

Addis Ababa
University
(Since 1950)



Addis Ababa University
College of Natural and Computational Sciences
School of Information Science

Big Data Governance Framework for Ethio Telecom

By:

Bereket Shewangizaw

ID Number: GSR/0124/12

Advisor: Wondwossen Mulugeta (Phd)

Jan, 2022

Addis Ababa, Ethiopia

Addis Ababa
University

(Since 1950)



Addis Ababa University
College of Natural and Computational Sciences
School of Information Science

Big Data Governance Framework for Ethio Telecom

**A Thesis Submitted to the College of Natural and Computational
Sciences of Addis Ababa University in Partial Fulfilment of the
Requirements for the Degree of Master of Science in Information
Science and Systems (Information Systems)**

By

Bereket Shewangizaw

Advisor: Wondwossen Mulugeta (PhD)

Jan, 2022

Addis Ababa, Ethiopia

Addis Ababa
University
(Since 1950)



Addis Ababa University
College of Natural and Computational Sciences
School of Information Science

Big Data Governance Framework for Ethio Telecom

**A Thesis Submitted to the College of Natural and Computational
Sciences of Addis Ababa University in Partial Fulfilment of the
Requirements for the Degree of Master of Science in Information
Science and Systems (Information Systems)**

By:

Bereket Shewangizaw Shumete

Name and Signature of Members of the Examining Board

Name	Title	Date	Signature
_____	Examiner	_____	_____
_____	Examiner	_____	_____
Wondwossen Mulugeta (PhD)	Advisor	_____	_____

DECLARATION

I, Bereket Shewangizaw, hereby declare that the research proposed in this thesis paper titled "Big data governance framework for Ethio Telecom" is my original work. It was completed and presented under the supervision of my advisor, Dr. Wondwossen Mulugeta. This thesis has not been submitted for any academic achievement at the university level, and all other resources used in this study have been fully acknowledged.

Bereket Shewangizaw Shumete

Jan, 2022

This thesis has been submitted for examination with my approval as a university advisor.

Wondwossen Mulugeta (PhD)

ACKNOWLEDGEMENT

This thesis was made with the help and guidance of several individuals and organizations. Hence, I am overjoyed to express my appreciation to those who have played important roles in my graduate career.

First and foremost, I want to thank God for assisting me with everything. Next, I'd like to thank Addis Ababa University for providing me with the opportunity to study for free through female scholarships.

Next, I would like to express my sincere gratitude to Dr. Wondwosen Mulugeta, who was a very good advisor and mentor to me.

Moreover, it is my pleasure and honour to thank Ethio Telecom's IT staff for their efforts during the data collection and for their willingness to participate in the survey and interview. The staff of the data warehouse team, in particular, deserve special thanks for providing me with all of the information and documents I need for my paperwork.

Finally, I'd like to express my deepest thanks to my respected family members, Ato Shewangizaw Shumete, W/ro Birzegen Abi, and Elsa Shewangizaw, for their unconditional love and support. Apart from that, the good times and words of encouragement from all of the people around me have been a great help to me.

ABSTRACT

With the digitization of most operations, the popularization of various social media channels, blogs, deployment of various types of devices, sensors, hand-held smart devices, wearables, and the explosion in Internet usage, large volumes of data are being generated on a regular basis. Today, businesses are searching for ways to successfully manage and optimize these large datasets due to a massive amount of information being swapped every day and the growing need to make better data-driven decisions.

Big data refers to huge data sets with large, more varied, heterogeneous, and complex structures that are expensive to retrieve, examine, and visualize using conventional data processing technologies. One of the most pressing issues in dealing with big data is the adoption of suitable big data governance frameworks to customize big data in a sufficient manner to enable quality data access for effective knowledge extraction using machine learning techniques. It also aspires to outline the blueprint that governs the storage and processing of data from owners and consumers in a truthful manner within the applicable regulatory landscape. Ethio telecom is one of the major big data custodians in Ethiopia that lacks such insight. The main objective of this work, therefore, is to explore and propose a big data governance framework for Ethio telecom based on governance activities related to data handling in Ethio telecom networks. The proposed framework consists of three domains and within the domains, there are nineteen components identified to formulate the structure.

Towards that end, the research adopted Design Science Research as a research approach coupled with a mixed methodology for data collection and analysis. Towards achieving the objective, a review of literature on big data, big data analytics, big data management, big data governance, and previously designed big data governance frameworks were explored with the aim to develop a suitable framework and identify common characteristics and shortcomings in the available big data governance frameworks. Primary data were obtained through survey questionnaires and key informant interviews to maintain organizational context and domain-specific big data governance specifications. As governed by the chosen research strategy, these two contributions (qualitative and quantitative data) are integrated to design a detailed big data governance framework suitable for the Ethio telecom setting. Expert validation was then used to assess the proposed big data governance framework. As a result, the research process and study results are thought to be acceptable, indicating the usability and applicability of the proposed framework.

Keywords: Big data, big data analytics, big data governance, big data governance framework

TABLE OF CONTENTS

DECLARATION	IV
ACKNOWLEDGEMENT	V
ABSTRACT.....	VI
LIST OF TABLES	X
LIST OF FIGURES	XI
LIST OF ACRONYMS	XII
CHAPTER ONE.....	1
INTRODUCTION	1
1.1 Background of the research	1
1.2 Statement of the problem.....	3
1.2.1 Research questions.....	6
1.3 Objectives	7
1.3.1 General objective	7
1.3.2 Specific Objective.....	7
1.4 Significance of the study	7
1.5 Scope and Limitation of the Study	8
1.6 Organization of the Thesis.....	8
CHAPTER TWO	10
LITERATURE REVIEW	10
2.1 Big data.....	10
2.2 Characteristics/ Features of Big Data	11
2.3 Big data in telecom industry	13
2.4 Big data analytics.....	16
2.4.1 Big data analytics applications in telecom industry.....	16
2.4.2 Big data challenges in telecom	18
2.5 Data Governance	20
2.6 Big Data and Ethio telecom.....	22
2.6.1 Overview of Ethio telecom	22
2.6.2 Big data Governance Current Practice in Ethio telecom Industries	23
2.7 Big data governance framework.....	24
2.7.1 Top Frameworks and Software Tools for Handling Big Data.....	25
2.7.1.1 Apache Hadoop	25
2.7.1.2 Hive	27
2.7.1.3 Hbase	28

2.7.1.4	Flume	28
2.7.1.5	Kafka	28
2.7.1.6	Apache spark	29
2.7.1.7	ELK Stack.....	30
2.7.1.8	Apache Metron	30
2.8	Related Works	30
CHAPTER THREE		36
RESEARCH METHODOLOGY.....		36
3.1	Introduction	36
3.2	Research design	36
3.2.1	Problem identification and motivation.....	37
3.2.1.1	Study area	37
3.2.1.2	Source of Data and Data Collection Technique	37
3.2.1.2.1	Source of Data.....	38
3.2.1.2.1.1	Primary Sources of Data	38
3.2.1.2.1.2	Secondary Sources of Data	39
3.2.1.3	Population size and Sampling technique	40
3.2.1.3.1	Population size	40
3.2.1.3.2	<i>Sampling technique</i>	40
3.2.1.4	Data Analysis and Interpretation	42
3.2.1.5	Validity and Reliability	42
3.2.2	Objectives of a solution	43
3.2.3	Design and development.....	43
3.2.4	Evaluation	44
3.2.5	Communication.....	44
CHAPTER FOUR.....		45
DATA ANALYSIS AND PRESENTATION		45
4.1	Introduction	45
4.2	Data Analysis-Part I - General category	45
4.2.1	Demographic Profile of Respondents	45
4.2.2	Use of Big Data Analytics	47
4.2.3	The type of BDA technology/Software adopted.....	49
4.2.4	Total Amount of Data managed.....	50
4.2.5	Data Type collected and analyzed by Ethio telecom	51
4.2.6	Rate of BDA Usage	52

4.2.7	Challenges affecting usage of big data analytics Technology in Ethio telecom	53
4.3	Data Analysis-Part II -Governance.....	54
4.3.1	Strategy	54
4.3.2	Technology	57
4.3.3	People.....	60
4.3.4	Environment.....	61
4.3.5	Ethio telecom specific BDGF requirements	64
4.4	Summary.....	65
CHAPTER FIVE		68
THE PROPOSED BIG DATA GOVERNANCE FRAMEWORK.....		68
5.1	Introduction	68
5.2	Framework Development	68
5.3	Components of the proposed BDGF	69
5.4	The proposed BDG framework and discussions	70
5.4.1	Goals	70
5.4.2	Drive domain	70
5.4.3	Big data governance Implementation Domain	72
5.4.4	Guiding Principles	74
5.4.5	Outcomes	75
5.5	Evaluation of the Proposed Framework	77
5.5.1	Expert Validation	77
5.6	Result and discussion.....	79
CHAPTER SIX.....		82
CONCLUSION AND RECOMMENDATION.....		82
6.1	Introduction	82
6.2	Conclusion.....	82
6.3	Recommendations	83
REFERENCES		85
APPENDIX A: AAU LETTER		93
APPENDIX B: QUESTIONNAIRE.....		94
APPENDIX C: SEMI-STRUCTURED INTERVIEW GUIDE		101
APPENDIX D: EVALUATION CRITERIA FOR THE PROPOSED BDGF.....		104

LIST OF TABLES

List of Tables	Page
Table 1: Potential Data Availability	15
Table 2: Ethio telecom position as of July, 2021	23
Table 3: Summary of related works.....	35
Table 4: Target population for a survey questionnaire	41
Table 5: Target population for interview	42
Table 6: Reliability Statistics	42
Table 7: Distribution of respondents by Role level, work position, education and work experience	47
Table 8: Total Amount of Data managed	50
Table 9: Rate of BDA Usage	53
Table 10: challenges affecting usage of big data analytics Technology in Ethio telecom	53
Table 11: Strategy towards big data governance at Ethio telecom	57
Table 12: Technology towards big data governance at Ethio telecom	59
Table 13: People aspect towards big data management at Ethio telecom	61
Table 14: Environmental factor for BDG at Ethio telecom	64
Table 15: Reliability Statistics for Evaluation of the Proposed Framework	77
Table 16: Mean and standard deviation of the Framework Evaluation Survey.....	78

LIST OF FIGURES

List of Figures	Page
Figure 1: Growth of Data.....	11
Figure 2: 3v model.....	12
Figure 3: 5 v model.....	13
Figure 4: The number of mobile, internet and active social media users in the world.....	14
Figure 5: Big data challenges in telecom	20
Figure 6: Map-Reduce Architecture	27
Figure 7: Design science research cycles by.....	37
Figure 8: Distribution by Use of Big Data Analytics	48
Figure 9: The type of BDA technology/Software adopted	49
Figure 10: Data Type collected and analyzed.....	52
Figure 11: The proposed big data governance framework for Ethio telecom	69
Figure 12: an implementation framework for the proposed BDG framework	76

LIST OF ACRONYMS

Acronym	Description
API	Application Program Interface
BD	Big data
BDA	Big data analytics
BDG	Big data governance
BDGF	Big data governance framework
BDM	Big data management
CDR	Call detail record
CSPs	Communications service providers
DBMS	Database management system
DevOps	Development and Operations
DIL	Data Integration Layer
DSRM	Design science research Methodology
DW&BI	Data warehouse and business intelligence
ELK Stack	Elastic search, Log stash, Kibana
GDPR	General data protection regulation
HDFS	Hadoop Distributed File System
IBM	International Business Machines
ICT	Information communication technology
IoT	Internet of Things
ISDR	Information Systems Design Research
JDBC	Java Database Connectivity
LA	Lambda Architecture
OLAP	Online analytical processing
OLTP	Online transaction processing
OTT	Over-the-top
RTAP	Real time analytics processing
SQL	Structured Query Language
WFH	Work-from-home
XDRs	Extended data records

CHAPTER ONE

INTRODUCTION

This chapter contains the background of the research, statement of the problem, research questions, objective of the research, significance of the research, scope and limitation of the research and organization of the research.

1.1 Background of the research

With the automation of most processes, the rise in popularity of various social networking platforms, blogs, the deployment of various types of sensors, the adoption of hand-held digital devices, wearable devices, and the explosion in internet usage, massive amounts of data are produced on a daily basis (Kumar & Jothimani, 2017). This continuous, multi-dimensional and versatile production of data is called big data (Zafar, Ul Haq, Haroon- Ur-Rashid, & Khalid, 2021). Big data is a term that refers to the massive amounts of unstructured and structured data that can influence a company's operations (Zafar, Ul Haq, Haroon- Ur-Rashid, & Khalid, 2021). In 2012, Gartner retrieved and offered a more detailed definition: “*Big data is high volume, high velocity and high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*” (Glossary, 2012). Here the researcher concentrates on three characteristics of big data, which are: 1) Volume: that refers to the vast scale of big data, which ranges from terabytes to petabytes, 2) Velocity: that refers to the rate at which data is produced and processed (both batch and real-time data), and 3) Variety: that refers to various types of data from various data sources, which are known as structured, unstructured, or semi-structured data (Katal, Wazid, & Goudar, 2013).

Every day, 2.5 quintillion bytes of data are created, and 90 percent of the world's data has been created in the previous two years alone (The-Next-Tech, 2021). In particular, there has been a significant increase over the past 12 months due to covid-19 (datareportal, 2021). That is, the covid-19 pandemic, as well as the quarantines and stay-at-home restrictions, set a new record for data usage (Kastouni & Lahcen, 2020). In this regard, surprisingly, the usage of telecom customers has improved significantly. Remote work, telework, eLearning, virtual worlds, streaming, and other practices that were once assumed exceptional have now become extremely common (Kastouni & Lahcen, 2020). The bandwidths are becoming increasingly congested, and the amount of data generated is rising dramatically every minute.

As described above, in the age of digitalization, information is so overwhelming that we are immersed in it (Paschalidi, 2015). Therefore, organizations saw how critical big data is and, they are changing their concentration from programming and equipment to the information they measure, to acquire a competitive advantage in knowing more than their rivals (Paschalidi, 2015). A couple of years prior, companies did not use data for everything (Paschalidi, 2015). That is, they either did not have access or did not know what to do with it (Desalegn, 2016). These days, organizations use it to uncover new experiences in the business processes, market growth, or customer sentiment, as well as to predict and respond to changes (Paschalidi, 2015). However, getting data isn't enough (Paschalidi, 2015). Data itself is useless. What matters is what you do about it and how you manage it that has value. To investigate the value of data, we need to investigate the data first (LI et al., 2019). This requires the acknowledgment of data as corporate resources and the execution of some type of data governance system for viable data management and data analytics (Al-Badia, Tarhinia, & Khana, 2018).

On the other hand, transforming data into actionable analytics is the most ideal approach to support effectiveness, quality, and profitability (Yılmaz, 2019). During this time, some companies use big data to increase customer experience and create better decision making using data analytics. Big data analytics helps to make sense of vast volumes of data by turning it into intelligent information for decision-making (Hogarth & Soyer, 2015). And also, allows companies to better understand the knowledge found in the data and helps to determine the most important business-related data (Gaurav, Goyal, & Kaliyar, 2018). But, companies cannot compete purely on the basis of analytics (Feki & Boughzala, 2016). It is the governance that permits to return BDA and data management more successful (Feki & Boughzala, 2016).

Big data governance is a subset of a wider information governance program that develops policies for the optimization, privacy, and monetization of big data by harmonizing the goals of numerous activities (Al-Badia, Tarhinia, & Khana, 2018). It concentrates on concise, efficient, valuable, and adequate data services (Al-Badia, Tarhinia, & Khana, 2018; O'Neal, 2012). In addition, big data governance helps coordinate people, policies, processes, strategies, standards, and technologies to enable organizations to use data as one of their key business assets (Yang, Li, Elisa, Prickett, & Chao, 2019). It illustrates the information that helps to set the decision-making structures around data. Hence, the governance of big data is essential to the success of any organization that processes massive amounts of data (Al-Badia, Tarhinia, & Khana, 2018; Kim & Cho, 2018; Yang, Li, Elisa, Prickett, & Chao, 2019).

In the case of Ethiopia, Ethio telecom is one of them and is one of the leading producers of big data, which is used for a variety of purposes. Owing to the use of mobile phones and the explosion of social media and IoT, along with the emergence of next-generation communication networks, Ethio telecom faces an abundance of data on a daily basis. The data appears in batch and real-time mode. Remarkable data examples are call detail records, mobile network usage, user call data, user online data, user location data, network performance data, geographical user data, network monitoring, customer/subscriber profiles, hardware, and so on. Even though Ethio telecom is one of the big companies that produce a lot of data, it also faces technical difficulties in handling, storing, organizing, collecting, managing, processing, and analysing this enormous amount of data. Big data offers new possibilities for sustainable growth, but it also introduces several difficulties. These are primarily related to concerns about privacy, access, technology and ability. These issues together encompass the need to develop a framework for the organization to assist the use of big data for current decision-making and to guarantee the availability and efficiency of big data for future use. In the decision-making processes of organizations, big data governance has recently become very relevant (Kim & Cho, 2018). As companies try to implement big data analysis projects, they must develop well-formed big data governance frameworks to support the management of big data projects (Paschalidi, 2015).

Therefore, it is important to implement BDGF that helps for efficient storage and analysis of structured and unstructured data to carry out the company mission or to generate detailed, timely, accurate, and standardized information.

1.2 Statement of the problem

The problem is obtained from the growing significance of organizations to concentrate much more on their data. It is common for the majority of firms to realize the significance of their data but have no idea what to do with it. In the present era, when our digital lives are as powerful and essential as our regular daily lives and the exposure of hills of data keeps growing, firms need to understand what their big data is and how to use it effectively.

(Asressa & Beshah, 2016) point out that most organizational significant challenges and issues include the collection and processing of large volumes of structured and unstructured data, the management of high-speed data streams, the analysis of data, the discovery of quality or complete data, the cleaning of noise and data irregularities. On the other hand, according to (Akal, Beshah, Sackmann, & Negash, 2019) investigation, even though big data has been

produced by potential industries, they do not use it extensively for decision-making purposes. Some of the possible difficulties that may come are lack of trained experts in the field, lack of leadership and management skills, lack of BDA awareness, lack of accuracy and completeness of data, lack of structured data registry, data privacy concerns, as well as infrastructure issues, such as a massive amount of storage device limitations (Akal, Beshah, Sackmann, & Negash, 2019). Furthermore, (Yang, Li, Elisa, Prickett, & Chao, 2019; Al-Badia, Tarhinia, & Khana, 2018) noted also, lack of big data governance frameworks and digitization of businesses remains a big challenge for the majority of organizations.

Generally, the combined effect of big data and business tasks may result in more frequent and higher-risk data breaches (Kim & Cho, 2018). Hence, it is crucial to examine and tackle big data management problems (Yang, Li, Elisa, Prickett, & Chao, 2019). Therefore, a framework is essential for the effective execution of big data services to allow the implementation of a big data task as a guideline and framework (Kim & Cho, 2018). Big data governance helps organizations to make constructive use of organizational data and also, helps to build new analytics and results, along with the organization's strategic goals (Kim & Cho, 2018).

In the age of information, nearly every big company is confronted with big data issues (Madhlangobe, 2018). Telecom industry is one of the leading industries that produce a lot of data and has big data management problems (Zahid, Mahmood, Morshed, & Sellis, 2019). The fast expansion in the use of smartphones, internet broadband, peer-to-peer traffic, social media chatter, IoT (Internet of Things), and other connected mobile devices has created a sharp rise in the amount of data flowing through telecommunications companies' networks (Yayah, Ghauth, & Ting, 2017). The telecom industry is largely the cornerstone of a modern digital world (Kastouni & Lahcen, 2020). Because information passes on the telecoms backbone, the telecoms industry has become an important contributor to big data (Kastouni & Lahcen, 2020). In Ethiopian context, Ethio telecom is one of the big governmental organizations that produce big data and has big data problems. Depending on the large number of subscribers connected to networks on daily basis, Ethio Telecom was positioned with the chance of collecting a large amount of data. The company is now gathering ever more data by expanding its voice of customer service into the company. This pattern contributes to the characteristics of broad data from device logs, internet activities, and customer cell phone use. However, the perspective and potential of how big data governance frameworks should be implemented within the Ethio telecom are still vague. The company encountered numerous hurdles in establishing a BDA project, most notably in terms of selecting a technological solution from a myriad of available

tools and selecting governance approaches for managing the task and its data. Due to the lack of proper data management strategy and big data governance framework, the company has trouble with: network performance monitoring, customer segmentation, customer churn detection, fraud management & prevention, call drop analysis, network analytics, predictive campaign, and credit risk analysis. As a result of the aforementioned challenges, major issues arise, such as the lack of making smart investment decisions, improper risk management, loss of potential consumers, loss of profit, high capital expenditure and competitiveness, are among the most insignificant. Hence, the need for big data governance is now more essential than before. Currently, there is no such framework in place to govern bug data, and there is considerable confusion within the organization about who is responsible for big data, who is aware of large datasets, and what should they do with it.

Furthermore, there are very few studies that can fill this gap. Of these, the study's by (Al-Badia, Tarhinia, & Khana, 2018), (Yang, Li, Elisa, Prickett, & Chao, 2019), and (Fikiru, 2020) are exemplary. (Al-Badia, Tarhinia, & Khana, 2018) investigates existing big data governance frameworks and their weaknesses and proposes a new conceptual framework. The framework was developed through an examination of the existing governance issues in big data. This study describes the organization's structure identification, setting the policies and standards, stakeholders' identification, optimizing and computing, identifying the scope of big data, measuring and monitoring quality, data storage, communication and data management without any detail of a framework. The study lacks clarification and a telecom background.

A research project by (Yang, Li, Elisa, Prickett, & Chao, 2019) recommends a big data governance framework that assists organizations to make bigger data-informed business decisions within the applicable framework, with a focus on data privacy, confidentiality, availability, security, and accessibility. To prove this workflow, the work also includes an illustration framework execution case study approach of big data governance in cybersecurity. However, several of these studies endorsed the concept of using conventional data governance attributes for big data such as organizational benefits and challenges, top management, metadata, business process integration, IT tools and infrastructure, and business end-users without addressing the details of big data analytics lifecycle management in associated with big data governance framework. Therefore, there is a gap in finding in-depth research that incorporates both big data governance framework and details of big data analytics lifecycle management method (covering all the steps from business case evaluation to visualizations of data) in a single central framework. To address the limitations and research gaps presented in

this section, the researcher recommends big data analytics framework (capable of collecting, storing, analysing, and querying both streaming and historical (batch) data in a centralized single data framework) along with BDGF. Because the goal of the big data governance framework is to effectively manage big data projects and produce value from data by employing a new analytical method (Al-Badia, Tarhinia, & Khana, 2018).

To sum up, existing big data governance frameworks are contextual and are not customized for Ethio telecom. “Contextual factors such as organizational, national and environmental” (Bogale, 2018) perspective affect the design of big data governance framework. Even if, there is one research that has a big data management framework for Wegagen Bank S.C, while the study seems applicable for the financial sector, it lacks a telecom context to address the issues under our study. The financial services sector is a sector that provides financial services to individuals and organizations and is one of the sectors that need and use the Internet, telephone services and telecom infrastructure. On the other hand, Ethio Telecom is an Ethiopian telecommunication company serving as the major internet and telephone service provider (Getachew, 2018). Ethio Telecom has a plan to be a world-class provider of telecom services and to provide world-class, modern, and high-quality telecom services for all citizens equitably to transform the multifaceted development of the country to the highest level. Thus, assessing the current state and designing big data governance framework help Ethio Telecom to increase organizational success and ensure risk, satisfy user's needs, profitability growth, and performance.

Therefore, BDGF standard procedure recommended to other companies cannot be directly used for Ethio Telecom unless it is explored, investigated, and studied in the context of Ethio telecom.

1.2.1 Research questions

Based on the problem statement, the research seeks to answer the following questions:

1. What is the existing big data governance practice in Ethio telecom?
2. What workable big data governance framework can be designed for Ethio telecom?
3. How the proposed framework can be integrated seamlessly with the main business process of Ethio telecom?

1.3 Objectives

1.3.1 General objective

The main objective of this study is to assess and design a big data governance framework for Ethio telecom, to assist the company to achieve bigger data-driven business decisions.

1.3.2 Specific Objective

The following specific objectives are formulated to meet the general objective of this study:

- Assess the existing big data governance practices in Ethio telecom
- Identify major factors that impact the practices of big data governance in Ethio telecom,
- Review available big data governance frameworks that can be used as a basis to build big data governance framework for Ethio telecom.
- Formulate a suitable big data governance framework for Ethio telecom.
- Evaluate the proposed framework, and
- Recommend further works in the subject

1.4 Significance of the study

Aside from serving as an academic exercise to satisfy the requirement of (a master of science in Information systems), the purpose of this research is to produce results that can be used to improve big data management culture in the telecommunications industry. Therefore, this research is important for appropriately understanding and managing both structured and unstructured data, getting maximum value from big data, and empowering and developing good big data practice.

Furthermore, it guides Ethio telecom to make better decisions around data usage and boosting economic growth. But not only guide organizations to make smart decisions about data usage based on their critical success factors but also assist organizations in efficiently achieving the desired results in innovative data utilization, which frequently leads to a culture change in organizations for deep data-driven processes.

And also, allowing them to be used by someone with negligible technical skills. For example, data warehouse experts, data analysts, big data researchers, big data developers, business analysis specialists, system administrators, and information architecture specialists, as well as database administrators, will benefit significantly from this centralized big data governance architecture.

This research also opens the way for future academic research in the area of big data-related activities especially data governance/ management tasks by serving as a reference material and starting point for further investigation.

1.5 Scope and Limitation of the Study

As the title implies, the scope of this study is designing big data governance architecture for Ethio telecom. The research process involves an assessment of current big data governance practices/experience in Ethio telecom, which encompasses the awareness aspect of big data, big data analytics, big data governance, and the overall management of data in Ethio telecom. And also, major factors that impact the practices of BDA, BDG, and BDM have been assessed. Finally, numerous different BDG frameworks from the business and telecom scenarios are explored, and a BDG framework is formulated and proposed that can be best suited to the Ethio telecom context. The data collection process of this research is also limited to Ethio Telecom's Legehar branch in Addis Abeba. This branch is a place where major data services and devices, as well as an IT manpower office, are situated. Such employees may be more knowledgeable about the concepts in this study than other employees who work in HR, finance, and more or other districts. Even though the selected staff had a better understanding than the others, there was still a gap in the respondents' knowledge of big data and related concepts.

In addition, due to the difficulty of obtaining sample data sources from Ethio telecom to measure the benefits of the proposed BDG framework and process, it is not possible to illustrate the demonstration of the proposed framework. Therefore, the study's limitation is the evaluation of the proposed big data governance framework, which would be finished after implementation by providing feedback mechanisms to evaluate the effectiveness of the suggested BDG framework and allow it to evolve and change iteratively as needed.

1.6 Organization of the Thesis

This research is organized into six chapters. The first chapter is an introduction chapter in which the overview and background of the study, statement of the problem, objective, research questions, significance, scope, and limitation of the study are discussed.

In the second chapter literature review of BD, BDG, related concepts and models, and directly related works have been discussed.

The third chapter presents the methodologies used by the researcher in the research study. In this chapter, the research approach, research design, study area, population size, sampling

methods, method of data collection, ethical consideration, analytical techniques, and strategies, as well as methodologies for validating the thesis and the research's reliability, are addressed. The fourth chapter discusses the findings and data analysis based on the evidence accumulated. Chapter five describes the proposed big data governance framework for Ethio telecom based on the findings of the analysis of the research and reviewing existing big data governance practices. And the last chapter gives a synopsis and suggestion dependent on the discoveries.

CHAPTER TWO

LITERATURE REVIEW

2.1 Big data

The term big data has recently become a buzzword and prominent (Mauro, Greco, & Grimaldi, 2014), particularly with the rapid advancement of technology, the digital revolution and the increasing overdependence of the human race on social media and the internet. Almost everyone, which includes academicians and industry professionals, uses it (Bhadani & Jothimani, 2016).

There are various definitions available in the literature. Of this, many researchers have cited the definition by Gartner Glossary; which describes big data as “high volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation” (Glossary, 2012). The attention here is on big data's attributes, which are large volume, greater velocity, and a wide variety. On the other hand, according to Laney, the concept of big data was defined with a 3Vs model (Wang, Wei, Zhan, & Sun, 2017). 3Vs, also known as the characteristics of big data, represent the increasing volume, variety, and velocity of data (Yayah & Ghauth, 2017).

Today's enterprises rely on complex business processes, and the volume of data generated by diverse business activities and operations is expected to expand at an exponential rate (Madhlangobe, 2018). So, it is unthinkable, under all situations, not to use the principle of big data (Madhlangobe, 2018). Because of there has been a substantial rise in data over the last two decades (Doshi, Agrawal, Kanani, & Padole, 2020). This trend can be observed in almost every field and every industry. For this display, the volume of data increases exponentially per day at an awkward angle from Terabytes to Petabytes (= 10^{15} bytes) and Zettabytes (= 10^{21} bytes) (Doshi, Agrawal, Kanani, & Padole, 2020). This growing dissemination of data is known as big data (Doshi, Agrawal, Kanani, & Padole, 2020).

Big data is a collection of data sets (or a blend of data sets) that is growing daily (Panimalar.S, Shree.S, & Kathrine. A, 2017). Because data is generated by everyone and for everything from portable devices, call centres, network devices and social media outlets (Panimalar.S, Shree.S, & Kathrine. A, 2017). However, the issue is that it is too huge, too quick, and also difficult to manage for conventional systems and existing technologies (Al-Badia, Tarhinia, & Khana, 2018). Numerous organizations collect huge amounts of data produced by high-volume

transactions such as call centres, sensors, weblogs, and digital files (Kastouni & Lahcen, 2020). For their business to achieve great things, they must overcome big data challenges while continuously enhancing productivity and improving operational efficiency (Panimalar.S, Shree.S, & Kathrine. A, 2017).

Every second, a massive amount of data sets are created from all over the world, implying that the volume of data will never decrease but will only grow (Panimalar.S, Shree.S, & Kathrine. A, 2017). Figure 1 clearly illustrates the tremendous development of big data volume over time.

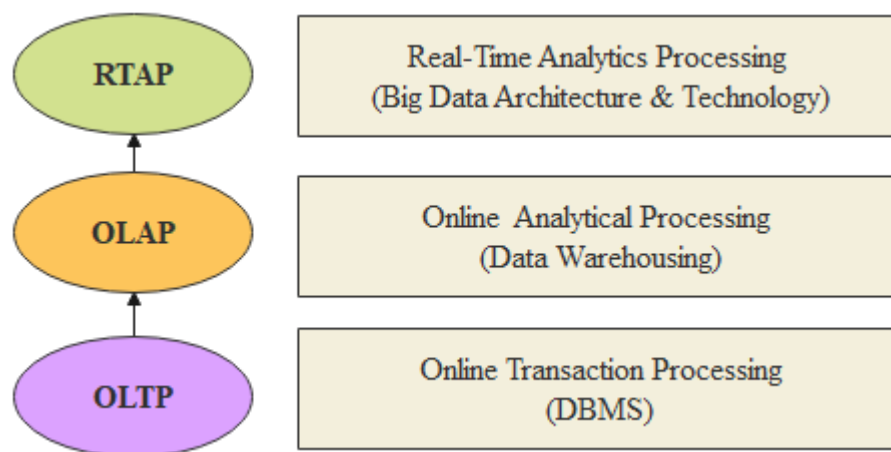


Figure 1: Growth of Data (Panimalar.S, Shree.S, & Kathrine.A, 2017)

To summarize, big data is a collection of massive or complicated datasets that surpass the processing capabilities of traditional data processing systems, and the problems include collecting, processing, managing, generating, storing, sharing, and visualizing data. It is not a single method or tool; rather, it has become a comprehensive and complex subject that includes a variety of tools, techniques, and frameworks.

2.2 Characteristics/ Features of Big Data

There are many interpretations about big data, but many authors claim that big data best describes, primarily, the following variables, collectively known as the 3V model (Bhadani & Jothimani, 2016).

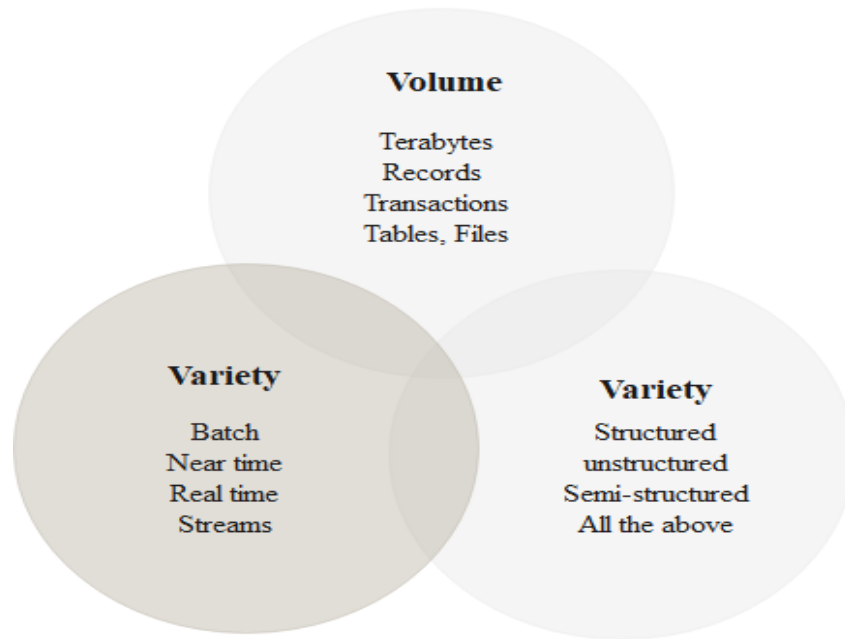


Figure 2: 3v model (Riahi & Riahi, 2018)

- 1) **Volume:** volume is the quantity defined by the data size. The massive amount of data requires a reconstruction of data processing pipelines to facilitate the use of traditional analytical tools (Yayah, Ghauth, & Ting, 2017). There were also implications for the storage and processing of such data due to the huge data volume.
- 2) **Velocity:** velocity is characterized as the amount of data flow from different sources such as sensors and logs files of fresh data events in real-time (Yayah, Ghauth, & Ting, 2017). As the data speed of arrival grows, new constraints are placed on storage and processing frameworks.
- 3) **Variety:** big data comes from different sources which demonstrate the diversity of these sources. It can be structured, unstructured, or semi-structured data. There are different formats and different structures of big data (Sassi, Anter, & Bekkhoucha, 2019). Which are listed below:
 - Structured data: it can be handled, stored, retrieved, processed and interpreted in the form of supervised manner/ fixed format. They are used in relational databases.
 - Unstructured data includes data that either doesn't even have a pre-defined database schema and/or doesn't function well in relational tables. Most of the data generated are unstructured, so it cannot be addressed by conventional database management tools. Examples of such data are E-mail, books, scripts,

pictures, word files, documents, assets for multimedia, digital records, papers, and records.

- Semi-structured data is a hybrid of the previously mentioned types of data, typically represented by XML. It is a form of structured data that does not adhere to a defined and predetermined schema.

Later, a few additional dimensions were added, which are shown below (Bhadani & Jothimani, 2016):

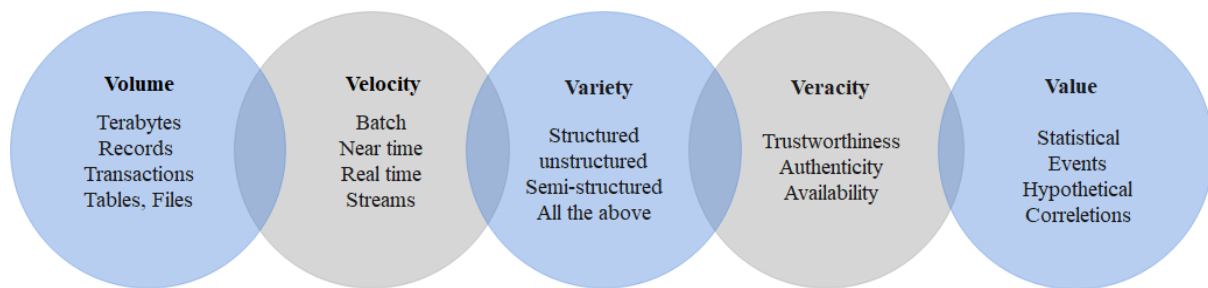


Figure 3: 5 v model (Riahi & Riahi, 2018)

- 4) Veracity: level of quality, accuracy and uncertainty of data and data sources (Riahi & Riahi, 2018).
- 5) Value: the value and potential derived from data (Riahi & Riahi, 2018). This means, data usability and the practical possibilities of their use in decision-making and, consequently, to the possibility of generating important value for an organization.

To sum up, based on the key characteristics mentioned above, big data is always confronted with many problems, some of which are created by the nature of big data at the level of information acquisition, storage, processing, transmission and management.

2.3 Big data in telecom industry

Regarding, big data in telecoms companies, (datareportal, 2021) discovered that roughly 60% of the world's population now has internet access. This indicates that they have increased by 20% since 2016 (Khan et al., 2016). The number of Internet users has grown massively from 0.4 billion in 2000 to 4.72 billion in 2021 (Khan et al., 2016) & (datareportal, 2021). According to Datareportal annual internet trend report for 2021, 4.72 billion people worldwide use the internet in April 2021, accounting for more than 60 of the population of the world (datareportal, 2021). This number is still growing, with the most recent data indicating that 332 million new

users joined the internet in the last year. Every day, over 900,000 new users register to the network (datareportal, 2021).

In addition to this, social media and unique mobile phone user continues to grow too, with global users reaching there are 4.33 billion social media users and 5.27 billion unique mobile phone users in the world today (datareportal, 2021).

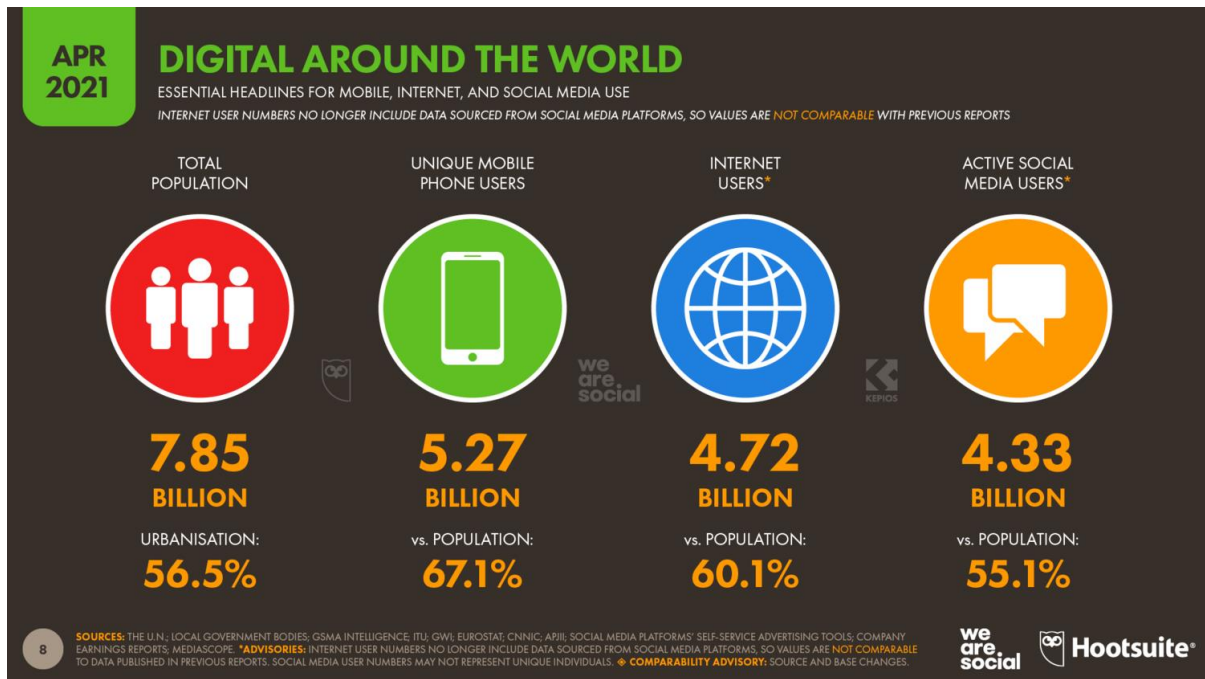


Figure 4: The number of mobile, internet and active social media users in the world (datareportal, 2021)

The main reason for this is the Covid-19 pandemic and the digital revolution (Kastouni & Lahcen, 2020) & (datareportal, 2021). But especially in the last 12 months, the Covid-19 pandemic has played a major role (datareportal, 2021). Because of the ongoing Covid-19 epidemic, the worldwide population has become heavily dependent on residential Internet connectivity for employment, education, social activities, and entertainment (Kastouni & Lahcen, 2020). That means, due to social distancing rules and nationwide lockdowns, the Covid-19 pandemic has ultimately led to an increase in the usage of digital technologies. People and organizations around the world have had to adapt to new work and lifestyles (Kastouni & Lahcen, 2020). As a result of increased digitalization, more organizations and educational institutions are implementing work-from-home policies (WFH) (Kastouni & Lahcen, 2020). So, Total internet traffic has increased enormously in the last twelve months (datareportal, 2021). This raises questions about how traffic characteristics changed over time and whether these changes posed a threat to the Internet infrastructure's ability to function (Kastouni &

Lahcen, 2020). All of this movement or change reflects the fact that the telecom industry is dealing with a tsunami of data on a daily basis as a result of smartphone usage, the rise of social media and the Internet of Things, as well as the availability of next-generation communication networks (Zahid, Mahmood, Morshed, & Sellis, 2019). Data occurs in both batch and real-time modes (Delotte, 2015). In a telecom corporation, there are numerous forms of data, such as user call data, user online data, user location data, network traffic data, and so forth (Tata Tele Business Services, 2021).

Network infrastructure	Product	Marketing and sales	Customer care	Billing
<ul style="list-style-type: none"> ➤ Network events ➤ Call records (on and off network) ➤ Number of text and multimedia messages ➤ Volume of data traffic ➤ Location-specific data ➤ User handset data ➤ Technical fault data 	<ul style="list-style-type: none"> ➤ Product catalog ➤ Product life-cycle data ➤ Product and platform costs ➤ Innovation road map ➤ Product usage ➤ Critical products ➤ Product delivery management 	<ul style="list-style-type: none"> ➤ Customer devices ➤ Option preferences ➤ Sales channel data ➤ Response rate of marketing campaigns ➤ Segmentation data ➤ Usage patterns 	<ul style="list-style-type: none"> ➤ Order data ➤ Contact data ➤ Fault handling data <ul style="list-style-type: none"> ✓ Problem type ✓ Resolution time and rates ✓ Repeated faults ➤ Call center logs ➤ Termination reasons 	<ul style="list-style-type: none"> ➤ Call duration records ➤ Tariff data ➤ Usage history ➤ Customer account data

Table 1: Potential Data Availability (Tata Tele Business Services, 2021)

However, in general data in the telecom market can be divided into three categories, as shown below (Hafez, 2016).

- **Call detail data** refers to “information about the call, which is stored as a call detail record (CDR). For every call placed on the network, a call detail record is generated to store the call details” (Hafez, 2016). CDRs contain information about phone numbers that originated and terminated calls, as well as call duration and period (Joseph, 2013; Weiss, 2005). In addition to this, the growing use of smartphones has resulted in the emergence of a new type of transaction record known as extended data records (XDRs)

(Hafez, 2016). Other activities, such as mobile payments and the buying and downloading of a music or video clip, are captured by the XDRs (Dam, 2013).

- **Network data**, such as log files, error generation and status messages, is created in real-time and in massive quantities (Hafez, 2016). These records are used to assist overall network management functions (Hafez, 2016).
- **Customer data** contains personal information including the customer id, name, address, subscription type, account details and payment history (Hafez, 2016).

Billions of customers' data stored in data warehouses system must be managed to recognize key customer attributes (Hafez, 2016). Therefore, it is important to analyse the big data in the telecommunications companies as described above to provide information that would help them improve income and reduce expenses (Kastouni & Lahcen, 2020).

2.4 Big data analytics

BDA refers to a new breed of technologies and systems that enable high-velocity data gathering, discovery, and analysis to cheaply generate profit from huge volumes of a wide range of data (Côte-Real, Oliveira, & Ruivo, 2016). The authors view BDA as the ability (people, processes, and technology) to process large data sets to discover insights, correlations, and useful information. As defined by (Abbasi, Sarker, & Chiang, 2016), big data analytics (BDA) employs advanced analytical techniques to generate models from large data sets through structured modeling processes. Modeling process, machine learning, and big data are all included in this concept.

The development methodology of BDA solutions includes data extraction, store, and analysis, present and decide (Madhlangobe, 2018). However, without an organized data gathering, storage, and analysis procedure, decision-making might be misleading or inaccurate (Poletto, De Carvalho, & Seixas Costa, 2015). That is, big data analytics software platforms can assist businesses in making data-driven decisions and improve organizational productivity. More effective marketing, additional income possibilities, customer personalization, and improved operational efficiency are all potential benefits (Zahid, Mahmood, Morshed, & Sellis, 2019). These potential benefits can provide a competitive advantage over competitors when combined with a successful strategy/data governance framework (Al-Badia, Tarhinia, & Khana, 2018).

2.4.1 Big data analytics applications in telecom industry

There are various application areas of big data in telecom industry, of this

- Customer experience
 - ✓ Customer Segmentation
 - ✓ Customer churn prevention
- Network Management & Optimization
- Fraud detection

1. Customer experience

Today's telecom is improving and optimizing customer experiences, which are vital to building market distinction and churn lowering (Tata Tele Business Services, 2021). The best marketing strategy of all is a delighted customer (Tata Tele Business Services, 2021). Organizations must follow the new advanced technologies that will help them grow to sustain competitive advantage and stay successful in this ever-changing and highly competitive world (TechVidvan, 2021). To expand the number of clients, the right products at the right moment can be provided to them (Tata Tele Business Services, 2021).

Telecom would then perform effective micro-segmentation of their client base based on the detailed customer information, provide a compelling user experience, make individualized offer predictions, and forecast churn based on the detailed customer history (Tata Tele Business Services, 2021).

- 1) **Customer segmentation:** refers to the practice of dividing a market or customer base into groups that behave similarly (Delotte, 2015). This technique is becoming increasingly popular as telecom companies expand to serve a bigger number of consumers, and it is a vital component that allows telecom operators to make strategic decisions (TechVidvan, 2021). A few of the advantages that telecom firms may gain through segmentation are stated below (Delotte, 2015):

- Customer Value Segmentation – “Identify loyal customers who have a high potential lifetime value, enabling targeted marketing and retention activities to reduce the churn rates” (Delotte, 2015)
- Create tailored products for customers
- Identifying high-value and long-term customers
- Identify potential customers

- 2) **Customer churn prevention:** to avoid customer churn, it is essential to analyse the customer's behaviors and take appropriate actions (TechVidvan, 2021). Big data enables intelligent platforms that combine customer transaction data with data from real-time communication streams (TechVidvan, 2021). Insights into this data

can help to reveal the client's opinion on a particular service. This enables telecom organizations to address client satisfaction problems as soon as possible, thus aids to prevent churn (Chen, 2016).

2. Network Management and Optimization

Any telecom industry's network traffic bandwidth is one of its most significant assets (Tata Tele Business Services, 2021). These industries are monitoring and regulating their network capacities through big data analytics (TechVidvan, 2021). This allows for better decisions to be made regarding the expansion of their network data platforms (Delotte, 2015). The network expansion solution can be found simply by analysing real-time network traffic in heavily populated regions (Tata Tele Business Services, 2021). This will help to optimize the network's cumulative expansion expense (TechVidvan, 2021). Championing suitable options in areas where network activity is low will help to increase traffic and generate more revenue (Delotte, 2015).

3. Fraud detection

The telecom industry is the one that entertains the majority of customers on a daily basis, and it is like a cream on the cake of fraudsters (Tata Tele Business Services, 2021). This massive data set, which includes millions and billions of individuals, is ripe for a variety of scams (Tata Tele Business Services, 2021). Illegal access, hacking, false profiles, cloning, and other fraudulent actions are all frequent in the telecom industry (TechVidvan, 2021). Fraud has a direct impact on the relationship that has been formed between both the firm and the client (TechVidvan, 2021). That is why big data analytics is so important in the world of telecommunications industry (Chen, 2016). Thus big data analytics aids in real-time controlling to detect fraud (TechVidvan, 2021). This method is extremely efficient because it enables a near-real-time response to any illegal behaviour (TechVidvan, 2021).

2.4.2 Big data challenges in telecom

Big data has the potential to put communications service providers (CSPs) in a strong position to gain clients and develop new income streams (Delotte, 2015). It gives companies a plethora of data on their clients' habits, preferences, and activities (Delotte, 2015). Despite this, numerous CSPs are having problems getting the maximum value from big data (Bhadani & Jothimani, 2016). Along with its variety, velocity, and complexity, big data analytics presents many challenges to CSPs on a global level (Delotte, 2015; Chen, 2016).

Variety: telecoms companies are struggling to deal with the onslaught of data generated by connected devices, customer behaviours, social media networks, call data records, government sites, payment information, and other sources (Chen, 2016; Delotte, 2015). At this point, the majority of data gathered from various sources is unstructured (Bhadani & Jothimani, 2016). To standardize the data for business intelligence platforms, telecom companies must enhance their Call Data Records (CDR) with additional information such as location-based services, financial records, and more (Delotte, 2015).

Velocity: as explained above, 2.5 quintillion bytes of data are created every day (The-Next-Tech, 2021). By 2020, the digital universe was around 40 ZETTABYTES (Tata Tele Business Services, 2021). That's 5,247 GB of machine-generated data for every person on the planet (Tata Tele Business Services, 2021). All of this indicates that the rate at which data is generated is extremely fast and that to extract value from it, it must be handled in a timely manner (Al-Badia A. A., 2018; Yang, Li, Elisa, Prickett, & Chao, 2019). This large amount of data necessitates new real-time operating capabilities for a variety of purposes, which necessitates more data storage for compliance and potential future uses, as well as new tools for filtering, managing, organizing, and storing data within available time constraints (Delotte, 2015).

Complexity: because there is no common format for storing data, user-generated data is typically unstructured and complex (Delotte, 2015). The existing network and storage devices do not have a standardized template for storing data that could be useful for advanced analytics (Yang, Li, Elisa, Prickett, & Chao, 2019). The data varies based on demographic factors, geographic location, way of living, and more (Delotte, 2015). If the data isn't adequately filtered, analytics can produce undesired findings (Al-Badia, Tarhinia, & Khana, 2018).

On the other hand, telecom businesses confront additional challenges when they implement big data for advanced analytics (Delotte, 2015).

1.	<ul style="list-style-type: none"> ➤ Determine a strategy how to leverage on the benefits of big data ➤ Define criteria for evaluating return on investments ➤ Determine business drivers and if big data can play a role in better insight
2.	<ul style="list-style-type: none"> ➤ Identifying and acquiring the skill sets required to understand and leverage big data ➤ Organizing business analysts team with strong knowledge of company ecosystem
3.	<ul style="list-style-type: none"> ➤ Flexibility of infrastructure to interact with various data formats and large volumes ➤ Cost and effort associated with scalability
4.	<ul style="list-style-type: none"> ➤ Identifying the best software and hardware solutions and determining the best overall Infrastructure solution ➤ Transitioning from legacy systems to newer technology
5.	<ul style="list-style-type: none"> ➤ Compromise of quality due to volume and variety of data ➤ Cost of maintaining all data quality dimensions: completeness, validity, integrity, Consistency, timeliness, accuracy
6.	<ul style="list-style-type: none"> ➤ Re-evaluation of internal and external data policies and regulatory environment ➤ Privacy issues related to direct and indirect use of big data sources

Figure 5: Big data challenges in telecom (Delotte, 2015)

2.5 Data Governance

The digital revolution offers tremendous opportunity for the public and private sectors and organizations to capture, store, process, and transfer massive quantities of data; as a result, they face growing issues related to data security, data structure management, data privacy protection, and data quality creation (Yang, Li, Elisa, Prickett, & Chao, 2019; Asressa & Beshah, 2016). Data governance is a concept or structure that organizations can use to resolve such difficulties in digital asset management (Al-Badia, Tarhinia, & Khana, 2018; LI et al., 2019). People, processes, policies, and technologies are all part of data governance, which allows an organization to use data as a digital asset (Yang, Li, Elisa, Prickett, & Chao, 2019). It establishes a broad framework for managing and maintaining data quality, security, availability, usability, relevance, and integrity (Yang, Li, Elisa, Prickett, & Chao, 2019). It also guarantees that accurate data is used to define corporate goals, maintain business processes, and make key decisions (Al-Badia, Tarhinia, & Khana, 2018).

On the other hand, “Data governance refers to the exercise of authority and control over the management of data” (Abraham, Schneider, & Brocke, Data Governance: A conceptual framework, structured review, and research agenda, 2019). Its goals include establishing a corporate-wide data agenda, maximizing the value of data assets, and addressing data-related risks (Kim & Cho, 2018). Although data governance used to be wonderful to have in the previous, it is now becoming more important in businesses and government agencies (Haneem, Kama,, Taskin,, Pauleen,, & AbuBakar, 2019). This is because of a few key trends. The total volume of data generated annually on the universe is forecast to expand from 4.4 zettabytes in 2013 to 44 zettabytes in 2020 (IDC, 2014). The increasing volume of data from various sources triggers data inconsistencies, which must be defined and understood before decisions are made based on wrong data (Al-Badia, Tarhinia, & Khana, 2018). Companies are implementing more self-service reporting and analytics, which necessitates a mutual understanding of data throughout the organization (Abraham, Schneider, & Brocke, 2019). The ongoing impact of regulatory requirements such as the General Data Protection Regulation (GDPR) puts additional pressure on businesses to have a clear understanding on what data is stored where and how the data is used (Abraham, Schneider, & Brocke, Data Governance: A conceptual framework, structured review, and research agenda, 2019).

Traditional data governance frameworks as reviewed by (Yang, Li, Elisa, Prickett, & Chao, 2019) “usually only consider data principles, data quality, and metadata management for traditional structured and moderately small-sized data sets, rather than high-volume, high-variety, and high velocity live data”. In a big data scenario, the vast amount of data processed within or outside organizations necessitates an additional level of management for data quality, security, and ethical data processing (Akal, Beshah, Sackmann, & Negash, 2019; Yang, Li, Elisa, Prickett, & Chao, 2019). Furthermore, the combined effect of big data and business tasks may increase the frequency and severity of data breaches (Yang, Li, Elisa, Prickett, & Chao, 2019).

According to the researchers (Katal, Wazid, & Goudar, Big Data: Issues, Challenges, Tools and Good, 2013; Akal, Beshah, Sackmann, & Negash, 2019) illustration, the major challenges that the existing data governance frameworks face for big data are:

- The absence of frameworks on big data governance.
- There is a scarcity of qualified big data experts.
- Big data security and privacy.
- Lack of required tools to generate insight from the data in a timely manner.

- The unwillingness of organizations to use data in goal-setting and decision-making.
- Corporate managers have insufficient knowledge about big data.
- Most firms are still finding it difficult to digitize.
- The complexity of the data collected and stored (e.g. unstructured data).

2.6 Big Data and Ethio telecom

2.6.1 Overview of Ethio telecom

Ethio telecom, formerly known as Ethiopian Telecommunications Corporation (ETC), is an Ethiopian telecommunications firm that serves as the country's primary internet and telephone service supplier. It is Africa's oldest Public Telecommunications Operator (PTO). Telecommunications services were first developed in Ethiopia in 1894, when Menelik II, King of Ethiopia, launched telephone technology. But, the first Ethiopian founder of telephony was his cousin Ras Mekonnen, who returned from Italy with telephone machinery in 1889 and formed a corporation.

Following the federal government's determination to focus on enhancing telecom services as a cornerstone to national development, Ethio telecom has had its present status since November 29, 2010 as part of Ethiopia's 2005/06 – 2009/10 GTP. As a result, the country's telecom services and infrastructure have been upgraded to global levels, facilitating national growth and resulting in a significant paradigm change in the sector's progress (ethiotelecom.et).

The following table depict Ethio telecom's position as of July 2021.

Fixed Broadband subscribers	374K+
Fixed Service Subscriber	900K+
Data and Internet Users	25M+
Mobile Service Subscribers	54.3M+
Total Customer	56.2M+
Services	Mobile Fixed-line Broadband Tele berr
Revenue	36 billion berr
Operating income	12.4 billion berr
Net income	5.7 billion berr

Total assets	10 billion birr
--------------	-----------------

Table 2: Ethio telecom position as of July, 2021

2.6.2 Big data Governance Current Practice in Ethio telecom Industries

Currently, big data governance practices in telecoms are relatively low. The organization has accumulated a lot of data over many years but the stored information is not used properly in value-added approach. On the other hand, Ethio Telecom has become one of the biggest firms in terms of data volume, but they still don't know how to effectively monetize the data itself. The concern should be more on how the data can be utilized to reduce operating expenses, provide a more personalized client experience, lower churn, establish new sources of revenue, and more.

In addition to this, Ethio telecom generates a massive amount of data, up to billions of CDRs/events every day. Customer, usage, and network data are examples of this type of information. Gathering, storing, interpreting, and analysing this volume of data successfully might be a serious issue for Ethio telecom. Therefore, the infrastructure needs to have a strong capacity for computing and storage. As a result, it's critical to create the best architecture possible to manage these technological challenges while also meeting commercial needs.

Moreover, when it came to launching a BDA project, Ethio telecom will face a number of hurdles, including a decision on a technical solution from a variety of available tools and determining governance approaches for managing the data.

To sum up, at a local scale, Ethio telecom businesses have difficulties when it comes to implementing big data for advanced analytics. Of this:

- Establish a strategy for capitalizing on big data's advantages.
- Finding and obtaining the necessary skill sets to comprehend and utilize large data.
- Organizing a group of business analysts who are well-versed in the company's ecosystem.
- Infrastructure agility to interface with a variety of data formats and massive volumes.
- Choosing the greatest software and hardware solutions, as well as the highest performance infrastructure solution.
- Shifting from older systems to newer technology and more.

2.7 Big data governance framework

The goal of implementing data governance is to ensure that there is a promising solution of employing data to achieve the organization's business goals and purposes (Al-Badia, Tarhinia, & Khana, 2018; Yang, Li, Elisa, Prickett, & Chao, 2019). Data governance procedures are often regulated by a framework for data acquisition, storage, management, and usage (Yang, Li, Elisa, Prickett, & Chao, 2019). The framework is created to ensure data privacy, usability, availability, reliability, and integrity (Berson, Dubov, Plagman, & Raskas, 2011). It should enable the flow of data and business processes within an organization and even the culture and structure of an organization (LI et al., 2019). And also aids in directing workers as they carry out their data management responsibilities (Yang, Li, Elisa, Prickett, & Chao, 2019). A well-developed data governance framework typically includes data management approaches, organizational drivers, data management structure, technological advances, and strategies (Berson, Dubov, Plagman, & Raskas, 2011).

A data governance framework is a collection of procedures that guarantees that critical data assets are officially handled across the organization (LI et al., 2019; Al-Badia, Tarhinia, & Khana, 2018). Data governance ensures that the correct data sets are made available to the right people at the right time, allowing for the best decisions to be made (Al-Badia, Tarhinia, & Khana, 2018). Data governance needs a framework for managing information. According to (Yang, Li, Elisa, Prickett, & Chao, 2019) investigation Data governance in a big data environment mainly focuses on three areas, as listed below:

- 1) **Data architecture:** Big data is gathered from a wide range of sources, including structured, semi-structured, and unstructured data. Data management becomes more complex with such a diverse structure of data. Furthermore, various organizations keep their data structure standards, which hamper information exchange across systems across different enterprises, reducing the effectiveness of information extraction and use substantially.
- 2) **Data quality:** Data quality can be an actual problem for big data management because most existing measures cannot be properly adjusted to big data. For example, the real-time nature of big data necessitates that organizations focus on improving data access performance, thereby reducing data transmission delays.
- 3) **Data security, privacy, and ethical considerations:** Data sharing and analysis techniques can effectively raise the value of data for a company by improving data transparency, but they can also inappropriately reveal sensitive data or leave doors open for cybercriminals

to manipulate data without permission. In fact, one of the most significant difficulties of data governance in the era of big data is big data security and privacy. Besides this, the ethical considerations of data management have become an important concern, as have the rising tensions of data providers and the increased participation in data donations.

On the other hand, as mentioned above in the introduction part, Big Data is a good opportunity for organizations but at the same time, it has put big problems in cleaning, governing, and managing the data. The Big Data problem needs to be managed (Al-Badia, Tarhinia, & Khana, 2018). Big Data implementations necessitate the adoption of new governance policies (Abraham, Schneider, & Brocke, 2019). A robust Big Data governance strategy is needed, as this is the cornerstone of information management (Soares, 2013). Stewardship, information governance, data definition and usage standards, master data management, metadata management, data lifecycle management, risk, and cost containment must all be considered while developing Big Data governance (Zarate Santovena, 2013). Big data governance covers the data governance program that includes optimization policies, privacy and big data monetization in accordance with the set objectives (Soares, 2013).

Big data governance framework is essential for defining rules, policies, strategies, methods, processes, procedures, and guidelines for properly managing and assuring big data accessibility, usability, reliability, accuracy, accountability, and confidentiality of big data (Al-Badia, Tarhinia, & Khana, 2018). The governance of big data is vital to the development of any organization that processes large volumes of data (Abraham, Schneider, & Brocke, 2019; Charikleia, 2015). In order to handle big data in an organization, a big data governance framework is necessary (Al-Badia, Tarhinia, & Khana, 2018; LI et al., 2019; Yang, Li, Elisa, Prickett, & Chao, 2019).

2.7.1 Top Frameworks and Software Tools for Handling Big Data

Since the emergence of big data, a number of tools/frameworks have been developed by experts and researchers to assist in the process of data analysis (Ms. Komal, 2018). This section presents some of the most prominent and commonly used big data frameworks.

2.7.1.1 Apache Hadoop

Apache Hadoop is a free and open-source java-based software framework for storing data and running applications on clusters of commodity hardware (ApacheSoftwareFoundation, 2021). It offers distributed storage as well as computational capabilities, HDFS for storage and Map-

Reduce for computation (Kumar & Valcin, 2020). Aside from that, it is not support OLAP (online analytical processing), as well (Kumar & Valcin, 2020). This means it is used for batch processing. Hadoop is often run in a cluster, which means it is always run in a homogeneous environment (Singh & Singla, 2015). Furthermore, a homogeneous environment means that all of the systems in the cluster's elements, such as RAM and CPU, must be the same (Singh & Singla, 2015).

The traditional method of dealing with data is storing them in a database to be processed through SQL queries, however, this approach becomes inefficient when dealing with large and unstructured data, HDFS, as well as the Map Reduce, developed to address this problem (Perera, 2013). Basically, Hadoop consists of two main components (Kumar & Valcin, 2020), one is HDFS (Hadoop Distributed File System) and another is Map Reduce.

1. HDFS

HDFS is a file system that is built on the master-slave model (Kumar & Valcin, 2020). It divides huge files into default 64 MB blocks and efficiently stores them in large clusters (Singh & Singla, 2015). A Hadoop cluster consists of two basic nodes, the data node and the name node, according to this architecture (Singh & Singla, 2015). On the Hadoop cluster, there is a secondary name node as well (Kumar & Valcin, 2020). The Name Node is the master node that controls all of the data nodes and stores Metadata (Kumar & Valcin, 2020). It maintains all file operations such as read, writes, and so on (Vozábal, 2016). The slave nodes in a Hadoop cluster are known as data nodes (Singh & Singla, 2015). All file operations are conducted on these nodes, and data is actually stored on these nodes, as determined by name nodes (Vozábal, 2016). The secondary name node serves as the name node's backup (Vozábal, 2016). Because the name node is the master node, it is critical that it be backed up. If the primary name node fails, the secondary name node will be used as the name node (Singh & Singla, 2015).

2. Map Reduce

Map-reduce is a core technology developed by Google, and Hadoop is an open-source implementation of it (Singh & Singla, 2015). It is a critical component of Hadoop and is extremely useful when dealing with large amounts of data (Vozábal, 2016). The fundamental concept of map-reduce is to divide a large task into smaller chunks and then deal with them accordingly (Kumar & Valcin, 2020). Map Reduce is made up of four main parts: input, mapper, reducer, and output (Singh & Singla, 2015).

- *Input*: Input refers to data that is received for processing and is then divided into sub small pieces that are assigned to mappers.
- *Mapper*: Individuals assigned to the smallest unit of work for some computation are known as mappers.
- *Reducer*: Reducers use the data aggregated by the mappers as input to create the final output.
- *Output*: The jobs of reducers are finally gathered in the form of aggregated output.

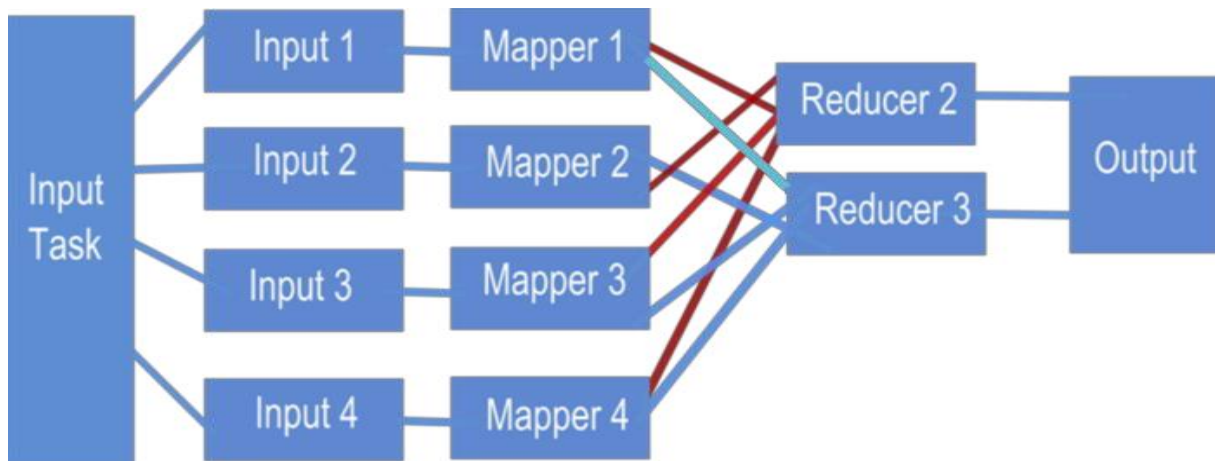


Figure 6: Map-Reduce Architecture (Singh & Singla, 2015)

2.7.1.2 Hive

Hive is a SQL-like language known as HiveQL (Singh & Singla, 2015). That is, Hive offers a SQL interface as well as a relational model for big data processing on Hadoop. In addition, as a data warehouse application framework, Hive offers summarization, querying, and analysis on top of Hadoop (Zahid, Mahmood, Morshed, & Sellis, 2019). (Wiley & Sons, 2015) mentions a few instances when it is better to consider using Hive:

- *Data is easily incorporated into a table structure.*
- *Data is already in HDFS*
- *There is a desire to divide datasets into time segments (For example, daily updates are added to the Hive table).*
- *Batch processing is acceptable.*

Hive, on the other hand, is not designed for real-time querying (Wiley & Sons, 2015). A hive query is first converted into a Map-Reduce job, which is then executed on the Hadoop cluster (Zahid, Mahmood, Morshed, & Sellis, 2019). It is preferable to use HBase for real-time querying, as mentioned below (Singh & Singla, 2015; Vozábal, 2016).

2.7.1.3 Hbase

Currently, Hadoop can only process data in batches, and data can only be accessed in sequentially (Kumar & Valcin, 2020). Even for the simplest of jobs, this means searching the entire dataset (Kumar & Valcin, 2020). When a massive dataset is processed, it generates another massive dataset, which should also be handled in a sequential manner (Vozábal, 2016). Therefore, In order to access any data point in a single unit of time, a new solution is required (random access) (Singh & Singla, 2015). Hence, HBase is one of the applications of the databases that store huge amounts of data and access the data in a random manner (Patel, 2017). HBase is a column-oriented non-relational database that runs on top of HDFS (Hadoop distributed file system) (Singh & Singla, 2015). It enables real-time data reading and writing on HDFS (Hadoop distributed file system) (Singh & Singla, 2015). It can deal with petabytes of data (Patel, 2017).

2.7.1.4 Flume

Flume is a distributed, freely available, and durable framework for capturing, aggregating, and transmitting massive amounts of log information from various sources to a centralised database/ repository (Singh & Singla, 2015). HDFS, HBase, Solr, and other centralized repositories are examples. Apache Flume enables us to capture data from different sources in both real-time and batch modes (Nagdive, Tugnayat, Regulwar, & Petkar, 2019).

On the other hand, Hadoop is well-known for processing large amounts of data (Kumar & Valcin, 2020). However, it is questionable how data produced by various web servers is reassigned to the Hadoop file system (Kumar & Valcin, 2020). Apache Flume is the solution for this confusion (Nagdive, Tugnayat, Regulwar, & Petkar, 2019). Flume is intended for high-volume incident loading data into Hadoop (Nagdive, Tugnayat, Regulwar, & Petkar, 2019). Imagine a case in which a large number of web servers generate log files that must be transmitted to the Hadoop file system (Singh & Singla, 2015). Flume parses those files into events and sends them to Hadoop (Nagdive, Tugnayat, Regulwar, & Petkar, 2019).

2.7.1.5 Kafka

Apache Kafka is a highly throughput, free and open-source stream processing framework which is a distributed publish-subscribe messaging system and robust queue capable of handling large amounts of real-time data and transferring messages from one point to another

(Wu, Wolter, & Zhihao, 2019). And also, Offline, as well as online message consumption, are possible with Kafka (Khan, 2021).

To prevent data loss, Kafka messages are stored on the disk and replicated within the cluster to prevent data loss and corruption (Khan, 2021). When integrated with Apache Storm, Apache HBase, and Apache Spark, Apache Kafka productively processes real-time, streaming data (Wu, Wolter, & Zhihao, 2019).

2.7.1.6 Apache spark

It is a free and open-source Big Data processing software platform for performing data analytics on distributed computing clusters like Hadoop (Vozábal, 2016). Spark supports in-memory processing, which speeds up and simplifies data processing compared to Map Reduce (Kumar & Valcin, 2020). It's being talked about as a more effective Hadoop analysis suitable alternative (Pointer, 2020). Spark is built on top of the previous Hadoop cluster, allowing it to access data through HDFS (Kumar & Valcin, 2020). It can also handle structured data through Hive and stream data from HDFS (Vozábal, 2016).

Spark, as an alternative to the conventional bash Map-reduce framework, introduces a new data processing model that is said to be more productive than Hadoop in aspects such as real-time streams and iterative algorithms used in graph applications and machine learning (Pointer, 2020). It is also thought to be a good atmosphere for interactive business intelligence and data mining tools (Bell, 2020).

In contrast to Hadoop, Spark makes greater use of RAM (Random Access Memory) rather than network and disk I/O (Pointer, 2020). Spark commonly requires a high-end physical machine to create efficient results due to the higher memory utilization (Vozábal, 2016). To acquire high availability, Hadoop employs replication, whereas Spark employs a different data storage model known as Resilient Distributed Datasets (RDD) (Naik & Naik, 2020). RDD saves additional data for lost block rebuilding in order to avoid failure and ensure effectiveness (Naik & Naik, 2020). In addition, spark supports multiple languages such as Java, Scala, and Python (Bell, 2020).

Spark has four different modules that are distinct from its core (Vozábal, 2016).

- Spark SQL – it provides access to spark datasets via the JDBC (Java Database Connectivity) API and permits SQL queries to be operating on Spark data. This also enables users to retrieve, transform, and load (ETL) data in multiple formats.
- Spark Streaming is an element used to handle real-time streams of data.

- MLib is a distributed machine learning framework that contains various machine learning algorithms. These include popular learning algorithms and utilities such as correlations and hypothesis testing, classification, regression, clustering, principal component analysis, and so on.
- GraphX – is a distributed graph processing framework that is used to manipulate graphs and perform graph-parallel computations.

2.7.1.7 ELK Stack

The ELK stack is an abbreviation that refers to a grouping of three open-source productions: Elastic search, Log stash, and Kibana (Bavaskar, Kemker, & Sinha, 2020). Elastic search is a search and analytics engine that can handle large amounts of data (Erik & Joel, 2012). Log stash is a log aggregator that gathers data from a variety of sources, applies various transformations and improvements to it, and then sends it to one of the several supported output destinations (Bavaskar, Kemker, & Sinha, 2020). And finally, Kibana provides a visualization layer that interacts and works on top of Elastic search engine, providing users with the ability to analyse and visualize the data (Prakash, Kakkar, & Patel, 2016).

2.7.1.8 Apache Metron

Apache Metron is an open-source platform for security monitoring and analysis that incorporates a number of other open-source big data technologies (Verheijde, Gnanavarotheyan, Moelchand, & Stam, 2019). Because of this variability, Metron's functionality can be divided into four categories: a platform for capturing, storing, and normalizing any sort of data at extremely high rates, real-time processing and application of enrichments, effective information storage, and a user interface for viewing data and notifications that have passed through the device (Verheijde, Gnanavarotheyan, Moelchand, & Stam, 2019). Apache Kafka, Apache Storm, and Apache Hadoop are the three major open-source resources at the heart of Metron (Pranita, Onkar, Aditya, & Dr M, 2020).

2.8 Related Works

Existing data governance models differ in scope and focus due to the vast nature of the subject and the lack of data governance standards. Academic research on big data governance is still in its beginning stages. In this case, the study presents the available literature on best practices linked to the establishment of big data governance frameworks relevant to the current study and industry under consideration.

Study conducted by (Al-Badia, Tarhinia, & Khana, 2018) presents a detailed examination on existing big data governance frameworks and their shortcomings. As a result, a conceptual model for big data governance is offered. The framework was developed following a thorough examination of the existing governance problems in big data is presented. There are eight major components in the suggested frameworks. These components are the organization's structure identification, setting the policies and standards, stakeholders' identification, optimizing and computing, identifying the scope of big data, measuring and monitoring quality, data storage, communication and data management. The proposed approach is lacks a telecom background. A research project by (Yang, Li, Elisa, Prickett, & Chao, 2019) recommends a big data governance framework that assists organizations to make bigger data-informed business decisions within the applicable framework, with a focus on data privacy, confidentiality, availability, security, and accessibility. To prove this workflow, the work also includes an illustration framework execution case study approach of big data governance in cyber security. They began their investigation by outlining the shortcomings of existing data governance mechanisms for dealing with big data governance. The existing conventional data governance problems, as per the study, are: lack of big data governance mechanisms, a scarcity of individuals with the necessary skills in big data, big data security and privacy, lack of necessary tools for extracting insights from data in a timely, organizations resistance to use data in goal setting and decision-making, inadequate understanding of big data by company managers, corporate digitization remains a hurdle for most firms, and the volume of data produced and stored (e.g. unstructured data). Therefore, based on this fact, data governance for big data environments, according to the researchers, should concentrate on three areas: data architecture, data quality, and data security, privacy, and ethical issues. Accordingly, the researcher proposed a big data governance framework for cyber security with three domains: (Planning, Governance, and Evaluation) and twelve components (data risk management, policies, stewardship, distributed data processing, distributed data storage, integration and metadata management, data quality management, data security and privacy, accountability, transparency, integrity and availability). The framework has the power to lead big data management in many companies, allowing for knowledge exchange and cooperative decision-making.

Research performed by (LI et al., 2019) offers a big data governance framework for healthcare data based on the governance activities related with the processing of RHINs in China. As a strategical technique, the researcher employs a combination of qualitative methods such as

personal experience in countrywide consulting, a review of literature, professional consultation, and interpretative structural modeling tools. Accordingly, the healthcare big data governance practices in China were articulated using an analysis of ten typical RHIN case studies. Depending on Chinese practice, a big data governance framework with three domains and twelve elements was proposed, which include a drive domain (big data strategy planning, laws and regulations, open transaction, and industry support), capability domain (healthcare big data organization, collection, storage, process and analysis, and usage), and support domain (healthcare big data resource planning, standards system, and privacy and security protection) and The researcher also derived 12 guidelines for healthcare big data governance, which could serve as useful references for the cross-dimensional growth of RHINs, offer additional overall direction for the feasible development of regional health automation, and add value to recognizing the competitive advantages of healthcare big data.

Study conducted by (Charikleia, 2015) presents a conceptual framework for applying data governance into the data lake in the sense of a big financial institution in the Netherlands. The execution of data governance was described at both the low and high levels, in accordance with the agile methodologies. Accordingly, the use of the conceptual Framework has assisted the DIL team in approaching specific steps that were previously unclear in order to effectively utilize data governance in the data lake. Besides this, the team members, particularly the DevOps team, are no longer perplexed and understand the logic behind every action they take. Having a clear image of who is responsible for each step of the procedure improves team performance and collaboration within and between teams.

The study conducted by (Abraham, Schneider, & Brocke, 2019) design a conceptual framework for data governance, summarise the literature and present a research agenda based on a structured literature review. Accordingly, for this study, the researcher conducted 145 research papers and practitioner publications published during 2001-2019. The researcher then identifies the main construction frameworks of data governance and classifies them in six components. The study contributed to upcoming data governance research by providing five research topics and presenting a total of 15 research questions. Besides this, the conceptual framework gives an outline of antecedents, scoping dimensions, and governance processes to support practitioners in approaching data governance in an organized format.

The study conducted by (Fikiru, 2020) explore the existing big data management frameworks and proposed a new framework for Wegagen Bank S.C based on the Information Systems Design Research (ISDR) approach. The primary goal of this research is to evaluate Wegagen

Bank S.C's present big data management practices and to create a big data management framework for the bank. As a result, this study uses a quantitative and qualitative survey to analyse the challenges with the current big data management method in order to address the issues that have been found. The proposed framework consists of three domains namely planning domain, data management implementation domain and evaluation and optimization domain and within the domains, there are sixteen components (data management goal, identify organizations structure, awareness creation, big data scope determination, risk management planning, policies development, stakeholders selection, infrastructure development, data collection, data cleaning, data quality, data security and privacy, banking big data standards system, data processing, evaluation mechanism and optimization). While the study appears to be relevant to the financial industry, it lacks a telecom context to address the challenges addressed in our study.

Existing big data governance frameworks are contextual and are not found suitable for Ethio telecom context. Contextual factors such as organizational, national and environmental affect the design of the framework for a specific industry and organization are not addressed.

The following table presents the summary of the above-related studies for better visualization and understanding of the existing gap and focus is of this work.

Author [year]	Objective	Methodology	Key Finding
(Al-Badia, Tarhinia, & Khana, 2018) Exploring Big Data Governance Frameworks	To explore the existing Big Data governance frameworks and their shortcomings, and propose a new framework.	Literature review	Proposes a new framework which has eight components
(Yang, Li, Elisa, Prickett, & Chao, 2019) Towards Big data Governance in Cyber security	To assess the challenges of the existing data governance mechanism for big data management	Literature review	Proposed a big data governance framework for cyber security which has three domains and thirteen components

Author [year]	Objective	Methodology	Key Finding
(LI et al., 2019) A Framework for Big Data Governance to Advance RHINs: A Case Study of China	To propose a big data governance framework for healthcare data based on the governance activities associated with the processing of RHINs in China.	personal experience in nationwide consulting, literature review, expert consultation, and interpretative structural modelling methods,	Proposed a big data governance framework for advanced RHINs, with 3 domains and 12 elements was presented based on Chinese practice and 12 guidelines for healthcare big data governance.
(Charikleia, 2015) Data Governance: A conceptual framework in order to prevent your Data Lake from becoming a Data Swamp	To propose a conceptual framework for implementing Data Governance into the Data Lake in the context of a Big Financial Institution in the Netherlands.	Action research	Introduce a conceptual framework for implementing Data Governance into the Data Lake in the context of a Big Financial Institution in the Netherlands.
(Abraham, Schneider, & Brocke, 2019) Data governance: A conceptual framework, structured	Develop a conceptual framework for data governance, synthesize the literature, and provide a research agenda.	literature review	Introduce a conceptual framework for data governance comprising six Dimensions: governance mechanisms, organizational scope, data scope, domain scope, antecedents, and consequences of data governance. And also, The

Author [year]	Objective	Methodology	Key Finding
review, and research agenda			study contributed to upcoming data governance research by providing five research topics and presenting a total of 15 research questions.
(Fikiru, 2020) A Big Data Management Framework for Wegagen BankS.C	The major objective of this study is, to assess current Big Data management practices at Wegagen Bank S.C, and to develop Big Data management framework for the bank.	Design Science Research Methodology	Proposed a big data governance framework for Wegagen BankS.C which has three domains and sixteen components

Table 3: Summary of related works

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This study used the design science research approach to answer the stated research questions. The study identified the lack of big data governance framework in Ethio telecom as a problem and its objective was to design a big data governance framework as an artifact that enables Ethio telecom to protect big data whilst obtaining the highest value from big data, thereby empowering and promoting appropriate data-usage practices. And also assist organizations in efficiently achieving the desired results in innovative data utilization, which frequently leads to a culture change in organizations for deep data-driven processes.

Design-science research paradigm is a prescription-driven and problem-solving paradigm that seeks to explore the possibilities of human and organizational capabilities by introducing new and innovative artifacts (Hevner, March, Park, & Ram, 2004). That is, knowledge, understanding, and solution of the problem area have been found in the construction and implementation of the designed artifact (Hevner & Chatterjee, 2010). Such artifacts may include constructs, models, methods, and instantiations (Hevner & Chatterjee, 2010).

3.2 Research design

The design science research methodology (DSRM) process method introduced by (Peffer et al., 2006) was used to conduct this research. It is because of the compatibility of the model that is relevant for the context of this study. The researcher can simply understand the mental model for what design science research in IS should look like and know how to apply those mental models when and where he will use them and in which circumstance of the research context. It is a sufficiently complete, robust and clear process model that provides a mental model that clearly outlines how to design science research in information systems (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007). (Peffer et al., 2006) Has stated that a common framework is necessary for DS research in IS and mental model or template for readers and reviewers to recognize and evaluate the results of such research.

This process model consists of the following sequential steps, which are described below and illustrated graphically in the following figure:

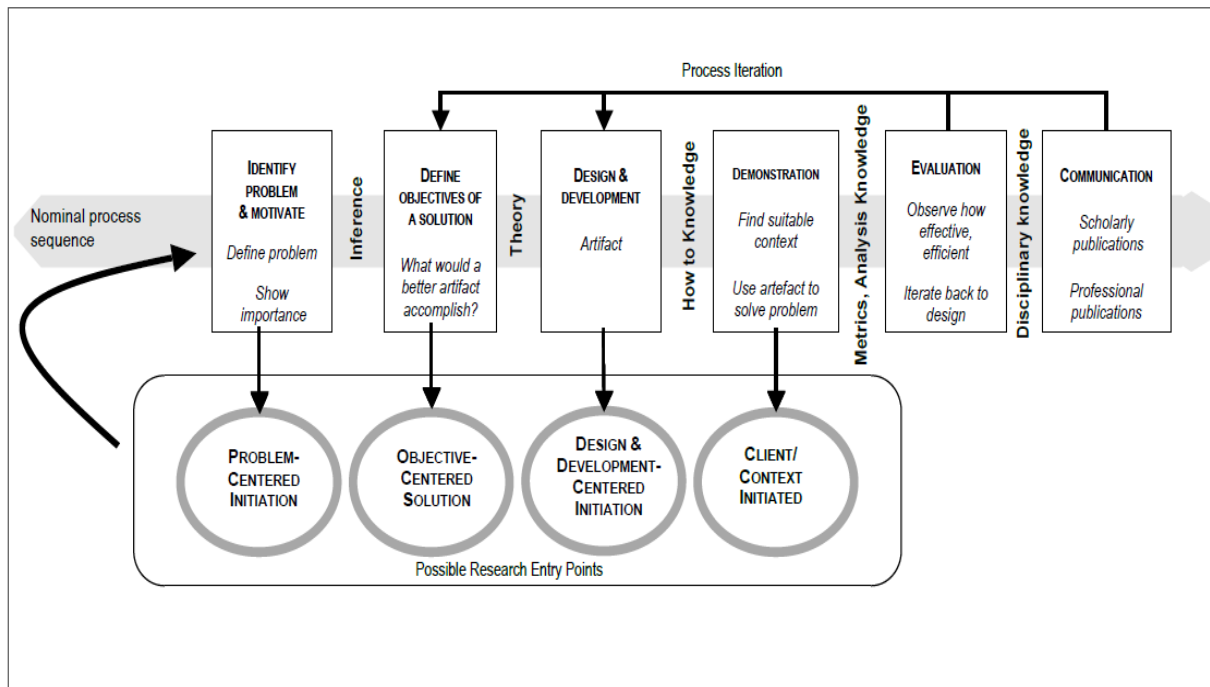


Figure 7: Design science research cycles by (Peffers et al., 2006)

3.2.1 Problem identification and motivation

This activity focuses on defining a problem and motivating the importance of a solution. In defining the study problem, developed an artifact that can effectively provide a solution. These activities include knowledge of the state of the problem and the importance of its solution. In accordance with this step, chapter one of the thesis report defined the problem and justified the value of a solution. As a result, the exact details of this step are mentioned in chapter one.

3.2.1.1 Study area

The study was conducted on employees of Ethio telecom located in Addis Ababa legehar branch. Because this branch contains main data services and equipment, as well as an IT manpower office.

3.2.1.2 Source of Data and Data Collection Technique

For conducting this research, the researcher used a mixed methodology (qualitative and quantitative methods). This method was selected for this particular study because the design of the research questions and the objectives of the study necessitates the use of both approaches. It is also relevant to gather detailed descriptions of existing conditions and practices of data governance in the organization.

Quantitative approach involves the generation of data in a quantitative form which can be subjected to rigorous quantitative analysis in a formal and rigid fashion (Getachew, 2018). It generates statistics through the use of survey research, using methods such as questionnaires and structured interviews. Whereas the qualitative approach tries to explore attitudes, behaviour, and experiences through such methods as interviews or focus groups (Getachew, 2018). The qualitative approach also attempts to get an in-depth opinion from participants. Accordingly, the researcher used a quantitative research approach to express employees' awareness level in terms of quantity. Mainly qualitative study is used to deeply examine the current state of Ethio Telecom data-related activities.

3.2.1.2.1 Source of Data

The study used both primary and secondary data sources. The primary data was collected from sample respondents through questionnaires and interviews. In addition, it is widely explained as follows.

3.2.1.2.1.1 Primary Sources of Data

a) Questionnaire

Based on the objective of the study, survey questionnaires are used to collect information on the organization's current big data governance experience, processes and tools/techniques being used, employee's attitude towards using BDA, BDM and BDGF, and issues that influence data governance/management in Ethio telecom. The questionnaire was adapted from Marta (2019) cited in (Fikiru, 2020) and (OCHIENG, 2015) and moderately adjusted to fit the context. It was also customized by using simple and straightforward language to reduce respondents' confusion and misconceptions about the questions.

The questionnaire had two parts. The first part gathers respondent's demographic information (Respondent Information) such as respondent's education status, position (respondents working staff), work experience and other general information about big data. The second section consists of the 5 point Likert-scale questionnaires (1=strongly disagree, 2 = disagree, 3 = neutral, 4 = agree and 5 = strongly agree) that seeks to understand the perception towards different inquiries of the research, such as strategy, technology, people, and environment. The survey questioner was then conducted using the prepared survey instrument, as shown in Appendix B.

b) Key Informant Interview

For a qualitative study, a semi-structured interview is employed for data collection. Semi-structured interviews, according to (Kabir, 2016), are appropriate to address what and how open-ended questions, which are common in the interview guide and useful for obtaining participant views, concepts, ideas, perceptions and experiences. Thus, semi-structured interviews helped to understand current Ethio Telecom BDG practices from participant experience. The interview guide is designed based on various literature and thesis.

The qualitative interview is conducted with IS division team member who is found to participate on big data-related tasks. In addition, corporate application operational manager, IT service design infrastructure manager and IT service design application manager were included in the interview in order to get insight about BDG strategy, standards, policy and the overall practices of BDG in Ethio telecom.

The procedure began with a phone call to the corporate application manager, IT service design infrastructure manager, and IT service design application manager and concisely defining the study's objective and the interviewees were then arranged for a suitable time for the meeting. The purpose and significance of the study were briefly described to the interview participants. The interview was then conducted using the prepared guideline, as shown in Appendix C. A mobile phone and a notebook were used to record the interview. After that, the audio recording was translated into a written transcript.

After qualitative data was collected and well organized, it was finally analysed using the content analysis (human sense) technique with an interpretative form to make appropriate generalizations and presentation accordingly to design the framework.

3.2.1.2.1.2 Secondary Sources of Data

The goal of this information is to better understand the level of adoption of BDGF in the study organization and to identify the existing literature relevant to big data governance framework and their components. And also, authenticate and increase proof from other sources (Kabir, 2016). According to (Gorschek & Wohlin, 2003) Usage of multiple data sources is not only used to decrease the possibility of missing vital information, acquire different perspectives and get more ideas but rather to have a confirmation of the identified process by checking against the other data sources. Therefore, the document review claims to support patterns evolved from the data gathered through questionnaires and interviews to maximizing the accuracy of the findings. Some of the data will be obtained mainly from records and reports of the organization, from the website, book chapters, Newspapers, and Internet articles.

3.2.1.3 Population size and Sampling technique

3.2.1.3.1 Population size

Ethio telecom's IS division employs has more than 350 people, including contract workers. However, for this study, the population size was drawn from four directorates such as corporate application operation, IT service strategy and program, IT service design and IT solution application. The target populations from the corporate application operation directorate include (data warehouse specialists and corporate application operational manager), from the IT service design directorate (IT service design infrastructure manager, IT service design application manager, application support specialist, business analysis specialist, IT service design requirement manager, information architecture specialist, customer operation and technical design specialist), from the IT solution application directorate (IT solution application rollout specialist and manager) and from IT service strategy and program directorate (IT service strategy and program manager).

3.2.1.3.2 Sampling technique

Sampling is a method that allows researchers to obtain information about an entire population by examining only a part of it (Kabir, 2016). It can be grouped under two distinct heads, that is, probability and non-probability sampling. Probability sampling is accomplished through random selection of the sample from all individuals from the population. While non-probability sampling is a non-random and subjective sampling approach in which the sampler's personal opinion or decision is used to pick the population elements that comprise the sample (Kabir, 2016). Hence, for this research the researcher applied purposive sampling, the sample units must have the potential and richness in information to be key informants for the study. Purposive sampling technique is a type of non-probability sampling that is most effective when one needs to study a certain cultural domain with knowledgeable experts within.

Purposive sampling techniques are more appropriate for studies of this type as it allows the researcher to capture and interpret individual judgment forwarded by key informants who will provide adequate information to address the research problem (Getachew, 2018).

Accordingly, the qualitative study consists of corporate application operational manager, IT service design infrastructure manager, IT service design application manager, IT service design requirement manager, IT solution application rollout manager and IT service strategy and program manager because all these key participants can give detailed information about Ethio Telecom big data governance activities. Based on the data from the manager level, for the

qualitative study, there are six potential data sources and the researcher selects only convenient participants. Regarding sample size determination on qualitative study (Getachew, 2018) noted that “you are guided by your judgment as to who is likely to provide you with the best information”. In addition, Saturation point determines the sample size in qualitative research as it indicates that adequate data has been collected for a detailed analysis. It is when the researcher reaches a point where no new information is obtained from further data. That is enough information has been gathered to achieve their research purpose, further data collection would yield similar results, so it is sufficient that the researcher is limited to 3 participants to confirm emerging themes and conclusions.

On the other hand, the quantitative study consists of four directorates such as corporate application operation, IT service strategy and program, IT service design and IT solution application directorates. Accordingly, 24 individuals out of 36 were selected as respondents for the survey questionnaire. Because as described earlier Purposive sampling techniques allows the researcher to capture and interpret individual judgment forwarded by key informants who will provide adequate information to address the research problem. Therefore those individuals are rich in information on the topic. In addition, for a homogeneous population, sample size can be relatively small (Kabir, 2016). Therefore, the sample size of the survey questionnaire is twenty four.

Respondents	Total population	Target population
Application support specialist	6	2
Business analysis specialist	4	3
IT service design (infrastructure)	1	1
IT service design (requirement)	1	1
Data warehouse specialist	4	4
IT solution application rollout	6	4
Information architecture specialist	4	3
Technical design specialist	4	2
Customer operation	4	2
IT service design (Application)	1	1
IT service strategy and program manager	1	1
Total target population		24

Table 4: Target population for a survey questionnaire

Respondents	Total population	Target population
Corporate application operational manager	1	1
IT service design infrastructure manager	1	1
IT service design Application manager	1	1
Total target population		3

Table 5: Target population for interview

3.2.1.4 Data Analysis and Interpretation

As depicted above in the methodology part, blended technique or mixed-method data analysis provides qualitative and quantitative data analysis that are combined, connected, or integrated. In such a manner, both qualitative and quantitative data have been examined and analysed separately and afterward associated and coordinated in a manner to give data to respond to the research questions.

Accordingly, the data collected via questionnaires were analysed and summarized with descriptive statistics using a statistical package for social scientists (SPSS) V. 26.0 and Microsoft excel. On the other hand, many scholars suggested that qualitative research is fundamentally interpretative. Based on this, the data that was collected from the interview and reviews of documents were interpreted manually.

Overall, the results from both methods have been triangulated with the secondary data to maximize the reliability and validity of the findings. And also the draft findings of this research have been returned to the key informants who mostly participated in the primary data collection for validation.

3.2.1.5 Validity and Reliability

Validity and reliability are the two fundamental elements in the evaluation of a measurement instrument. The researcher used Cronbach's alpha from the SPSS tool to determine internal consistency or reliability, and an acceptable (0.713) CA was observed.

Reliability Statistics	
Cronbach's Alpha	N of Items
.713	59

Table 6: Reliability Statistics

The researcher also analysed big data governance-related files, as well as related rules, policies, procedures, and standards, to enhance the validity and reliability of data collected through questionnaires and interviews.

3.2.2 Objectives of a solution

The second activity of the DSRM process model entails determining the objectives of the proposed solution, while the first activity encompasses the general identification of the problem. “These objectives can be quantitative, such as terms in which a desirable solution would be better than current ones, or qualitative, such as a description of how a new artifact is expected to support solutions to problems not hitherto address” (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007). From these interpretations, the study has qualitative objectives, as Ethio Telecom currently does not have any data governance framework to manage data. According to (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007), the objectives should be inferred rationally from the problem specification. Based on this, the objective of big data governance framework is to assist Ethio telecom to achieve bigger data-driven business decisions. That will help to ensure the integrity, usability, and value of their data by reducing data risk in the organization. In addition, the exact details of this step are mentioned in chapter five. That is, chapter five explores the benefits of the big data governance framework and how it can help Ethio telecom with its operations.

3.2.3 Design and development

This activity is defined as “create the artifact” (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007). This activity entails determining the required functionalities of the artifact as well as its framework, followed by the actual creation of the artifact. Big data governance framework is presented as a solution to the problem. According to scholars (Hevner & Chatterjee, 2010), these artifacts could be constructs, models, methods, or instantiations. In this study, the artifact (big data governance framework) is a model that is designed and formed based on the aforementioned objectives.

Accordingly, this part is carried out by taking input from the environment (people, organizational structure and technical system). And also, drawn from the knowledge base (scientific theory and methods, and design product). To illustrate well, the following information has been used in the iterative design process: the first one is, several resources were used for input, including (organization strategic planning, goals and vision of an organization, values of the organization, existing business culture, IT and data architecture

principles and data governance requirements outline from stakeholder) as well as input (reflections during previous projects). In addition, literature on big data governance frameworks and related work is conducted to aid in the development of the framework. Finally, the information gathered is translated and added to the set of framework being developed.

Then, the proposed big data governance framework for Ethio telecom is formed after identifying the big data governance framework. The developed framework demonstrates each component and its functionalities to solve the firm's current problems.

3.2.4 Evaluation

To verify innovation (to ensure that the designs created are research contributions rather than routine designs based on the use of known design processes and the appropriation of known design artifacts), this process employs previous knowledge bases such as scientific theories and methods, experience and expertise, design products and processes (Hevner, March, Park, & Ram, 2004; Peffers et al., 2006).

Accordingly, the proposed big data governance framework is evaluated by experts with practical interaction in information technology/ data experts. It was evaluated by twelve ICT staff who participated in the questionnaire and interview process and two administrative staff (director level).

3.2.5 Communication

Communication in design science research is one of the most important tasks in research (ÖZARSLAN, 2015). Scholarly research publications are one example of this type of communication.

This research will be communicated and presented to the relevant stakeholders. That is, communication is achieved through the publication of the thesis work and making it available to all interested particulars and mainly to Ethio telecom professionals. In addition, the findings of this research will be presented in the journal papers.

CHAPTER FOUR

DATA ANALYSIS AND PRESENTATION

4.1 Introduction

This chapter presents the basic data and analysis of the quantitative and qualitative data gathered through surveys questioner, interviews, and document analysis. The main purpose of this chapter is to present the finding with respect to the current big data governance practice in Ethio Telecom and see if the current data governance strategy is efficient to handle or manage data growth in the organization. The analysis also puts light and identify major factors that impact the practices of big data governance in Ethio telecom and the specific requirements of the organization to develop BDGF in the context of the organization. Accordingly, a surveys questioner is conducted with 24 individuals from different roles and a Semi-structured interview is conducted with 3 individuals at the managerial level.

Finally, an interpretation and summary of the results will be presented. The framework was created using the results of the quantitative and qualitative analysis.

4.2 Data Analysis-Part I - General category

4.2.1 *Demographic Profile of Respondents*

This section focuses on presenting the general characteristics of the survey participants. The variables include role level, work position, education and work experience. The demographic profile of the respondent (role level, work position, education status and work experience of the respondents) is needed to show that the key informants are knowledgeable and prosperous, as well as to show professionals with a certain cultural background. In addition, these participants are key informants who provide sufficient information to the researcher to solve the research problem. Accordingly, members of the Telecom IS division team with a bachelor's degree and a master's degree, from senior professionals to managers as well as staff with 5 to 20 years of experience, especially data-related staff, are involved in this research.

About the role level, work position, education status and work experience of the respondents, Table 7 shows that the majority of respondents are 17 (70.8%) staff/experts and the rest are 2 (8.3 %) supervisors and 5 (20.8%) managers. Concerning education, the majority of the respondents, 15 (62.5%), are master's degrees and 9(37.5%) have first degree graduates. What this shows is that almost all participants have a good level of educational background. In terms

of work position, there are 2(8.3%) application support specialists, 3(12.5 %) are business analysis specialists, 1(4.2%) are IT service design (infrastructure), 1(4.2%) are IT service design (requirement), 4(16.7%) are data warehouse specialist, 4(16.7%) are IT solution application rollout, 3(12.5%) are information architecture specialist, 2(8.3%) are technical design specialist, 2(8.3%) are customer operation, 1(4.2%) are IT service design (Application) and 1(4.2%) are IT service strategy and program management. When work experience is concerned, 6 (25%) of the respondents have 5-10 years of experience in the Ethio telecom. In addition, 18(75%) of the respondents serve in the Ethio telecom for 11-20 years. This indicates that Most of the respondents 18(75%) had worked for more than or equal to 11 years. On the other hand, with the role level, work position, education status and work experience of the respondents, the interview result shows that, work position: Corporate application operational manager, IT service design infrastructure manager and IT service design Application manager, education: all are masters and work experience: 16, 15, 15 respectively.

Overall, these findings here indicate that in Ethio telecom most of the IT-related tasks are managed by a well-educated class of people and high-level experts.

		Frequency	Percent	Cumulative Percent
Role Level	Staff/Expert	17	70.8	70.8
	Supervisor	2	8.3	79.2
	Manager	5	20.8	100.0
Work position	application support specialist	2	8.3	8.3
	Business analysis specialist	3	12.5	20.8
	IT service design (infrastructure)	1	4.2	25.0
	IT service design (requirement)	1	4.2	29.2
	Data warehouse specialist	4	16.7	45.8
	IT solution application rollout	4	16.7	62.5

		Frequency	Percent	Cumulative Percent
	Information architecture specialist	3	12.5	75.0
	Technical design specialist	2	8.3	83.3
	customer operation	2	8.3	91.7
	IT service design (Application)	1	4.2	95.8
	IT service strategy and program management	1	4.2	100.0
Education	First Degree	9	37.5	37.5
	Master's Degree	15	62.5	100.0
Work Experience	From 5-10 years	6	25.0	25.0
	From 11-20 years	18	75.0	100.0
	Total	24	100.0	

Table 7: Distribution of respondents by Role level, work position, education and work experience

4.2.2 Use of Big Data Analytics

Regarding the availabilities of big data analytics application area in Ethio telecom one of the interviewee from IT service design infrastructure department said that, “big data analytics has a wide range of applications area in the Ethio telecom industry, including:

- ✓ Customer Relationship and Experience
- ✓ Customer Profiling
- ✓ Marketing /Product recommendation
- ✓ Data monetization
- ✓ Network Fault Isolation
- ✓ Churn analysis and prediction
- ✓ Fraud Detection ”

However, among the various BDA application areas in the Ethio Telecom market, the researcher also investigated whether Ethio telecom had adopted big data analytics technology. Accordingly, the results in figure 8 below show that the majority of the respondents 20(83.3%)

had not known whether big data analytics is adopted while only 4(16.7%) had assumed otherwise.

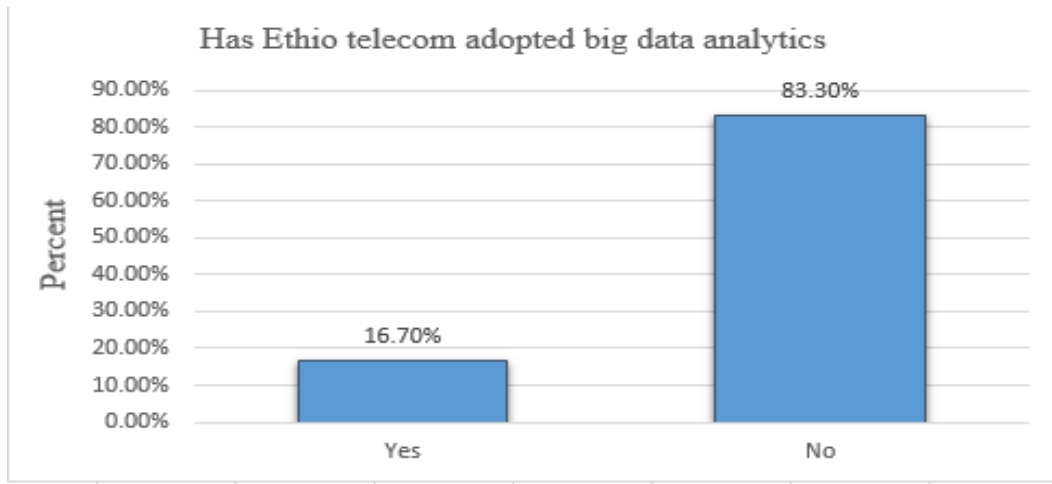


Figure 8: Distribution by Use of Big Data Analytics

Overall, big data analysis has many application areas in Ethio telecom industry. However, the perception of how big data analytics are implemented within the Ethio telecom industry still seems to be vague. This would imply that Ethio Telecom has become one of the richest companies in terms of data volume, but still doesn't know how to capitalize on it efficiently. The problem is how they can use this data to reduce operating costs, provide a personalized customer experience, reduce churn, and drive profitability.

4.2.3 The type of BDA technology/Software adopted

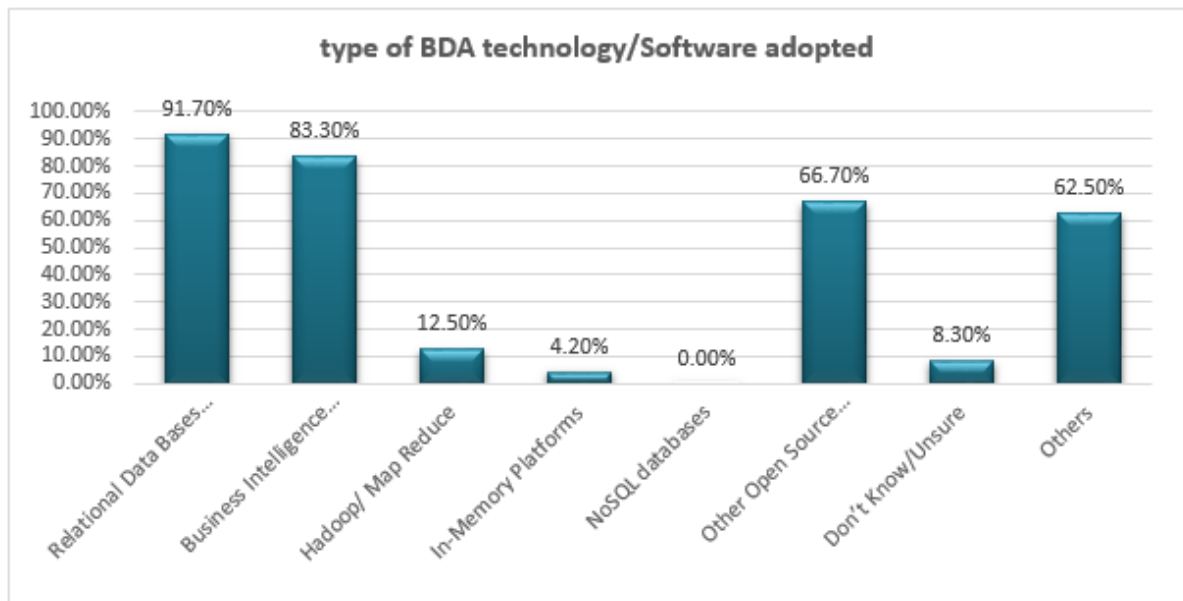


Figure 9: The type of BDA technology/Software adopted

As shown in figure 10 above, there is a confirmation of the predominant use of relational database software and business intelligence tools as a means of collecting, storing, and processing information, as this came first in figure 9 above at 91.7% and 83.3% respectively, with a large majority of participants indicating the use of other open-source technology and other software (66.7% and 62.5% respectively). A significant minority of respondents (12.50%) indicated they use Hadoop/ map-reduce and another small proportion of respondents indicated they use in-memory platforms (4.2%). Only 8.3 of the respondents indicated they were not sure of the software or technology being used.

A large number of respondents specified the use of relational databases, business intelligence, and other open source technologies. Relational databases and business intelligence tools only handle structured data as opposed to big data which can store and analyse huge amounts of unstructured data from multiple sources/channels (OCHIENG, 2015). In addition, the interview respondents describe the type of BDA technology/Software adopted as follows, there are many ways to analyse data but the following are the technologies used to analyse data in Ethio Telecom.

- Internally developed dashboard system (Reporting tool): This tool involves presenting analysis results in a visual or graphical format that is easier for the viewers to understand.

- BI data warehouse project (Oracle solution):- Now, Ethio telecom has an ongoing partial BI and data warehouse project. This project provides data from various sources such as from the network, IT service, internal solutions, ERP, etc., and presents the data on a dashboard and reporting basis.
- Mediation tool (Huawei solution)
- BICP (Business Intelligence Core Platform): it is used to analyse and present only business-related tasks. Ethio Telecom has an NGBSS (new generation business supporting system) service, and there is a system for subscribing to customer service, tariff presentation and registration, and customer unsubscribe. Therefore, BICP is a module for presenting and analysing data coming from NGBSS solutions.

On the other hand, regarding the question ‘Did you believe the tools/ technology currently used for data analytics/management are sufficient for yielding the big data utilization benefits?’ one of the interviewees said that, “*not at all. The tools currently used in Ethio telecom lack the analytics and interpretation ability. The current tools only provide structure data, so they will not yield the benefits of big data management*”. This shows that Ethio telecom is not working towards the analysis of data in terms of its data-generating capacity. Because of lack of integrated systems that manage all dimensions of information for a given customer infrastructure, such as storage device challenges, data analytics and processing challenges, and data privacy challenges.

Overall, these results imply that big data analytics adoption in Ethio telecom is still in its early stages.

4.2.4 Total Amount of Data managed

		Frequency	Percent	Cumulative Percent
Specify the total amount of data managed by Ethio telecom	100 TB +	13	54.2	54.2
	10 TB to 100 TB	8	33.3	87.5
	5 TB to 10 TB	1	4.2	91.7
	1 TB to 5 TB	1	4.2	95.8
	100 GB to 1 Terabyte (TB)	1	4.2	100.0
	Total	24	100.0	

Table 8: Total Amount of Data managed

The respondents were asked to state the amount of data managed by Ethio telecom and over half of the respondents 13(54.2%) as shown in Table 8 above indicated they managed a huge amount of data which is more than 100 TB+. On the other hand, some participants (about 33.3%) stated that it would be between 10 TB to 100 TB with a significant proportion also indicating they manage data less than 10TB (4.2%).

On the other hand, regarding the question ‘Is there big data in Ethio telecom? How to characterize data in Ethio telecom? Explain it in terms of volume, velocity and variety?’ the respondents said that Ethio telecom has been subjected to their bits of data as a result of a vast subscriber base that is connected to networks on a daily basis. This event caused big data characteristics from device logs, network activities, internal solutions, ERP systems, call detail records and customer mobile phone usages. Data occurs in both batch and real-time modes. Call detail records, mobile network usage, network performance, network monitoring, and customer/subscriber profile data are just a few examples. Generally speaking, big data in Ethio telecom can be defined by the standard 3Vs: volume, variety, and velocity. That is, Ethio telecom generates a massive amount of data at an alarming rate and in a variety of formats.

Accordingly, these findings imply that the amount of data managed by Ethio telecom qualify to be referred to as big data. Big data is measured in terabytes, petabytes and exabytes and also measured by the three characteristics of big data: volume, velocity and variety.

4.2.5 Data Type collected and analyzed by Ethio telecom

With regard to the data type available in Ethio telecom, interviewees acknowledged that, there are many types of data available in Ethio Telecom and one of the most important among all the data set is call detail data. The amount of call detail records generated and stored is extremely large. In addition, the data is also generated in real-time. Call detail records contain detailed information to characterize the important attributes of each call. Each call detail record includes, at least, the originating and terminating phone numbers, the date and time of the call, and the length of the call. The second one is Network data, like call detail data, which is created in real-time as a stream of data. Ethio telecom networks are highly complex device configurations consisting of a large number of interconnected components. Each network element can generate error and status messages, resulting in a massive amount of network data. And the last one is, Customer data Ethio telecom, like some other big corporations, can have millions of customers. Name and address information, as well as other details such as service

plan and contract information, payment history and income, will be included in this information.

However, among the various types of data available in the Ethio Telecom market, the researcher also examined which data type was collected and analysed by the firm/organization. Accordingly, most of the respondents (79.2%) indicated that they collected and analysed Archived/ Historical data in figure 10 below followed by other data at (66.7%), customer profile data (62.5%), network data (58.3%) and transactional data (54.2%), while a significant minority indicated they collected and analysed sensor data (45.8%). A small number also indicated they collected and analysed social media data (20.8%).

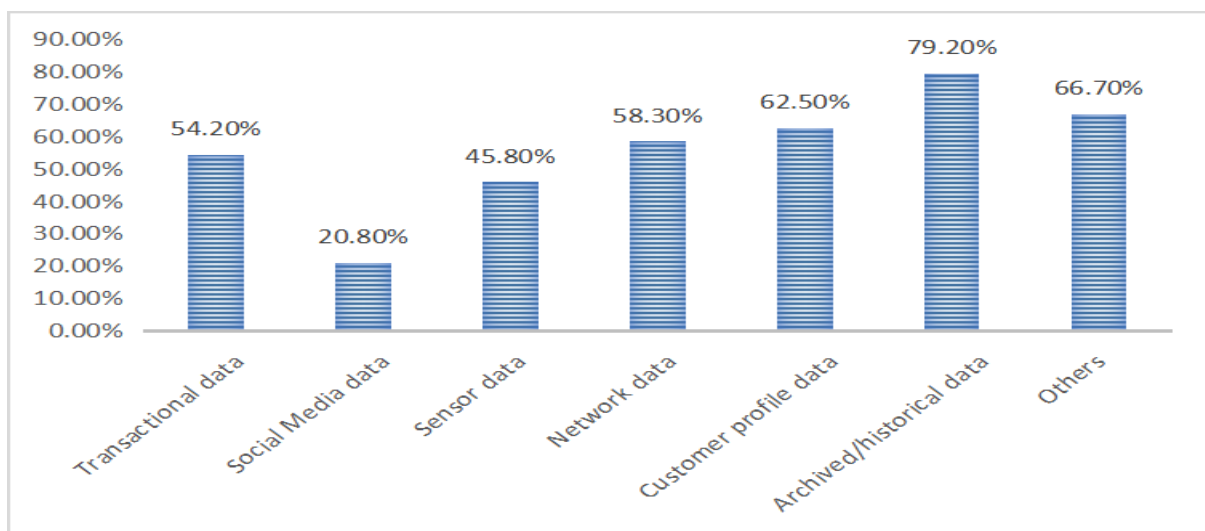


Figure 10: Data Type collected and analyzed

In general, the types of information available in Ethio Telecom market can be categorized as real-time, batch (historical/archived), and stream (near to real-time). These findings lead to the conclusion that Ethio telecom has seen the importance of collecting and analysing non-traditional data types to gain insight about the business and the customers even though quite a significant number are still relying on traditional data.

4.2.6 Rate of BDA Usage

		Frequency	Percent	Cumulative Percent
The rate of Ethio telecom usage of Big Data /Data Analytics is very high	Yes	1	4.2	4.2
	No	23	95.8	100.0
	Total	24	100.0	

Table 9: Rate of BDA Usage

Results from tables 9 above regarding the rate of big data analytics usage indicate that there is a low rate of big data analytics usage 23 (95.8) by the Ethio telecom. This would imply there is no infrastructure, IT platforms, and big data governance framework to enable the high rate of big data analytics usage by Ethio telecom.

4.2.7 Challenges affecting usage of big data analytics Technology in Ethio telecom

Challenges	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean
	N (%)	N (%)	N (%)	N (%)	N (%)	
Complexity of the amount of data in terms of volume, variety and velocity	12(50)	12(50)	0	0	0	4.50
Hosting and maintaining the technological infrastructure	3(12.5)	19(79.2)	1(4.2)	1(4.2)	0	4.00
Lack of capacity to store non-traditional data types	3(12.5)	20(83.3)	1(4.2)	0	0	4.08
It is too costly	0	11(45.8)	8(33.3)	4(16.7)	1(4.2)	3.21
Internal resistance to adopting BDA as a new way of doing business	0	5(20.8)	15(62.5)	4(16.7)	0	3.04
Lack of skill	3(12.5)	18(75)	0	2(8.3)	1(4.2)	3.83
Data security and privacy laws	1(4.2)	7(29.2)	16(66.7)	0	0	3.38

Table 10: challenges affecting usage of big data analytics Technology in Ethio telecom

The results in the table revealed the complexity of amount data in terms of volume, variety and velocity, lack of capacity to store non-traditional data types, and hosting and maintaining the technological infrastructure as major challenges (mean = 4, mean = 4.08 and mean= 4.00 respectively) considering a five-point Likert scale where 5 means that the respondents strongly agreed that they were a challenge followed by lack of skill and data security and privacy laws (mean = 3.83 and mean = 3.38 respectively). The results also revealed that Internal resistance

to adopting BDA as a new way of doing business and is too costly as other challenges (mean = 3.04 and mean = 3.21 respectively).

On the other hand, the answer I received from the interview on ‘what are the main roadblocks in effective implementation of big data analytics project /governance strategy and big data management in Ethio telecom?’

“In our organizations, there is no understanding of the significance of big data. However, there were already massive datasets available. More than a Terabyte of data is obtained and recorded daily. When I look at other countries' telecom sectors, I see that they use big data for various applications/business advancements in the industry. But here Decision-makers, legislators, stakeholders, top management, data collectors, and users are all unaware of the implementation of big data analytics. In addition, do not have proper data storing, analysing, capturing and handling procedures. This noted that, despite generating a huge amount of data from customers, they are not storing and implementing BDA due to a lack of infrastructure facilities and expert awareness”.

On the other hand, (Akal, Beshah, Sackmann, & Negash, 2019) also acknowledged that, challenges of big data analytics implementation in Ethio telecom such as there is a challenge on awareness, lack of integrated systems that handle all-rounded information for a given customer infrastructure including storage device challenge, data privacy challenges, organizations have not their standard in data handling and collection, lack of management and leadership in the area of big data analytics, lack of qualified experts in the area of big data analytics, lack of data correctness and completeness, and lack of stable infrastructure availability.

Overall, these findings imply that Ethio telecom is still facing many challenges that have hindered them from adopting BDA technology.

4.3 Data Analysis-Part II -Governance

4.3.1 Strategy

The second part of questionnaire starts by raising questions about Ethio telecom’s big data governance strategic approach.

A strategy is a well-defined framework for an organization. It describes an organization's core objective, mission, vision, and direction (Ferreira, Raposo, & Mainardes, 2014). Strategy, in a summary, bridges the gap between "where we are" and "where we want to be" (Ferreira,

Raposo, & Mainardes, 2014). Accordingly, big data strategy supports companies in creating a roadmap to know and implement big data technology. The strategy is very significant in big data governance; without preplanning and preparedness, working with big data technology will be a hard process (Fikiru, 2020).

The nine (9) questions incorporated in the strategy section are aimed at assessing Ethio telecom's strategic direction, plan and vision for big data governance. The questions are designed to evaluate Ethio telecom's big data technology strategic plan, stakeholder selection, and screening of employees.

The results on table 11 indicate 45.8% of the respondents agree and 29.2% of the respondents strongly agree with Ethio telecom had a vision and mission about implementing big data technology. Only 20.8% of the respondent neutral and 4.1% of them disagree with this statement. Concerning the policies, direction and plan to promote and manage big data and also to implement big data governance, Table 11 shows that the majority of respondents (the mean score value of 3.96, 3.88 and 3.96 respectively) are near to agree.

In terms of Ethio telecom planning and direction for big data governance, the respondents said that Ethio telecom had plans and direction for big data analytics and they also stated that the organization is developing a data analytics road map and that Ethio telecom has a responsible division (IS) dedicated to this task. In addition, Regarding the interview question ‘Does Ethio telecom has a BDG policy, procedure, strategy, technology, guideline or any kind of formal document that governs data in the organization?’ one of the respondents stated that *"there is still no experience actively associated with big data governance in Ethio Telecom. As I first tried to explain, the current experience is not satisfactory. That is to say, Ethio Telecom has started to focus on big data opportunities, but it is still planning. Overall, big data governance is the plan for next year"*.

Regarding future development plan, the results in the table revealed that 12.5% of the respondent strongly agree and 66.7% of the respondent agree on Ethio telecom has a future development plan and respond towards the dynamic ICT advancement. And also, 20.8% of the respondents strongly agree and 58.3% of the respondents agree on Ethio telecom has a future development plan for basic ICT and information infrastructure to implement Big Data Governance.

When staffing of ICT experts is concerned, the results show that 66.7% of the respondents agree, and also 8.3% of the respondents strongly agree on the Ethio telecom hiring strategy of ICT experts to promote the strength and efficiency of the firm. Moreover, the mean distribution

of responses was found to be 4.38 which is near to the value of ‘Strongly Agree’ about the importance of strategy. Thus, most of the respondents in the survey have believed that to compete Ethio Telecom in the marketplace, strategically necessary to implement /practice big data governance. This question is also confirmed on the interview part.

Strategy	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean
	N (%)	N (%)	N (%)	N (%)	N (%)	
Ethio telecom has a clear mission, vision about implementing big data governance framework	7(29.2)	11(45.8)	5(20.8)	1(4.1)	0	4.00
Ethio telecom policies are in place to promote and manage use of Big Data Governance in the organization	6(25.0)	12(50.0)	5(20.8)	1(4.2)	0	3.96
Ethio telecom has a direction, towards Big Data Governance	3(12.5)	16(66.7)	4(16.7)	1(4.2)	0	3.88
Ethio telecom has a plan towards Big Data Governance	5(20.8)	14(58.3)	4(16.7)	1(4.2)	0	3.96
Ethio telecom has a future development plan for basic ICT and information infrastructure to implement Big Data Governance	5(20.8)	14(58.3)	4(16.7)	0	1(4.2)	3.92
Ethio telecom implements strategic plan related to ICT provisioning/ supply and support	4(16.7)	14(58.3)	5(20.8)	1(4.2)	0	3.88
Ethio telecom has future development plan and respond towards the dynamic ICT advancement	3(12.5)	16(66.7)	4(16.7)	1(4.2)	0	3.88
Ethio telecom has ICT HR strategy related to hiring experienced and professional experts’ strength the efficiency of the organization	2(8.3)	16(66.7)	5(20.8)	1(4.2)	0	3.79

In order to compete Ethio Telecom in the marketplace, strategically necessary to implement /practice big data governance	15(62.5)	6(25)	1(4.2)	1(4.2)	1(4.2)	4.38
--	----------	-------	--------	--------	--------	------

Table 11: Strategy towards big data governance at Ethio telecom

4.3.2 Technology

The purpose of these constructs is to analyse the technological readiness and IT infrastructural facilities of Ethio telecom. Big data covers a broad spectrum of technology and processes that enable big data technology to perform as well as provide services. The eight (8) questionnaires in the technology section are intended to evaluate the technology's benefits, the firm's IT infrastructural facilities, and the security capabilities to adopt the technology.

Accordingly, table 12 show the result that 20(83.3 %), 2(8.3%), 1(4.2%), and 1(4.20%) of the respondents strongly agreed, agree, neutral, and disagree respectively that Using big data analytics/big data governance has provided many benefits to the firm/organization.

On the other hand, most of the respondents disagreed on the firm having the right technical skills, infrastructure and IT platforms to implement big data analytics technology/big data governance framework, (54.2%, 79.2% and 62.5% respectively) select disagree.

Moreover, according to the data from the interview, Regarding the approaches and tools used for managing the Ethio telecoms data one of the respondents answered, “our organization is currently using Centrally developed dashboard system (reporting tool), BI data warehouse project (oracle solution), a big data integration tool/ Mediation tool (Huawei solution) and BICP (business intelligence core platform). In addition, there are relational databases which are oracle and SQL database servers and they said these all tools are helpful for managing the Ethio telecoms data”. However, the tools and approaches currently used for data management are not sufficient for yielding the big data utilization benefits. The reason was Traditional data management tools/technology typically consider data management only for traditional structured and snapshotted moderately sized data sets, rather than high-volume, high-variety, and high-velocity live data. The large volume of data processed within organizations in a big data environment necessitates an additional level of management to ensure data integrity, confidentiality, and ethical data processing. Using 994 (call support to assist all customers) as an example, 994 customers alone generate over 2 million call detail records nearly every day. Given that several months' worth of call detail data is typically kept online, this implies that

tens of billions of call detail records must be managed record at any time. All customer data is collected and stored for a set period of time. Regarding the data policy, this can range from 4 to 6 months. After that, it can be automatically dismissed. In most cases, this is due to a lack of storage systems. Besides this, from a network standpoint, each network system can produce error and status messages, leading to a massive amount of network data. This data must be saved and analysed in order to aid network management functions such as fault isolation. Due to the enormous number of network messages generated, operators are unable to handle every message. In addition to this, Traditional data techniques and platforms, on the other hand, are inadequate in the context of Big Data (Oussous, Benjelloun, Lahcen, & Belfkih, 2018). They demonstrate a slow ability to respond as well as a lack of scalability, effectiveness, and accuracy (Oussous, Benjelloun, Lahcen, & Belfkih, 2018).

Most of the respondents 70.8% were neutral on answering It is easy to implement and use big data analytics technology/big data governance framework. 70.8 % of the respondents strongly disagree and 16.7 % of the respondents disagree on the existing ICT tools and technologies that would help in storing and analysing Big Data in a better way. This question is also confirmed in the interview part.

On the other hand, 83.3% of the respondents strongly agreed, 1(4.2%) of the respondent agreed, 6(25%) of the respondents neutral and 8(33.3%) of the respondents disagreed on Big data analytics has been tried and accepted by the firm/organization. The mean score on many firms/organizations that are already using big data analytics/big data governance to drive their business processes is 3.25, which is nearly neutral, so it is assumed the respondents have no information.

In summary, Ethio telecom produces enormous amounts of data, which are embedded into its storage systems, and this enormous amount of data is not being used productive way.

Technology	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean
	N (%)	N (%)	N (%)	N (%)	N (%)	
Relative advantage: Using big data analytics/big data governance has	20(83.3)	2(8.3)	1(4.1)	1(4.1)	0	4.71

provided many benefits to the firm/organization						
Compatibility: The firm has the right technical skills to implement big data analytics technology/big data governance framework	0	6(25)	5(20.8)	13(54.2)	0	2.71
Compatibility: The firm has the right infrastructure to implement big data analytics technology/big data governance framework	0	1(4.2)	4(16.7)	19(79.2)	0	2.25
Compatibility: The firm has the right IT platform to implement big data analytics technology/big data governance framework	0	4(16.7)	5(20.8)	15(62.5)	0	2.54
Complexity : It is easy to implement and use big data analytics technology/big data governance framework	0	3(12.5)	17(70.8)	4(16.7)	0	2.96
the existing ICT tools and technologies would help in storing and analysing Big Data a better way	0	3(12.5)		4(16.7)	17(70.8)	1.54
Triability: Big data analytics has been tried and accepted by the firm/organization	1(4.2)	9(37.5)	6(25)	8(33.3)	0	3.13
Observability: Many firms/ organizations are already using big data analytics/big data governance to drive their business processes	1(4.2)	5(20.8)	17(70.8)	1(4.2)	0	3.25

Table 12: Technology towards big data governance at Ethio telecom

4.3.3 People

The purpose of this question is to analyse the employees' level of education and knowledge of big data governance, as well as the organization's change acceptance behaviour of the employees.

The results in the table revealed that Most of the respondents 58.3% disagree and 29.2% are strongly disagreed on the staff's level of education and training for big data in the Ethio telecom. In addition, regarding with question 'Did the staff have a basic knowledge of big data technology?' The participants mentioned that, a few specialists, along with high-level management teams, may be familiar with big data technology, but only theoretically. Big data analytics have been tested and received by the organization. However, the current experience is not satisfactory. That is to say, Ethio Telecom has started to focus on big data opportunities, but it is still planning. Furthermore, concerning with the question 'how is Ethio telecom engaging in big data governance/data related activities? Who is in charge of BDG related tasks?' one of the interviewees stated, *"There seems to be no Big Data Team that is entirely responsible for Big Data-related tasks, however, we do have it in BI format, not Big Data format. Therefore, the work of Big Data is done entirely by BI as well as data warehouse experts"*.

On the other hand, 62.5% are disagree and 8.3 are strongly disagree on the Ethio telecom program to train the staff on big data governance/analytics. Regarding with the interview question 'Did Ethio telecom have programs in place to train the staff on big data technology?' The participant explained, *"This is something I'm not sure about"*.

Furthermore, Most of the respondents 83.3% are disagree on the staff have knowledge on big data governance/analytics. This query is also confirmed by the interview part. On the other hand, 4.2% agree and 58.3% are strongly agreed on that employees are fast learners to learn and apply new technologies. Regarding with technical and managerial skills, half of the respondents (50.0%) Saied that Ethio telecom has technical and managerial skills on the use of technological innovation. Furthermore, the mean score 3.46 which is nearly agree indicates that previous organizational change is accepted by the majority of the employees.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean

	N (%)	N (%)	N (%)	N (%)	N (%)	
People						
The level of training and education for ICT and Big Data in Ethio telecom is adequate.	1(4.2)	2(8.3)	0	14(58.3)	7(29.2)	2.00
Ethio telecom has programs in place to train the users for Big Data Governance /Management /analytics	0	2(8.3)	5(20.8)	15(62.5)	2(8.3)	2.29
Ethio telecom employees have basic knowledge about Big Data Governance /Management /analytics	0	0	3(12.5)	20(83.3)	1(4.2)	2.08
Ethio telecom IT employees have the ability to quickly learn and apply new technologies	1(4.2)	14(58.3)	7(29.2)	2(8.3)	0	3.58
Ethio telecom has technical and managerial skills on the use of technological innovation	0	12(50.0)	11(45.8)	1(4.2)	0	3.46
Previous organizational change has been accepted by the majority of employees	1(4.2)	11(45.8)	10(41.7)	2(8.3)	0	3.46

Table 13: People aspect towards big data management at Ethio telecom

4.3.4 Environment

The purpose of this portion is to analyse the benefits and challenges of big data governance in Ethio telecom in regard to the telecommunication company and the current market environment. The inquiries concern the organizational framework, competitive advantage, privacy rules and regulations, and market opportunities for adopting and effectively implementing big data governance.

As presented in table 14 response majority of the respondents 20(83.3%) strongly agree and 2(8.3%) agree with the adoption of big data will give Ethio telecom a competitive advantage. Only 2(8.3%) of them are neutral with this statement. On the other hand, regarding with the interview question ‘Do you believe the adoption of a big data governance framework/ technology gives Ethio telecom a competitive advantage? If yes, how?’ The respondent said that *“Absolutely yes. For example, some other Companies use big data in their systems for improving business functions, offering better customer service, creating personalized experiences, and taking other measures that, in turn, can increase profitability. Businesses that use it successfully have a tremendous opportunity over those that do not because they can make quick and better business decisions”*. In addition to this, regarding on ‘why big data governance is important for Ethio telecom?’ one of the participants mentioned that *“In my personal view, Ethio Telecom requires big data governance framework to promote and maintain information quality and use, as well as to ensure that the right information is used properly to set business goals, maintain business processes, improve customer satisfaction, improve operational efficiency, and make important decisions while maintaining profitability and managing risks”*. The mean score on Ethio telecom closely follows the competitor’s big data initiatives to determine its strength and weakness, is 3.08, which is nearly neutral, so it is assumed the respondents have no information. When infrastructure reliability and efficient is concerned, the results show that 1(4.2%) of the respondents strongly agree, 6(25%) of the respondents agree, 11(45.8%) of the respondents neutral and also 6(25%) of the respondents disagree on the telecommunications infrastructure is reliable and efficient in Ethio telecom.

According to the respondent, the result shows that 16.7%, 62.5%, 16.7% and 4.2% of the respondents were agree, neutral, disagree and strongly disagree respectively that There are firms in the market who provides support for use of this technology. The result shows 12.5 %, 66.7%, 12.5% and 8.5% of the respondents were strongly agree, agree, neutral, disagree that It is not difficult for Ethio telecom to protect data privacy for big data initiatives

And also, the mean score on it is easy for Ethio telecom to comply with privacy-related regulation to implement big data analytics technology/ big data governance framework is 3.17, which is also nearly neutral, so it is assumed the respondents have no information. On the other hand, 83.3% of the respondents strongly agreed and 1(4.2%) of the respondent agree on Ethio telecom can adopt the framework given the resources, opportunities and knowledge currently available in the organization.

These findings come to the fact that adopting big data governance will give Ethio telecom a competitive advantage, and Ethio telecom will be ready to adopt it if given the necessary resources, opportunities, and knowledge to do so.

Environment	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean
	N (%)	N (%)	N (%)	N (%)	N (%)	
Adoption of big data governance will give Ethio telecom a competitive advantage	20(83.3)	2(8.3)	2(8.3)	0	0	4.75
Ethio telecom closely follows the competitor's big data initiatives to determine its strength and weakness	1(4.2)	4(16.7)	15(62.5)	4(16.7)	0	3.08
The telecommunications infrastructure is reliable and efficient in Ethio telecom	1(4.2)	6(25.0)	11(45.8)	6(25.0)	0	3.42
There are firms in the market who provides support for use of big data governance	0	4(16.7)	15(62.5)	4(16.7)	1(4.2)	2.92
It is not difficult for Ethio telecom to protect data privacy for big data initiatives	3(12.5)	16(66.7)	3(12.5)	2(8.3)	0	3.83
It is easy for Ethio telecom to comply with privacy related regulation to implement big data analytics technology/ big data governance framework	2(8.3)	2(8.3)	19(79.2)	0	1(4.2)	3.17

Given the necessary resources, opportunities and knowledge to use big data, Ethio telecom will be ready to adopt it	20(83.3)	1(4.2)	2(8.3)	0	1(4.2)	4.63
---	----------	--------	--------	---	--------	------

Table 14: Environmental factor for BDG at Ethio telecom

4.3.5 Ethio telecom specific BDGF requirements

According to key informant interviews and a review of literature, there are many components and approaches to BDGF in the Ethio telecom industry, each with its own set of tools, methodologies and techniques. According to the information collected from interviews, Ethio Telecom does not currently have any big data governance framework. According to one respondent:

“Currently, we do not have any data management framework in place. However, in the future, we will need a much more individualized data analytics solution and prediction. Therefore, big data analytics, as well as governance solutions, are issues that need to be addressed strategically. Big data governance is our plan for next year. In addition to what I described earlier, the other problem is the market situation/competitiveness. That is, in a situation where there is no competitive and matured customer, there is nothing to enforce. For example, competitive environment, customer buying capability (purchasing power) and market demand are increase, it needs to focus on designing and developing cheaper data collection and analysis tools that are more customer-oriented”.

Moreover, according to the data from the interview, Regarding the critical requirements that must be covered in the practice of BDGF in Ethio telecom the other respondent stated that *“According to my opinion, a critical requirement that must be covered in the practice of BDGF in Ethio telecom, It should be able to solve the problems I mentioned above and be an organizational structured-centered and big data-based (The scope of Big Data being considered). And also, BDGF processes within the context of Ethio telecom operational environment at least require emphasis on the policies, rules, and standards corresponding to data capture, analysis, management, process, privacy, security, risk, and data usage requirements. A central framework containing all of the ideas I mentioned above is needed”.*

4.4 Summary

The main objective of this chapter is reviewing the current big data governance experience/practice at Ethio Telecom and compiling specific requirements, with the aim of developing a big data governance framework to assist Ethio telecom in properly manipulating both structured and unstructured big data, maximizing the power of big data, and empowering and promoting good big data practice.

Based on this fact, the study tried to investigate the current state of Ethio telecom in relation to big data governance. As a result, members of the Telecom IS division team with a bachelor's degree and a master's degree, from senior professionals to managers as well as staff with 5 to 20 years of experience, especially data-related staff, are involved in this research.

Accordingly, the analysis of the survey questioner and semi-structured interview result indicated that big data analytics has numerous application areas in the Ethio telecom industry such as

- Customer Relationship and Experience
- Marketing/Product recommendation
- Data monetization
- Customer Profiling
- Network Fault Isolation
- Churn analysis and prediction
- Fraud Detection

On the other hand, big data in Ethio telecom can be defined by the standard 3Vs: volume, variety, and velocity. That is, Ethio telecom generates a massive amount of data at an alarming rate and in a variety of formats. Therefore, the amount of data managed by Ethio telecom qualify to be referred to as big data. In general, the types of information available in Ethio telecom market can be categorized as real-time, Bache (historical/archived) and stream (near to real-time). However, Ethio telecom is not working enough towards analyse of data in contrast to the data generating capacity and the availabilities of big data application area. This would imply that Ethio telecom has become one of the richest companies in terms of data volume, but they still don't know how to capitalize on it efficiently. That is, BDA/BDG at Ethio telecom, on a trial basis, not to the point of being fully used, that is, Ethio telecom has started to focus on big data opportunities, but it is still planning. It is observed that Ethio telecom did put in place a mechanism to apply big data analytics/ governance for decision making, policy

formulation and for future strategic planning. Rather, they are using traditional data analysis and simple office package software for data processing in some situations such as centrally developed dashboard system (reporting tool), BI data warehouse project (oracle solution), Mediation tool (Huawei solution) and BICP (business intelligence core platform). In addition, there are relational databases which are deployed on Oracle and SQL database servers. And these tools are found to be helpful for managing the Ethio telecoms data. However, those tools and approaches that are currently in use for data management are not sufficient for yielding the big data benefits. The reason was Traditional data management tools/technology typically consider data management only for traditional structured and moderately small size of data sets, rather than high-volume, high-variety, and high-velocity live data.

The challenges behind not fully applying big data analytics by Ethio telecom were identified in the study. The key challenges were mentioned in the study are:

- Lack of awareness
- Absence of big data governance framework
- Lack of availability of big data expert people
- Big data confidentiality (privacy and security challenge)
- Availability of necessary tools to generate insight from data in a timely manner;
- Absence of data validity and completeness
- Absence of integrated systems that handle all aspects of information for a given customer infrastructure, such as storage device challenges, data analysis challenges, and data privacy challenges
- Internal resistance to implementing BDA as a new way of doing business

This implies that Ethio telecom has a better chance to gain new revenue streams through big data analytics (BDA) options, but they do not have a clear roadmap on how to capitalize on it effectively. The challenge is on understanding how they can use this data to reduce operating expenses, provide a personalized customer experience, make smart decisions, and establish new revenue streams.

On the other hand, currently, there is no big data governance framework at Ethio telecom and, the main reason that prevents the firm from formulating the framework are a lack of knowledge, lack of awareness, shortage of the required skilled personnel on big data, and a lack of exposure to the advancement in technology.

The researcher learns that Ethio telecom is still in the premature stages of implementing data analytics/ big data governance practices. Therefore, it is important to develop a big data

governance framework for Ethio Telecom based on the current data management problem as well as the specific requirements that may be included in the framework based on Ethio Telecom's current big data management experience. Management commitment, strategic goals, technology capability, business environment, and people's awareness and change acceptance behaviour will all play a significant role in the utilization of the data analytics/data governance initiative.

CHAPTER FIVE

THE PROPOSED BIG DATA GOVERNANCE FRAMEWORK

5.1 Introduction

In this chapter, the proposed big data governance framework will be discussed. Using findings from the study and related literature, a framework has been developed to address the third and fourth research questions. In general, this chapter has two sections. The proposed framework is introduced in the first section and the second section of the chapter explores the evaluation of the proposed framework.

5.2 Framework Development

As it is seen in the course of the study, Ethio telecom is one of the big governmental organizations that generate a huge volume of data, which has the capability of reaching millions of CDRs/events per day. This data may include usage, customer and network-related data. This data is generated in both batch and real-time modes. Quickly and effectively capturing, recording, processing, and interpreting this volume of data can pose serious problems for Ethio telecom. The infrastructure must have powerful computational performance as well as plenty of storage space. It must also be able to analyse various data formats. Therefore, the researcher noted that the framework suggested will greatly assist in the monitoring of this embedded data in a value-added way.

Accordingly, the framework development is guided by design science principles, some of the framework components derived from previous frameworks in the business sector and adjusted to meet the Ethio telecom setting. Depending on the specifications outlined in the data analysis, additional components particular to the Ethio telecom sense are also provided. Aside from that, the information governance principles have been implemented into the proposed framework. According to (Dai, et al., 2016), the seven foundational pillars of information governance principles “(i.e. organization, metadata, privacy, data quality, business process integration, master data integration, and information lifecycle management)” are also significant to big data governance. But primarily, the researcher adopts the existing (Yang, Li, Elisa, Prickett, & Chao, 2019) proposed framework to the Ethio telecom context and discipline. As per the questionnaire and interview findings, the proposed framework is customized to the sense of Ethio telecom, and adjustments were made to the adopted framework. Elements on the adopted framework domains are added (organizations structure, big data scope determination, working

mechanism clarification, business case evaluation, stakeholders selection, data identification, big data collection and big data visualization), modified (strategy development and activity plan and big data processing and analysing) and removed (integration and metadata management, and stewardship) based on the requirements identified in the data analysis.

On the other hand, the expected outcome of IS design science research is a meaningful IT artifact that solves a critical corporate problem (Hevner, March, Park, & Ram, 2004). As a result of this investigation, the proposed framework to Ethio telecom is a useful IT artifact that can help the organization solve big data management issues, make better use of its data, and pave the way for a peaceful transition to data-informed business.

5.3 Components of the proposed BDGF

The proposed big data governance framework is organized into three domains and nineteen components (Figure 11). The domain and components of the framework will be briefly presented in the following section.

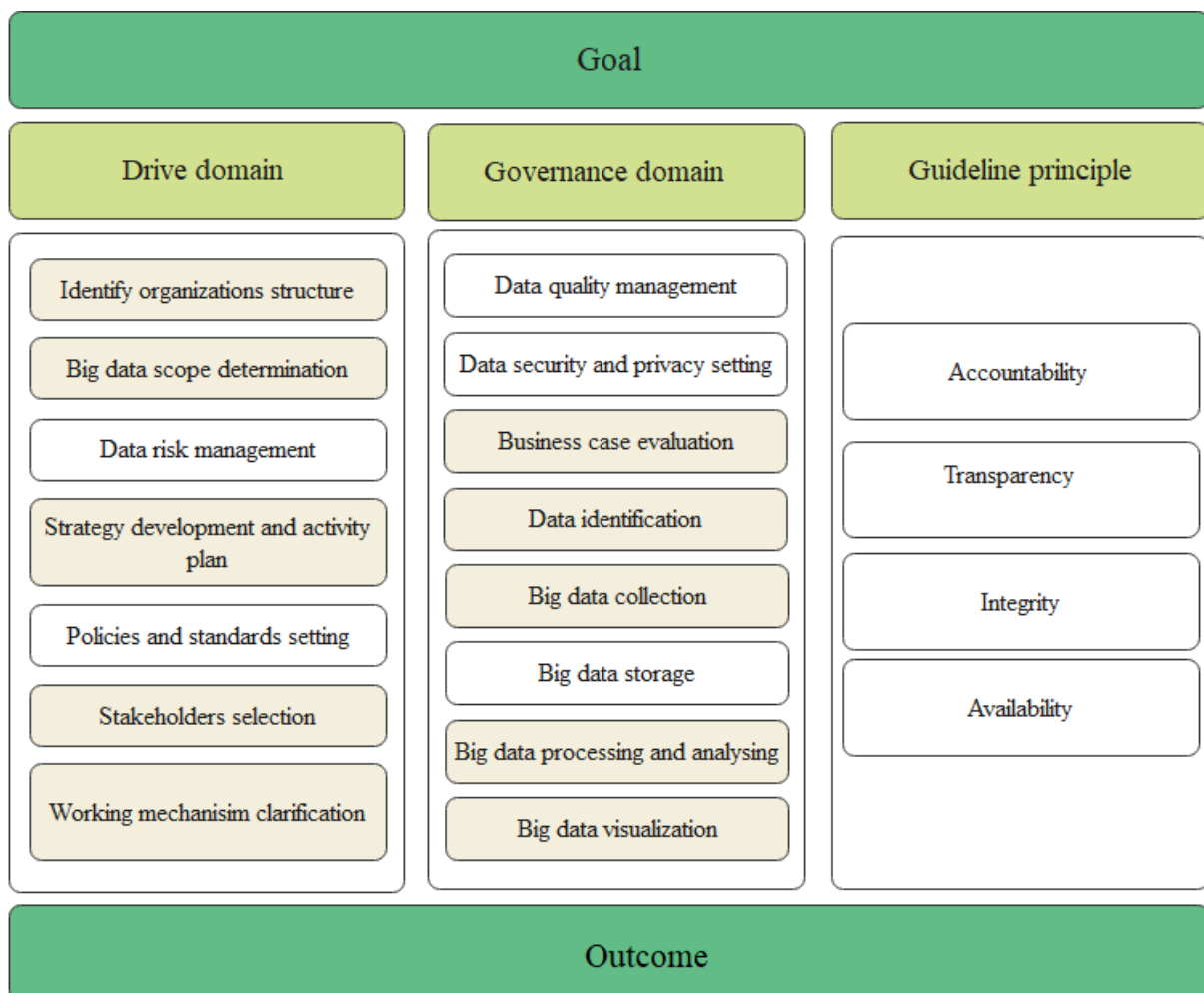


Figure 11: The proposed big data governance framework for Ethio telecom

5.4 The proposed BDG framework and discussions

5.4.1 Goals

The proposed big data governance framework aims to successfully execute big data projects and generate value from data using a new analytical method. And also enables Ethio telecom to protect big data whilst obtaining the highest value from big data, thereby empowering and promoting appropriate data-usage practices. The goals are typically at the heart of those organizations' strategic objectives that can be realized through the use of advanced information and communications technologies to utilize big data.

5.4.2 Drive domain

For the proposed big data governance framework the driving domain had seven components (identify organizations structure, big data scope determination, data risk management, strategy development and activity plan, policy setting, stakeholders selection, and working mechanism clarification) and each component will be explained in more detail below.

1. **Identify organizations structure** - As per (Al-Badia, Tarhinia, & Khana, 2018), big data governance decisions are influenced by the organization and its structure. Therefore, the structure of the organization necessitates further investigation. Big data governance should be consistent with the organization's goals and vision (Kim & Cho, 2018). As a result, the organizational structure has been included as a component.
2. **Stakeholder selection** – Stakeholders are the players in the big data management scheme, so identifying them is essential (Fikiru, 2020). Therefore, The key stakeholders: such as data scientists, data engineers, data analysts, big data researchers, big data developers, business steward leads, data stewards, advisory committees, and so on (Al-Badia, Tarhinia, & Khana, 2018) need to be framed.
3. **Big data scope determination** – previous to beginning the big data governance activity, the scope of the big data governance must be outlined because contradictions between the scope and the organization's infrastructural facilities will lead to the failure of the framework or methodology (Fikiru, 2020). Several issues come up with big data as a result of insufficient technology solutions to handle it in an effective manner (Adler-Milstein & Jha, 2013). This scope determination contains the following components: project budget, schedule, work plan, staffing guidelines, and structures (technical resources) (Fikiru, 2020). The scope determination process must also

consider whether the project will have an impact on the organization's operational performance, business development, and income. Therefore, those things need to be framed.

4. **Data risk management** – Regarding data risks during the big data governance procedure can be recognized based on the aforementioned objective and corresponding strategies, techniques, policies, standards, and guidelines, such as data storage methodologies, data processing mechanism, and information utilization standards can be developed to address the identified risks (Yang, Li, Elisa, Prickett, & Chao, 2019).
5. **Big data strategy development and activity plan** – Strategy establishes a plan of action for the advancement of an organization (Zepa & Ribickis, 2015). The major objectives are included in the core strategy document, but a more detailed description can be discovered in the activity plan (Zepa & Ribickis, 2015). As per (Radomska, 2014) description, internal processes, structures, systems, people, and the environment should all be considered when developing a strategy to ensure its successful implementation. Therefore, these things need to be framed.
6. **Policies development** – The policy development stage articulates the issues related to data capture, data storage, data governance, data usage, data confidentiality, security, risk, retention, regulatory issues, data optimization, data extraction and clustering techniques (Al-Badia, Tarhinia, & Khana, 2018). At this stage, all the policies, laws and regulations, standards, procedures, and principles should be framed.
7. **Working mechanism clarification** – Job clarification is a written document that encompasses the roles, responsibilities, and capabilities of a specific position. Job descriptions, in particular, are relevant because they: enable training and development programs by presenting written documentation outlining what a specific role entails as well as the training and activities required. Establish clear guidelines for employees' authority and responsibility. Ensure that all departments have a clear understanding of how various job roles and positions contribute to the growth of the organization. Therefore, the job identification, specification, summary, working conditions, accountabilities, and responsibilities need to be framed.

5.4.3 Big data governance Implementation Domain

The governance domain defines the data governance objectives that Ethio telecom should concentrate on while performing data governance operations, which are primarily made up of eight components, as illustrated in figure 11.

1. Data quality management

Monitoring and evaluating the quality of big data should be a primary concern (Al-Badia, Tarhinia, & Khana, 2018). In the big data analytics pipeline, the big data managers should look for and correct issues of inconsistent or invalid data. They should keep track of all changes from the initial concept to the final visualization. To answer the questions, the data must be sanitized before analysis (Al-Badia, Tarhinia, & Khana, 2018). “There are seven factors that play a huge role in determining data quality, which is accuracy, availability, completeness, granularity, relevance, reliability and timeliness (Ayswarrya)” (Fikiru, 2020).

To verify the quality of the data, Ethio telecom can take several steps to accomplish the above features of the data: initially, initiate with the source of data by ensuring its reliability, utilizing data quality control tools, and following best business practices such as teaching everyone in the organization on data quality, assigning roles such as data owners, data stewards, and data custodians within your organization, and establishing suitable procedures to make sure high data quality (Fikiru, 2020). Aside from that, preparation and analysis are critical for quality data (Al-Badia, Tarhinia, & Khana, 2018).

2. Business case evaluation

Each big data analysis workflow should begin with a well-defined business case that provides a clear understanding of the justification, validity, motivation, and goals of conducting the analysis (Erl, Khattak, & Buhler, 2015). Accordingly, business case evaluation requires the creation, evaluation, and approval of a business case before proceeding with real handicraft analysis. Hence, before moving on to real handicraft analysis, Ethio telecom must first create, evaluate, and approve a business case.

A big data analytics business case evaluation assists decision-makers in analysing the business resources that will be considered necessary and the organization issues that the analysis will resolve. Accordingly, as mentioned in the previous chapter, the analysis results indicate that big data analytics has numerous business areas in the Ethio telecom industry such as customer relationship and experience, marketing/product recommendation, data monetization, customer profiling, network fault isolation, churn analysis and prediction and fraud detection. Besides

this, another output of this workflow is the identification of the underlying financial plan needed to complete the analysis project. Therefore, any required investments, such as devices, tools, and training, must be acknowledged in advance at Ethio telecom.

3. Data identification

The data identification workflow is devoted to understanding the datasets and sources needed for the analysis project (Erl, Khattak, & Buhler, 2015). The required datasets and their sources can be internal or external to the enterprise, depending on the business scope of the analysis project and the nature of the business problems being addressed.

Accordingly, as mentioned in the previous chapter, the analysis results indicate that in the case of Ethio telecom internal datasets, a list of available datasets, such as call detail data, network, customer profile data, device data, product data, marketing and sales data and billing data. Besides this, external sources may be incorporated within blog sites or other various kinds of content-based internet sites, in which case it may be necessary to extract it using automated systems.

4. Big data privacy and security setting

Data security, privacy, and other ethical considerations are primary concerns while collecting, transferring, storing, and analysing large amounts of data. To prevent data breaches, all operations related to data processing in Ethio telecom networks must be appropriately protected. Looking at network data, for example, may raise privacy concerns owing to the inclusion of personal and organizational sensitive data. As a result, privacy-protection methods and data security technologies must be implemented.

In this regard, the researcher suggests GDPR as a guideline and Apache Metron as a framework in order to mitigate risks and protect data assets. Because recently implemented EU general data protection regulation (GDPR) is the most significant shift in data privacy regulation over the last 20 years, providing detailed guidelines for organizations related to data processing (Yang, Li, Elisa, Prickett, & Chao, 2019). On the other hand, Apache Metron is a cyber-security software platform that allows firms to collect, analyse, and store a variety of security data streams at scale in order to recognize cyber breaches and respond quickly (Verheijde, Gnanavarothayan, Moelchand, & Stam, 2019).

5. Big Data Collection: This component is in charge of gathering data from various sources inside the telecom networks. Data is taken either through the execution of an application

programming interface (API) or through the use of big data collection tools (apache Kafka and apache Flume).

6. **Big Data Storage:** This component is in charge of combining the collected data and storing the results in a central database. The planned logic is to store each data source in its appropriate database (Hbase, Hive and HDFS), and then use API Controller to coordinate between the various stores. The author suggests HDFS, Hbase and Hive because it is an effective tool to extract real-time (Hbase) and batch (Hive) data and help designers to work quickly.
7. **Big Data Processing and Analysis:** This allows the invention of information and knowledge from data through the application of specific techniques and algorithms depending on the data's characteristics. The Hadoop (map-reduce) and Spark analysis clusters were adopted in this study. The concept is to use Hadoop for batch processing and Spark for the streaming process.
 - Batch processing: Batch processing of Ethio telecom data collected from the data repository is handled by this layer. The author suggests that tasks be classified as per their speed requirements. The ones that require faster processing will be handled by Spark, while the rest will be handled by Map Reduce.
 - Real-time/ Streaming process: This layer is in charge of processing real-time streams. The author suggested that data be collected using Apache Kafka, stored in Hbase and processed using the Apache Spark Streaming feature.
8. **Big Data Visualization:** It allows the user to visualize the data that has been ready by the four previous components. The technologies applied were Kibana.

5.4.4 Guiding Principles

The guidelines principle for big data governance assists Ethio telecom in maintaining and handling big data by providing additional directions for any uncovered facets stated in data management regulatory requirements, as described in the previous part. As illustrated in Fig. 11, the guiding principles inside this job are data integrity, accountability, availability, and transparency. The guiding principles must be clear and easy to understand for them to be useful.

1. Data Integrity

According to (Yang, Li, Elisa, Prickett, & Chao, 2019), to ensure data integrity, which is critical to successful data usage, organizations should regularly assess any changes that may occur to the data over time. Data integrity is accomplished by ensuring that data within an

organization is properly delineated, regulated properly, and correctly accessed (Yang, Li, Elisa, Prickett, & Chao, 2019). Applying the guidelines principle (data integrity) allows data to be strongly aligned with organizational objectives as well as cross-business specifications.

2. Transparency

To prevent an Ethio telecom from possible data theft while ensuring that data be used strategically, data management practices must be transparent. Transparency aids in revealing how critical information was managed during the assessment so that an organizational or third-party auditor, or any other data stakeholders, can comprehend data-related guidelines. Moreover, data-related actions, operations, workflows, and decisions should be auditable, with documentary evidence supporting them in compliance-based and organizational auditing regulations.

3. Accountability

Accountability refers to the provision of data as well as the creditability of data operators (Yang, Li, Elisa, Prickett, & Chao, 2019). To improve data ownership and accountability, all divisions of an Ethio telecom must work collaboratively. Data breaches become less of a worry within an organization if all divisions are accountable and responsible for data.

4. Availability

Data availability refers to the accuracy and dependability with which data may be accessed and used (Yang, Li, Elisa, Prickett, & Chao, 2019). As a result, while managing data availability at Ethio Telecom, it is critical to evaluate the following: whether data must be kept available, whether storage techniques and locations are appropriate, and if data can be quickly re-acquired or reproduced if deleted, stolen, or damaged.

5.4.5 Outcomes

Finally, the information is made available and distributed to the clients.

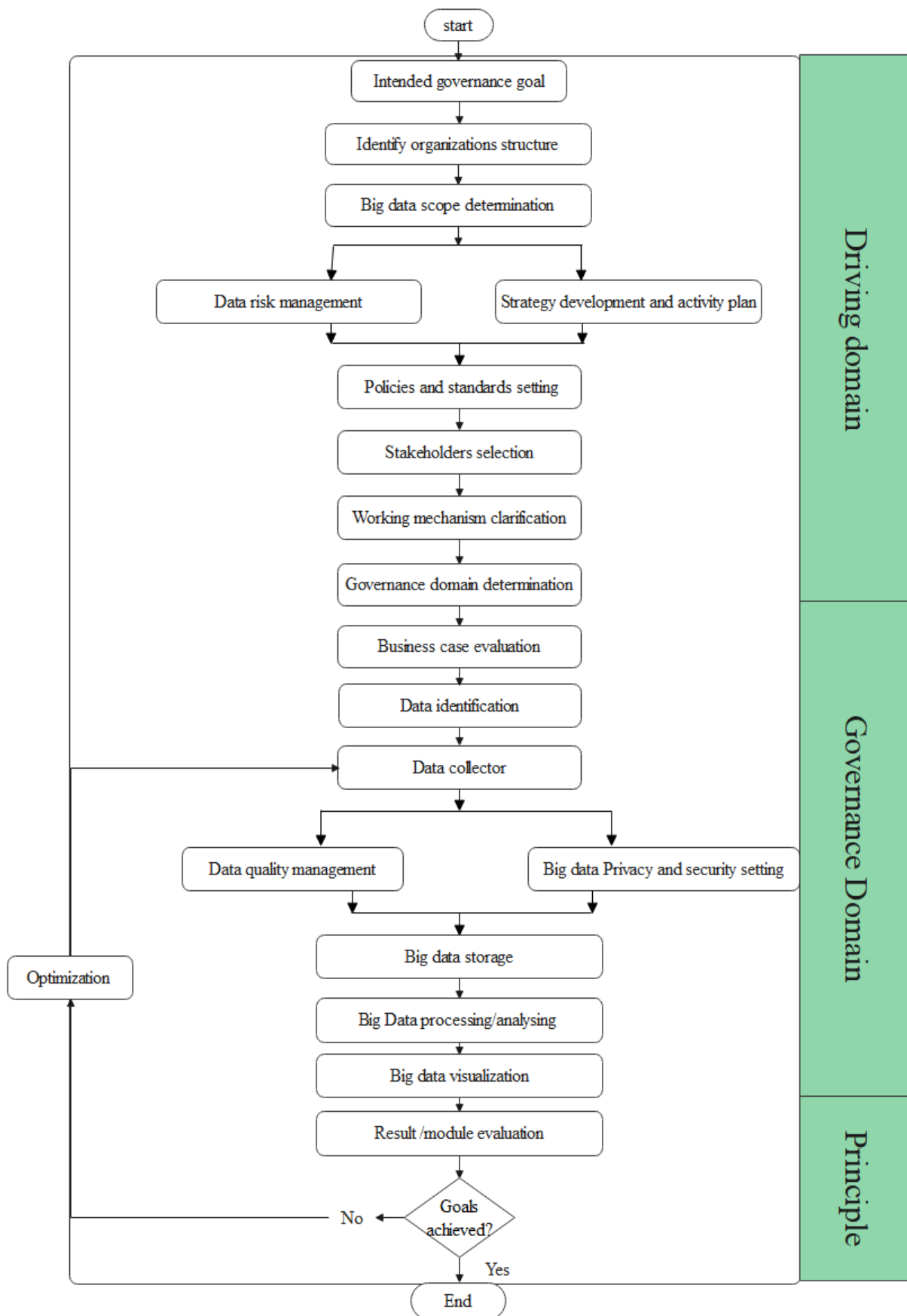


Figure 12: an implementation framework for the proposed BDG framework

5.5 Evaluation of the Proposed Framework

5.5.1 Expert Validation

Expert validation was used in accordance with descriptive method to evaluate the proposed framework in this study. The expert validation was carried out by fourteen professionals with many years of data related experience and high-level IT managers using the evaluation matrix which can be seen in Appendix D to collect individual evaluations of the optimized framework in terms of completeness, clarity, and correctness of the revised framework.

Following the identification of the fourteen experts, they are given the proposed BDG framework and the evaluation matrix and invited to speak on its completeness, clarity, and correctness. The content of the questionnaire is derived from the evaluation criteria recommended by (Hevner, March, Park, & Ram, 2004) which consists of fit to the organization, comprehensiveness, reliability, clarity, correctness, and usability quality attributes and the evaluation checklist is adopted from (TIGIST, 2018) cited in (Fikiru, 2020), because the researcher finds the checklist appropriate for this study.

In order to evaluate the consistency of the evaluation of the study, Cronbach's alpha reliability test is applied and the value of the coefficient alpha (0.815) indicates that the survey is reliable since it is greater than 0.7.

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.815	.810	10

Table 15: Reliability Statistics for Evaluation of the Proposed Framework

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
The proposed framework is comprehensive in terms of coverage	14	3	5	4.36	.745
The organization and presentation of the framework is suitable for Ethio telecom	14	3	5	4.43	.646
The objective of the framework is clear	14	3	5	4.29	.726

The content of the proposed framework is complete	14	3	5	4.36	.633
The content of the proposed framework is relevant	14	4	5	4.57	.514
The content of the proposed framework is clear.	14	3	5	4.14	.535
The content of the proposed framework is Scalable	14	4	5	4.57	.514
The proposed framework is easy to be applicable	14	4	5	4.36	.497
The applicability of the proposed framework can improve resource utilization, scalability, data management and yield data management benefits	14	4	5	4.64	.497
The implementation of the proposed framework fits with the organization problems	14	4	5	4.71	.469
Valid N (listwise)	14				

Table 16: Mean and standard deviation of the Framework Evaluation Survey

From the evaluation result table 16 revealed the proposed framework is comprehensive in terms of coverage, the organization and presentation of the framework is suitable for Ethio telecom, the objective of the framework is clear and the proposed framework is easy to be applicable with the mean score is (mean = 4.36, mean = 4.43, 4.29 and 4.36 respectively) considering a five-point Likert scale where 5 means that the respondents strongly agreed.

On the other hand, the results in the table 16 revealed the content of the proposed framework is complete, relevant, clear and scalable as the evaluation result shows a mean of 4.36, 4.57, 4.14 and 4.57 respectively, so we can conclude that the evaluators strongly agreed on complete, scalability, clarity and relevance of the proposed framework. The applicability of the proposed framework can improve resource utilization, scalability, data management and yield data management benefits (4.64) and the implementation of the proposed framework fits with the organization problems (4.71) revealed the validity of the framework to be implemented at Ethio telecom.

The descriptive analysis (mean and standard deviation) of the survey result is computed and as it can be seen in table 16 the mean result of the evaluation variables is found to be greater than 3 which indicated that the respondents agreed on the clarity, completeness, usefulness, correctness of the proposed framework.

Therefore, based on the above analysis evaluation result confirmed the completeness, correctness, clarity, and applicability of the proposed framework.

However, it is recommended that the framework has to be evaluated through practical application and more research be conducted so that it can be modified and enhanced over time.

5.6 Result and discussion

As per the study findings, big data analytics has several application areas in the Ethio telecom business, and data in Ethio telecom may be described by the standard 3Vs: volume, variety, and velocity. However, in contrast to the data generation capacity and the availability of big data application areas, Ethio telecom is not working hard enough on the data analysis and also they don't use it fully for the purpose of decision making, policy formulation and for future strategic planning.

On the other hand, Ethio telecom uses traditional data analysis and simple office package software for data processing. However, those tools and approaches that are currently in use for data management are not sufficient for yielding the big data utilization benefits. This implies that Ethio telecom has a better chance to gain new revenue streams through big data analytics (BDA) options, but they do not have a clear roadmap on how to capitalize on it effectively. The key challenges were mentioned in the study are lack of awareness, absence of big data governance framework, lack of availability of big data expert people, big data confidentiality (privacy and security challenge), availability of necessary tools to generate insight from data in a timely manner, absence of data validity and completeness, absence of integrated systems that handle all aspects of information, internal resistance to implementing BDA as a new way of doing business and lack of exposure to the advancement in technology.

As a result, the value of designing BDGF inside Ethio Telecom is undeniable. Therefore, according to the data collected from the study, the special requirements included in the framework to fill the data management gap in Ethio Telecom and to design the big data governance framework are:

- Emphasizing the structure of the organization,
- Focus on developing big data-centric policies and standards
- Identifying stakeholders
- Determining the scope of big data in terms of organizational structure and purpose
- Big data strategy development and action plan
- Working mechanism clarification
- Big data privacy and security setting

- Big data Analytics lifecycle management (business case evaluation, data identification, data collection, data storage, data analysing, and visualization) and
- Who will be the ultimate beneficiary of this business is the most significant.

In addition, regarding the important needs that must be met in the practice of BDGF in Ethio telecom, one responder noted that *“A critical requirement that must be covered in the practice of BDGF in Ethio telecom, It should be an organizational structured-centered and big data-based (The scope of big data being considered). And also, BDGF processes within the context of Ethio telecom operational environment at least require emphasis on the policies, rules, and standards corresponding to data capture, analysis, management, process, privacy, security, risk, and data usage requirements”*.

Therefore, as per the study findings, the organization structure (environment), the strategic goals of the company, technology capability, business environment, people's awareness are all identified as special criteria in this study to design the new big data governance framework.

Accordingly, a big data governance framework for Ethio telecom based on the governance activities associated with the processing of data in Ethio telecom networks has been proposed. The proposed framework consists of three domains and within the domains, there are sixteen components. The framework development is guided by design science principles, some of the framework components derived from previous frameworks in the business sector and adjusted to meet the Ethio telecom setting.

However primarily, the researcher adopts the existing (Yang, Li, Elisa, Prickett, & Chao, 2019) proposed framework to the Ethio telecom context and discipline. The adopted big data governance framework has three domains: (planning, governance, and evaluation) and twelve components (data risk management, policies, stewardship, distributed data processing, distributed data storage, integration and metadata management, data quality management, data security and privacy, accountability, transparency, integrity, and availability). As per the questionnaire and interview findings, the proposed framework is customized to the sense of Ethio telecom, and adjustments were made to the adopted framework.

Accordingly, depending on the specifications outlined in the data analysis, additional components particular to the Ethio telecom sense are provided such as organizations structure, big data scope determination, working mechanism clarification, business case evaluation, stakeholders selection, data identification, big data collection, and big data visualization.

Besides this, elements on the adopted framework domains are modified such as strategy development and activity plan, and big data processing and analysing. On the other hand,

elements on the adopted framework domains are removed (integration and metadata management, and stewardship) based on the requirements identified in the data analysis.

Aside from that, the information governance principles have been implemented into the proposed framework. According to (Dai, et al., 2016), the seven foundational pillars of information governance principles “(i.e. organization, metadata, privacy, data quality, business process integration, master data integration, and information lifecycle management)” are also significant to big data governance.

As a result, depending on Ethio telecom big data practice, a big data governance framework with three domains and nineteen elements was proposed, which include a drive domain (identify organizations structure, big data scope determination, data risk management, strategy development and activity plane, policies and standards setting, stakeholder selection, and working mechanism clarification), governance domain (data quality management, data security and privacy setting, business case evaluation, data identification, big data collection, big data storage, big data processing and analysing, and big data visualization), and guideline principle (accountability, transparency, integrity, and availability). It helps Ethio telecom in efficiently achieving the desired results in innovative data utilization, which frequently leads to a culture change in organizations for deep data-driven processes. Finally, expert validation was then performed to evaluate the proposed big data governance framework. As a result, the research procedure and study findings are deemed satisfactory, demonstrating the usefulness and applicability of the proposed framework.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Introduction

The latest chapter provides the study's conclusion as well as recommendations based on the research findings. Besides this, the chapter suggests possible future research concepts.

6.2 Conclusion

Due to the enormous growth in data, which is often structured, unstructured, or semi-structured in origins, big data management has become a challenging task in the IT and business fields. Big data governance has recently gained prominence in organizational decision-making and policy formulation.

From an Ethiopian perspective, Ethio telecom is one of the significant government organizations that generate big data. As seen throughout the study, this data is produced in both batch and real-time forms. Big data analytics has a wide range of applications in the telecom industry. According to the study, Ethio telecom generates a large volume of data with varying types and speeds. However, Ethio telecom has not used big data analytics/ governance for decision-making, policymaking, or long-term strategy development. Rather they are using state-of-the-art of traditional computing methods to manage, store and process big data. For example, for storage purposes, experts use database providers such as Oracle, Huawei solution, etc. The study discussed the challenges that have prevented Ethio telecom from fully implementing big data analytics. These are challenges on awareness, absence of big data governance framework, lack of availability of big data expert people, big data confidentiality (privacy and security challenge), availability of necessary tools to generate insight from data in a timely manner, absence of data validity and completeness, absence of integrated systems that handle all aspects of information for a given customer infrastructure, such as storage device challenges, data analysis challenges, data privacy challenges, and internal resistance to adopting BDA as a new way of doing business. On the other hand, currently, there is no big data governance framework at Ethio telecom and, the main reason that prevents the firm from formulating the framework are there is a challenge on awareness, shortage of the required skilled people on big data, and a lack of exposure to big data technology.

Therefore, this study presents a big data governance framework that can fill the gaps mentioned earlier in this section, as well as a detailed understanding of big data governance processes.

There has been very little research on big data governance frameworks. In this study, a big data governance framework for Ethio telecom based on the governance activities associated with the processing of data in Ethio telecom networks has been proposed, which could serve as a useful reference and give latecomers a chance to participate in the multi-dimensional growth of Ethio telecom. The proposed framework consists of three domains and within the domains, there are nineteen components. The proposed governance framework greatly helps in the coordination of people, policies, processes, strategies, standards, and technologies to allow Ethio telecom to use data as a key business asset. Furthermore, it demonstrates the information that aids in the establishment of decision-making structures based on data.

6.3 Recommendations

The following recommendations for practice and future research are made depending on the study's main findings and conclusions.

The researcher strongly recommends Ethio telecom put the proposed framework into action. This is a framework for assisting Ethio telecom in effectively controlling both structured and unstructured large datasets, enhancing value from big data, and allowing and promoting appropriate big data practices by maintaining the integrity, usability, and value of their data and reducing data risk in the organization. It also helps to guide staff to perform their roles in data management. Because data governance is like the process of designing and creating the road map for new construction on a building, and data management is the process of actually constructing the building. And, while you can build a house without a road map (data governance), it will be inefficient and ineffective, with a higher risk of failure in the future. Therefore, Big Data governance is critical to the success of a company that processes massive amounts of data.

While this framework is critical to Ethio telecom's success in managing big data, the work can be improved in many ways.

Firstly, Future research should emphasize the present implementation of the framework based on the case study of big data governance in various scenarios to demonstrate the work more efficiently and assess the framework's success or impact.

Secondly, it is desirable to successfully validate and test the proposed framework by implementing it in a real-world situation such as fraud detection, customer relationship, and experience, marketing/product recommendation, data monetization, customer profiling, network fault isolation, and churn analysis and prediction.

On the other hand, with the launch of innovative platforms and technologies like networking sites, smartphones, and digital marketing and advertising, the data created no longer has a uniform format or structure like traditional ones and cannot be analysed using relational models. Text, XML, emails, photos, weblogs, videos, and other sorts of data are among them, resulting in a flood of new data formats. This non-structured data is either semi-structured or unstructured, making searching and analysis difficult. Therefore, future research should concentrate on the design and development of big data analytics for processing unstructured data in real-time.

Furthermore, the proposed work is only provided in an Ethio telecom frame of reference; therefore, it would be very attractive to broaden the work towards other data-driven areas such as healthcare, transportation, energy, pharmaceuticals, payment systems, agricultural transformation agency, and educational provider companies by adjusting to their business settings, which would aid in developing an understanding of the topic and may produce additional insights.

REFERENCES

- Tata Tele Business Services. (2021). Big Data and the Telecom Industry. Retrieved 05 01, 2021, from <https://www.tatatelebusiness.com/downloads/WhitePapers/resources/Big-Data-and-the-Telecom-Industry.pdf>
- Abbasi, A., Sarker, S., & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17(2). Retrieved from <http://ezproxy.lib.monash.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=113533298&site=ehost-live&scope=site>
- Abraham, R., Schneider, J., & Brocke, J. v. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*. Retrieved from <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Abraham, R., Schneider, J., & Brocke, J. v. (2019). Data Governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*. doi:10.1016/j.ijinfomgt.2019.07.008
- Adler-Milstein, J., & Jha, K. A. (2013). Healthcare's "big data" challenge. *he American journal of managed care*, 19(7), 537-538.
- Akal, T. D., Beshah, T., Sackmann, S., & Negash, S. (2019). Challenges of Identifying and Utilizing Big Data Analytics in a Resource-Constrained Environment: In the Case of Ethiopia. *IFIP International Federation for Information Processing*, 234–254. Retrieved from https://doi.org/10.1007/978-3-030-20671-0_17
- Al-Badia, A. A. (2018). *Exploring Big Data Governance Frameworks*.
- Al-Badia, A., Tarhinia, A., & Khana, A. I. (2018). Exploring Big Data Governance Frameworks. *ScienceDirect*. doi:10.1016/j.procs.2018.10.181
- ApacheSoftwareFoundation. (2021). Apache Hadoop. Retrieved from <http://hadoop.apache.org/>
- Asressa, B. M., & Beshah, T. (2016). An Architecture for Big Data Analysis in the Context of Social Media: The Case of Twitter. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 17, 234-244.
- Bavaskar, P. P., Kemker, O., & Sinha, A. K. (2020). A SURVEY ON: "LOG ANALYSIS WITH ELK STACK TOOL". *SSRN Electronic Journal*, 6(4). Retrieved from <https://www.researchgate.net/publication/343775739>
- Bell, J. (2020). *Apache Spark*. doi:10.1002/9781119642183.ch13
- Berson, A., Dubov, L., Plagman, B., & Raskas, P. (2011). Master Data Management and Data Governance. *McGraw-Hill*.
- Bhadani, A. K., & Jothimani, D. (2016). Big Data: Challenges, Opportunities and Realities.

- Bhargava et al., M. G. (2017). Potential of Big Data Analytics in Bio-Medical and Health Care Arena: An Exploratory Study. *Global Journal of Computer Science and Technology: C Software & Data Engineering*, 17(2).
- Bogale, M. (2018). PROPOSING INFORMATION SECURITY AWARENESS. *Addis Abeba University*.
- Charikleia, P. (2015). Data Governance: A conceptual framework in order to prevent your Data Lake from becoming a Data Swamp. *Luleå University of Technology*.
- Chen, c.-m. (2016). Use cases and challenges in telecom big data. Retrieved from <https://doi.org/10.1017/ATSIP.2016.20>
- Chen, c.-m. (2016). Use cases and challenges in telecom big data analytics. *APSIPA Transactions on Signal and Information Processing*, 5. doi:10.1017/ATSIP.2016.20
- Christiaens, J. (2020, may 14). Key Architecture Principles for a viable and evolvable IoT architecture. *Antwerp Management School*, 5. Retrieved from https://www.researchgate.net/publication/341370668_Key_Architecture_Principles_for_a_viable_and_evolvable_IoT_Architecture
- Côrte-Real, N., Oliveira, T., & Ruivo, p. (2016). Assessing the business value of Big Data Analytics in European firms. *Journal of Business Research*, 379-390. Retrieved from <https://doi.org/10.1016/j.jbusres.2016.08.011>
- Cothari, C. (2010). *Research Methodolgy: Methods and techniques*.
- Creswell, J. W. (2014). Research design: qualitative, quantitative, and mixed methods approaches. 4.
- Dai, W., Wardlaw, I., Cui, y., Mehdi, K., Li, Y., & Long. (2016). Data profiling technology of data governance regarding big data: Review and rethinking. *In Information Technology: New Generations* Springer, 439-450.
- Dam, R. (2013). Big data a sure thing for telecommunications: telecom's future in big data. *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*.
- DAMAInternational. (2009). The DAMA Guide to The Data Management Body of Knowledge. *New Jersey: Technics Publications, LLC*.
- datareportal. (2021, apr). *datareportal*. Retrieved from DIGITAL AROUND THE WORLD: <https://datareportal.com/global-digital-overview>
- Debele, M. L. (2018). KNOWLEDGE MANAGEMENT FRAMEWORK FOR THE MINISTRY OF DEFENSE (MOD) OF THE FEDERAL DEMOCRATIC REPUBLIC OF ETHIOPIA (FDRE). *Addis Ababa University*.
- Delotte. (2015). Opportunities in Telecom Sector: Arising from Big Data. *Aegies school of bussiness*.

- Desalegn, T. (2016). Big Data Processing and Visualization in the Context of Unstructured data set.
- Doshi, Z., Agrawal, R., Kanani, P., & Padole, M. (2020). Big Data, Big Challenges.
- Erik, K. D., & Joel, T. (2012). Industrial Network Security: Second Edition. 2.
- Erl, T., Khattak, W., & Buhler, P. (2015). *Big Data Fundamentals Concepts, Drivers & Techniques*. (M. Taub, Ed.)
- ethiotelecom.et. (n.d.). *BRIEF HISTORICAL REVIEW OF TELECOM SECTOR IN ETHIOPIA*. Addis Abeba .
- Feki, M., & Boughzala, I. (2016). BIG DATA GOVERNANCE: A LITERATURE REVIEW AND RESEARCH AGENDA. *ResearchGate*. Retrieved from <https://www.researchgate.net/publication/338677320>
- Ferreira, J. J., Raposo, M., & Mainardes, E. W. (2014). Strategy and Strategic Management Concepts: Are They Recognised by Management Students? doi:10.15240/tul/001/2014-1-004
- Fikiru, Z. (2020, sep). A Big Data Management Framework for Wegagen. *Addis Ababa University*.
- Gaurav, D., Goyal, A., & Kaliyar, R. (2018). An Outline on Big Data and Big Data Analytics. *International Conference on Advances in Computing, Communication Control and Networking* .
- Getachew, E. (2018). Software Requirement Engineering Process Improvement for Outsourced Projects in the case Ethio Telecom:. *Addis Ababa University*.
- Glossary, G. (2012). *Big Data*. Retrieved from gartner.com: from <https://www.gartner.com/it-glossary/bigdata/>
- Gorschek, T., & Wohlin, C. (2003). Identification of improvement issues using a lightweight triangulation approach. *European Software Process Improvement Conference*.
- Hafez, H. A. (2016). Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity.
- Haneem, F., K. N., T. N., P. D., & AbuBakar. (2019). Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management*, 45, 25-43.
- Hevner, A. R., March, S., Park, J., & Ram, S. (2004). DESIGN SCIENCE IN INFORMATION. *MIS Quarterly*, 28, 75-105.
- Hevner, A., & Chatterjee, S. (2010). Design Research in Information Systems. (R. Sharda, & S. Voß, Eds.) *Springer Science and Business Media*. doi:10.1007/978-1-4419-5653-8
- Hogarth, R. M., & Soyer, E. (2015). Using Simulated Experience to Make Sense of Big Data. *MIT Sloan Management Review*, 5-10.
- IDC. (2014). The DIGITAL UNIVERSE of OPPORTUNITIES.

- Joseph, M. (2013). Data mining and business intelligence applications in telecommunication Industry. *International Journal of Engineering and advanced Technology (IJEAT)*.
- Kabir, S. M. (2016). Basic guidelines for research: An introductory Approach for all Disciplines . In S. M. Kabir, *Basic guidelines for research: An introductory Approach for all Disciplines* (1 ed., pp. 201-275). Bangladesh: Book Zone Publication.
Retrieved from https://www.researchgate.net/publication/325846997_METHODS_OF_DATA_COLLECTION
- Kabir, S. M. (2016). SAMPLE AND SAMPLING DESIGNS. *Research Gate*.
- Kanjilal, J. (2021). An Introduction to The Lambda Architecture. Retrieved from <https://www.developer.com/design/intro-to-lambda-architecture/#:~:text=The%20Lambda%20Architecture%20is%20a%20data%20processing%20architecture,new%20paradigm%20for%20handling%20vast%20amounts%20of%20data.>
- Kastouni, M. Z., & Lahcen, A. A. (2020). Big data analytics in telecommunications: Governance, architecture and. *Journal of King Saud University – Computer and Information Sciences*. Retrieved from <https://doi.org/10.1016/j.jksuci.2020.11.024>
- Katal, A., Wazid , M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good. *Sixth International Conference on Contemporary Computing (IC3) (IEEE, 2013)*, (pp. 404-409).
- Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good. *Sixth International Conference on Contemporary Computing (IC3) (IEEE, 2013)*, (pp. 404-409).
- Katkar, J. (2015). STUDY OF BIG DATA ARHITECTURE LAMBDA ARHITECTURE. *San Jose State University*. doi:<https://doi.org/10.31979/etd.rja5-w7my>
- Khan et al. (2016). Impact of Big Data over Telecom Industry. *Pakistan Journal of Engineering Technology and Science (PJETS)*, 6.
- Khan, F. R. (2021). Apache kafka with real-time data streaming. Retrieved from <https://www.researchgate.net/publication/348575301>
- 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 (JBRMR)*, 12(3).
- 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 (JBRMR)*, 12(3).
- Kumar, A., & Jothimani, D. (2017). Big Data: Challenges, Opportunities and Realities. Retrieved from

https://www.researchgate.net/publication/316956144_Big_Data_Challenges_Opportunities_and_Realities

- Kumar, K., & Valcin, F. (2020). Analysis of Big Data Tools and Algorithms. *International Journal of Computer Trends and Technology (IJCTT)*, 68(6). doi: 10.14445/22312803/IJCTT-V68I6P101
- Kumar, Y. (2020). Lambda Architecture – Realtime Data Processing. *Zukovsky State University*. doi: 10.13140/RG.2.2.19091.84004
- LI et al., Q. (2019). A Framework for Big Data Governance to Advance RHINs: A Case Study of China. *research gate*, 7. doi:10.1109/ACCESS.2019.2910838
- Madhlangobe, W. (2018). Assessment of Factors Influencing Intent-to-Use Big Data Analytics in an Organization: A Survey Study. *Nova Southeastern University*. Retrieved from https://nsuworks.nova.edu/gscis_etd/1054
- Mauro, A. D., Greco, M., & Grimaldi, M. (2014). What is Big Data? A Consensual Definition and a Review of Key Research Topics.
- Ms. Komal. (2018). A Review Paper on Big Data Analytics Tools. *International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)*, 4(5). Retrieved from <https://www.researchgate.net/publication/326111866>
- Nagdive, A. S., Tugnayat, R., Regulwar, G., & Petkar, D. (2019). Web Server log Analysis for Unstructured data Using Apache Flume and Pig. doi:10.26438/ijcse/v7i3.220225
- Naik, P., & Naik, G. (2020). *Big Data Tools -Which, When and How? (Volume-III) (Apache Flume and Apache Spark)*. Shashwat Publication. doi:ISBN: 978-81-945989-4-7
- OCHIENG, G. F. (2015). THE ADOPTION OF BIG DATA ANALYTICS BY. *University Of Nairobi*.
- O'Neal, K. (2012). Big Data: Governance is the Critical Starting Point. Retrieved from [http://www.b-eyenetwork.com/blogs/oneal/archives/2012/11/big_data_govern.php]
- Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University – Computer and Information Sciences*.
- ÖZARSLAN, S. (2015). MobileCDP: A MOBILE FRAMEWORK FOR THE CONSUMER DECISION PROCESS. *Middle East Technical University*.
- Panimalar.S, A., Shree.S, V., & Kathrine. A, V. (2017). The 17 V's Of Big Data. *International Research Journal of Engineering and Technology (IRJET)*, 04(09).
- Paschalidi, C. (2015). Data Governance: A conceptual framework in order to prevent your Data Lake from becoming a Data Swamp. *Luleå University of Technology*.
- Patel, H. (2017). HBase: A NoSQL database. Retrieved from <https://www.researchgate.net/publication/317399857>

- Peppers et al. (2006). THE DESIGN SCIENCE RESEARCH PROCESS: A MODEL FOR PRODUCING AND PRESENTING INFORMATION SYSTEMS RESEARCH. *Journal of Management Information Systems*, 24.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24, 45-77. Retrieved from <http://www.jstor.org/stable/40398896>
- Perera, S. (2013). *Hadoop MapReduce Cookbook* (1st ed.). Packet Publishing. doi:ISBN: 1849517282
- Pointer, I. (2020). What is Apache Spark? The big data platform that crushed Hadoop.
- Poleto, T., De Carvalho, V., & Seixas Costa, A. (2015). The roles of big data in the decision-support process: An empirical investigation. *Lecture Notes in Business Information Processing*, 10-21. Retrieved from https://doi.org/10.1007/978-3-319-18533-0_2
- Prakash, T., Kakkar, M., & Patel, K. (2016). Geo-identification of web users through logs using ELK stack. doi:10.1109/CONFLUENCE.2016.7508191
- Pranita, B., Onkar, K., Aditya, S., & Dr M. (2020). A SURVEY ON : LOG ANALYSIS WITH ELK STACK TOOL. 6.
- Radomska, J. (2014). Operational risk associated with the strategy implementation. *Management*, 18, 31-43.
- Rakhman, R. A., Widiastuti, R. Y., Legowo, N., & Rober, E. (2019). Big Data Analytics Implementation In Banking Industry – Case Study Cross Selling Activity In Indonesia’s Commercial Bank. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, 8(9).
- Riahi, Y., & Riahi, S. (2018). Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering*, 5(9). doi: <http://dx.doi.org/10.21276/ijre.2018.5.9.5>
- Sassi, I., Anter, S., & Bekkhoucha, A. (2019). An Overview of Big Data and Machine Learning Paradigms.
- Saunders et al. (2009). *Research Methods for Business Students*. 5th edn.
- Sholaye, M., & Teferi, D. (2021). Designing Big-data Analytics Platform for Real-time Log Management.
- Singh, J., & Singla, V. (2015). Big Data: Tools and Technologies in Big Data. *International Journal of Computer Applications* (0975 – 8887), 125.
- Soares, S. (2013, October 17). Big data governance. Retrieved from file:///C:/Users/user/Downloads/Documents/damanyc_bigdatagovernance17_october_2013_2.pdf

- Tan et al. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph,. *International Journal of Production Economics*, 165, 223-233.
- TechVidvan. (2021). Role of Big Data in Telecom Industry with Case Studies. Retrieved from <https://techvidvan.com/tutorials/big-data-in-telecommunication/>
- The-Next-Tech. (2021). *How much Data is Produced every Day 2021?* (The -Next-Tech) Retrieved from <https://www.the-next-tech.com/blockchain-technology/how-much-data-is-produced-every-day-2019/>
- Tongco, M. D. (n.d.). Purposive Sampling as a Tool for Informant Selection. *A journal of plants, people and applied research*.
- Verheijde, J., Gnanavarotheyan, K., Moelchand, P., & Stam, J. v. (2019). Real Time Threat Detection Through Network Analysis. *Delft University of Technology*. doi:DOI: 10.13140/RG.2.2.32802.17607
- Von Alan, R. e. (2004). Design science in information systems research. *MIS quarterly*, 75-105.
- Vozábal, M. (2016). Tools and Methods for Big Data Analysis. *University of West Bohemia*.
- Wang, Z., Wei, G., Zhan, Y., & Sun, Y. (2017). Big data in telecommunication operators: data, platform and practices. *Journal of Communications and Information Networks*, 2. doi: 10.1007/s41650-017-0010-1
- Weiss, G. (2005). Data mining in Telecommunication. (O. Maimon, & L. Rokach, Eds.)
- Wiley, J., & Sons. (2015). Data Science & Big Data Analytics. *EMC Education Services*. doi: 978-1-118-87613-8
- Wu, H., Wolter, K., & Zhihao, S. (2019). System, Performance Prediction for the Apache Kafka Messaging. Han Wu. doi:10.1109/HPCC/SmartCity/DSS.2019.00036
- XenonStack. (2018, oct 23). *XenonStack*. (XenonStack) Retrieved from www.xenonstack.com: <https://www.xenonstack.com/insights/big-data-governance/>
- Yang, L., Li, J., Elisa, N., Prickett, T., & Chao, F. (2019). Towards Big data Governance in Cybersecurity.
- Yayah, F. C., & Ghauth, K. I. (2017). Adopting Big Data Analytics Strategy in Telecommunication Industry. *Journal of Computer Science & Computational Mathematics*, 7(3). doi:10.20967/jcscm.2017.03.002
- Yayah, F. C., Ghauth, K. I., & Ting, C.-Y. (2017). Adopting Big Data Analytics Strategy in Telecommunication Industry. *Journal of Computer Science & Computational Mathematics*, 7(3).
- Yayah, F. C., Ghauth, K. I., & Ting, C.-Y. (2017). Adopting Big Data Analytics Strategy in Telecommunication Industry. *Journal of Computer Science & Computational Mathematics*, 7(3). doi:10.20967/jcscm.2017.03.002

- Yilmaz, S. K. (2019). BIG DATA AND BIG DATA APPLICATIONS IN THE WORLD.
- Zafar, S., Ul Haq, H. B., Haroon- Ur-Rashid, & Khalid, I. (2021). Big Data: Challenges, Popular Tools Of Big Data -Benefits And Applications. *International Journal of Scientific & Technology Research*, 10(5). Retrieved from <https://www.researchgate.net/publication/351820412>
- Zahid, H., Mahmood, T., Morshed, A., & Sellis, T. (2019). Big data analytics in telecommunications: literature review and architecture recommendations. *IEEE/CAA Journal of Automatica Sinica*. doi:10.1109/JAS.2019.1911795
- Zarate Santovena, A. (2013). Big data: evolution, components, challenges and opportunities. *a PhD thesis at Massachusetts Institute of Technology*.
- Zepa, A., & Ribickis, L. (2015). Strategy development and implementation – process and factors influencing the result: Case study of Latvian organizations. *20th International Scientific Conference Economics and Management (ICEM)*. Elsevier Ltd. doi: 10.1016/j.sbspro.2015.11.507

APPENDIX A: AAU LETTER

አዲስ አበባ ዩኒቨርሲቲ
የተፈጥሮ ሳይንስ ኮሌጅ
የኢንፎርሜሽን ሳይንስ ት/ቤት



Addis Ababa University
College of Natural Science
School of Information Science

Date: March 3, 2021
Ref No. SIS/23/2021/13

To: Ethio telecom,
Addis Ababa

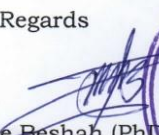
Subject:- Student Bereket Shewangizaw Shumete

Dear Sir /Madam,

Student Bereket Shewangizaw Shumete (ID.No GSR/0124/12) is graduate student at the School of Information Science, Addis Ababa University. She is currently conducting a M.Sc. Thesis research under the title “Big Data Governance Framework for Ethio telcom Company: The case of Ethio telecom, ETRA”.

I would like to thank you in advance for all the assistance that you would provide to the student.

With Regards


Tibebe Beshah (PhD)
Head, School of Information Science



Date: March 3, 2021

☒: 1176 Email: information_cci_cns@aau.edu.et ☎: +251-(11)-122-91-91

APPENDIX B: QUESTIONNAIRE

Dear Sir/Madam:

I am Bereket Shewangizaw, a graduate student in school of information science (Information system) at Addis Ababa University, Ethiopia. Currently I am conducting a research which aims to design big data governance (BDG) framework for Ethio telecom (ET).

The main purpose of this study is to enrich the Big Data governance framework according to the structure of the company by review the existing Big Data governance experience and implementation at Ethio Telecom to overcome the challenges surrounding the implementation of big data governance practice, to address the current operational limitations within the organization and to contribute to the knowledge base. Based on this, this questionnaire is prepared to collect the necessary data on the practices/experience of big data governance, processes and tools/techniques being used, employee's attitude towards using big data analytics, big data management and big data governance framework, and factors that may influence the effectiveness of managing data in Ethio telecom, so an organization-specific big data governance framework can be created. Depending on the information you provide, the final output is a standardized big data governance strategic approach for Ethio telecom.

Your candidate answers to each question and comment are critical to the research's success. The collected data can be used only for academic work. As a result, all respondents will be kept anonymous and will have no impact on anyone in any way.

After you have received this questionnaire in your hands, the analyst may collect it in 5 days. Your commitment is greatly admired and respected. I'd like to take this opportunity to express my gratitude in advance for your thoughtful engagement, genuineness, and promptness in responding to the survey.

If you have any inquiry, please feel free and contact me at

Name: Bereket Shewangizaw

Phone: 0921140131

Email: bereket.shewangizaw07@gmail.com

Part one: General information

1. Demographic Profile

Please provide your answer by putting (X) sign

Personal Data	
<p>1. Role Level</p> <p>Director <input type="checkbox"/></p> <p>Manager <input type="checkbox"/></p> <p>Staff/Expert <input type="checkbox"/></p> <p>Supervisor <input type="checkbox"/></p>	<p>2. Your role at your company</p>
<p>3. Level of education</p> <p>College Diploma <input type="checkbox"/></p> <p>First Degree <input type="checkbox"/></p> <p>Master’s Degree <input type="checkbox"/></p> <p>Above Master’s Degree <input type="checkbox"/></p>	<p>4. Work experience in Ethio telecom</p> <p>from 1-5 years <input type="checkbox"/></p> <p>From 5-10 years <input type="checkbox"/></p> <p>From 11-20 years <input type="checkbox"/></p> <p>More than 20 years <input type="checkbox"/></p>

2. In this section you are expected to respond to the following items related to Ethio telecoms usage of Big Data Analytics/big data governance

1. Has Ethio telecom adopted Big Data Analytics/governance?

Yes No

2. Which of these Data Analytics technologies/software have been adopted by Ethio telecom you may choose more than one item.

Relational Data Bases e.g. (ORACLE, DB2, etc.) Business Intelligence Tools ,
 Hadoop/ Map Reduce In-Memory Platforms Other Open Source
 Technologies NoSQL databases Don’t Know/Unsure Others Specify

3. Specify the total amount of data managed by Ethio telecom

Less than 100 Gigabytes (GB) 100 GB to 1 Terabyte (TB) 1 TB to 5 TB
 5 TB to 10 TB 10 TB to 100 TB 100 TB + others, specify -----

4. Specify which of the following data types are being collected and analysed by Ethio telecom. You may choose more than one item.

Transactional data Social Media data Sensor data Multimedia data []
 Spatial data Archived/historical data Mobile data Voice data others,
 specify -----

5. The rate of Ethio telecom usage of big data /data Analytics is very high

Yes No

6. The following are some of the challenges of using big data analytics technology. Please tick [✓] in the appropriate boxes to indicate the extent to which the challenge has hindered the usage of big data analytics in Ethio telecom

Challenges	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1) Complexity of the amount of data in terms of volume, variety and velocity					
2) Hosting and maintaining the technological infrastructure					
3) Lack of capacity to store non-traditional data types					
4) It is too costly					
5) Internal resistance to adopting BDA as a new way of doing business					
6) Lack of skill					
7) Data security and privacy laws					

Part two. Instruction: Please select your answers by putting (X) sign on the scale ranging strongly Agree through strongly Disagree in the appropriate space provided.

Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Strategy					
1) Ethio telecom has a clear mission, vision about implementing Big Data governance framework					
2) Ethio telecom policies are in place to promote and manage use of Big Data Governance in the organization					
3) Ethio telecom has a direction, towards Big Data Governance					
4) Ethio telecom has a plan towards Big Data Governance					
5) Ethio telecom has future development plan for basic ICT and information infrastructure to implement Big Data Governance					
6) Ethio telecom implements strategic plan related to ICT provisioning/ supply and support					
7) Ethio telecom has future development plan and respond towards the dynamic ICT advancement					
8) Ethio telecom has ICT HR strategy related to hiring experienced and professional experts' strength the efficiency of the organization					

9) In order to compete Ethio Telecom in the marketplace, strategically necessary to implement /practice big data governance					
Technology					
1) Relative advantage : Using big data analytics/big data governance has provided many benefits to the firm/organization					
2) Compatibility: The firm has the right technical skills to implement big data analytics technology/big data governance framework					
3) Compatibility: The firm has the right infrastructure to implement big data analytics technology/big data governance framework					
4) Compatibility: The firm has the right IT platform to implement big data analytics technology/big data governance framework					
5) Complexity : It is easy to implement and use big data analytics technology/big data governance framework					
6) the existing ICT tools and technologies would help in storing and analyzing Big Data a better way					
7) The software required for the building Big Data Governance is readily available					
8) The hardware required for the building Big Data Governance is readily available					

9) Triability: Big data analytics has been tried and accepted by the firm/organization					
10) Observability: Many firms/organizations are already using big data analytics/big data governance to drive their business processes					
People					
1) The level of training and education for ICT and Big Data in Ethio telecom is adequate.					
2) Ethio telecom has programs in place to train the users for Big Data Governance/Management/analytics					
3) Ethio telecom employees have basic knowledge about Big Data Governance/Management/analytics					
4) Ethio telecom IT employees have the ability to quickly learn and apply new technologies					
5) Ethio telecom has technical and managerial skills on the use of technological innovation					
6) Previous organizational change has been accepted by the majority of employees					
Environment					
1) Adoption of big data will give Ethio telecom a competitive advantage					
2) Ethio telecom closely follows the competitor's big data initiatives to determine its strength and weakness					

3) The telecommunications infrastructure is reliable and efficient in Ethio telecom					
4) There are firms in the market who provides support for use of this technology					
5) It is not difficult for Ethio telecom to protect data privacy for big data initiatives					
6) It is easy for Ethio telecom to comply with privacy related regulation to implement big data analytics technology/ big data governance framework					
7) Given the necessary resources, opportunities and knowledge to use big data, Ethio telecom will be ready to adopt it					

APPENDIX C: SEMI-STRUCTURED INTERVIEW GUIDE

Dear Sir/Madam:

I am conducting a research which aims to develop a big data governance (BDG) framework for the Ethio Telecom (ET), as part of the partial fulfilment of the Master's Degree in Information System, at Addis Ababa University.

This interview is engineered to obtain the necessary information on BDG practices and factors that determine the effectiveness of data management in ET, in order to develop a BDG framework that is appropriate for the organization. Accordingly, I would like to ask some questions about Ethio Telecoms big data. General information will be shared with the university.

This interview is strictly confidential. No personally identifiable data will be obtained, and all data will be analyzed and reported in aggregate. None of the queries raised will be used to identify you or your unit. Your responses will be stored private and confidential and will only be used for the purposes of this study.

This interview will last approximately 1:30-2:00 hours. Therefore, you are kindly requested to share as accurate and up-to-date information as possible.

A. Personal information

- 1) Could you please describe your academic credentials, role within Ethio telecom, and work experience?

B. Assessment of existing big data governance practice

- 1) Is there big data in Ethio telecom? How to characterize data in Ethio telecom? Explain it in terms of volume, velocity and variety
- 2) What kind of data is available in Ethio Telecom market? Example:-
 - 13.1 Real time /batch/ streaming
 - 13.2 User call data, network traffic information, etc.
- 3) What are the big data analytics application areas in Ethio telecom?
- 4) In what way does Ethio Telecom analyse data?
 - 2.1 Would you explain Ethio telecom experience in using, accessing, storing, capturing, ingesting, processing, analysing, visualizing techniques of big data and the availability of existing ICT tools?
 - 2.2 Did you believe the tools/ technology currently used for data analytics/management are sufficient for yielding the big data utilization benefits?
- 5) How do you describe the level of understanding regarding big data governance and how big data governance is considered within Ethio telcom?
- 6) How is Ethio telecom engaging in big data governance/data related activities? Who is in charge of BDG related task?
- 7) Why big data governance is important for Ethio telecom?
- 8) Did Ethio telecom employees feel it is a strategic necessity to use big data technology/big data governance framework to compete in the market place?
- 9) Do you believe adoption of big data technology gives Ethio telecom a competitive advantage? If yes, how?
- 10) Will big data technology be compatible with Ethio telecom existing operating practices and infrastructure?
- 11) Did the staff have a basic knowledge of big data technology?
- 12) Did Ethio telecom have programs in place to train the staffs on big data technology?
- 13) Does Ethio telecom has a BDG policy, procedure, strategy, technology, guideline or any kind of formal document that governs data in the organization? If the answer is yes

13.1 What tools/strategy/methods/system/framework are used for the management of data in Ethio telecom?

13.2 Did you believe the tools/system/strategy and approaches currently used for data governance/management are sufficient for yielding the big data utilization benefits?

C. Variables that affect big data governance/management

- 1) What are the main roadblocks in effective implementation of big data analytics project /governance strategy and big data management in Ethio telecom?

D. Specifications/Requirements for the proposed big data governance framework

- 1) If you have developed a big data governance framework (BDGF) in Ethio telecom, what kind of standard or framework the organization is employed in the process of BDGF development? If you have not developed BDGF, What are the challenges to design big data governance system framework in Ethio telecom?
- 2) What would you suggest if you were asked to recommend a critical requirement that must be covered in the practice of BDGF in Ethio telecom?

E. Conclusion

- 1) Is there something else that you might want to remark on that I haven't effectively gotten some information about?

APPENDIX D: EVALUATION CRITERIA FOR THE PROPOSED BDGF

Dear Sir or Madam:

In partial fulfillment of the requirements for the Degree of Master of Science in Information Science, I am undertaking research on a big data governance framework for Ethio telecom at Addis Ababa University. Based on the study findings, I have amended the proposed framework and accordingly prepared this questionnaire. The objective of the questionnaire is to evaluate the proposed framework concerning its comprehensiveness, clarity, completeness, correctness, and applicability.

This research is believed to produce results that can improve Ethio telecom data management tasks. Thank you for your dedication to providing your genuine feedback regarding the proposed framework.

Thank you again!

Bereket Shewangizaw

Please select your answers by putting (X) sign on the scale ranging strongly Agree through strongly Disagree in the appropriate space provided.

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
General					
The proposed framework is comprehensive in terms of coverage					
The organization and presentation of the framework is suitable for Ethio telecom					
The objective of the framework is clear					
The content of the proposed framework is complete					
Regarding the content of the Framework					
The content of the proposed framework is relevant					
The content of the proposed framework is clear.					
The content of the proposed framework is Scalable					
Regarding utility and applicability of the framework					
The proposed framework is easy to be applicable.					
The applicability of the proposed framework can improve resource utilization, scalability, data management and yield data management benefits					
The implementation of the proposed framework fits with the organization problems					