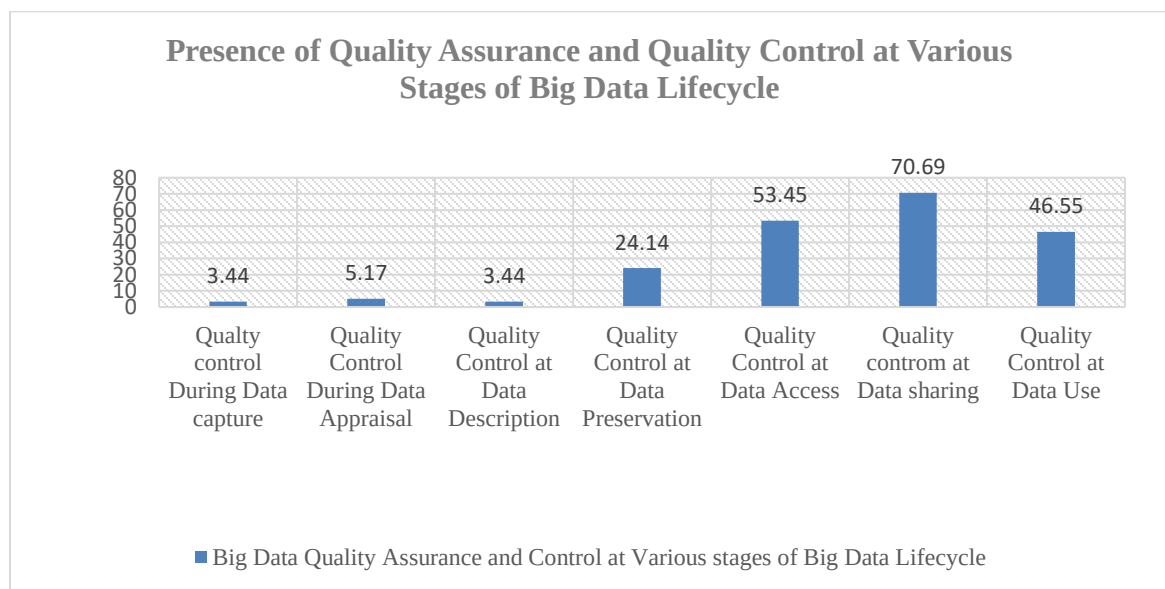


Figure 4. 15: Presence of Quality Assurance and Quality Controls at various Stages of Big Data Management (n=58)

Based on their feedback as presented in fig 4.15, 2(3.44%), 3(5.17%), 3(5.17%), 2(3.44%), 14(24.14%) 31(53.45%), 41(70.69%) and 27(46.55%) respondents acknowledged the existence of quality assurance and control measures at stages of data capture, appraisal, description, preservation, access, sharing, and use respectively. The respondents acknowledged interacting often with huge volumes of data which are fast-moving and are retrieved from various sources including unstructured data from websites, raw data feeds, and social platforms among others. This data is used to optimize the company's processes and to make objective decisions effectively. However, from the findings, Ketraco's data is not subjected to quality checks, and this is majorly attributed to the unavailability of proper mechanisms that guide the same. According to Gao, Xie, and Tao (2016), inadequate mechanisms for quality control often result in duplicated, inaccurate, ambiguous, hidden, too much, and inconsistent data as well as data downtime. Therefore, it is more likely that failure to have control mechanisms will expose Ketraco's big data to the mentioned challenges that, thereafter, affect internal organizational processes and strategic decisions that rely on accurate and authentic data. As a result, the company is not only exposed to risks related to managing data across its various stages but also to issues compliance with the existing legal and regulatory frameworks.

However, respondents stated the existence of significant improvements in quality control in areas of data access, sharing, and use with a representation of 31(53.45%), 41(70.69%), and 27(46.55%) respectively. The improvements were attributed to internal ICT Mechanisms with User access control mechanisms to manage access to computer systems and data. On the use of computer systems to manage access to systems and data, the system analysts in charge of various system application modules had the following to say:

System Analyst [SAP Finance and Controlling]

'...we normally receive requests from departments (of course) addressed to the head of the ICT department requesting a particular user to be granted access to certain functionalities within the system. This is then brought to my action...'

System Analyst [SAP Human Capital Management]

'Our Enterprise Resource Planning (SAP) is so dynamic, supporting various workflows across the departments. However, the system provides access depending on the role of each user within the workflow. Nevertheless, there are areas where everyone can access (Employee Self Service) where any staff can log in to apply for Leave, access pay slips, make a travel request, or apply for an internal job.'

System Analyst [Business Applications]

Data access and system use is controlled, you will always see people talk about rights, and this is the norm for any system.

System Analyst [Supply Chain]

.... the company subscribes 3rd parties to have its data hosted on various e-platforms. And even these licenses are paid by the company to grant a new user permission to perform certain tasks. Even for me, there are areas within the system where I don't have access.

It is evident that Ketraco's systems and data access had continuous improvements in investments in ICT Systems. Other factors that have contributed to Ketraco's improvement in data control include the implementation of ISO 9001:2015, the implementation of a quality management system (QMS), and ISO 2700 2018. Before these ISO certifications, certain parameters are required as a fulfilment for certification. All the interviews commended this ISO Certification as it had improved the classification of data, which had reduced the challenges of data mix-up, security, access control, sharing, and unavailability. Furthermore, it was recorded from interviewees that the Company introduced the risk management and quality assurance department that specifically oversees risk management, and quality assurance and also monitors compliance of company procedures with various

statutory regulations. Thus, it is likely that going forward, Ketraco through this new department can improve performance reviews, assessment, monitoring, and quality controls for its data.

4.6 Skills, Knowledge, and Training for Big Data Management at Ketraco

Big Data has grown into an important component by helping organizations achieve operational efficiency and position strategically for the future. However, this requires prerequisite skills and knowledge for staff managing big data and users of the data to capture, interpret, and use this data to make meaningful inferences (Song et al., 2015). According to the Community Capability Model (CCM), human capability factors including skills, knowledge, and training play a key role in a community's capability to accomplish a given task (Lyon et al, 2012). Thus, the researcher sought to evaluate the existing skills, knowledge, and training appropriately to enhance the employees' capability to capture and apply big data in their routine activities.

4.6.1 Existing Skills for Big Data Management at Ketraco

First, the researcher assessed the level of respondents' skills in critical aspects that constitute an effective data management program. In lieu of this, the respondents were required to indicate the existence of various skills on a five-point Likert Scale where 1 denoted strongly disagreed, 2= Disagreed, 3= Not sure, 4= Moderate, 5=Agreed; and 6=Strongly Agreed. The findings were computed and presented in Table 4.8.

Table 4. 8: Employee skills for Data Management at Ketraco (n=58)

Employees' Skills	1 Strongly Disagree	2 Disagree	3 Not sure	4 Moderate	5 Agree	6 Strongly Agree	N
Metadata skills	23(39.65 %)	28 (48.28%)	3 (5.17 %)	2 (3.45%)	1 (1.72 %)	1 (1.72%)	58
Search and Retrieval skills	1 (1.72%)	7 (12.07%)	29 (50%)	10 (17.24%)	8 (13.7 9%)	3 (5.17%)	58
Data Capture skills	0	3 (5.17%)	17 (29.3 1%)	12 (20.69%)	15 (25.8 6)	11 (18.97%)	58
Data Analysis skill	35 (60.34%)	14 (24.14%)	3 (5.17 %)	2 (3.45%)	3 (5.17 %)	1 (1.72%)	58
Collaboration and communication skills	0	4 (6.90%)	12 (20.6 9%)	21 (36.21%)	14 (24.1 3%)	7 (12.07%)	58

Table 4.8 presents the skills that Ketraco employees possess in relation to Big Data management. Respondents agreed that employees at Ketraco possessed both data capture and collaboration and communication skills with a representation of 27 (46.55%) and 35 (60.34%) respectively. The existence of at least two sets of skills conforms with respondents' information provided in the demographic section to possess the minimum skills required to perform their respective data-related duties. However, a good number of

respondents, 29 (50%) were not sure while 10 (17.24%) remained neutral regarding their search and retrieval skills. Further, majority of the respondents disagreed with the statements of skills relating to data analysis and metadata skills as majority of them 51(87.93%) disagreed to possess these skills.

On the other hand, the interviewees also had varying opinions regarding their skills vital for big data management. For instance, the Manager Human Resource Services {MHRS} stated:

I am not good at analyzing data, and in most cases, I seek help from experts. Maybe I can do better with communication and collaboration capabilities. To add to this, the Senior Manager Supply Chain [SMSCM]

May be communication skills but I am not sure about the others you have mentioned. It was evident from the findings that gaps in skill sets existed especially on key aspects of metadata capture and data analysis. These skills according to Zhang et al. (2018) are critical in the management of big data in the organization to guarantee the accuracy, reliability, and integrity of data. Thus, failure to possess such skills will expose the company to risks related to managing the life cycle of data, and this affects critical functions of data creation, description, use, preservation, and sharing.

4.6.2 Knowledge and Training for Big Data Management at Ketraco

The study sought to explore employees' knowledge and training on data capture, Appraisal, description, security, storage, and sharing. The respondents were required to indicate the existence of various knowledge on a five-point Likert Scale where 1 denotes strongly disagree, 2= Disagree, 3= Not sure, 4= Moderate, 5=Agree; 6=Strongly Agree.

The findings were computed and the mean presented in Table 4.9

Table 4. 9: Knowledge and Training for Big Data Management at Ketraco (n=58)

Knowledge and Training	1 <i>Strongly Disagree</i>	2 <i>Disagree</i>	3 <i>Not sure</i>	4 <i>Moderate</i>	5 <i>Agree</i>	6 <i>Strongly Agree</i>
<i>Knowledge on Data capture</i>	1 (1.72%)	3 (5.17%)	9 (15.52%)	23 (37.93%)	19 (32.76%)	3 (5.17%)
<i>Data Appraisal</i>	1 (1.72%)	29 (50%)	22 (37.93%)	3 (5.17%)	2 (3.45%)	1 (1.72%)
<i>Data description</i>	0	1 (1.72%)	17 (29.31%)	18 (31.03%)	14 (24.14%)	5 (8.62%)
<i>Data Security</i>	27 (46.55%)	20 (34.48%)	7 (12.07%)	3 (3.45%)	1 (1.72%)	1 (1.72%)
<i>Data Preservation and storage</i>	43 (74.14%)	6 (10.34%)	3 (5.17%)	3 (3.45%)	3 (5.17%)	1 (1.72%)
<i>Data sharing</i>	1 (1.72%)	47 (81.03%)	2 (3.45%)	2 (3.45%)	3 (5.17%)	3 (5.17%)

The study revealed that majority of the respondents possessed basic knowledge of data capture and description as this was denoted with 23 (37.93%) for moderate and another 19 (32.76%) who agreed to possess the said skills. The study also noted the existence of a knowledge gap on data appraisal and sharing amongst Ketraco staff as most of them 29 (50%) disagreed and confessed not having an understanding of data appraisal. Equally, another 22 (37.93%) respondents were not sure whether their knowledge of data appraisal was adequate to support big data initiatives.

Similarly, inadequate knowledge for implementation of the Big data program was realized when 27(46.55%) respondents strongly disagreed and another 20(34.48%) disagreed with understanding and training data security. Similar responses were recorded when

43(74.14%) and 6(10.34%) disagreed and confessed not to possess knowledge or undertaken training on data preservation and storage.

From the findings, the company has not taken into consideration incorporating big data training needs to complement career progression training in its annual training plans. This was further explained by the Ag Manager Legal Services [Ag. MLS] who stated:

I have not heard any recommendations on training needs regarding Big Data. The common training courses are those that lean towards the area of specialization of an employee.

Similarly, it was observed from the questionnaire that the respondents with the requisite knowledge in data management and who did not require training at every stage of big data management were staff from the IT and the Records Management departments. This could be attributed to their career background and frequent interaction with data assets and platforms. However, the knowledge gap exhibited amongst other Ketraco staff can be attributed to a lack of awareness and mechanisms that incorporate data needs in annual training plans. According to Gharaibeh et al. (2017), the current information age demands employees to possess the capability of identifying key information from large data sets. In reference to this, Ketraco employees need appropriate training to enable them to assess and weed out less vital data and focus on essential ones that will support decision-making and problem-solving.

Further, the study assessed the respondents' level of expertise and capability to communicate, collaborate, analyze, and share data amongst themselves. According to Raptis, Passarella, and Conti (2019), expertise in Big Data management refers to the ability of an individual to possess and apply advanced knowledge in managing data across its

lifecycle from identifying sources of data, modelling, and analysis to storage. The respondents moderately agreed to possess the capabilities of performing basic processes in big data management such as data naming, uploading, sharing, and storing in different media. However, they were not sure whether they could design strategies and processes that make data understandable and accessible to policy and decision-makers. Perhaps, having dedicated officers perform big data activities could help implement the big data program at Ketraco. For instance, the General Manager for Technical Services {Ag. GM, TS} in an interview stated that;

Perhaps we need to consider hiring data officers or just dedicate some members of staff to champion all these issues you are raising.

The Ketraco scenario of Big Data demonstrates that employees were moderately equipped with the knowledge to perform basic big data assignments rather than perform complex tasks that require an expert. Besides, Gharaibeh et al. (2017) posited that an expert in data management requires a combination of analytical, technical, and business knowledge. Thus, Ketraco should consider training staff, and engage the services of a data expert to support the development of data management strategy. Similarly, training in data literacy programs is essential to increase employee awareness and help them embed big data processes into their workflows.

Moreover, Poleto et al. (2015) illuminated that data management strategies are integral to all aspects of a digital work environment. Therefore, data literacy programs as part of the data management strategy can help employees acquaint themselves with the basics of the whole concept of big data. Having employees with an understanding of big data can help a company to seamlessly work with its data and adopt the best practices and technologies.

It is evident that Ketraco does not have adequate and relevant big data management strategies and metrics that are causing data skill and knowledge gaps, and this directly affects its operations and decision-making processes. Similarly, the findings published in Gartner's publication in a study carried out worldwide on 496 chief data officers on the seventh Chief data officers' survey revealed that although big data is essential in driving business value, poor data literacy remains the second major hindrance to the success of big data adoption in many organizations globally (STAMFORD, Conn, 2022). To address these skill gaps, the same publication recommended that companies should establish data literacy programs to create an environment where learning and acquisition of skills form part of the organizational culture (STAMFORD, Conn, 2022).

4.6.3 Proficiency of Employees to Various Big Data Skills at Ketraco

The study further examined the level of expertise and proficiency of Ketraco employees, specifically in the management of big data. The skills evaluated included SQL database, Hadoop, Microsoft Office, Quantitative Data analysis, Data mining, and analysis. The skill level was evaluated in terms of basic level, intermediate level, advanced level, planned in the future, and not sure. The findings were presented in Table 4.10.

Table 4. 10: Expertise and Proficiency in Various Big Data Skills (n=58)

	Basic	Intermediate	Advanced	Planned	Not sure
SQL	2 (3.44%)	1 (1.72%)	1(1.72%)	10 (17.24%)	44(75.86%)
Hadoop	1 (1.72%)	2 (3.44%)	1(1.72%)	6 (10.34%)	48(82.76%)
Word	20 (34.28%)	34 (58.62%)	1(1.72%)	1(1.72%)	2 (3.44%)
Statistical	5 (8.62%)	16 (27.59%)	4(6.90%)	20 (34.28%)	13 (22.41%)
Data mining	6 (10.34%)	17 (29.31%)	3 (5.17%)	16 (27.59%)	16 (27.59%)
Data Analysis	7 (12.07%)	18(31.03%)	5 (8.62%)	13 (22.41%)	15 (25.86%)

Based on the findings illustrated in Table 4.10, most employees at Ketraco have attained intermediate skills in areas of Data Analysis, data mining, statistics, and analysis as captured by 7 (12.07%), 6 (10.34%) 16 (27.59%) and 34 (58.62%) respondents respectively.

However, Database management skills such as Hadoop and SQL, the key skills for data analysis and storage were the most obscure as only 15 (25.86%) of the participants had attained the advanced level of expertise. It can therefore be concluded that Ketraco can fail to derive maximum benefits from its data resources due to inadequate skills in the utilization of big data management software. Similar observations were made by Kibe, Kwanya and Owano (2019) in a comparative study on the Influence of Data analytics on the performance of organizations between the Technical University of Kenya and Strathmore University employing a mixed method approach established that many

organizations in Kenya lack adequate personnel with expertise in Big Data, thus, hindering its adoption and use.

Similar observations were also highlighted by Onyango and Ondiek (2021), who conducted a study on digitalization and integration of Sustainable development goals in public companies in Kenya. The researchers mapped the role of digital platforms, ICT, and personnel skills vis-à-vis the performance of public organizations in Kenya and found that many organizations in Kenya experience limited technological skills and capacities to disaggregate, manage, and use big data. Inadequate skills in big data thereafter affect the implementation of other key systems such as Records and document management systems and workflows that rely on data.

In contrast, in a study to determine the extent of adoption of Big Data technologies in supermarkets in Kisumu County, Ochieng (2015) established the presence of skills to use ICT tools and software in the county, and this played a key role in the performance of the chain stores. The study attributed this to higher computer usage at 68% and the ability to integrate new enterprise systems into the existing skills that employees possessed. This could be interpreted to mean that the acceptance of a new system in an organization is greatly influenced by the users' ability to operate or integrate the system into its business processes.

Regardless, the low level of adoption and implementation of the right and advanced big data technologies as observed in the reviewed studies is an indication that big data adoption in many companies in Kenya is still in the early stages, and these companies are challenged

with inadequate human capital, skills, culture, and other resources needed in organizations for big data implementation.

4.7 ICT Infrastructure for Big Data Management

This section examined the availability of the ICT infrastructure at Ketraco and its ability to support the corporation's big data strategy. Respondents identified various ICT infrastructures and systems and rated the extent to which the systems and infrastructure were adequate to support big data. According to the Community Capability Model (CCM) (2012) and Data Curation Centre (DCC) lifecycle model (2008), technological infrastructure is an enabler of a community's capability to successfully undertake an intensive project such as the implementation of the Big data management initiative (Lyon et al., 2012). The technological factors include software for data capture and processing, the network and cloud platforms for data sharing and storage. Table 4.12 highlights various ICT infrastructures that exist at Ketraco.

Table 4. 11: ICT Infrastructure at Ketraco for Big Data (n-58).

ICT infrastructure at Ketraco for Big data.	Yes	No
Reliable internet/network infrastructure	58 (100%)	
Computer software and programs	58 (100%)	
Servers, Personal computers, and other hardware	58 (100%)	
Storage platforms	58 (100%)	
Databases	58 (100%)	

Table 4.11 above indicates that all 58 (100%) respondents indicated existence of various ICT Infrastructure that support big data management. This observation demonstrates that Ketraco has invested in various ICT infrastructures including computer software and programs, reliable internet and network infrastructure (servers, personal computers, and other hardware), storage platforms, and databases to support big data management. The respondents were also required to evaluate the efficacy of the various ICT infrastructure to support big data. Therefore, the respondents were asked to the extent to which they agreed with the efficacy of the big data management infrastructure and were to indicate on a five-point Likert Scale where 1 denotes strongly disagree, 2= Disagree, 3= Not sure, 4= Moderate, 5=Agree; 6=Strongly Agree. The findings were computed in Table 4.12.

Table 4. 12: Efficacy of ICT Infrastructure at Ketraco (n=58)

Efficacy of ICT Infrastructure at Ketraco	1 Strongly Disagree	2 Disagree	3 Not sure	4 Moderate	5 Agree	6 Strongly Agree
There is a network infrastructure for data communication and sharing	0	0	8 (13.79 %)	11 (18.97%)	21 (36.21 %)	18 (31.03%)
There are storage platforms/server spaces to store your data	0	1 (1.72%)	10 (17.24 %)	26 (44.81%)	17 (29.31 %)	4 (6.90%)
There is a digital institutional repository for archiving data	1 (1.72%)	2 (3.45%)	14 (21.14 %)	18 (31.03%)	16 (27.58 %)	7 (12.07%)
There exists Software and hardware for capturing and describing data	1 (1.72%)	2 (3.45%)	23 (39.66 %)	8 (13.79%)	17 (29.31 %)	7 (12.07%)
There are systems that enhance the efficiency and flow of data within the workflows	0	1 (1.72%)	16 (27.59 %)	9 (15.52%)	17 (29.31 %)	15 (25.86%)

Table 4.12 presents the findings on the effectiveness of the ICT infrastructure at Ketraco essential for the effective management of Big Data throughout its lifecycle.

From the presentation, majority of the respondents expressed moderate to general agreement views regarding the existence of various ICT infrastructure tools that support Big Data management. For instance, 26 (44.81%) and 18 (31.03%) of the respondents moderately concurred on the existence of data storage platforms and e-archiving repositories. Equally, 39 (67.24%) respondents comprising 21 (36.21%) who agreed, and 18 (31.03%) who strongly agreed positively pointed out *the availability of a network infrastructure for data communication and sharing, a key ICT infrastructure for big data management*. Thus, Ketraco is moderately equipped with the right infrastructure that plays a significant role in data management. For instance, by having a strong and reliable internet and network infrastructure, users can easily communicate and share data amongst themselves and across systems. Therefore, the Company is in the right direction in implementing an effective data management program if the existing infrastructure is effectively integrated with other data management initiatives. The ICT infrastructure such as the networks, cloud solutions, and systems were also mentioned by the 10(100%) of the respondents interviewed affirming the existence of the ICT initiatives at Ketraco.

These findings correlate with the study by Boore, Mutiso, and Obonyo (2021) who examined the potential of utilizing big data in the healthcare sector in Kenya using a scope review research approach. In this research approach, the researchers mapped the types of evidence, key concepts, and gaps in the defined area of research. The study established that Kenya being a lower middle-income nation experiences numerous challenges in

implementing national programs across several sectors. It further revealed that the implementation of Information Technologies in Kenyan public sectors has majorly favoured sectors in Finance (Integrated Finance management systems), Energy (Metering), and transport and citizen services (Huduma). However, sectors such as agriculture, health, and education are often faced with budgetary constraints. Therefore, it can be concluded that Ketraco, being in the Energy sector enjoys government support in terms of funding to finance and implement its projects and can easily incorporate Big Data into its annual work plans.

4.7.1 Security of Big Data Management

The researcher, through open-ended questions on both the surveyed and interviewed respondents, shared security measures put in place to enhance the security of Big Data. The findings were presented in fig 4.16.

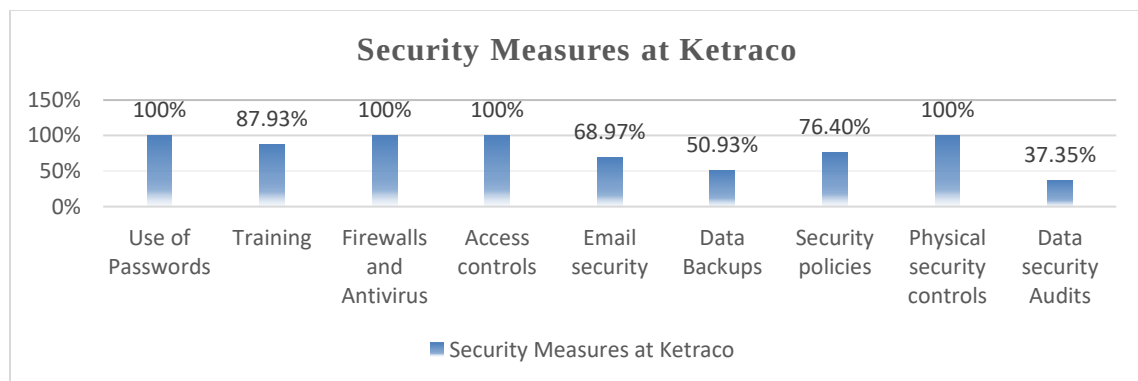


Figure 4. 16: Security of Big Data Management (n=58)

The findings as illustrated in Figure 4.16 revealed that all the respondents agreed that Ketraco has established strategies that ensure security of its data assets. For instance, 58 (100%), 51 (87.93%), 58 (100%), 58(100%), 40(68.97%), 30(50.93%), 45(76.40%), 58(100%), 22(37.35%) reported that the Company uses passwords, trains staff, has

installed firewalls and Antivirus, Access controls, Email security, Data Backups, Security policies, Physical security controls and Data security Audits respectively.

The findings revealed by the survey as illustrated in Figure 4.15 were reiterated by the System Analyst in charge of all Business Application modules [SABA] in the interview who pointed out:

The office runs regular data backups, and this is a good disaster management plan to help recover/restore Company data in case of disaster or data breach. Further, the system Analyst stated that Authentication, encryption, and use of passwords are enhanced in the sense that Ketraco staff are required to set up a strong password to access their laptops and printers. In addition, they are prompted to change these passwords monthly'. The challenge could be when staff carry laptops home or access data remotely. In this case, it becomes difficult to control what staff share with external parties especially when they choose to use their personal emails.

All the interviewees 10(100%) agreed that physical security was also enhanced by locking laptops on workstations and data centres and server rooms are properly secured with locks and alarms, and Staff ID badges are issued to staff to control unauthorized access.

According to Talesh (2018) on data security concerns, arguably stated that 95% of cyber security breaches result from human error. Thus, it is vital to train employees and other information users on the action plan relating to identification and reporting signs of data breach or leak. It is likely for data security to be enhanced in an environment where employees have requisite information and support best practices such as password policies, email best practices, and disaster recovery initiatives. Kumar, Raj, and Jelciana (2018) found that data security is a significant segment of any data management initiatives in an organization to the extent that it should be part of the strategic plan.

The divergent security measures observed at Ketraco correlate with the findings in the study by Chen and Zhao (2012) on Data privacy and security protection issues in cloud computing who found that it is difficult to develop and deploy unified data security measures in an organization setup. This study by Chen and Zhao (2012) explored secondary literature on how different firms in the US safeguarded their data against threats. For instance, IBM developed an encryption scheme in 2009 that facilitates data to be processed without necessarily being decrypted, thus preventing privacy leakage during data processing and circulation. On the other hand, Talesh (2018) recommended a client and cloud-based privacy management tool that places users at the centre of data usage and control. The model allows users to control the usage and storage of their sensitive data in the cloud. Consequently, Abouelmehdi et al. (2017) conducted an empirical review of privacy and security issues of Big Data in the healthcare sector in Indiana State in the United States and found that various measures were put in place by various private and public healthcare facilities to protect individual data included authentication, Access control, Encryption, and data masking, which involves replacement of sensitive data items with an unidentifiable value.

It is observed that Data security concerns exist in any organization operating in the current digital environment. The fact that many organizations have taken multiple initiatives to protect their data implies that data security has attracted global attention regarding the protection of data. Therefore, divergent approaches that keep up with the changing trends are encouraged to help protect individual and organizational data.

4.8 Collaborative Partnerships in Big Data Management at Ketraco

Companies across the world work alongside development partners, service providers, and technological partners among others to share resources, skills, and knowledge (Li and Nguyen, 2017). Ketraco, being a competitive energy company and a key stakeholder in Kenya's energy sector does not work in isolation. Instead, the company collaborates with partners to help achieve its mandate. It was therefore appropriate to explore big data collaborations and how they contribute to the performance of the Company. The Community Capability Model (CCM) assesses the capability of a community to perform a certain task collectively and suggests ways of improving that capability. The community capability elements include the elements and resources possessed by a community including the skills, information, physical and financial assets that can be used for its empowerment (Lyon, et al, 2012). In relation to big data, it was necessary to explore the collective approach employed by Ketraco to create, manage, and mutually benefit from the big data community with other stakeholders. Thus, the researcher posed several questions to the senior management team to shed light on different forms of collaboration at Ketraco, the impact of these collaborations, the benefits and the challenges, that the Company derived from these partnerships if they exist. The management team through the interviews, highlighted several forms of collaborative partnerships with service providers, partners, and research institutions, among others. One of the notable partnerships identified was with the information technology service providers who get into formal contracts with Ketraco, especially in offering services and licenses in data storage and System Application Integrations. According to Raptis, Passarella, and Conti (2019), many technology providers are providing big data solutions to help Companies address business challenges. A growing

number of technology providers are offering big data- and analytics-powered solutions that help organizations solve business challenges. Thus, Ketraco is tapping and integrating the already available business applications into their operations to achieve efficient workflows and business processes. Besides this form of partnership that was highlighted from the interviews, the study also identified other forms of partnerships that were elaborated in figure 4.17.

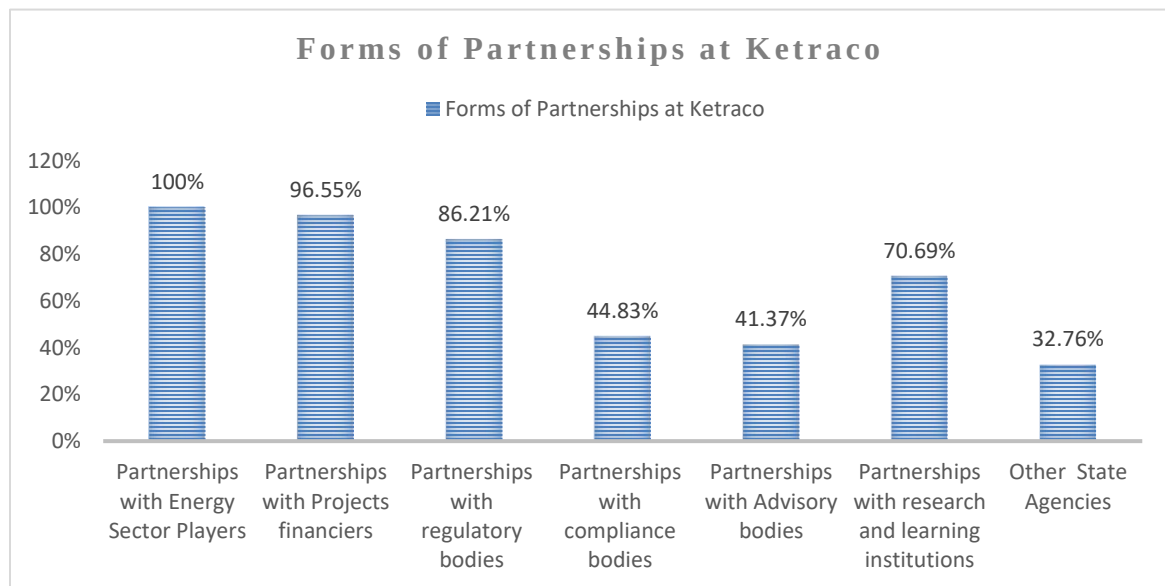


Figure 4. 17: Forms of Partnerships at Ketraco (n=58)

From the study, as illustrated in Figure 4.17, the respondents surveyed 58 (100%) indicated that Ketraco has partnered with stakeholders in the Energy Sector to share data relating to the sector. In addition, 56(96.55%), 50(86.21%), 26(44.83%), 24(41.37%), 41(70.69%) and 19(32.66%) of the surveyed respondents indicated to share data with Financiers and Service providers, regulatory bodies, Compliance Bodies, Advisory Agencies, Research and Learning Institutions, and other state Agencies respectively. Similar observations were made with the senior management team who highlighted the key role of other State

Agencies in enabling Ketraco to perform its Mandate. For instance, the Manager of Corporate Planning [MCP.] testified:

Ketraco does not operate in Isolation; we exchange data and collaborate with our sister agencies even in developing key documents such as strategic plans and performance contracts, and in developing regulations around the Energy sector.

From the findings, it is evident that Ketraco achieves its goals through a collaborative effort with other key energy sector players such as Kengen, Kenya Power, independent power producers, regulatory agencies, government parastatals and agencies, research institutions, and financial institutions like World Bank and private partners. Data sharing at Ketraco has resulted in collaborative growth. For instance, when the Company intends to undertake a national project in a certain part of the Country, it gets data, especially on Land ownership from the National Land Commission and the Survey of Kenya which facilitate the tabulating of accurate data relating to land ownership in that region traversed by the project. This data is critical to helping the company compensate the rightful landowners, eliminate fraudsters, limit conflicts with landowners, and foster mutual coexistence within communities affected by Company projects. According to an interview with the head of project development services; in return, this results in project acceptance in a community, as well as project completion within the specified period.

It is evident from the findings that all directorates at Ketraco appreciate data sharing and collaboration with other Agencies to unlock unprecedented success and growth. In addition, the types of data exchanged vary greatly from raw data to more refined types of data. Irrespective of the data sets and types, Ketraco has leveraged data collaboration to foster innovation and optimize decisions, implying that collaboration is essential to enhance sustained operations while streamlining the company to operate with the highest

standards of transparency and accountability. The case at Ketraco is in line with the findings of the study by Yego (2014) who revealed that public institutions in Kenya are promoting engagement and collaboration on data exchange especially on citizen data. The study administered questionnaires to 50 selected public institutions and explored how these entities shared data amongst themselves, and analysis done using descriptive statistics utilizing the SPSS Software revealed that departments of Registrations of Bureau, Immigration, the Kenya Revenue Authority, Higher Education Loans Board, as well as Criminal investigation offices and the National Police Service work collaboratively to exchange citizen data majorly on obligation or request.

However, there are several challenges related to big data management during the collaboration such as confidentiality issues, corporate restrictions on data shared, incompatibility across systems, and restrictions on data protected under the Data Protection Act (2019). Similar observations were made by Yego (2014) who pointed out the lack of clear standards, policy, communication protocol, data security, and incompatibility and interoperability issues as the major challenges affecting Big Data during collaborative partnerships across public institutions.

For Ketraco's case, the system analyst in charge of Business applications indicated that the Company had established controls and measures to protect its data assets. For instance, the already developed password policy, user access controls in place, policy for use of the company's computers, physical access policy in its offices, and restriction for downloading and installing programs onto the Company's laptops. Therefore, there is an opportunity for improving information security through the creation of big data awareness, regular review

of guidelines and compliance, training, and feedback collection on areas that require improvement.

The biggest challenges that Ketraco faces in big data management relate to process, negligence, and lack of resources that limit investment in advanced systems. The management teams were of the view that additional budgetary allocations, exchange programs and improved training on various aspects of Big Data are essential to support the big data agenda.

4.8.1 Challenges Facing Big Data Management at Ketraco

This section examined possible challenges that affect big data management at Ketraco. Respondents were asked to rate the level of Agreement on various Big Data management at Ketraco on a five-point Likert Scale whereby 1 denotes strongly disagree, 2= Disagree, 3= Not sure, 4= Moderate, 5=Agree; 6=Strongly Agree. The findings were computed and presented in Table 4.13 below.

Table 4. 13: Challenges Facing Big Data Management at Ketraco (n=58)

Big Data Management Challenges at Ketraco	<i>1 Strongly Disagree</i>	<i>2 Disagree</i>	<i>3 Not sure</i>	<i>4 Moderate</i>	<i>5 Agree</i>	<i>6 Strongly Agree</i>	<i>N</i>
Privacy and confidentiality issues in Big Data sharing	0	0	2 (3.45%)	2 (3.45%)	11 (18.97%)	43 (74.14%)	58
Insufficient understanding of Big Data	3 (5.17%)	2 (3.45%)	11 (18.97%)	3 (5.17%)	30 (51.72%)	9 (15.52%)	58
Absence of Organization support in advocating for Big Data practices	4 (6.90%)	3 (5.17%)	3 (5.17%)	2 (3.45%)	18 (31.03%)	28 (48.26%)	58
Complexity in handling large volumes of data	1 (1.72%)	2 (3.45%)	5 (8.62%)	3 (5.17)	12 (20.69%)	35 (60.34%)	58
Inadequate ICT tools that support data storage and sharing	0	1 (1.72%)	5 (8.62%)	1 (1.72%)	30 (51.72%)	21 (36.21%)	58
Difficulty in ascertaining elements such as orderliness, accuracy, completeness, trustworthy in data	1 (1.72%)	1 (1.72%)	2 (3.45%)	3 (5.17)	21 (36.21%)	30 (51.72%)	58
Absence of clear policies and procedures that guide data administration	0	1 (1.72%)	3 (5.17)	5 (8.62%)	7 (12.07%)	42 (72.41%)	58

The findings in Table 4.13 above indicate that the majority of respondents represented by 11 (18.97%) agreed and 43 (74.14%) strongly agreed that privacy and confidentiality issues are major hindrances to big data management. In addition, and by combining respondents who agreed and strongly agreed, the study revealed that lack of clear policies and procedures, difficulty in ascertaining data reliability, elements such as orderliness, accuracy, completeness, trustworthy in data, complexity in handling large sets and Inadequate ICT infrastructure were also found to hinder Big Data as this was represented by 49 (84.48%), 51 (87.93%), 47 (81.03%) and 51 (87.93%) respectively. Similarly, respondents 46 (79.31%) and 39 (67.24%) mentioned that lack of organization support as well as an inadequate understanding of Big data to an extent affected Big data implementation at Ketraco. The interviews provided divergent views regarding challenges facing Big Data management at Ketraco. For instance, while the Human Resources Services {MHRS} pointed out the diversity of big data and its voluminous nature as a major challenge, the Manager Corporate Planning {MCP} affirmed inadequate policies as the major hindrance to Big Data adoption at Ketraco.

The observations at Ketraco were similar to those recorded by Kwanya (2014) in his study on Big Data in the management of land records in Kenya, who conducted an exploratory study on the users of the lands information management system. Despite the researcher pointing out the high viability of performance and transactional management that comes with Big Data in Kenya, the study also noted that the viability of an information management system is hampered by inadequate skills and infrastructure, organizational structure, as well as culture. Similarly, Hersi (2022) in a study to establish possible challenges affecting the adoption of big data analytics among the telecommunication firms

in Kenya established that managerial challenges, technological challenges, process challenges, and technological challenges as key factors affecting the adoption of Big data analytics. From this study, and those of Kwanya (2014), and Hersi (2022), it can be concluded that establishing robust solutions to these challenges through capacity building, institutional strengthening, and essential infrastructure development will enhance the viability of any data management program.

4.9 Chapter Summary

Companies today operate in a highly dynamic business environment characterized by intense competition, rapid innovation, and changing consumer needs and preferences. Firms therefore operate under intense pressure to streamline their processes and pursue efficiency and sustainable performance. In the current digital world, organizations are bombarded with millions of structured and unstructured data. Data has become a huge resource that every organization, large or small, strives to gather, analyze, and transform into meaningful use. Big data is a new phenomenon that allows companies to gather millions of data from different sources including search engines, emails, social media, customer service, and other platforms. Companies that can gather, analyze, and store big data have been positioned for success through understanding the changing market needs, consumer preferences, and industry opportunities, among others.

The study sought to examine the application of big data at Ketraco. The data was collected through the administration of questionnaires to employees across all departments at the Company including ICT, procurement, Finance and accounts, technical, human resource management, communications, legal, and corporate planning. The researcher also

interviewed management leaders to shed light on the organization's course and roadmap toward adopting big data and leveraging its functionalities to achieve operational efficiency and competitiveness. This section presents a detailed analysis of the qualitative and quantitative data for inference. The study results are also discussed in the context of previous research to identify patterns of similarities and differences. Finally, the study reviews key challenges that hinder Big Data adoption at Ketraco with a view to proposing possible solutions for its effective implementation.

CHAPTER FIVE

SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter aims to summarize the findings, conclusions, and recommendations based on the study's research questions and findings. The study aimed to investigate Big Data Management at Ketraco with the view of proposing interventions to improve the management, sharing, and reuse of big data to facilitate proper planning, improved organizational processes, and better decision-making. The study adopted a concurrent mixed method approach collecting both quantitative and qualitative data using questionnaires and interviews as sources of data collection.

The research questions were used as the organizing themes for chapter five under the following subject headings: summary of findings, conclusion, recommendations, contribution and originality of the study, and suggestion for future research.

5.2 Summary of the Findings

The summary of findings covers the current state of Big Data management at Ketraco, analyses the existing infrastructure for big data, the influence of collaborative partnerships on big data management, and proposes strategies for effective big data management.

5.2.1. Demographic Information of Respondents

The study found out that the majority of the respondents 32 of the surveyed respondents were male, while 26 were female, while the ten interviewed had an equal representation of 5 from each gender. The findings further indicated that the respondents had education qualifications ranging from certificate (5), diploma (13), Bachelor's (34) Masters (14), and finally PhD (2).

Each department was represented in the study from Finance (4), Technical (29); with Supply Chain, Corporate Planning, and Corporate Communications each having a representation of 2 respondents in the survey. Further, the Department of Human Resource & Administration had a representation of 13 respondents, while the legal and ICT departments were each represented by 3 respondents in the survey. On the other hand, the composition of the 10 interviewed respondents included System analysts (4), Senior Executive Principal Office Administrator from the Executive Office of the MD (1), and the senior management team (5).

Majority of the respondents were from low-level management (35) while the top management and the middle-level management had a representation of 5 and 28 respondents respectively.

The study also revealed that majority of the respondents (38) had worked in the company for a period of 5 to 8 years and had the requisite understanding of the company's workflows and operations to give reliable and credible information.

5.2.2 Capture, Appraisal, Description, Preservation, Accessibility, Re-use, and Sharing of Big Data.

This section provides a summary of the findings of the first research question which sought to find out how Ketraco captures, appraises, describes, preserves, accesses, and uses its big data. The research question was addressed by the empirical part of the study with findings recorded from the survey conducted on Ketraco employees. The findings presented in this section cover the current state of big data from creation/receipt through use to ultimate disposal.

There was a consensus from the responses that at Ketraco, data entry and scanning were the common techniques for data capture as revealed by the majority (45) of the respondents. The captured data is thereafter managed in PDF, word, and JPEG formats as this was confirmed by all (68) respondents. Other formats highlighted by respondents included Database (XML, CSV) (26) and Geospatial formats (21). The utilization of various data formats indicated that the company is on a positive trajectory toward the adoption and implementation of big data management initiatives.

Further, the study revealed that data appraisal is rarely undertaken since a substantive number of respondents (41) confessed not to subject data received/created to any form of appraisal. This was also noted by the management team who anonymously agreed that inadequate policy framework and guidelines were major contributors to poor subjection of data to appraisal. Equally, the majority (41) of the respondents remained pessimistic and unaware of the company's data appraisal mechanisms and few (17) participants who understood the appraisal policy could not point out whether the process follows procedural mechanisms such as inventory worksheets and appraisal checklists. It was evident therefore that Ketraco had not established proper mechanisms and guidelines regarding data appraisal. According to Patil and Yogi (2011), data appraisal and validation are important processes in managing big data as they ensure that the collected data is evaluated and subjected to examination for accuracy, completeness, and truthfulness. Since Ketraco gathers data from multiple sources, it was therefore necessary to subject this data to appraisal before it is validated for storage.

Regarding data description, it was vital to define how data description is undertaken as this informs other processes such as access, sharing, and use. The researcher incorporated the Data Consistency Constraint model to understand how data description is undertaken at Ketraco. The model shows coherence and relationships among elements to capture underlying patterns, dependencies, and relationships in data. For instance, the coherence in data subject, date, title, author, and reference is crucial in deriving relationships and hence informing the description of these sets of data. The findings revealed that the description approach is mostly at the discretion of the author or originator of data, implying that the Company lacks a standard approach to data description. Even though all the respondents agreed to use subjects as a primary element for data description, there were no restrictions for incorporating other metadata elements in the data description. For instance, whereas 40 of the surveyed respondents pointed out that they describe data using document reference numbers, a substantive number of respondents also indicated other elements such as Author's name (37), title (32), and dates (25) for data description. Therefore, for the company to effectively utilize its existing data, there is a need to establish tools and models that systematically describe the metadata of its data sets, and this eases the subsequent stages of data storage, preservation, access, use, and re-use.

The study further explored existing data storage platforms and examined the person responsible for the day-to-day storage of this data. From the findings, Ketraco utilizes its own servers and database to manage and store data. A shared online (SharePoint) is also available for seamless sharing of data by users across the organization. SharePoint is a cloud-based platform that allows Ketraco employees to access, store, and share data with their colleagues. The company also utilizes Microsoft Office 365 as a platform for internal

collaboration, storage, and online sharing of data. On the same note of data storage, other respondents indicated storing data on corporate emails, external storage drives, and personal laptops.

The existence of multiple and uncoordinated platforms for data storage as revealed in the study poses a risk to the confidentiality of the data as employees may be targeted by hackers seeking to infiltrate the company's systems (Dezi et al., 2018). Therefore, there is a need for a prudent storage system at Ketraco that can offer adequate storage capacity with minimum latency for large volumes of users. Deb and Choudhury (2021) recommend the use of data warehouses and cloud storage. A data warehouse works like a typical warehouse for the storage of goods. This is a building facility that serves to store and process data at the enterprise level. It also offers a valuable tool for the analysis of data and supports other functions such as reporting, data analytics, business intelligence, data mining, and cyber monitoring among others (Deb and Choudhury, 2021). On the other hand, Srinivasan et al. (2015) offer a contrary solution that involves the use of cloud-based solutions. Unlike data warehouses, cloud-based storage offers large-scale storage capabilities for the storage of data in the Cloud. Cloud-based data storage is important as it allows seamless access to the data from anywhere eliminating the need for direct access to the office, computers, or hard drives.

Regarding data backup, majority of the respondents (44) were aware of the existence of data back up and they undertook both manual and automated backups. It was noted that the manual backup was routine-based and was performed regularly, while the automated one was undertaken by automated software to minimize human intervention. However, half of

the respondents 34 indicated that they were not aware of the person responsible for making decisions relating to data preservation and archiving. Others were divided on whose responsibility to undertake backups. For instance, some respondents (14) claimed it is the responsibility of the ICT to undertake backups while others (21) stated that it is the responsibility of the records management staff to ensure data backup and preservation.

The study also established that Ketraco has set up Access controls to authenticate and authorize staff to access and use the information as this was supported by a 100% response rate. However, it was also established that external agencies have access to company data though their access is limited to data classified as not confidential, and only relevant to the targeted external audience. The existence of access controls implied that the Company had considered emerging issues relating to data security and privacy. Regarding the users of data, all the respondents (68) regarded Ketraco staff as the users of big data, and none rated Ketraco's big data usability as poor. For instance, 42 respondents rated the usability of Big Data as very good, followed by Good (28), fair (16), and (14) rated it as excellent. The major uses of company data as revealed by all respondents in the study included their pivotal role in supporting company operations, supporting tactical, operational, and strategic decisions, measuring performance, and evaluating its operations. Further, the respondents unanimously agreed with the existence of multiple channels for data sharing by indicating that emails (Outlook), One Drive, and SharePoint were the common channels for data sharing. The key partners in data sharing included the internal staff as well as external partners such as development partners, partners in the Energy sector, government agencies, suppliers, and members of the public, especially those affected by Ketraco projects.

The study also revealed the key challenges hindering data sharing, including the employees' perception of seeing information as a source of power and hence holding onto it. Also, respondents highlighted the aspect of fear of breaching the confidentiality of information as one of the key challenges that hindered data sharing. Other challenges included the issue of confidentiality and secrecy, the reporting structure that requires employees to seek authorization before sharing data, and the limited document size that limited data that could be shared using the company software platform.

5.2.3 Policies and Regulation for Big Data Management at Ketraco

The study sought to find out if there were any policies that guide Big Data management throughout the life cycle from creation, description, appraisal, preservation, access, use, and share. Regarding the legal framework, the study established that there were no internally approved data policies, as this was confirmed by majority of respondents (48). The absence of data policies as pointed out by respondents (48) makes it impossible to assign roles, responsibilities, and decision-making within Ketraco's data management architecture and structure. The responsibility for the management of data rests on various stakeholders within the organization, mostly the creators of those data sets, hence confidentiality of this data is at the creators' discretion. Therefore, the inability to define and assign roles and responsibilities in big data management hinders accountability in big data applications, especially in workflows.

Even though a lot of data is created at the individual level, the study revealed other vital data sets managed centrally by the IT and records management departments. This data includes contracts, procurement records, finance records, HR (Human Resources) records,

performance guarantees, Easements, and other legal documents. Similarly, the records management team is tasked with the transfer and updating of this data onto SharePoint, which is the company's online storage. However, there were no established guidelines that outline the categories of data that should be managed centrally by the IT and Records Management departments. Therefore, as revealed in this study, the state of the big data policy framework at Ketraco is an indication of the missing tight regulations at every level of big data management from the levels of data creation, use, user profile, and semi-automated decision-making processes.

5.2.4 Skills, Knowledge, and Training for Big Data Management

The study sought to explore employees' skills, knowledge, and training on Data capture, Appraisal, description, security, storage, and sharing since the ability to tap, and effectively utilize big data depends on the existing employee skills and Knowledge.

In relation to skills possessed by Ketraco staff, the study revealed that the majority of the respondents lacked vital big data management skills. For instance, respondents lacked metadata skills (51), data analysis (49), and collaboration skills (16). Equally, the respondents (37) were not sure whether their search and retrieval skills were adequate to help make a generalization on whether their skills would contribute effectively to Big data management. The study's revelation of the existence of skill gaps has a direct impact on the management of big data at Ketraco.

On the proficiency of using various big data software/tools, the study revealed that most employees at Ketraco have attained basic to intermediate skills in areas such as Microsoft Office Applications (54), data mining (21), and data analysis (25). However, the capability

of using various database management software critical for data analysis and storage was the most obscure as only 3 respondents had attained the minimum to advanced level of expertise in Hadoop and SQL database platforms.

The study recorded that Ketraco employees had a moderate understanding and basic knowledge of data capture, description, and security. However, the respondents agreed to possess inadequate knowledge, especially on data appraisal, security, storage, and data sharing. Further, the researcher also observed that the respondents who possessed requisite knowledge in data management functions and did not require training at every stage of big data management were mainly staff from the IT and the Records Management departments. This could be attributed to their career background and frequent interaction with data assets and platforms.

The study also revealed inconsistencies in training needs that lean towards Big Data. Despite annual training plans, there was no staff training on big data. For instance, the Ag. legal services manager noted in the interview that there were no training proposals either from management or staff themselves on big data. Therefore, lack or inadequate training leaves Ketraco employees less equipped with the skill and knowledge to perform assignments ranging from basic to complex data-related tasks.

5.2.5 Level of ICT Infrastructure Preparedness for Big Data Management

The study revealed that the ICT infrastructure at Ketraco is key to supporting the corporation's big data strategy as all the respondents (58) agreed that Ketraco has Reliable network infrastructure as well as software and hardware tools such as storage platforms that support Big Data management. The key infrastructure tools already established at

Ketraco include network infrastructure, software and programs, servers, personal computers, and databases. Consequently, the majority of respondents (39) agreed that the existing ICT infrastructure is reliable and efficient in supporting internal workflows in various Company directorates, hence allowing data to flow and be accessed within the systems seamlessly. From the findings, it can be noted that Ketraco is well-equipped with the right infrastructure that if utilized well can enhance the management of Big Data.

5.2.5.1 Security Measures for Big Data

The study established that security measures had been established to guarantee the safety and reliability of company data; this was confirmed with a 100% response rate. The security measures put in place as captured by the respondents included the use of passwords, installed firewalls and Antivirus, Access controls, Email Security, Data Backups, Security policies, Physical security controls, and regular Data Security Audits. The fact that authentication, encryption, and use of passwords are enhanced for Ketraco staff to access laptops, printers, registries, and strong rooms is an indication that the company has recognized data security and confidentiality as key issues in managing data. However, the freedom of carrying office laptops home or the ability to use personal emails on office laptops as captured by the interview response of the system analyst in charge of Business applications exposes the company data to threats of data breaches that could have a negative impact on the company especially if the data is accessed by unauthorized persons. Addressing this challenge will enhance the security of company data. Equally, the divergent security measures observed at Ketraco demonstrate the company's goodwill in promoting best practices regarding Big Data Management, and it is in the right direction to ensure it keeps up with the changing trends to protect individual and Company data.

5.2.6 Collaborative Partnerships for Big Data Management at Ketraco

The study through the interviews with the senior management team shed light on different forms of collaborative partnerships that exist at Ketraco, the impact of these collaborations, the challenges, as well as the benefits that the Company derived from these partnerships. There was a 100% response rate that Ketraco has partnered with several stakeholders within and outside the Energy sector. The partners include the Project financiers, Service providers and consultants, regulatory bodies, Compliance Bodies, Advisory Agencies, Research and Learning Institutions, and other state Agencies. The partnerships are mutually beneficial and help the company unlock unprecedented success and growth through timely decision-making. Through partnerships, Ketraco collaboratively works with other agencies in developing master plans, strategic plans, and performance contracts and collectively undertakes national government projects. Therefore, the company benefits immensely by utilizing big data to collaborate and make data-driven improvements.

However, the study revealed that this form of partnership comes with challenges that include a lack of clear standards, policies, communication protocols, data security concerns, systems incompatibility and interoperability, and restrictions that come with the Data Protection Act. In addition, there are other challenges arising from processes, negligence, and lack of resources limiting investment in advanced systems that protect company data.

5.2.7 Challenges Facing Big Data Management

There was a 100% response rate on the existence of challenges that affected Big Data management at Ketraco. The challenges identified in the study were summarized as follows:

The study revealed that Privacy and confidentiality were the main challenges as recorded by (54) respondents, lack of clear policies and procedures (49), complexity in handling large sets of data (37), insufficient understanding of big data (39), and difficulties in ascertaining the accuracy and completeness of data (30). The challenges pointed out were the major hindrance to enforcing big data management initiatives at Ketraco. The management team in the interviews also mentioned complexity in handling large data sets, unavailability of key skills required to utilize big data management tools, and inadequate policy framework as major setbacks for big data management at Ketraco. Precisely, managerial challenges, technological challenges, process challenges, and technological challenges were identified as key factors affecting the adoption of big data within Ketraco.

5.3 Conclusion

This section provides conclusions informed by the key findings of the study. Generally, the findings revealed that Ketraco had not adopted effective strategies for Big Data management, hence hindering its capability of deriving maximum value from its data assets. Also, there were inconsistencies and weaknesses in managing data across its lifecycle, thus data was not managed professionally as required by the Data Centre Curation model which dictates that every distinct stage should be guided by relevant policies to ensure successful curation and preservation of data. There were no procedures or policies that govern data management at the individual, directorate, or corporate level. For instance, there were no procedures for data creation, description, appraisal, preservation, access, use, and sharing data at Ketraco. Equally, the study revealed the unavailability of proper mechanisms that ensure regular review and subsection of company

data to quality assurance and checks. Thus, most of Ketraco's data could be duplicates, inaccurate, too much, or inconsistent. Therefore, much more is needed regarding data management as a strategic asset for Ketraco.

The study revealed the existence of limited development of human capital and physical capital to expand Ketraco's ability to realize the potential of big data. For instance, the majority of Ketraco staff lacked vital skills and basic knowledge, especially on key aspects of metadata capture, data analysis, search and retrieval skills, collaboration skills as well as skills of utilizing big data management applications and software. The skill gap can be attributed to inconsistencies and lack of staff training that lean towards big data; this leaves Ketraco employees less equipped with the skill and knowledge to perform assignments ranging from basic to complex data-related tasks.

The study found that Ketraco is well equipped with the right ICT infrastructure such as network infrastructure, software and programs, servers, personal computers, and databases. The existing ICT infrastructure is reliable and efficient in supporting internal workflows across various directorates, thus allowing data to flow and be accessed by users. The study findings also revealed the existence of collaborative partnerships on matters of Big Data with other stakeholders. Despite these partnerships being beneficial, all interviewees highlighted several challenges that arise while exchanging data with other parties. For instance, challenges such as confidentiality, lack of clear guidelines on data exchange, data privacy, and the need to adhere to communication channels and protocols were pointed out by the respondents.

Based on the study findings, there is a need for Ketraco to put flexible measures in place to support the full integration of big data into its operations and support the seamless sharing of data amongst employees and other agencies to create a solid foundation on which big data insights can be achieved. Many big data environments go beyond the traditional data warehouse and relational database platforms to accommodate innovative technologies that are suitable for processing and storing non-transactional types of data. Ketraco should follow suit and implement modern technologies that provide flexible storage platforms.

5.4 Recommendations

Based on the findings of the study, big data is inevitable to any organization in the current technological environment. An effective Big Data management strategy needs to be implemented to ensure the company derives maximum value from its data assets. Thus, the study recommended as follows based on the study themes;

5.4.1 Data curation for Big Data Management

Creation and Capture are among the first phases in managing big data according to the Data Centre Curation Lifecycle model (DCC) (Haggins, 2008). It is paramount for Ketraco's top leadership to initiate a comprehensive framework that includes policies and procedures that guide the entire process of data creation. The policy will help define the scope of data creation by outlining key elements relating to the type of data to be captured, sources of data, when and how data should be captured. In addition, Pouchard (2016) argued that having a structured approach to data creation provides a wide scope of data entry points

ensuring clearly defined methodology and systems are used in data creation, as well as clearly defined roles in data creation.

According to Pramanik et al. (2021), the lack of a unified approach affects the classification and storage of those data sets in appropriate formats and platforms. Therefore, this study recommends that Ketraco should establish appropriate standards that provide technical, descriptive, administrative, preservation, and structural metadata needed to describe corporate data adequately. Ketraco through the Records and ICT departments should undertake regular data appraisals as a way of maintaining information appropriately based on the value and usability of that information asset. Appraisal is vital in any information or data management program as it helps the organization select and preserve information with long-term value (Kimble and Milolidakis, 2015). Therefore, the appraisal of data for long-term curation and storage should be guided by well-documented policies, guidelines, and legal requirements.

Data reusability comprises the use of the existing data for multiple purposes while reducing redundancy and maximizing its value (Toshniwal, Dastidar, and Nath, 2015). Data use and reuse can also be enhanced when Ketraco develops and enforces data quality control measures. Quality control and Assurance are essential for managing company data. According to Gao, Xie, and Tao (2016), quality control measures help an organization maintain data integrity, adhere to regulatory standards, and implement robust processes.

5.4.2 Policies and Regulation for Big Data Management at Ketraco

Therefore, the study recommends that Ketraco management should establish a data governance framework that provides the required accountability, structure, and policies to

maintain high data quality standards. The policies should define the accepted standards such as data completeness, consistency, accuracy, and timeliness in addition to aligning with existing statutory regulations. The management should nominate data champions to oversee and reinforce governance policies and quality issues to ensure compliance and data quality. In addition, the data management platforms should be automated for alerts to notify quality assurance officers and champions when quality issues arise. This will notify and help data administrators respond to any challenges that arise. The data champions should also ensure regular data reviews and audits are undertaken to ensure compliance with governance policies and maintain data quality.

Data policies should be comprehensive to ensure scalability, compliance, security, and effective utilization of data assets (Kemp, 2014). Ketraco through the ICT should develop a data privacy and security policy and ensure the policy framework complies with the existing standards and legal requirements such as data privacy regulations. This implies that the policy framework should be tied with other governance policies to ensure data is effectively managed and utilized. The governance policy should define classification schemes to classify data based on criticality, sensitivity, and compliance requirements.

5.4.3 Skills, Knowledge, and Training for Big Data Management

There is a need to enhance skills, knowledge, and training for Big Data management at Ketraco. Tabesh, Mousavidin, and Hasani (2019) opined that employee must stay updated on the latest tools, trends, and best practices to manage vast amounts of data effectively. Therefore, Ketraco's Human Resource directorate should equip employees with skills vital to the management of big data across its lifecycle in their respective departments from data

acquisition skills, description, classification, appraisal, and storage through to final destruction. Sabaityte, Davidaviciene, and Karpoviciute (2020) are of the view that companies must try to equip their staff with data management skills to allow them to demonstrate and apply best practices consistently. The human resource directorate should also promote training programs that are tailored to the needs of every department.

In addition, the Human Resource directorate should encourage employees across the directorates to take part in soft skills development workshops, such as leadership, teamwork, and effective communication which are vital when working in cross-functional teams. Sabaityte, Davidaviciene, and Karpoviciute (2020) suggest that employee participation in industry certifications, such as those offered by Google and Microsoft, can further deepen and validate their expertise. The top leadership of Ketraco should promote an organizational culture that realizes the value of big data and ensure big data training programs are included in annual staff training plans.

A solid grasp of the big data lifecycle, from data collection to disposal, is necessary for managing big data throughout its pipeline (Kemp, 2014; Surbakti et al., 2020). Therefore, Ketraco should promote a big data management culture through internal collaborations, effective communications, best practices, and training to equip employees with the necessary knowledge for effective big data management.

5.4.4 ICT Infrastructure for Big Data Management

Investing in Information and Communication Technology can significantly enhance big data management for a company. According to Zhang et al. (2018), a comprehensive ICT approach focused on tools, systems, processes, infrastructure, and talent development can

enhance efficient data collection, analysis, storage, and utilization across the company. Therefore, Ketraco should invest in ICT infrastructure with a key focus on scalable cloud-based data storage solutions such as Google Cloud that allow for expansion as data volume increases.

Secondly, to ensure organizations derive much value from their data assets, Cui, Yu and Yan (2016) recommend the use of high-performing systems and tools that easily analyze and process large volumes of data. Therefore, the study recommends that Ketraco's ICT department should invest in high-performance computing solutions with high-performance data processing and analysis capabilities, and high-speed secure networks to ensure secure and efficient data communication, sharing, and transfer across the company.

5.4.5 Security Measures for Big Data Management

Data security, privacy, and confidentiality are critical issues for companies handling big data and sensitive information. Addressing security issues requires a multi-faceted approach that incorporates employee training, comprehensive policies and robust security measures, and employee training (Cuzzocrea, 2014; Jeske and Calvard, 2020). The ICT should develop a well-designed security infrastructure to protect sensitive data from breaches and unauthorized access. These measures include access controls and encryption to prevent unauthorized access to data both in transit and at rest, role-based access controls, multi-factor authentication, intrusion detection systems, and firewalls for network security (George et al., 2016).

The ICT should establish robust data security measures, including firewalls, encryption, and intrusion detection systems; according to Sheng et al. (2017), this is essential to protect

data against breaches and unauthorized access. This can be achieved by developing comprehensive privacy policies in compliance with relevant regulations such as data protection, access to information regulations, and the Constitution of Kenya 2010. Further, the policy should be in line with the data curation theory by integrating elements for data description, formats, interoperability, and integration (Pouchard, 2016).

5.4.6 Collaborative Partnerships for Big Data Management at Ketraco

There is a need for collaborative partnerships in Big Data management at Ketraco. Strategic partnerships can help the company manage large-scale data projects, facilitate innovation, and offer access to specialized expertise. It is vital to enhance collaborative partnerships, especially with key partners who directly or indirectly contribute and support Ketraco in undertaking its mandate. Chakphet et al. (2020) recommend companies to focus on building trust, sharing goals, and clearly communicating to enhance effective big data management. On their part, Akhtar, et al. (2022) and Surbakti et al. (2020) recommend several collaborative strategies for the end-to-end exchange of data across partners and companies, with the key focus on the scope, communication, sharing, and technology integration techniques for effective collaboration.

Any form of partnership for mutual benefit requires regular assessment to determine the success of desired goals (Chakphet et al., 2020). Therefore, for partnership success, the study recommends the need to regularly evaluate and measure the success of its partnerships to confirm if they meet the desired objectives. This can be achieved by Ketraco's top management to establish key performance indicators to measure outcomes with a focus on desired outcomes, conducting regular reviews, and relying on feedback to

assess the effectiveness and identify areas for improvement. The key partners the study identified included the stakeholders in the energy sector such as the state Agencies within the sector, the independent power producers, energy regulators, investors in the private-public partners (PPP), the contractors and suppliers, and the members of the public especially the ones Ketraco's projects traverse their lands.

By implementing these measures, Ketraco will manage its big data assets effectively, ethically, and in line with legal requirements, thus ensuring compliance, scalability, and security. The robust data framework fosters data-driven decision-making and supports business goals.

5.5 Contribution and Originality of the Study

From this study's findings, Big Data is an emerging field that requires a comprehensive approach to its effective utilization. However, little research has systematically explored data management from creation, description, classification, preservation and storage, security, access, use, and re-use, through to disposition. The results of this study contribute to theory and knowledge by adding to this grey area and enriching the field further by outlining the inputs required to manage data at distinct stages of its lifecycle. By using the premises of the data curation model together with the findings of this study, the study provides insights into the effective management of Company data assets by outlining how to clearly receive data, describe, and assign appropriate metadata, for easy storage and retrieval for use. The key practices pointed out in this study include uniformity in data capture by establishing standard formats for data creation, description, classification, and storage. These best practices support organizations in the systematic management of data

and ensure effective accessibility and utilization of data to support business processes and enhance productivity.

The body of knowledge is further enriched by the study findings which have revealed gaps in key data management practices that directly impact big data management. For instance, the study revealed inconsistencies or lack of employees with requisite skills in data appraisal and knowledge in developing policies and implementing information technology tools for processing and storing big, hence directly impacting big data management. The study's recommendations can be applied by organizations to formulate a big data management model that addresses these gaps and ensures maximum benefits from data assets.

The study serves as the basis for policymakers for future design and formulation of amendments that govern big data management. The study affirms the need for a comprehensive policy framework that outlines how data can effectively be utilized without compromising privacy and confidentiality issues. Policy makers can use this study to formulate realistic guidelines on how data should be processed and used internally within organizations, and to an extent how cross-organizational data collaboration can be enhanced without compromising confidentiality and privacy issues.

5.6 Suggestion for Further Research

Big Data management is a rapidly evolving field, with new methodologies, tools, and technologies emerging regularly. Therefore, future research in this area can help respond to the current challenges and derive massive opportunities that come with effective big Data management. The study recommends future research in data governance and

compliance since data governance regulations have become critical in the current business environment. The studies in this area can focus on establishing models that ensure compliance with existing local, national, and international regulations.

Another key area that requires further research is cloud computing and hybrid infrastructure for managing data. A hybrid infrastructure is a computing and data storage model that combines both private storage (clouds) and on-premises data centers. The approach allows companies to leverage the benefits of both the cloud and on-premises environments.

Data ethics and social responsibility are equally emerging areas for further research. Data ethics is basically the data-related practice that preserves and guides information usage in the digital age. It is a set of norms of behaviour that encourages suitable accountability and judgment when collecting, processing, using, or storing data with the objective of minimizing risks to the company, individuals, and society while maximizing the public good. Therefore, studies in this area will help companies establish frameworks that provide a systematic approach to ethical and responsible data usage.

Data sharing and collaboration are also interesting areas that should be enhanced with new studies that recommend effective approaches that can be adopted by companies to promote data exchange. Thus, future studies can explore new models for collaborative partnerships, with a key interest in cross-functional teamwork, secure data exchange, and data sharing protocols.

REFERENCES

- Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., & Saadi, M. (2017). Big data security and privacy in healthcare: A Review. *Procedia Computer Science*, 113, 73-80.
- Akhtar, P., Osburg, V. S., Kabra, G., Ullah, S., Shabbir, H., & Kumari, S. (2022). Coordination and collaboration for humanitarian operational excellence: big data and modern information processing systems. *Production Planning & Control*, 33(6-7), 705-721.
- Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, 60(3), 285-292.
- Ali, I., Sabir, S., & Ullah, Z. (2019). Internet of things security, device authentication and access control: a review. *arXiv preprint arXiv:1901.07309*.
- Allen, M. (2017). *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications.
- Al-Najran, N., & Dahanayake, A. (2015). A requirements specification framework for big data collection and capture. In *New Trends in Databases and Information Systems: ADBIS 2015 Short Papers and Workshops, BigDap, DCSA, GID, MEBIS, OAIS, SW4CH, WISARD, Poitiers, France, September 8-11, 2015. Proceedings* (pp. 12-19). Springer International Publishing.
- Ash, R. B. (2012). *Information theory*. Courier Corporation
- Asnawi, N. S., Abd Kadir, I. K., Ab Rahman, A., & Yunus, A. M. (2022). Records Management and Big Data Environment: The roles of records professional in managing big data. *Environment-Behaviour Proceedings Journal*, 7(S110), 213-217.
- Belov, V., Tatarintsev, A., & Nikulchev, E. (2021). Choosing a data storage format in the apache hadoop system based on experimental evaluation using apache spark. *symmetry*, 13(2), 195.
- Bengtsson, S., & Johansson, S. (2022). The Meanings of Social Media Use in Everyday Life: Filling Empty Slots, Everyday Transformations, and Mood Management. *Social media + Society*, 8(4). <https://doi.org/10.1177/20563051221130292>
- Bernard, H. R. (2017). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman & Littlefield.
- Bertino, E., Ghinita, G., & Kamra, A. (2011). Access control for databases: Concepts and systems. *Foundations and Trends® in Databases*, 3(1-2), 1-148.
- Bertot, J. C., & Choi, H. (2013, June). Big data and e-government: issues, policies, and recommendations. In *Proceedings of the 14th annual international conference on digital government research* (pp. 1-10).

- Betz, S., & Hall, R. (2015). Self-archiving with ease in an institutional repository: microinteractions and the user experience. *Information Technology and Libraries*, 34(3), 43-58.
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of b questionnaires used in social and health science research. *Nigerian Postgraduate Medical Journal*, 22(4), 195.
- Booker, P., Barnett, J., & Cribbin, T. (2016). Doing social media analytics. *Big Data & Society*, 3(2). <https://doi.org/10.177/2053951716658060>
- Boore, C., Mutiso, V. M., & Obonyo, V. (2021). Use of ehealth and big data in management of non-communicable diseases in developing Countries; the Kenyan healthcare sector.
- Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. Sage.
- Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, 70, 328-337.
- Brannen, J. (2017). Combining qualitative and quantitative approaches: an overview. *Mixing methods: Qualitative and quantitative research*, 3-37.
- Broeders, D., Schrijvers, E., van der Sloot, B., Van Brakel, R., De Hoog, J., & Ballin, E. H. (2017). Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. *Computer Law & Security Review*, 33(3), 309-323.
- Chakphet, T., Saenpakdee, M., Pongsiri, T., & Jermstittiparsert, K. (2020). The role of big data analytics in the relationship among the collaboration types, supply chain management and market performance of Thai manufacturing firms. *International Journal of Supply Chain Management*, 9(1), 28-36.
- Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, 314-347.
- Chen, D., & Zhao, H. (2012, March). Data security and privacy protection issues in cloud computing. In *2012 international conference on computer science and electronics engineering* (Vol. 1, pp. 647-651). IEEE.
- Chen, P-L., Cheng, Y., & Chen, K. (2018). 'Analysis of Social Media Data: An Introduction to the Characteristics and Chronological Process.' In *Big Data in Computational Social Science and Humanities* (pp 297-321). DOI: 10.1007/978-3-319-9546-3_16
- Colombo, P., & Ferrari, E. (2019). Access control technologies for Big Data management systems: literature review and future trends. *Cybersecurity*, 2(1), 3.
- Cox, A. M., & Tam, W. W. T. (2018). A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, 70(2), 142-157.

- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks: SAGE.
- Cricelli, L., & Grimaldi, M. (2010). Knowledge-based inter-organizational collaborations. *Journal of Knowledge Management*, 14(3), 348-358.
- Cui, L., Yu, F. R., & Yan, Q. (2016). When big data meets software-defined networking: SDN for big data and big data for SDN. *IEEE network*, 30(1), 58-65.
- Cuzzocrea, A. (2014, November). Privacy and security of big data: current challenges and future research perspectives. In Proceedings of the first international workshop on privacy and security of big data (pp. 45-47).
- Cyr, J. (2016). The pitfalls and promise of focus groups as a data collection method. *Sociological Methods & Research*, 45(2), 231-259.
- Data Protection Commission (2019). *Quick Guide to the Principles of Data Protection*. Retrieved from: https://www.dataprotection.ie/sites/default/files/uploads/2019-11/Guidance%20on%20the%20Principles%20of%20Data%20Protection_Oct19.pdf
- Davenport, T. H., Barth, P., & Bean, R. (2012). *How 'big data' is different*. MIT Sloan Management Review.
- Deb, M., & Choudhury, A. (2021). Hybrid cloud: A new paradigm in cloud computing. *Machine learning techniques and analytics for cloud security*, 1-23.
- Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A. S., & Satapathy, S. C. (Eds.). (2018). *Internet of Things and big data analytics toward next-generation intelligence*. Springer International Publishing.
- Dezi, L., Santoro, G., Gabteni, H., & Pellicelli, A. C. (2018). The role of big data in shaping ambidextrous business process management: Case studies from the service industry. *Business Process Management Journal*, 24(5), 1163-1175.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.
- Esser, F., & Vliegenthart, R. (2017). Comparative research methods. *The international encyclopedia of communication research methods*, 1-22.
- Esteva, M., Sweat, S., McLay, R., Xu, W., & Kulasekaran, S. (2016, June). Data curation with a focus on reuse. In *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)* (pp. 45-54). IEEE.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1), 1-4.
- Fang, Z., Chen, J., Yi, M., Wu, Z., & Qian, H. (2010, November). Cloud computing business model based on value net theory. In *2010 IEEE 7th International Conference on E-Business Engineering* (pp. 462-469). IEEE.

- Flick, U. (2014). Mapping the field. *The SAGE handbook of qualitative data analysis*, 1, 3-18.
- Francis, H. M., Osborne-Crowley, K., & McDonald, S. (2017). Validity and reliability of a questionnaire to assess social skills in traumatic brain injury: a preliminary study. *Brain injury*, 31(3), 336-343.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 2007(2012), 1-16.
- Gao, J., Xie, C., & Tao, C. (2016, March). Big data validation and quality assurance--issues, challenges, and needs. In *2016 IEEE symposium on service-oriented system engineering (SOSE)* (pp. 433-441). IEEE.
- Gathu, P., & Kahindi, H. (2015). Access to information in Kenya. *Adili-Transparency International Kenya*, 1-12.
- George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research.
- Gharaibeh, A., Salahuddin, M. A., Hussini, S. J., Khreishah, A., Khalil, I., Guizani, M., & Al-Fuqaha, A. (2017). Smart cities: A survey on data management, security, and enabling technologies. *IEEE Communications Surveys & Tutorials*, 19(4), 2456-2501.
- Ghiringhelli, L. M., Vybiral, J., Ahmetcik, E., Ouyang, R., Levchenko, S. V., Draxl, C., & Scheffler, M. (2017). Learning physical descriptors for materials science by compressed sensing. *New Journal of Physics*, 19(2), 023017.
- Griffin, A. J. (2021). *A seat at the table: A phenomenological study of the gap in African American/Black women with nonprofit executive leadership roles*. Seattle University.
- Hancock, D. R., Algozzine, B., & Lim, J. H. (2021). Doing case study research: A practical guide for beginning researchers.
- Haq, M. I. U., Li, Q., & Hassan, S. (2019). Text mining techniques to capture facts for cloud computing adoption and big data processing. *IEEE Access*, 7, 162254-162267.
- Hathcoat, J. D., & Meixner, C. (2017). Pragmatism, factor analysis, and the conditional incompatibility thesis in mixed methods research. *Journal of Mixed Methods Research*, 11(4), 433-449.
- Hersi, M. (2022). Challenges Affecting Adoption of Big Data Analytics in Telecommunication Firms in Kenya (Doctoral dissertation, University of Nairobi).

- Higgins, S. (2008). The DCC curation lifecycle model. *International journal of digital curation*, 3(1), 134-140.
- Höchtel, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 147-169.
- Horta, E. G., Castro, C. L. D., & Braga, A. P. (2015). Stream-based extreme learning machine approach for big data problems. *Mathematical Problems in Engineering*, 2015.
- Hussain, T., Sanga, A., & Mongia, S. (2019, October). Big data hadoop tools and technologies: A review. In *Proceedings of International Conference on Advancements in Computing & Management (ICACM)*.
- Hussein, H. (2022). Interview Method. In *Principles of Social Research Methodology* (pp. 207-219). Singapore: Springer Nature Singapore.
- Information Today Inc. (2020). *Video Will Grow to 82% of Internet Traffic by 2022: InterDigital/Futuresource Report*. Retrieved from: <https://www.streamingmedia.com/Articles/ReadArticle.aspx?ArticleID=144177>
- Jeske, D., & Calvard, T. (2020). Big data: lessons for employers and employees. *Employee Relations: The International Journal*, 42(1), 248-261.
- Kabata, V., & Garaba, F. (2019). Leadership and political will for implementation of the access to information (ATI) Act (2016) in Kenya. *Records Management Journal*, 29(1/2), 117-133.
- Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In *2013 Sixth international conference on contemporary computing (IC3)* (pp. 404-409). IEEE.
- Katuu, S. (1999). Appraisal of electronic records: The path for African archivists. *African Journal of Library Archives and Information Science*, 9, 69-74.
- Kaur, K., & Sachdeva, M. (2017, January). Performance evaluation of NewSQL databases. In *2017 International Conference on Inventive Systems and Control (ICISC)* (pp. 1-5). IEEE.
- Kemp, R. (2014). Legal aspects of managing Big Data. *Computer Law & Security Review*, 30(5), 482-491.
- Kenyalaw.org (2021). *Data Protection Act, 2019*. Retrieved from: https://kenyalaw.org/kl/fileadmin/pdfdownloads/LegalNotices/2021/LN263_2021.pdf
- KETRACO (2018) Website. Retrieved from: <https://www.ketraco.co.ke/about/history.html>
- Kibe, L. W., Kwanya, T., & Owano, A. (2019). Techniques and Tools of Big Data Analytics at the Technical University of Kenya and Strathmore University.

- Kim, H. Y., & Cho, J. S. (2018). Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of Business and Retail Management Research*, 12(3).
- Kim, J. (2013). Data sharing and its implications for academic libraries. *New Library World*.
- Kimani, k. A. (2015). *Impact of information technology on organizational performance: case of population services Kenya* (doctoral dissertation, school of business, University of Nairobi).
- Kimble, C., & Milolidakis, G. (2015). Big data and business intelligence: Debunking the myths. *Global Business and Organizational Excellence*, 35(1), 23-34.
- Kitchin, R., & McArdle, G. (2016). What data librarians. *Journal of Librarianship and Information Science*, 49(1), 3-14.
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., ... & Wächter, J. (2006). Data publication in the open access initiative. *Data Science Journal*, 5, 79-83.
- Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3-14.
- Krosnick, J. A. (2018). Questionnaire design. In *The Palgrave Handbook of Survey Research* (pp. 439-455). Palgrave Macmillan, Cham.
- Kshetri, N. (2014). The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*, 1(2), 2053951714564227.
- Kshetri, N. (2016). Big data's role in expanding access to financial services in China. *International journal of information management*, 36(3), 297-308.
- Kumar, P. R., Raj, P. H., & Jelciana, P. (2018). Exploring data security issues and solutions in cloud computing. *Procedia Computer Science*, 125, 691-697.
- Kumar, R. (2019). *Research methodology: A step-by-step guide for beginners*. Sage Publications Limited.
- Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. *J Big Data* 6(111) <https://doi.org/10.1186/s40537-019-0268-2>
- Kumar, V. (2018). A model for content enrichment of institutional repositories using Linked Data. *Journal of Web Librarianship*, 12(1), 46-62.
- Kwanya, T. (2014). Big data in land records management in Kenya: a fit and viability analysis. In *International Conference on Knowledge Management in Organizations* (pp. 15-24). Springer, Cham.
- Kwanya, T. (2014). Big data in land records management in Kenya: A fit and viability analysis. In *Knowledge Management in Organizations: 9th International*

- Conference, KMO 2014, Santiago, Chile, September 2-5, 2014, Proceedings 9* (pp. 15-24). Springer International Publishing.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032-2033.
- Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care*, 4(3), 324.
- Lewis, M. (2010). Libraries and the management of. *Envisioning future academic library services: Initiatives, ideas and challenges*, 145.
- Li, M., & Nguyen, B. (2017). When will firms share information and collaborate to achieve innovation? A review of collaboration strategies. *The Bottom Line*, 30(1), 65-86.
- Lynch, C. (2008). Big data: How do your data grow?. *Nature*, 455(7209), 28.
- Lyon, L., Ball, A., Duke, M., & Day, M. (2012). Community Capability Model Framework. 2015-06-22]. [https://community-model, sharepoint, com/Documents/CCMDIRWhitePaper-24042012. pdf](https://community-model.sharepoint.com/Documents/CCMDIRWhitePaper-24042012.pdf).
- Macorr Online Research Calculator (2019). Retrieved from; <https://www.macorr.com/sample-size-calculator.htm>
- Magina, A. (2019). *Access to Information in Kenya: a Critical Analysis of the Access to Information Act 2016* (Doctoral dissertation, University of Nairobi).
- Malhotra, N. K., Nunan, D., Birks, D. F. (2020). *Marketing Research: Applied Insight*. Pearson.
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: guided by information power. *Qualitative health research*, 26(13), 1753-1760.
- Mao, Z., Wu, J., Qiao, Y., & Yao, H. (2022). Government data governance framework based on a data middle platform. *Aslib Journal of Information Management*, 74(2), 289-310.
- Mariani, M., & Baggio, R. (2022). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 34(1), 231-278.
- Marsh, C. (2025). *The survey method: The contribution of surveys to sociological explanation*. Taylor & Francis
- Mburu P (2023) Queries over Sh12bn paid by Ketraco for stalled projects. Daily Nation publication dated Monday, January 02, 2023. Retrieved from Queries over Sh12bn paid by Ketraco for stalled projects | Nation
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.

- McNeely, C. L., & Hahm, J. O. (2014). The big (data) bang: Policy, prospects, and challenges. *Review of Policy Research*, 31(4), 304-310.
- Meru, A. K., & Kinoti, M. W. (2022). Digitalisation and public sector service delivery in Kenya. In *Digital service delivery in Africa: Platforms and practices* (pp. 229-248). Cham: Springer International Publishing.
- Momani, A. M. (2020). The unified theory of acceptance and use of technology: A new approach in technology acceptance. *International Journal of Sociotechnology and Knowledge Development (IJSKD)*, 12(3), 79-98.
- Mwere D (2020); Kenyans to pay more over Ketrco delayed projects. *Daily Nation publication dated Tuesday October 20, 2020*. Retrieved from Kenyans to pay more over Ketraco delayed projects | Nation
- Ndemo, B (2017). Let's optimize Big Data for revenue collection. *Business Daily*. Retrieved from; <https://www.businessdailyafrica.com/analysis/optimise-Big-Data-for-revenue-collection/539548-4236776-20x49u/index.html>
- Ngila F, (2020 Sept 20) Is Kenya Tapping its Vast Big Data Opportunities?'. *Business Daily*. <https://www.businessdailyafrica.com/bd/data-hub/is-kenya-tapping-its-vast-big-data-opportunities--2279936>
- O'Leary, Z. (2014). The essential guide to doing your research project (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc
- Ochieng, G. O. (2015). *The adoption of big data analytics by supermarkets in Kisumu County* (Doctoral dissertation, University of Nairobi).
- Office of the Auditor General, Kenya (2022). *Report of the Auditor-General on Kenya Electricity Transmission Company Ltd for the year ended 30 June, 2021*. Retrieved from: <https://www.oagkenya.go.ke/wp-content/uploads/2022/11/Kenya-Electricity-Transmission-Company-KETRACO-LTD-2020-2021.pdf>
- Office of the Data Protection Commissioner (ODPC) (2024). *Data Protection Act, 2019*. Retrieved from: <https://www.odpc.go.ke/data-protection-laws-kenya/>
- Onyango, G., & Ondiek, J. O. (2021). Digitalization and integration of sustainable development goals (SGDs) in public organizations in Kenya. *Public Organization Review*, 21(3), 511-526.
- Osanloo, A., & Grant, C. (2016). Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your "house". *Administrative issues journal: connecting education, practice, and research*, 4(2), 7.
- Paskaleva, K., Evans, J., Martin, C., Linjordet, T., Yang, D., & Karvonen, A. (2017, November). Data governance in the sustainable smart city. In *Informatics* (Vol. 4, No. 4, p. 41). MDPI.

- Patel, A. B., Birla, M., & Nair, U. (2012). Addressing big data problem using Hadoop and Map Reduce. In *2012 Nirma University International Conference on Engineering (NUICONE)* (pp. 1-5). IEEE.
- Patil, M. V., & Yogi, A. N. (2011). Importance of data collection and validation for systematic software development process. *International Journal of Computer Science and Information Technology*, 3(2), 260-278.
- Penn, I. A., & Pennix, G. B. (2017). *Records management handbook*. Routledge.
- Phiri, M. J., & Tough, A. G. (2018). Managing university records in the world of governance. *Records Management Journal*, 28(1), 47-61.
- Poleto, T., de Carvalho, V. D. H., & Costa, A. P. C. S. (2015, May). The roles of big data in the decision-support process: an empirical investigation. In *International conference on decision support system technology* (pp. 10-21). Springer, Cham.
- Pouchard, L. (2016). Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation*, 10(2), 176-192.
- Pramanik, P. K. D., Pal, S., Mukhopadhyay, M., & Singh, S. P. (2021). Big Data classification: techniques and tools. In *Applications of Big Data in Healthcare* (pp. 1-43). Academic Press.
- Quinlan, C., Babin, B., Carr, J., & Griffin, M. (2019). *Business research methods*. South Western Cengage.
- Raj, A., & D'Souza, R. G. (2019). Big data anonymization in cloud using k-anonymity algorithm using map reduce framework. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, 5(1), 50-56.
- Ramesh, G., Logeshwaran, J., & Aravindarajan, V. (2022). The Performance Evolution of Antivirus Security Systems in Ultra dense Cloud Server Using Intelligent Deep Learning. *BOHR International Journal of Computational Intelligence and Communication Network*, 1(1), 15-19.
- Raptis, T. P., Passarella, A., & Conti, M. (2019). Data management in industry 4.0: State of the art and open challenges. *IEEE Access*, 7, 97052-97093.
- Roberts, L. D. (2015). Ethical issues in conducting qualitative research in online communities. *Qualitative Research in Psychology*, 12(3), 314-325.
- Roh, Y., Heo, G., & Whang, S. E. (2019). A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1328-1347.
- Ronoh, E. K. (2019). *The Role of Digitization in the Preservation of Corporate Archives at Kenya Power and Lighting Company* (Doctoral dissertation, Kenya Methodist University).
- Rubinfeld, D. L., & Gal, M. S. (2017). Access barriers to big data. *Ariz. L. Rev.*, 59, 339.

- Sabaityte, J., Davidaviciene, V., & Karpoviciute, R. (2020). Learning Skills for Enhancing the Use of Big Data. *World Journal on Educational Technology: Current Issues*, 12(1), 23-36.
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE.
- Shen, D. R., Yu, G., Wang, X. T., Nie, T. Z., & Kou, Y. (2013). Survey on NoSQL for management of big data. *Journal of Software*, 8, 1786-1803.
- Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 191, 97-112.
- Siddiqua, A., Karim, A., & Gani, A. (2017). Big data storage technologies: a survey. *Frontiers of Information Technology & Electronic Engineering*, 18, 1040-1070.
- Song, X., Wu, Y., Ma, Y., Cui, Y., & Gong, G. (2015). Military simulation big data: background, state of the art, and challenges. *Mathematical Problems in Engineering*, 2015.
- Srinivasan, A., Quadir, M. A., & Vijayakumar, V. (2015). Era of cloud computing: A new insight to hybrid cloud. *Procedia Computer Science*, 50, 42-51.
- STAMFORD, Conn. (2022, April 20). Gartner Survey Finds Successful CDOs Link Data and Analytics to Prioritized and Measured Business Outcomes. *Gartner Publication*. <https://www.gartner.com/en/newsroom/press-releases/2022-04-20-gartner-survey-finds-successful-cdos-link-data-and-analytics-to-prioritized-and-measured-business-outcomes>
- Stieninger, M., Nedbal, D., Wetzlinger, W., & Wagner, G. (2018). Factors influencing the organizational adoption of cloud computing: a survey among cloud workers. *International Journal of Information Systems and Project Management*, 6(1), 5-23.
- Stockemer, D. (2018). *Quantitative Methods for the Social Sciences: A Practical Introduction with Examples in SPSS and Stata*. Springer International Publishing.
- Suguna, S., & Suhasini, A. (2014). Overview of data backup and disaster recovery in cloud. In *International Conference on Information Communication and Embedded Systems (ICICES2014)* (pp. 1-7). IEEE.
- Sullivan, D. (2015). *NoSQL for Mere Mortals*. Pearson Education.
- Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, 57(1), 103146.
- Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons*, 62(3), 347-358.

- Talesh, S. A. (2018). Data breach, privacy, and cyber insurance: How insurance companies act as “compliance managers” for businesses. *Law & Social Inquiry*, 43(2), 417-440.
- The National Treasury (2023). *Report on the Auditor-General on Kenya Electricity Transmission Company Ltd for the Year ended 30 June 2021*. Retrieved from: <https://www.treasury.go.ke/wp-content/uploads/2023/06/Kenya-Electricity-Transmission-Company.pdf>
- Toshniwal, R., Dastidar, K. G., & Nath, A. (2015). Big data security issues and challenges. *Complexity*, 2(2).
- Trifiro, B.M., & Gerson, J. (2019). Social Media Usage Patterns: Research Note Regarding the Lack of Universal Validated Measures for Active and Passive Use. *Social Media + Society*, 5(2). <https://doi.org/10.1177/2056305119848743>
- Vahed, A. (2014). Data intensive research initiative for South Africa (DIRISA): A reinterpreted vision. Presentation given at the eResearch Africa Conference.
- Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). Managing and Sharing Data; a best practice guide for researchers.
- Vilita V, Chore M (2024). An Overview of the Laws Governing Data Protection in Kenya. Retrieved from; AN OVERVIEW OF THE LAWS GOVERNING DATA PROTECTION IN KENYA - CFL Advocates (cfllegal.com)
- Villanti, A.C., Johnson, A.L., Ilakkuvan, V., Jacobs, M.A., Graham, A.L., & Rath, J.M. (2017). Social Media Use and Access to Digital Technology in US Young Adults in 2016. *J Med Internet Res*. 19(6): e196.doi: 10.2196/jmir.7303.PMID: 28592394; PMCID: PMC5480010
- Vliegthart, R., & Damstra, A. (2019). Parliamentary questions, newspaper coverage, and consumer confidence in times of crisis: A cross-national comparison. *Political Communication*, 36(1), 17-35.
- Wamba, S.F., & Akter, S. (2023). *Handbook of Big Data Research Methods*. Edward Elgar Publishing.
- Wanekeya, E. (2023). *Effectiveness of Domestic Data Protection Laws in African Countries-a Case Study of the Data Protection Law in Kenya* (Doctoral dissertation, University of Nairobi).
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.
- Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34(1), 65.

- Wikström, E., Eriksson, E., Karamehmedovic, L., & Liff, R. (2018). Knowledge retention and age management–senior employees’ experiences in a Swedish multinational company. *Journal of Knowledge Management*, 22(7), 1510-1526.
- Woerner, S. L., & Wixom, B. H. (2015). Big data: extending the business strategy toolbox. *Journal of Information Technology*, 30(1), 60-62.
- Wood, T., Cecchet, E., Ramakrishnan, K. K., Shenoy, P., Van der Merwe, J., & Venkataramani, A. (2010). Disaster recovery as a cloud service: Economic benefits & deployment challenges. In *2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10)*.
- World Bank Report, 2024, The World Bank AFR RI-The Eastern Electricity Highway Project under the First Phase of the Eastern Africa Power Integration Program (P126579)
- Xiaofeng, M., & Xiang, C. (2013). Big data management: concepts, techniques and challenges [J]. *Journal of computer research and development*, 1(98), 146-169.
- Xue, C. T. S., & Xin, F. T. W. (2016). Benefits and Challenges of the Adoption of Cloud Computing in Business. *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, 6(6).
- Yan, L., & Ma, Z. (2019). *Emerging Technologies and Applications in Data Processing and Management*. IGI Global.
- Yego, P. K. (2014). *Data sharing among selected government institutions in Kenya* (Doctoral dissertation, University of Nairobi).
- Yeo, G. (2018). Records, information and data: exploring the role of record-keeping in an information culture
- Yildirim, O., Gottwald, M., Schöler, P., & Michel, M. C. (2016). Opportunities and challenges for drug development: public–private partnerships, adaptive designs and big data. *Frontiers in pharmacology*, 7, 461.
- Yin, S., & Kaynak, O. (2015). Big data for modern industry: challenges and trends [point of view]. *Proceedings of the IEEE*, 103(2), 143-146.
- Zachlad, C., Samuel, O., Ochsner, A., & Werthmuller, S. (2022). Analytics of social media data - State of characteristics and application. *Journal of Business Research*, 144:1064-1076. <https://doi.org/10.1066/jbusres.2022.02.016>
- Zellner, M., Tilahun, N., & Thakuriah, P. (2016). *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*. Springer International Publishing.
- Zhang, N., Yang, P., Ren, J., Chen, D., Yu, L., & Shen, X. (2018). Synergy of big data and 5g wireless networks: opportunities, approaches, and challenges. *IEEE Wireless Communications*, 25(1), 12-18.

APPENDICES

Appendix I: Survey Questionnaire

Michael Wechuli
P.O. Box 34942-00100,
Nairobi
E-mail: michaelwechuli@gmail.com
Mobile NO.: 0717915270

Dear Respondent,

I am a postgraduate student undertaking a Master of Science in Information sciences at Moi University. As part of the requirements to graduate with an award in the master's degree, I am conducting a study on ***“Big Data management at the Kenya Electricity Transmission Company Limited (KETRACO)”***.

In simple terms, Big Data refers to large and diverse sets of data received or created by an organization and increases in volume at a higher rate with time. Key areas that will be explored in this study include the status of Big Data Management at Ketraco, policies and legal framework for Big Data Management, Knowledge, skill and training of data managers, the ICT infrastructure for Big Data Management, collaborative partnerships in Big Data Management, and the contribution of Big Data in organizational functions and processes. By conducting this study, the researcher intends to propose interventions to improve the management, sharing and reuse of big data to support proper planning, improved organizational processes and decision making.

My study can only succeed with your support and participation in this survey to help me gather research data that addresses the research problem. The information you share will only be used for its intended academic purpose and treated with utmost confidentiality.

I will also welcome any additional information you may find relevant in making my study more conclusive. Thank you in advance.

Yours faithfully,

Michael Wechuli

Masters Student (Moi University, Kenya)

PART A: BACKGROUND INFORMATION

1. Gender

Male	
Female	

2. Level of Education

Secondary	
Certificate	
Diploma	
Undergraduate	
Masters/Postgraduate	

3. Department

4. Level of Management

Top Level Management	
Middle Level Management	
Lower-Level Management	

Years of Experience

0-4years	
5-8 years	
8- 12years	
Over12 years	

**PART B: CAPTURE, APPRAISAL, DESCRIPTION, PRESERVATION,
ACCESSIBILITY, RE-USE AND SHARING OF BIG DATA**

- 1. Creation;** (a). How do you capture/receive data that you use on a day-to-day basis
(*Tick all that applies*)

- ☐ Manual Data Keying
- ☐ Voice Recognition
- ☐ Scanning
- ☐ Bar coding
- ☐ Legacy Data Import
- ☐ Other (Please Specify)

2. (b), In what format do you capture/receive big data? (*Tick all that applies*)

- ☐ Text Documentation (Pdf, HTML, Plain Text)
- ☐ Still Images (JPEG, DNG)
- ☐ Geospatial (Geo TIFF)
- ☐ Audio (MP3)
- ☐ Video (MP4, MPEG-4)
- ☐ Database (XML, CSV)
- ☐ Other (Please Specify)

- 2 Appraisal:** (a) Do you appraise data received before choosing what to/not preserve?

- ☐ No
- ☐ Yes

- (b) If No, Please explain why

.....

.....

(c) If yes, what guides your appraisal on selecting data that deserves long term storage?

(Tick all that applies)

- ☐ Big data appraisal check list
- ☐ Big Data Appraisal policy
- ☐ An electronic data inventory worksheet
- ☐ Retention schedules and forms
- ☐ Other (Please Specify)

.....

3. **Description:** (a) Which of the following elements do you use to describe/give identity to data that you receive or create for the purpose of easier identification and retrieval when needed? **(Tick all that applies)**

Document Ref. Number	
Document Title	
Document Date	
Document subject	
Document creator	
Combination of the above elements	
Others (Please list)	

4. **Preservation/storage:** (a) Do you have any platform for storing your received/created data?

- ☐ Yes
- ☐ No
- ☐ Not sure

(b). If No, please

explain.....

.....

(c). If yes, how do you store data that you acquire/use on a regular basis? (***Tick all that applies***)

- ☐ In portable Flash/Hard drives
- ☐ In organizational database
- ☐ In shared drives on the server
- ☐ Institutional repository
- ☐ Other (please specify)

(d) Who is responsible for day-to day storage and back-up of big data? (***Tick all that applies***)

- ☐ Individual responsibility
- ☐ Records management staff
- ☐ ICT
- ☐ External/service provider
- ☐ Other (please specify)

e) How often do you make backups for the big data? (***Please Tick accordingly***)

- ☐ In less than 6 months,
- ☐ In less than one year,
- ☐ Between 1 and 2 years
- ☐ After 2 or more years
- ☐ Never conduct backups
- ☐ Once you get a notification to backup up
- ☐ Other (please specify).....

e) Who decides which of the big data are worth preserving? (***Tick all that applies***)

- ☐ Individual decision
- ☐ Head of the Section
- ☐ Records management staff

- ☐ ICT
- ☐ Other (please specify)

f) Where is your big data archived for long term storage? **(Tick all that applies)**

- ☐ In personal computers
- ☐ In organizational database
- ☐ In shared drives on the server
- ☐ Institutional repository
- ☐ In the platforms of the contracted 3rd party Vendor
- ☐ Other (please specify)

5. Access: (a). How would you rate the access to accurate, relevant and timely data in Ketraco?

Poor { } Fair { } Good { } Very Good { } Excellent { }

(b) Who currently accesses big data at Ketraco? *(Tick appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 strongly Agree)*

	1	2	3	4	5
Access to data is limited to the person who created/received					
Access to certain data is extended to members of the department					
There are restrictions on access to certain set of data within Ketraco					
Access to data is allowed to all Ketraco staff					
Access to data also extends to members of the public					
Access to data also extends to the Media, Partners in the Energy sector and independent offices such as office of the Auditor General, EACC, etc					

6. Uses and reusing Big Data; (a). What business functions in your organization are the key users of Big Data **(Tick all that applies)**.

Technical operations	
HR Services	
Customer service	

Information Technology	
Corporate planning	
Audit Management	
Legal services	
Finance Management	
Administration services	
Other (specify)	

(b). How would you rate your company on the ability to use Big Data to transform its operations and achieve its mandate?

Poor { } Fair { } Good { } Very Good { } Excellent { }

c) How would you rate the reasons for using and reusing big data at **Ketraco** (*Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 Strongly Agree*)

	1	2	3	4	5
Ketraco uses Big Data to engage customers					
Ketraco uses Big Data To make strategic decisions					
Ketraco uses Big Data To make operational decisions and transactions					
Ketraco relies on its data to improve and innovate its services					
Ketraco uses Data to reach out to new markets and access new business opportunities					
Ketraco uses data to assess its risks					
Ketraco relies on data to measure its performance and minimize on errors					

(d). How do you make your data available for use? (*Tick all that applies*)

- ☐ Through publications
- ☐ Through the existing institutional portal
- ☐ Through emails
- ☐ Through 3rd Party sites
- ☐ Through Licensing and subscription
- ☐ Through External hard disk drives
- ☐ Other (please specify)

7. **Sharing:** (a) With whom are you sharing your big data with? (*Tick all that applies*)

- ☐ Fellow colleagues in the department

- ☐ Other staff members
- ☐ Partners in the Energy Sector
- ☐ Government Agencies such as office of Auditor General
- ☐ Stakeholders such as contractors and Suppliers
- ☐ Private sector such as researchers and Banks
- ☐ Others (Please Specify)

(b) What are the challenges encountered in big data sharing?

.....

**PART C: LEGAL, POLICIES AND GUIDELINES THAT GUIDE THE
MANAGEMENT OF BIG DATA BY KETRACO.**

8. (a) Are there legal frameworks that govern big data management

- ☐ Yes
- ☐ NO

(b) If Yes, Please explain.....

.....

(c) If No, Please explain.....

.....

9 (a). Does Ketraco have the following policies that are key to Big Data Management?

(Tick all that applies)

Capturing policy	
Appraisal policy	
Description policy	
Preservation policy	
Access Policy	
Sharing policy	
Use and reuse policy	

(b). How do you rate the existing policy for Big Data Management?

Poor { } Fair { } Good { } Very Good { } Excellent { }

(c). How do you rate the existing in-house guidelines Big Data Management?

Poor { } Fair { } Good { } Very Good { } Excellent { }

(d). What would you recommend the organization to do in terms of formulating and implementing Big Data related policies? **(Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 strongly Agree)**

	1	2	3	4	5
Ensuring proper participation of users in developing the policy					
Ensuring that the policy reflects the varying needs of every department					
Properly communicating the policy to users					
Ensuring that the policy conforms and is consistency with other information related policies.					
Identifying those who lead the responsibility of developing the policy and delegating these responsibilities from a working group, sub groups, committees, and individuals					
Monitoring the policy to ensure proper implementation.					

(e). Are there quality assurance and control measures in place in the following stages of big data management? **Tick (✓) to accept or (X) to decline**

Big data capture	
Big data appraisal	
Big data description	
Big data preservation	
Big data access	
Big data Sharing	
Big data Use and reuse	

(f) Explain the relevancy of quality assurance and control measures in big data management in Ketraco.....

.....

.....

**PART D: SKILLS, KNOWLEDGE, AND TRAINING REQUIRED FOR
CAPTURING, DESCRIBING AND USING BIG DATA**

10. Do personnel at Ketraco possess the following knowledge on Big Data management
(Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 Strongly Agree)

	1	2	3	4	5
Knowledge on Data capture					
Data Appraisal					
Data description					
Data Security					
Data Preservation and storage					
Data sharing					

11. Do the personnel at Ketraco possess the following Key skills for Big Data (Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 Strongly Agree)

	1	2	3	4	5
Metadata skills					
Search and Retrieval skills					
Data Capture skills					
Data Analysis skills					
Collaboration and communication skills					

12. (a).Do you think Ketraco has employed experienced staff for big data management?

Yes { } No { } Not sure { }

- (b). How often do you seek professional support on Big Data?

Never { } Rarely { } Sometimes { } Often { } Always { }

13. (a) How do you rate your level of understanding of the following skills that contribute to big data management?

(a) SQL Databases

Basic Level { } Intermediate Level { } Advanced Level { } Planned in the future
 { } Not sure { }

(b) Hadoop

Basic Level { } Intermediate Level { } Advanced Level { } Planned in the future
 { } Not sure { }

(c) Word Processing

Basic Level { } Intermediate Level { } Advanced Level { } Planned in the future
 { } Not sure { }

(d) Statistical and Quantitative Analysis

Basic Level { } Intermediate Level { } Advanced Level { } Planned in the future
 { } Not sure { }

(e) Data Mining

Basic Level { } Intermediate Level { } Advanced Level { } Planned in the future
 { } Not sure { }

(f) Data Analysis

Basic Level { } Intermediate Level { } Advanced Level { } Planned in the future
 { } Not sure { }

(g) Any other skills

14. What areas in big data do you feel exist skill gaps? (*Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 5 Agree; 6 strongly Agree*)

	1	2	3	4	5
Metadata skills					
Search and Retrieval skills					
Data archiving					
Data Capture skills					
Data Analysis skills					
Collaboration and data sharing skills					

15. Rate the type of training areas that are highly needed in KETRACO (**Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 strongly Agree**)

	1	2	3	4	5
Capturing and appraisal					
Data description and preservation					
Data access and sharing					
Data literacy					
Use of institution repository					
Legal and policy framework for Big Data					
Data security and confidentiality					
Use of ICTs in Big Data					

16. How would you rate the quality of the following big data services offered by the Records management and IT departments at Ketraco? (**Tick Appropriately: 1 very satisfied, 2 satisfied, 3 neutral, 4 dissatisfied, 5 very dissatisfied**)

	1	2	3	4	5
Data Capturing and description					
Data curation/preservation/Archiving					
Search and Retrieval of data					
Data collaboration and Sharing					
Data Backup					
Data security and Confidentiality					

17. Have you undertaken any data literacy program that aims at enhancing your understanding of Big Data?

☐ No

☐ Yes

(a) If No, please explain

.....

(b) If yes, please explain briefly the nature of the program

undertaken.....

.....

.....

**PART E: LEVEL OF ICT INFRASTRUCTURE PREPAREDNESS FOR BIG
DATA MANAGEMENT**

18. (a) Does Ketraco have the following ICT infrastructure for Big Data Management?

	Yes	No
Reliable internet/network infrastructure		
Computer software and programs		
Servers, Personal computers and other hardware		
Storage platforms		
Databases		

(b). To what extent do you agree with the following ICT infrastructure related statements? *(Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 Strongly Agree)*

	1	2	3	4	5
There is a network infrastructure for data communication and sharing					
There are storage platforms/server spaces to store your data.					
There is a digital institutional repository for archiving data					
There exists software and hardware for capturing and describing data					
There are systems that enhance efficiency and flow of data within the workflows					

19. What are some of the software that support, data creation, description, analysis, storage, sharing and use in your department?

.....
.....

Briefly explain;

(a). The strengths of the software listed

.....
.....

(b). The weakness of these software

.....

.....

.....

20. What are some of the security measures put in place to protect the integrity and confidentiality of your data?

.....

.....

.....

21. In your own opinion, what type of tools and Technologies if installed can improve big Data Management?

PART F: INFLUENCE OF COLLABORATIVE PARTNERSHIPS ON BIG DATA MANAGEMENT AT KETRACO

22. (a) Does Ketraco have any collaborative partnerships with other players in the energy sector to facilitate data sharing?

Yes { } No { } Not sure { }

(b) If Yes in b above, can you please list some of these partnerships?

.....

.....

(c). If No, Please explain

.....

.....

23. To what extent do you agree with collaborations and partnerships at Ketraco with respect to big data? (*Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 Strongly Agree*)

	1	2	3	4	5
There is collaboration in data exchange within the department					

There is collaboration in data exchange across the departments					
There is collaboration and partnerships with the funding Agencies					
There is collaboration and partnerships with public-private sectors					
There is collaboration with the public					
There is collaboration in data sharing with other Partners in the energy sector					

PART G: CHALLENGES IN BIG DATA MANAGEMENT

24. In your View, are there challenges at Ketraco that affect Big Data Management practices?

Yes { } No { } Not sure { }

25. To what extent do you think the challenges listed below hinder Big Data management?

(Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 Strongly Agree)

	1	2	3	4	5
There are privacy and confidentiality issues with sharing of Data at Ketraco					
There is insufficient understanding of Big Data in the organization					
There is little organization support in advocating for Big Data practices					
It is very complex to handle large volumes of data at the workplace					
There are inadequate ICT tools that support data storage and sharing					
It is very difficult to ascertain elements such as orderliness, accuracy, completeness, trustworthy in the available data and this makes it difficult to use such kind of data					
There are no clear policies and procedures that guide data administration in the organization.					
Others (list any other associated challenge)					

PART H: STRATEGIES FOR EFFECTIVE BIG DATA MANAGEMENT

26. (a). What are the immediate strategies that can enhance Big Data Management in your organization? *(Tick Appropriately: 1 strongly disagree; 2 Disagree; 3 Not sure; 4 Agree; 5 strongly Agree)*

	1	2	3	4	5
Establishing a big data management culture					
A clear Company strategy					
Support from top level management					
Support systems and procedures					
Establish measures that enhance Data privacy and security					
Purchasing more ICT tools for Big Data					
Embedding big data in departmental procedures and processes					
Recruiting personnel trained specifically to handle big data.					
Establishing Collaborating partnerships on Big Data sharing with other stakeholders in the Energy Sector					
Establishing policies that guide on usage of organizational big data					

- (b) Is there any other strategy you find necessary to improve Big Data at Ketraco?

Yes { } No { } Not sure { }

- (c) Explain

.....

THANK YOU IN ADVANCE

Appendix II: Interview Schedule

Michael Wechuli
P.O. Box 34942-00100,
Nairobi
E-mail: michaelwechuli@gmail.com
Mobile NO.: 0717915270

Dear Respondent,

I am a postgraduate student undertaking a Master of Science in Information science at Moi University. As part of the requirements to graduate with an award in the master's degree, I am conducting a study on **“Big Data management at the Kenya Electricity Transmission Company Limited (KETRACO)”**.

In simple terms, Big Data refers to large and diverse sets of data received or created by an organization and increases in volume at a higher rate with time. Key areas that will be explored in this study include the status of Big Data Management at Ketraco, policies and legal framework for Big Data Management, Knowledge, skills and training of data managers, the ICT infrastructure for Big Data Management, collaborative partnerships in Big Data Management, and the contribution of Big Data in organizational functions and processes. By conducting this study, the researcher intends to propose interventions to improve the management, sharing and reuse of big data to support proper planning, improved organizational processes and decision making.

My study can only succeed with your support and participation in this survey to help me gather research data that addresses the research problem. Therefore, this letter seeks to an appointment with you at your most convenient time to participate in the study through an interview session. The information you share will only be used for its intended academic purpose and treated with utmost confidentiality.

I will also welcome any additional information you may find relevant in making my study more conclusive. Thank you in advance.

Yours faithfully,

Michael Wechuli

Masters Student (Moi University, Kenya)

PART A: BACKGROUND INFORMATION

1. Gender

Male	
Female	

2. Level of Education

Secondary	
Certificate	
Diploma	
Undergraduate	
Masters/Postgraduate	

3. Department

4. Level of Management

Top Level Management	
Middle Level Management	
Lower Level Management	

5. Years of Experience

0-4years	
5-8 years	
8- 12years	
Over12 years	

**PART B: CAPTURE, APPRAISAL DESCRIPTION, PRESERVATION,
ACCESSIBILITY, RE-USE AND SHARING OF DATA**

1. How do you capture, appraise, describe, preserve, access, use and share data at KETRACO?

.....

What kind of tools are available for big data capture, appraise, describe, preserve, access, use and share data at KETRACO

2. What subject areas of data are created in your organization?

.....

.....

What types of support are offered by the department to enhance big data management?

3. What do you think can be done to improve data sharing among users within Ketraco?

What importance does KETRACO attach data capture, appraisal, description, preservation, access, reuse and sharing.

.....

.....

**PART C: POLICIES AND GUIDELINES THAT GUIDE THE MANAGEMENT
OF DATA CREATED/RECEIVED BY KETRACO.**

4. What legal frameworks govern big data at Ketraco?
5. Does Ketraco have policies on capture, appraisal, description, preservation, access and sharing that are key to Big Data Management?
6. Many policies are formulated but fail at implementation. What do you think should be done to ensure that all the policies formulated on Big Data are successfully implemented?
7. Are there quality and control measures put in place to control the quality of data created or shared within and outside Ketraco?
8. In your own opinion, what else should be done in terms of policy formulation and implementation to support Big Data Management at Ketraco?
9. How is intellectual property (IP) managed with reference to big data at KETRACO

**PART D: SKILLS, KNOWLEDGE, AND TRAINING REQUIRED FOR
CAPTURING, APPRAISAL, DESCRIBING, PRESERVING, ACCESS, REUSE
AND SHARING BIG DATA**

10. Who is responsible for Big Data Capture, appraisal, description and preservation of data at Ketraco?
11. What are the competencies (knowledge and skill) required in big data capture, appraisal, description, preservation, access, use and sharing at KETRACO
12. What are the training needs for big data capture, appraisal, description, preservation, access, use and sharing at KETRACO
13. What measures can the organization adopt to ensure that the data management skills are trickled down to all data users, but not only IT staff?
14. What are the contributions of IT and Records management departments in driving big data initiative
15. Are there any Data literacy and advocacy programmes in place that support the management of data?

PART E: ICT INFRASTRUCTURE FOR BIG DATA MANAGEMENT

16. What range of ICT(s) tools and equipment does Ketraco use in big data capture, appraisal, description, preservation and access to big data?
17. What are some of the security measures put in place to protect the integrity and confidentiality of your data?
18. What is the level of ICT(s) infrastructure preparedness with regard to big data at KETRACO
19. In your own opinion, what other type of tools and Technologies if installed can improve Big Data Management?

**PART F: INFLUENCE OF COLLABORATIVE PARTNERSHIPS ON BIG DATA
MANAGEMENT AT KETRACO**

20. What forms of collaborative partnerships exist between Ketraco and other players in the Energy Sector and outside the energy sector with respect to Big Data sharing and use?
21. What is the impact of these collaborations on Big Data Management at Ketraco?
22. What are the benefits of these collaborative partnerships to Ketraco?
23. What are some of the challenges encountered when exchanging Big Data with other partners and stakeholders?

PART G: CHALLENGES IN BIG DATA MANAGEMENT

- (a) In your View, what are the challenges at Ketraco that affect Big Data Management practices such as data acquisition, sharing or preservation?

.....

.....

13. How do the challenges outlined above hinder Big Data Management?

.....

.....

PART G: STRATEGIES FOR EFFECTIVE BIG DATA MANAGEMENT

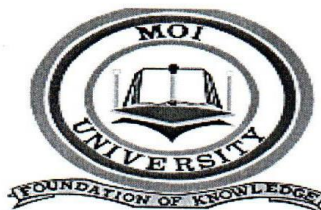
14. In your view, what are the strategies that can enhance Big Data Management in your organization?

.....

.....

THANK YOU IN ADVANCE

Appendix III: Moi University Introduction Letter



MOI UNIVERSITY
(ISO 9001:2015 CERTIFIED INSTITUTION)

Our Ref: MU/SIS/LRMIS/PGG/62

Date: 1st July 2020

To Whom IT May Concern

Ref: Data Collection by Michael W. Wechuli (IS/MS/SC/02/16)

The above named is a postgraduate student pursuing a Masters Degree programme in Records and Archives Management in the department of Library, Records Management and Information Studies, School of Information Sciences, Moi University. He is carrying out a research project entitled "Big Data Management at the Kenya Electricity Transmission Company (KETRACO), Kenya" under the supervision of Dr. Elsebah Maseh and Dr. Emily Ng'eno.

The purpose of writing is to request you to allow Mr. Wechuli to collect the necessary data. The information given will be for the purposes of writing the research project and will be treated with utmost confidentiality.

Any assistance accorded to her will be most appreciated

Thank You

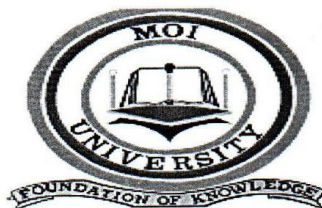
Yours Faithfully

Dr Elsebah Maseh

DEAN
SCHOOL OF INFORMATION
SCIENCES
MOI UNIVERSITY

Senior Lecturer and Head, Dept. of LRM&IS

Appendix IV: Moi University Letter to Nacosti



MOI UNIVERSITY
(ISO 9001:2015 CERTIFIED INSTITUTION)

Our Ref: MU/SIS/LRMIS/PGG/62

Date: 1st July 2020

Executive Director

NACOSTI

P.O Box 30623, 00100

Nairobi, Kenya.

Ref: Research Permit for Michael W. Wechuli (IS/MS/SC/02/16)

The above named is a postgraduate student pursuing a Masters Degree programme in Records and Archives Management in the department of Library, Records Management and Information Studies, School of Information Sciences, Moi University. He is carrying out a research project entitled "Big Data Management at the Kenya Electricity Transmission Company (KETRACO), Kenya" under the supervision of Dr. Elsebah Maseh and Dr. Emily Ng'eno.

The purpose of writing is to request you to issue Mr. Wechuli with a research permit. Any assistance accorded to her will be most appreciated

Thank You

Yours Faithfully

DEAN
SCHOOL OF INFORMATION
SCIENCES
MOI UNIVERSITY

Dr Elsebah Maseh

Senior Lecturer and Head, Dept. of LRM&IS

APPENDIX V: Ketraco Approval Letter


KETRACO

Kenya Electricity Transmission Company Limited
"Building a World Class National Grid"

DIRECTORATE OF HUMAN RESOURCE & ADMINISTRATION

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 | email: info@ketraco.co.ke

Our Ref: KET/5/1C/42/JEA/mw

24 August 2020

Michael Wechuli
 P.O Box 945 - 50100
 Bungoma

Dear *Sir*

RE: AUTHORIZATION TO CONDUCT RESEARCH

We acknowledge receipt of your letter dated 1st July 2020 to conduct research in KETRACO titled: BIG DATA MANAGEMENT AT THE KENYA ELECTRICITY TRANSMISSION COMPANY LIMITED as part of your Masters Degree program 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

For: GENERAL MANAGER, HUMAN RESOURCE & ADMINISTRATION

Appendix VI: Research License

 REPUBLIC OF KENYA	 NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION
Ref No: 940045	Date of Issue: 21/July/2020
RESEARCH LICENSE	
	
This is to Certify that Mr.. michael wechuli of Moi University, has been licensed to conduct research in Nairobi on the topic: BIG DATA MANAGEMENT AT THE KENYA ELECTRICITY TRANSMISSION COMPANY (KETRACO), KENYA for the period ending : 21/July/2021.	
License No: NACOSTI/P/20/5891	
940045 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.	