AN ANALYSIS OF DATA GOVERNANCE IMPLICATIONS ON BIG DATA

by

LOMSO TROM

Thesis in fulfilment of the requirements for the degree

Master of Technology: Information Technology

in the Faculty of Informatics and design

at the Cape Peninsula University of Technology

Supervisor: Prof JC Cronje

District Six Campus

September 2023

DECLARATION

I, Lomso Trom, declare that the contents of this dissertation/thesis represent my own unaided work, and that the dissertation/thesis has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.

08/09/2023

Signed   **Date**

ABSTRACT

This thesis analyses the data governance implications on Big Data, the probable challenges and prospects that Big Data adoption may present in various organizations. When the Big Data technologies landed, it came with a big buzz and disruption in both the public and private sectors. Big Data technologies have massively enhanced the ability of organizations to processes large quantities of data, analyse and visualize it in order to make sound business decisions. Globally, organizations have been able to identify the prospects of adopting Big Data and some have acted swiftly in formulating plans and strategies to build and develop their Big Data ecosystems.

Organizations that seek to leverage Big Data have been traversing different strategies whereas paying minimal or no attention towards the revision of the processes, methods and policies for effective governance of this gigantic data. The lack of attention towards the governance aspect of Big Data means that Big Data governance continues to be elephant in the room for various organization.

To gain an in-depth understanding, the researcher applied the multiple case study method to respond to the research questions and sub-questions. Semi- structured interviews were used to collect the relevant data and a greater emphasis on the role, experience and expertise was made to select the research participants. Furthermore, thematic analysis was used to identify, analyse and report themes emerging from the data.

This study provides an extensive analysis of data governance implications in Big Data and that is a massive contribution towards broadening the understanding of Big Data governance and the value it carries. The different existing views of the data governance concept in the literature also affords a greater insight into the governance dynamics that organization may encounter when adopting Big Data.

The findings of the study revealed that some organizations are lagging behind and struggling to deal with the governance complications that are presented by the adoption of Big Data and its technologies. The findings also revealed that having Big Data specific governing processes, method and policies remains a major challenge for some organizations. Characteristics such as large volume, high variety and velocity of Big Data demands a different approach to a range of data governance aspects such as data quality, security and policies.

Key words: data governance; governance; Big Data; traditional data; organization; implications.

ACKNOWLEDGEMENTS

Firstly, I would like to thank UQamata nezinyanya zamaMpemvu for granting me the desired strength, courage and the wisdom for this academic journey.

Furthermore, I would like to sincerely thank the following:

- Professor Johannes Christoffel Cronje, my God given supervisor, for his extended support and guidance from the beginning to the end this research journey. I am gratefully indebted to him for his immense contribution on this thesis.

- To my late mother whom I lost at a later stage of my research: Nongatsho Minah Trom, for your constant love, motivation, undivided support, continuous encouragements and persistent prayers throughout this journey.

- My daughters Unako, Melonisa and Imani for affording me time away from them and allowing me to complete my studies without complaining.

- I must express my very profound gratitude to my partner for caring, tolerating me when I could not make time and cheering me up when I was losing faith.

- I would like to acknowledge the support given to me by my former organization, manager and colleagues for all the patience and support throughout this period.

- Various organizations and staff for their maximum participation and the expertise they brought to this research.

DEDICATION

I dedicate this thesis to all those I have lost in the past years, namely: Nongatsho Minah Trom, Khethani Trom, Ndileka Trom, Thembeni 'Ngwenya' Trom, Thembinkosi 'Manoni' Trom, Lizwe 'Country' Trom, Bhabha Trom. I am certain that they would be so proud to see me complete this piece of work.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVATIONS

ACCA   Association of Chartered Certified Accountants

Abbreviation	Meaning
ACCA	Association of Chartered Certified Accountants
CCTV	Closed Circuit Television
CGMA	Chartered Global Management Accountant
CRM	Customer Relationship Management
DAMA	Data Management Association
DBMS	Database Management System
DML	Data Management Language
EDM	Enterprise Data Management
ENISA	European Union Agency for Cybersecurity
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GSMA	Global System for Mobile Communications
IBM	International Business Machines Corporation
IDC	International Data Corporation
IMA	Institute of Management Accountants
IT	Information Technology
ITGI	Information Technology Governance Institute
MIT	Massachusetts Institute of Technology
OECD	Organisation for Economic Cooperation and Development
POPI	Protection of Personal Information
POPIA	Protection of Personal Information Act
RFID	Radiofrequency Identification
SQL	Structured Query Language
TDWI	Transforming Data with Intelligence
WSNs	Wireless Sensor Networks

1. CHAPTER ONE: INTRODUCTION

1.1 Introduction

Big Data has unique characteristics such as large volume, high variety and high velocity, these characteristics are bound to pose new types of risks and dynamics that demand a comprehensive response by the organizations who wish to profusely exploit Big Data while managing the risks. These risks do vary from organization to organization and they may be related to various aspects such as data management, data governance, data architecture, data infrastructure, data usage, data quality management, data protection, privacy management and the skills.

According to Egan (2011) having access to the right data at the right time to make the right decision is not merely a business benefit, but rather a defining factor for success or failure of an organization. Organizations are investing more resources in efforts to explore various methods of fully leveraging big data technologies while resolving challenges faced by the business. Ricci and Kalyva (2015) claim that there is a rapid increase in the number of companies currently using big data to improve decision making, gain a deeper insight of their customer needs, for value enhancement and to achieve or maintain a competitive edge.

This research study strives to identify and analyse the rising implications of governing Big Data in various South African organizations. There is substantial deficiency of research in Big Data governance and this study will add value towards enhancing Big Data governance. This study investigates the dynamics associated with governing Big Data, how each of them is implicated and further provides recommendations on how to manage these dynamics.

LaValle, Lesser, Shockley, Hopkins, and Kruschwitz (2011) maintain that in the past years, organizations were highly interested in only understanding what is happening and the reasons behind its occurrence. This has shifted in the recent years and companies are now striving to gain an insight on what lies ahead or how the future will look like. The advent of big data has noticeably shifted how organizations conduct their business and utilize the data at their disposal (Hurwitz, 2013).

Global enterprises, small organizations and individuals use and store large amount of data and thus the scope of data is growing at a rapid pace. The data types and sources continue to expand and thus more challenges arise (Mathes, 2016). The recent trend, which is the adoption of Big Data is expected to further rise and is associated with the fact that massive amounts of data can be produced and stored at a very low cost (Fan, Han and Liu, 2014).

The size magnitude of data has grown to be so large that it becomes impossible for companies to manage it effectively and use it through the traditional data management tools (Abhyankar and Powar, 2016). Chen, Chiang and Storey (2012) confer that in the age of Big Data, companies are expected to unearth innovative technologies with the capability to process data that is high in velocity in order to yield business intelligence.

Chen et al. (2012) further affirms that Big Data technologies adoption in organizations tends to shift the data focus from being internal to external and more customer centric. Due to this shift, there is a growing concern of data privacy. Large amounts of data are obtainable and ready for collection, nonetheless, open access to consumer data may result to data privacy complications (Ricci and Kalyva, 2015).

Even though Big Data collection and use may yield an abundance of benefits for companies, citizens and the society, it demands extreme precautions, complete control and measures to safeguard privacy and human rights. These measures spell a massive shift in how organizations govern Big Data throughout all the phases of its lifecycle (Broeders, Schrijvers, van der Sloot, van Brakel, de Hoog & Hirsch Ballin, 2017).

1.2   Background of the Study

Tesfaye (2017) points out that certain Big Data characteristics such as (veracity, volume, variety, velocity, and value) present a different range of risks which are owed to poor management, governance, architecture, usage, quality, privacy and security.

A rapid increase in data exposes the governance shortfalls, infrastructure and organisational Big Data expertise required to use this data effectively. Big Data is still an emerging technology and that means a skills gap does exist due to a shortage of Big Data experts. Hagen, Evans, Thota, Wall, Seshadri and Khan (2014) claim that this leads to the loss of Big Data value altogether.

Luna, Mayan, García, Almerares, and Househ (2014) assert that it is common for organizations to face issues such as exponential data growth, unique infrastructure challenges, scarcity of trained personnel, interoperability shortfalls, privacy and security, the need for new processes, and policies but developing countries face the most difficulties on their Big Data implementation programs.

Data that is considered Big Data at this hour may not qualify to be called Big Data tomorrow due to the rapid increase in storage capacities, meaning that even bigger data sets might be collected (Gandomi and Haider, 2015). A question that needs to be addressed is "Will there be effective governance to oversee the ever growing volumes?"

Part of good governance is to guarantee data privacy and security compliance which is also good for business. According to Raschke, Krishen and Kachroo (2014), customer concerns for data privacy include the collection, unapproved use, inappropriate access, and misinterpretation. Höchtl, Parycek and Schöllhammer (2016) also mention that there is a thin line between the opportunities that come with Big Data and the potential violation of the right to privacy.

To gain trust from their consumers, organizations are now required to demonstrate good governance and stewardship. Hull (2015) claims that people have little or no information of how their data is used and that is an ethical concern. Among other issues, the lack of comprehensive policies and regulations to manage Big Data remains a challenge that organizations must deal with.

The focal point of this paper is on the emerging governance implications on Big Data. Notwithstanding the exponential growth of the data scope, the volume and velocity of Big Data poses significant challenges on the methods of governance, policies, regulations and roles. It is apparent that these challenges call for the need for novel approaches, policies, controls and methods that are tailored for Big Data.

Organizations are working tirelessly to be on the leading pack of Big Data implementations while minimal attention is given to the governance dynamics and risks carried by the new technology.

1.3   Research problem

Globally, most organizations have always regarded data as one of the most significant assets, however in the recent years, its significance has risen tremendously. Companies operating in several business sectors are making great strides towards the exploitation of the Big Data asset to gain sustainable competitive edge while implementing effective governance of data to alleviate the risks that accompany the benefits of this data (Gunasekaran et al., 2017).

Currently, many organizations and governments collect massive amounts of both structured and unstructured data and this is regarded as 'Big Data'. Data may present a lot of opportunities but it also presents several challenges (Mathur and Purohit, 2017). Unlike traditional data, Big Data comprises of unstructured data that is gathered from various sources and many organizations are not fully prepared to govern this influx of unstructured data.

Notwithstanding the fact that data is regarded as an asset, poor governance of the environment, lack of policies and controls may result to the organization

being prone to technical, legal, financial or reputational risks. The traditional data governance approach is custom-made for structured data and this implies that it is inappropriate for data with no structure (Gandomi and Haider, 2015).

In the organization's quest to implement Big Data solutions, the development of a solid Big Data governance framework must be made a priority (Myers, 2012). According to Peco (2018) data governance needs to be revised and transformed into data value driven governance, with a far-reaching focus than the traditional way of governance. The arrival of disruptive technologies such as Big Data has intensified the need for a shift towards new and innovative ways of governance.

With the increasing adoption of Big Data in various organizations, traditional data governance methods, processes and strategies are not effective to govern data with a large volume, high velocity and variety. It is therefore imperative to gain an understanding of how organizations are approaching Big Data governance and the complications of governing this data.

1.4   Aim and Objectives

The main aim of the study is to analyse the governance implications presented by Big Data on data governance.

The objectives are:

- To identify governance strategies, processes and methods that are currently employed by the organizations to govern Big Data
- To investigate the challenges encountered by the organizations in governing Big Data
- To analyse how organizations approach Big Data governance challenges
- To determine how the adoption of Big Data has affected traditional data governance

1.5   Research Question

What data governance implications are presented by the adoption of Big Data?

1.5.1   Sub-Questions

- What governance strategies, processes and methods are currently employed by the organizations to govern Big Data?
- What are the perceived Big Data governance challenges encountered by the organizations?
- How do organizations deal with the Big Data governance challenges?
- What major effects has the adoption of Big Data had on traditional data governance?

1.6 Significance of the study

This is an era where Big Data is extremely imperative to business growth and decision making. The hasty and exponential growth in data volume, variety and velocity presents both opportunities and challenges. To encounter these challenges, it is vitally important for organizations to seek a deeper insight of the governance implications and challenges.

In the past few years, a limited number of studies have explored traditional data governance and other elements of Big Data. Despite contributing a significant degree of insights, none of these studies have been able to deliberate on the governance implications in a Big Data environment. We can therefore agree that by outlining the Big Data governance dynamics and implications, a study of this nature will indeed make an immense contribution.

1.7 Key Concepts

Term	Clarification
Governance	Governance is the framework of authority and accountability that defines and controls the outputs, outcomes and benefits from projects, programs and portfolios (Murray-Webster & Dalcher, 2019).
Data Governance	Data governance is the overall management of the availability, usability, integrity, and security of the data used in an organization. In conclusion, data governance refers to how an enterprise is organized to handle its data and extract value out of it (Zhang & Yuan,2016).
Big Data	The term Big Data describes a data environment in which scalable architectures support the requirements of analytical and other applications which process, with high velocity, high volume data which may have a variety of data formats, and which may include high velocity data acquisition (Emmanuel & Stanier, 2016).
Data Management	Data Management (DM) – the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets (DAMA International, 2017).
Corporate Governance	Corporate governance is defined as the process and structure used to direct and manage the business and affairs of the company towards promoting business prosperity and corporate accountability with the ultimate objective of realizing long-term shareholder value while considering the interest of other stakeholders (Zinkin, 2019)
IT Governance	IT governance is the process by which decisions are made around IT investments. How decisions are made, who makes the decisions, who is held accountable, and how the results of

Term	Clarification
	decisions are measured and monitored are all parts of IT governance (Symons, 2005).
Regulatory compliance	Regulatory compliance is the act of keeping an organization in line with all international, federal, state, and local laws and regulations that apply (Mackenzie, 2020).

Table 1 :Clarification of key terms

1.8 Methodological consideration

This section provides a synopsis of the methodological considerations that were made throughout the study. In Chapter 3, a clear outline and detailed discussion of the methodologies that guided this study are presented.

After reviewing several philosophical assumptions, the study adopted the interpretivist paradigm. This study seeks to understand the governance implications presented by Big Data on data governance, which means the interpretivist paradigm is appropriate for the researcher to gain an insight on the experiences of the participants. As illuminated by Ponelis (2015) the interpretive paradigm is driven by a need to acquire an understanding of the world as it is from a subjective point of view and seeks that understating within the same frame of reference of the participant rather than that of an observer. Furthermore, the interpretivist research paradigm is crucial when seeking to understand the different dynamics that influence people's behaviours and involvement in specific things. Thanh and Thanh (2015) posit that interpretivist paradigm employs interviews, observations and analysis of existing literature to gain a meaningful reality.

This study used qualitative research method to address the main research question. Qualitative research seeks to explore and discover issues around the research topic. Since less is known about the topic, normally there is uncertainty about dimensions and characteristics of the problem (Domegan and Fleming, 2007). Creswell (2009) posits that qualitative research is conducted with an aim of gaining an understanding about the context or setting in which participants of a study address a complicated issue.

This study will combine findings from 3 case studies, as suggested by Baxter and Jack (2008) a case study is a research methodology that is most appropriate when intending to explore a phenomenon within a specific context using a variety of data sources. These case studies will comprise of semi-structured interviews with executive managers, Heads of Departments, Big Data Scientists and specialists and some observations by the researcher.

For case selection, the researcher used purposive non-probability sampling as the sampling method for data collection. According to Lewis and Sheppard

(2006) purposive sampling is a non-random sampling technique that affords the researcher the deliberate choice of participants by identifying certain qualities that a participant possesses. The research participants were selected according to their knowledge, experience, role and involvement in the governance of Big Data technologies. In total, 16 employees were selected as participants on the study.

As mentioned above, the data was collected through semi-structured interviews and the aim was to explore perceptions and understanding of the participants regarding Big Data governance. (Creswell, 2014) claims that one of the advantages of using semi-structured interview in a study is to allow the researcher to acquire in-depth insight and provide further explanation of the questions to the participants when necessary.

The qualitative data obtained from the case studies will be analysed using thematic analyses. According to Alhojailan (2012) thematic analysis is considered the most appropriate method for a study with an aim of making discoveries using themes. Furthermore, Alhojailan (2012) confers that thematic analysis offers a systematic element to the process of data analysis and it also provides the researcher with an advantage of associating the theme's frequency analysis with one of the complete contents.

This study will obtain themes from the semi-structured interviews and follow the six phases of conducting thematic analysis as guided by Braun and Clarke (2006) to pinpoint, examining, and recording themes or patterns within collected data.

1.9   Delineation of the research

This research study used three South African organizations operating in sectors such as telecommunication, financial service and insurance as cases. Moreover, the scope of the study was limited to companies that have implemented and are actively using Big Data technologies. Hence, they offer a relevant case for understanding implications of Big Data governance.

1.10 Outline of this study

This research paper seeks to report on the governance implications that are presented by the introduction of Big Data in various South African organizations. The nature of this study is qualitative in nature and data was collected through semi-structured interviews with various subject matter experts on data governance in order to gain a deeper insight on the implications of governing data in their respective Big Data environments. The interviews were recorded, then transcribed and later analysed using of thematic analysis as a method of analysis. The findings of this study suggest that it is necessary

for organizations to explore strategies, methods and processes to specifically govern Big Data while keeping up with the relevant laws and regulations.

Chapter two reviews existing literature in relation to the governance of Big Data and its implications. It further discusses what the literature about the characteristics of Big Data, its sources and the challenges of governing it.

Chapter three reports on the choice of research methodology, design, approach and philosophy employed in order to achieve the objectives of this study. For this study, a case study design with a qualitative research method was employed, and data was entirely collected using semi-structured interviews. Furthermore, the section gives a justification of what informed the choice of this methodology and sample and why is it relevant for this study.

Chapter four presents the results acquired from the semi-structured interviews that were conducted as means to collect data. This chapter further reports on how the research data was analysed through thematic analysis.

Chapter five presents and a discusses the findings on the implications of Big Data adoption on data governance. The study is concluded and recommendations are provided in relation to the key findings.

1.11 Summary of Chapter 1

This chapter gave an extensive introduction and background to the research problem and the topic of the study. The chapter further provided the research aims and objectives, research questions, sub questions, delineation of the study, methodological considerations and finally the significance of the study. The chapter concludes with the layout of the thesis, providing clear expectations for each section of the study.

The following chapter is Chapter 2, and it provides a detailed discussion existing literature that is relevant for this research.

2. CHAPTER TWO: LITERATURE REVIEW

The background for the term 'Big Data' is given, several connotations of the term from various authors are discussed in Section 2.5. A brief overview of several Big Data sources and how they are related is provided in Section 2.6. Section 2.7 outlines how Big Data has evolved over the years and the recent developments in the Big Data phenomenon. Section 2.8 discusses data governance in depth, explores different data governance methods and how data governance can be best implemented. Section 2.9 discusses the concept of data management, provides a detailed explanation of the process, its tasks and phases involved. An in-depth discussion of various challenges of governing data and the potential impact is found in Section 2.10. Section 2.11 outlines the theoretical frameworks used to guide this research.

2.1 Governance overview

The governance phenomenon has grown very rapidly and the literature on the phenomenon has developed at a very high rate (Aminuzzaman, 2010). Governance, either good governance or bad governance like other several buzzwords in the space of international development, the term means different things to different groups (World Economic Forum, 2016). There is no single accepted definition of governance, different people define it in different ways depending on their understanding and conception.

Adelaide (2015) defines governance as the framework of rules, their relationships, systems and processes to control and hold the enterprise accountable. It comprises of authority, accountability, stewardship, leadership, direction and control exercised in any organisation. In its present-day use, the word 'governance' is used to describe a new process of governing or an improved condition of ordered rule or the new manner by which society is governed (Rhodes, 2007).

Bruck (2017) suggests that the term "Governance" be clearly separated from "management" as they have two different definitions. Bruck (2017) further confers that governance focuses on high level decision that are very critical for organizations that seek to gain effective management and alignment of its resources with the company strategy.

Advanta Healthcare Partners (2014) states that governance is a method used by organizations when developing strategies, identifying goals and making critical business decisions. Advanta Healthcare Partners (2014) further claims that from one organization to another, the composition of a governance structure can differ and that includes the board of directors or an organizational structure.

Even though the responsibility resides with the board to decide and implement a governance structure, a major part of the daily activities of an organization are executed not only by board members, but also by executive management, staff members, and designated committees. This echoes the views of Advanta Healthcare Partners (2014) when stating that there is no single and standard governance structure for all organizations, hence it is always anticipated that over time the governance structure or model will be altered.

According to Oghojafor, Olayemi, Okonji and Okoli (2010) the governance structure stipulates the delegation of authority with the duties and responsibilities of all the individuals on the board of an organization. In order for the board and executive management to achieve their governance objectives, a clear governance model is needed. This model becomes a critical factor in the management's pursuit to implement an effective governance structure (Deloitte, 2013).

Governance is at the centre of any successful business or state. It is essential for any organisation that seeks to achieve its objectives, improve continuously, and maintain legal and ethical reputation in the wider community. It is not only large organizations that should be concerned about governance, but small businesses or organisations of any shape or size should be concerned too (National Computing Centre, 2005).

This growing emphasis on the importance of governance in general has led to this study analysing the governance processes, strategies and methods that are put in place to govern Big Data.

2.2   Corporate Governance

The popularity of the term corporate governance has grown in recent years from various fraternities. Its usage and definitions is loaded with different perceptions and understanding on the subjects.

King IV (2019) defines corporate governance as an effective leadership and ethical exercise performed by the governing body of the organization with an objective of achieving the following governance goals:

- Ethical Culture
- Good Performance
- Effective Control
- Legitimacy

According to the Organisation for Economic Co-operation and Development (2004) corporate governance determines the structure from which the organizational objectives are set and the efforts of achieving these objectives and monitoring performance are defined.

Wondra (2017) refers to corporate governance as a system that is used by the executive committee to steer and control the organization.

Wondra (2017) further states that for the organization to operate effectively both corporate boards and executive committees have a very critical role in achieving good governance through the development and implementation of the relevant policies.

Valenti, Luce and Mayfield (2011), proponents of corporate governance, assert that for an organization to witness an improvement in performance, much focus should be channelled towards the formulation of essential governance policies. According to the report by Gompers, Ishi and Metrick (2003), organizations that are well governed do register an improved market valuation with minimal expenditure.

Ammann, Oesch and Schmid (2011) also state that institutions with effective governing structures or models have relatively higher statistical and economic market values. When properly implemented, corporate governance generally drives a proficient monitoring process that is needed for an organization to better leverage its key resources while controlling the risk of possible data manipulation.

The cost of equity capital tends to drop when an organization is governed effectively governance and it also decreases for companies that are more dedicated towards practising business in a more ethical manner. This therefore serves as a recommendation to companies that seek to decrease their cost of equity whilst growing value by fully committing to effectively govern and direct its efforts towards achieving the highest standards of ethics (Pae and Choi, 2011).

Apart from the economic benefits or gains, a well-governed organization is guaranteed a transparent and easily trusted corporate setting and on top of that, the organization can be held accountable for its results and decisions. Nonetheless, a failure to govern an institution in an appropriate manner could result to a devastating situation that can severely impact the company reputation, the resources and performance (Youssef, 2010).

Corporate Governance is the larger whole of Data Governance that is executed by the organization's board. Corporate Governance entails overseeing the organization's activities from a holistic view and more often poor corporate governance may suggest poor Data Governance. In this study the implications of governing Big Data are investigated and in order to achieve the objectives of the study an in-depth understanding of how organizations are governed at all levels is vital before understanding how to govern data. Figure 1 below illustrates the relationship between the various elements of governance.
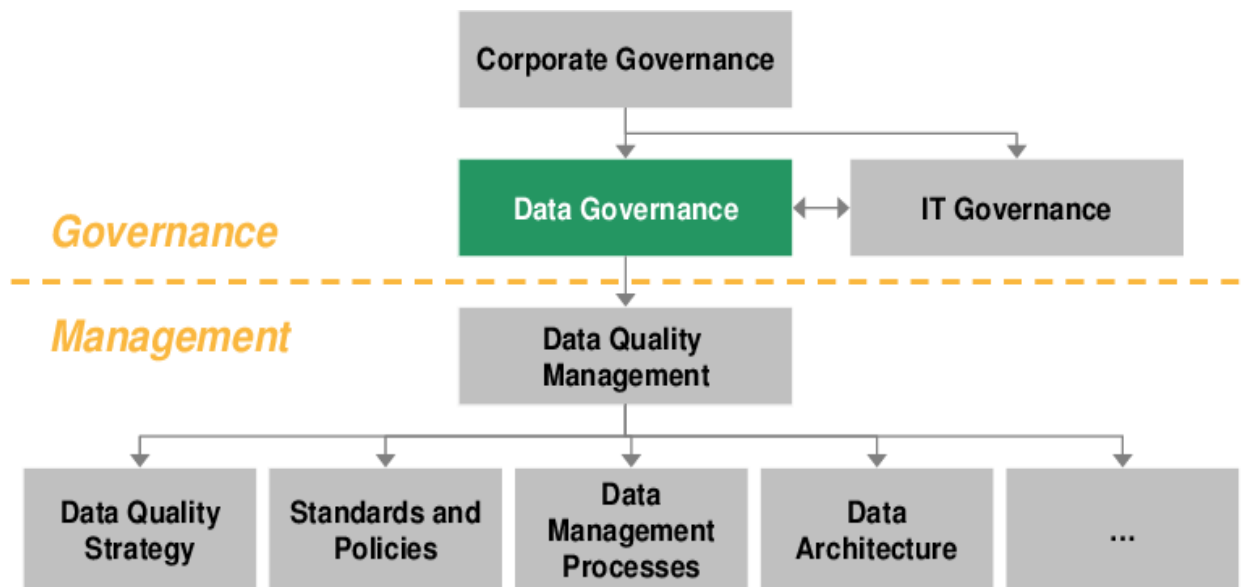
Figure 1 : Scope of Governance *(Source: Wende 2007)*

As mentioned above, the achievement of good data governance demands a great deal of synergy between corporate, data and IT leaders. It is vitally important that business leaders possess a substantial understanding of the data, the value it brings to the business and the risks associated with it.

2.3 IT Governance

Information technology has become a very fundamental part of modern organizations. Over the years, IT has changed dramatically, and it is one of the global economic growth's driving forces as it enables enterprises to be more productive, thereby encouraging growth. Effective governance of IT is vital for the organization's efforts to support business strategy and create shareholder value (Tiglias, 2014).

The Information Technology Governance Institute (2006) defines IT Governance as an array of high-level responsibilities and practices that are implemented by the corporate board and executive management with the main intent of providing strategic direction, while assuring that organizational objectives are achieved, ensuring that all the possible risks are managed properly and making sure that the organization's resources are used appropriately.

The National Computing Centre (2005) believes that IT Governance is not a one-moment exercise or something that can be achieved only by issuing instruction and setting of rules. Full commitment is required from the top

management or board of the organisation to instil a culture of dealing with the management and control of IT.

De Haes and Van Grembergen (2009) also points out that IT governance should be applied by the management of both the business and IT with an aim to control the formulation and implementation of IT strategy while ensuring that IT and business are more aligned. According to Spur Corporation (2017) The board of the organization has the responsibility to ensure that IT is governed effectively though the introduction of strategies and plans that clearly interpret how the governance of IT should be approached and conducted in the organisation.

Nolan and McFarlan (2005) acknowledge that corporate boards have a very critical role to play in IT and thus their involvement needs to be promoted in order to rip the benefits of the IT investments. Considering the strategic and operational role of IT, in most organization the boards fail to realize the value and importance of IT including its effect on the organizational strategy.

National Computing Centre (2005) mentions the alignment of business and IT, risk management, resource management and ultimately the delivery of IT value as the main pillars of an effective IT Governance. Governance Business (2019) agrees that IT Governance has the ability to resolve the long-standing business challenge of balancing risk and benefits when using IT to stimulate organizational transformation and support the achievement of the initial business objective.

Neirotti and Paolucci (2007) suggest that a continuous assessment of the relationship between IT governance efforts and the value that is gained by the business is very critical for the business to respond to a need for more robust and consistent link between IT investments and business profitability.

Tiglias (2014) asserts that due to the fast changing and evolving nature of technology, organizations are confronted with the governance related challenges and thus leadership has to ensure that the best decisions are taken regarding investments in IT. IT Governance is a consistent exercise that demands continuous improvement and the failure to govern IT assets effectively has a dire impact on how the data assets are governed. For this study to achieve its objectives a deeper insight of what it means to govern IT assets had to obtained.

2.4   Data as an Enterprise Asset

In the business fraternity data is currently a very precious necessity as it supports businesses to gain a competitive advantage when used effectively. Hicks (1993: 668) defines data as "a representation of facts, concepts or

instructions in a formalised manner suitable for communication, interpretation, or processing by humans or by automatic means".

According to Enofe (2017) the current era is driven by information, data is considered as an essential asset, thus data is the core of the present economy. Data has developed rapidly to an extent of becoming a key element of modern businesses and it has completely transformed the processes of several businesses. Stander (2015) posits that the hasty transformation inspires organizations to exploit their assets such as data to the maximum if they seek to remain ahead of their counterparts.

Data and information are key contributors towards improved performance, processes and productivity of businesses. One of the major resources for improving productivity, developing new technology and improving processes is data and information. GlueReply (2014) concurs that data is a highly valued commodity and many organizations have long realized the value that is carried by their data asset, although they are not doing much to leverage this asset.

The perception of data as a valuable asset is further echoed by the MIT Technology Review, when referring to data as a form of capital that is in the same range as monetary capital when it comes to producing novel digital services and products. In any organization, this kind of development has a dire impact on the business strategy and the relevant computing architecture (MIT Technology Review,2015).

Within the business industry millions of online transactions occur frequently, smart-phone applications increase daily, data is all over, resulting to businesses opting for data-driven decision making (Harris et al., 2010). The concept of data-driven business decisions has drawn a lot of attention from the business and researchers that seek to get an in-depth understanding of its prospects (Anastasopoulos, 2016).

Brynjolfsson, Hitt, and Kim (2011) revealed that institutions that have successfully incorporated data as an integral part of their decision-making processes are likely to grow their total productivity by 6% and they can also improve profitability and financial performance while experiencing growth in market share.

According to Jiang (2017) in the morden era businesses rely heavily on raw data to process and analyse in order to make informed business decisions that have a massive impact on the growth and success of the business. The business space has witnessed a major shift in the recent years as competition in the markets is now extremely intense and requires logical and ground-breaking decisions. The amount of pressure and desire for companies to increase productivity and minimize operational cost has driven businesses towards technology adoption in order to remain competitive.

Data analytics have come to aid in offering the much-needed reprieve to most companies. The technological advancements in analytical tools used by most businesses have been the pertinent determiners of the successes attained by these. In order for this study to understand the urgency behind effective governance of data, it was essential to understand how valuable data is to the success of many organizations.

2.5   Big Data defined

What is the ideal definition of "Big Data"? Going through the literature, one could realize that much has been written about the subject and that includes a variety of definitions and interpretations about Big Data. Big Data is one of the disruptive technologies that have unescapable impact on the business community and that justifies the different opinions and definitions. The composition of these opinions is discussed below:

Daily, enormous data chunks are produced from a variety of media sources such as images, videos, social media interactions, websites, sensors, transaction records, Global Positioning System signals and this is considered as Big Data (Zicari, 2012). Madden (2012) concurs that Big Data is essentially data that is too big, too fast, or too solid for the traditional data tools to process. Laudon and Laudon (2014) define Big Data as massive and complex data sets that cannot be captured and stored in a typical Database Management System.

Laney (2001) states that Big Data is an information technology term which refers to high volume, velocity and variety. This suggests that Big Data is a massive dataset with data that is produced at a very high velocity, in a variety of formats and types. Established organizations such as the International Business Machines Corporation, Google and Facebook have recognized the potential value that is carried by Big Data long ago. They have been leveraging Big Data for quite a while by making use of advance analytics or discovering new business models that are aligned to their Big Data objectives. On the other hand, the Association of Chartered Accountants and the US Institute of Management Accountants (2013) describe Big Data as a combination of complex processes that seek to collect, store and analyse massive volumes of data. Big Data is high in velocity and that means it needs to be captured, transformed and stored instantly. A common denominator for the several definitions becomes the recurring use of terms that describe the magnitude of this data and these are "large", humongous", "massive" and "complex".

Boyd and Crawford (2012) contend that Big Data is not simply an advancement in technology, but it is more of a popular belief that Big Data sets provide a high-level amount of intelligence and that Big Data can now yield the kind of insights that was deemed impossible over the years. Teradata (2019) agrees that organisations are now under enormous pressure and it is not only about

growing their customer base but also to gain a deeper insight of their needs and behaviour in order to deliver most relevant services and therefore improving relationships.

McAfee and Brynjolfsson (2012) assert that ever since the introduction of online shopping, the insight that businesses have about their customers has extremely improved. Actually, Big Data has uncovered new ideas and opportunities for business to gain an in depth understanding of their position in the market and therefore be able to make improvements (Chen, Chiang and Storey, 2012).

In the main, the term has a plethora of definitions and this can be attributed to the technical and sociological nature of the phenomenon. It is quite typical for individuals to use the term when describing the loads of data at their disposal, frequently produced by our digital interactions and lives (Kuurila, 2016).

Various definitions of Big Data inform us that it can never be processed through the use of traditional Structured Query Language queries and neither can it be stored using the relational database management system.

2.5.1 Big Data dimensions

All of these definitions have something similar as they all focus on the definition of the 3Vs (volume, velocity and variety) (see **Figure 2**) to best define the characteristics and technical attributes of Big Data
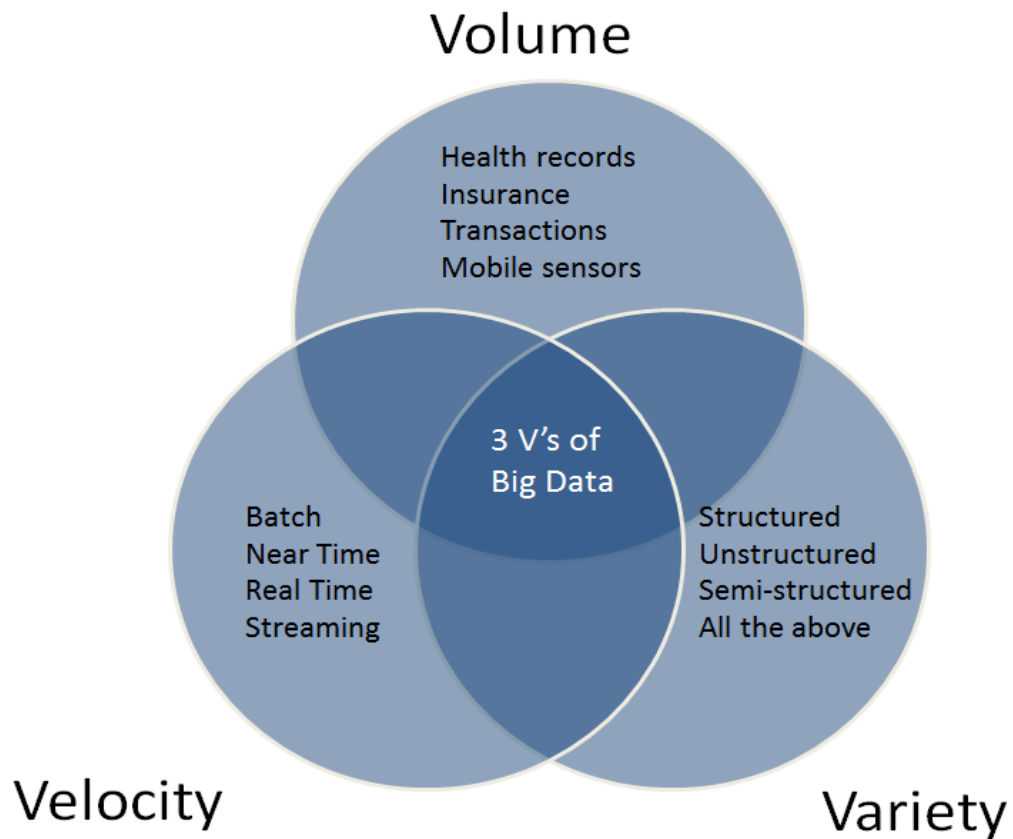
Figure 2: 3 Vs of Big Data *(Source: Austin and Kusumoto 2016)*

- **Volume** – This characteristic represents volume that is growing exponentially. The number of data sources has increased and massive amounts of data are generated.

- **Variety** – Big Data comes from a variety of sources and it is sometimes structured or unstructured depending on the source. Big Data comes in different formats and a diverse data source and thus working with multiple formats of data makes the process of storing and processing the data a complicated one.

- **Velocity** – This refers to the quick flow of data or the pace at which the data is produced. In a data driven business, data flow is expected to be fast and near real time as much as possible in order to make informed decisions.

Various authors have since extended the characteristics and descriptions of Big Data. Vossen (2013) mentions 4Vs of Big Data as depicted in Figure 3. The V's do include the 3 V's in Figure 2 but they have an extra V that stands for veracity. Oracle (2013) concurs by defining Big Data using four V's as well.
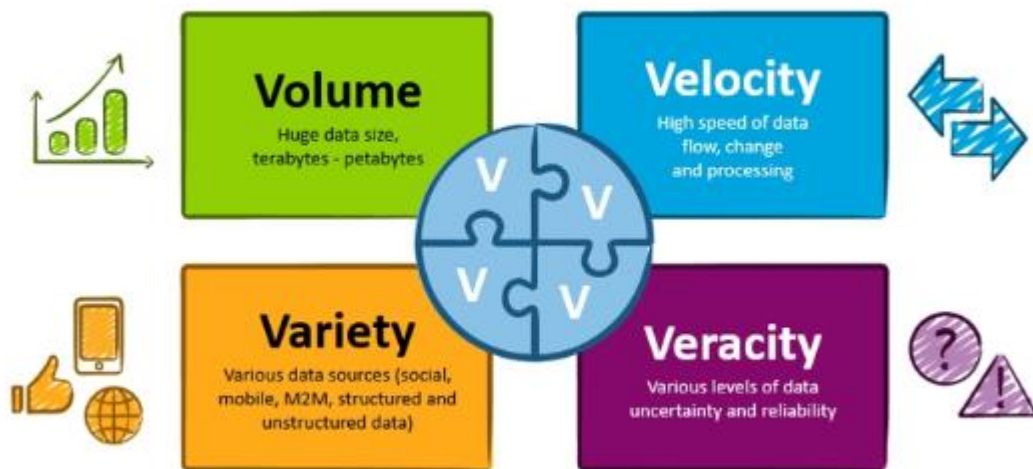
Figure 3: 4V's of Big Data (Source: Austin and Kusumoto 2016)

● **Veracity** – Datafloq (2019) defines veracity in Big Data as the level of accuracy or trustworthiness of the certain data. Veracity refers to the reliability of the data source, its accuracy, and how meaningful is the data for analysis. Lukoianova and Rubin (2014) recommends that veracity be defined using the three main dimensions which are objectivity, truthfulness, and credibility.

Even though the various attributes of Big Data are presenting a plethora of opportunities and a considerable value to business, it also carries several challenges that demand new technological solutions (Khan, Alsaqer, Shah, Badsha, Abbasi & Salehian, 2018). The rapid increase of heterogeneous data has motivated organizations to invest in Big Data clearly because they seek to gain a deeper understanding of their customers, products and business as a whole. The possibility of obtaining these insights is largely influenced by the technology, infrastructure, tools and skills they have at their disposal.

According to Ajah and Nweke (2019), several organizations have been less eager to adopt Big Data due to various concerns such as data accuracy, data consistency, data regulations and access. Some of these organizations do not have the required skills to exploit Big Data and those that tried to jump for the opportunity lacked the ability to translate the outcomes.

2.6 Sources of Big Data

The rapid and exponential growth of the data held by companies and on the internet has been growing due to the increased use of social media, increased access to the internet and the rapid increase of smart connected devices that are sharing data. Cloud Moyo (2019) states that the massive amount of Big Data generated comes from mainly three sources: social media data, machine data and transactional data.

Whitebookof Big Data (2019) concurs that there is a noticeable growth in the number of available data sources, especially in the volume of the social media data. The author further mentions other sources of Big Data such as data collected from cars, GPS data that is generated by smart devices, personal data that is collected by companies and governments and lastly streaming data that is collected for drawing real-time insights. Figure 4 illustrates the various sources of Big Data and how they are measured on velocity, variety and volume.
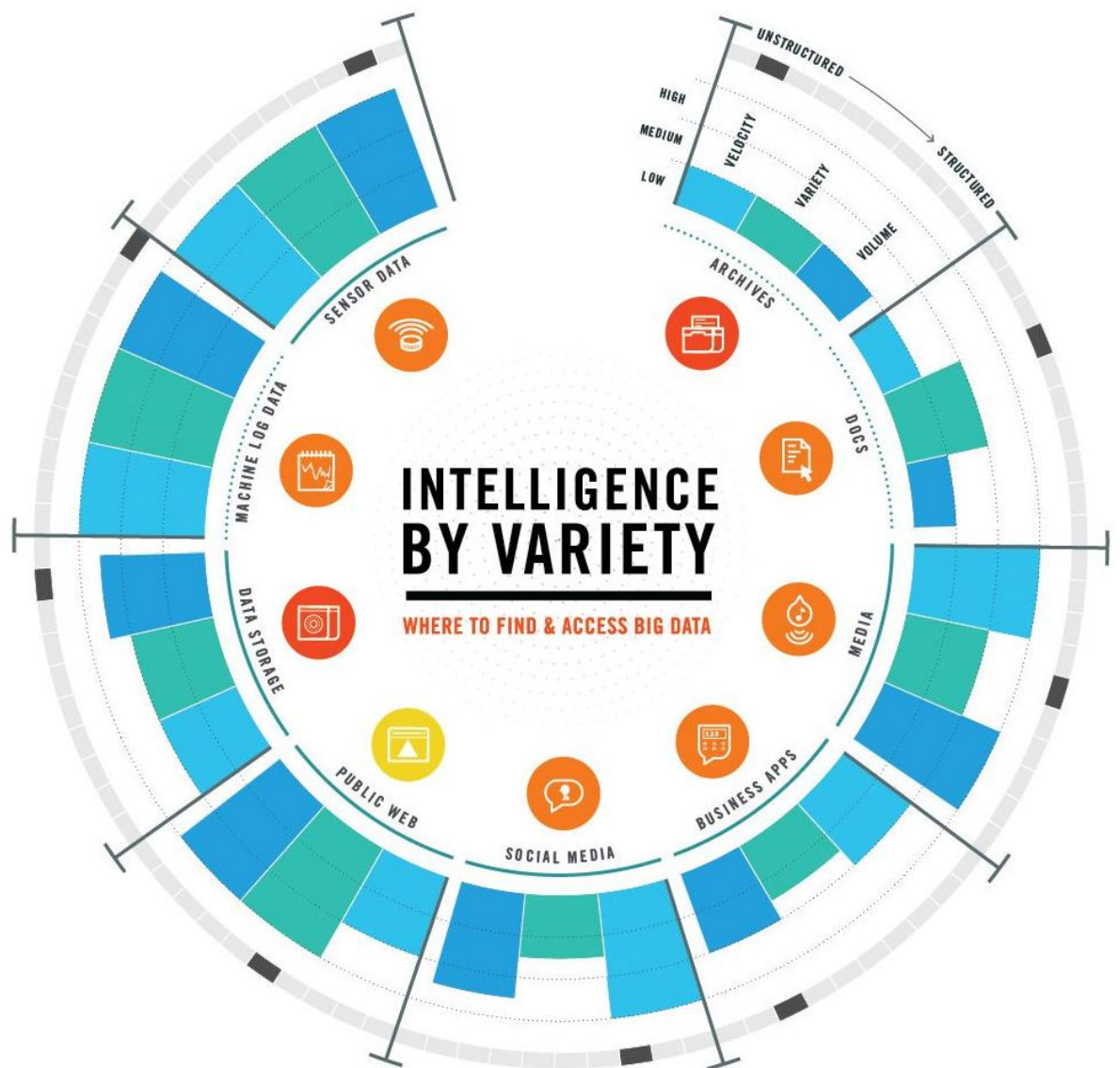
Figure 4: The Big 9 Big Data Sources (Source: Kapow 2014)

There is a major shift in how data is used and this has an effect on how it is collected. The steep rise in the population of internet users now guarantees companies sources of Big Data generation. There is a pool of new data sources and these include smart connected devices, radio-frequency identification tags (RFID), wearable health monitoring equipment, Power analysers, weblogs, meter readings and clickstream (Halwani, 2018). With such a variety of Big Data sources and increasing volumes at a fast pace or velocity, if utilized intelligently organizations can actually realize the concept of the information age. When this data is precisely processed and thoroughly analysed, much desired actionable insight can be obtained (Ali, Qadir, ur Rasool, Sathiaseelan, Zwitter & Crowcroft ,2016).

Agata, Miyata, Ishita, Ikeuchi and Ueda (2014) assert that for many organizations Big Data is considered to be a very valuable asset and we have witnessed its promises starting to come to fulfilment over the years. This can be attributed to the increasing discovery of the various data sources in our exposure and due to the rapid increase to the number of Big Data processing and analysis tools at our disposal.

In addition, many organizations collect data for many reasons but amongst all those reasons increasing revenue, minimizing costs and growing the business remain as common reasons. Below are other reasons why companies gather Big Data:

2.6.1 Sensor Data

Organizations use sensor data to improve the competitiveness of the business and the sustainability of the entire sector. According to Ernst & Young (2014), the automotive industry has been witnessing a huge increase when it comes to the use of sensor data in the recent years. The benefits of leveraging on this is not only receiving new revenue streams but also safety and convenience to users. Forrest (2015) points out that in the automotive industry, sensor data is used to minimize costs, to gain competitive advantage and reduce its environmental impact. Some of these sensors are used to monitor speed, fuel economy, mileage and to assess engine health.

The technological advancements have stimulated a transition from the old traditional processing methods to automated and advance processes. Sensor data has been collected and utilized for decades in other organizations and industries. Different types of sensors have always been used over the years but the disruption that was brought by the Internet of Things has resulted to a major paradigm shift. This development has now given rise to Wireless sensor networks that are becoming more popular in various industries (Zhong, Xu, Klotz, & Newman, 2017).

Li, He, and Fu (2008) define Wireless sensor networks (WSNs) as groups of spatially distributed self-configured sensor nodes. WSNs are used in various sectors of business such as farming and manufacturing to monitor and record physical or environmental conditions, such as temperature, volume, pressure, sound, motion or humidity. According to GSMA (2018) organizations are expected to consider elements such as openness and transparency when collecting data of this calibre. GSMA further states that the concerned individuals deserve to be informed about the data provided with adequate information on the purpose of data collection and how it will be used.

2.6.2 Archives

Big Data does not only consist of sources such as social media, website clicks, electronic transactions and GPS information but it is also collected from unstructured data sources such as digital documents and archives from digital libraries (Martinez-Uribe and Fernandez, 2015). According to Kumar (2018) Most of the data in the healthcare sector is unstructured and this includes archived documents such as patient records, handwritten physician's notes, patient prescriptions and laboratory test results that are not stored electronically.

Kumar (2018) further states that the nature of unstructured data is heterogeneous, varies from each source and it is in various formats. When it comes to growth, unstructured data is currently growing at a very rapid rate compared to the structured data. With the rising popularity of heterogeneous data formats comes explosive growth in volume and challenges in the Big Data sources, but the gain is the much sought-after insight and the ability to uncover business improving facts.

2.6.3 Machine Log Data

The present-day servers and applications generate large amounts of log data to process. This means that the increase in the quantity of servers or applications, the produced log data volume is quickly generated so it grows too big and therefore cannot be analysed manually. Within the large datasets only few records are meaningful, the sought-after meaningful records are not easily identified because there is relatively few of them (Kostjens, 2016).

Machine log data is collected and stored for a certain period of time with an intention of analysing it to detect or identify some trends or patterns on the network, application or system. Motadata (2019) confers that storing data logs is a positive and proactive approach for an organization to gain a holistic overview of systems and applications performance within their IT infrastructure. Motadata (2019) further states that even though logs are an overwhelming and multifarious chunk of data, they are a treasure to the organizations that seeks to gain insightful operational intelligence.

Considering the definitions of Big Data, it is defined as data that have non-traditional processing needs, when trying to explore what Big Data really epitomizes it becomes so obvious that machine log data has always been Big Data though it was not regarded as such (Beedgen, 2012). According to Elgendy and Elragal (2014) large corporates such as Amazon Web Services, IBM and Red Hat have been leveraging on the log data for quite some time, this type of data is very critical in the cybersecurity space. In the

telecommunications industry the quality of management can be improved by monitoring machine log data and performing real time analysis.

2.6.4 Media

Venter and Stein (2012) assert that 80% of the unstructured data in the public and private sector consists of images and videos. Verma, Agrawal, Patel and Patel (2016) concur that 80% of unstructured data is mainly videos and images and video is considered the most challenging source of Big Data due to its large size. Closed-circuit television cameras have become so prevalent and they are now sources of data. All this data is stored, processed and analysed for various reasons. Many organizations have been leveraging on video data for quite a while, using the video analytics to look for the purpose of monitoring, security and motion sensing.

Lipel (2016) claims that media is certainly one source of Big Data that changes more frequently with new discoveries and insights emerging very often, therefore presenting commercial opportunities. Varah (2014) adds that in a Big Data environment, video analysis is performed with an intention to detect specific objects in the scene, keep track of their movements and monitor behavioural patterns. Using video analytics one is able to monitor patterns of movement of not only one object, but also hundreds of other objects in the same scene with a clear provision of their size, shape, speed, and direction.

According to Dial and Storkey (2017), the healthcare sector has been a forerunner in storing meaningful medical imaging data and this type of data is mainly used by healthcare providers to gain a real-time insight to healthcare when diagnosis or treating patients. A major part of the data in the medical fraternity is collected mainly from the patients. This type of data comprises of signals, videos and images. These medical images are high definition, very large and they allow the doctors to monitor or diagnose the patients in order to provide a prescription (Kouanou, 2018).

2.6.5 Social Media

When used effectively Big Data supports data-driven decision making while simultaneously stimulating the innovation of new business solutions, creating new products & services and lastly improving how business operates (Lee, 2017). Dhawan and Zanini (2014) note that the Big Data phenomenon has fuelled new methods of keeping organizations in touch with their customers. Organizations seek to understand what the people are saying and what things are they sharing daily in their lives. This is one of the many reasons social media data is more of a treasure for many organizations.

At this present moment, social media is enjoying an overwhelming success in terms of usage growth and adoption. It has also provided a significant paradigm shift on how the people communicate globally, how the world does business, how people share their opinions or ideas and even how their services and products are rendered. Mallon (2018) refers to social media as a "chief generator" of Big Data and further notes that the enormous influence of social media presents an influx of data and a new experience for the organizations and consumers.

According to Funk (2013) customer impression is more likely to rise by approximately 74% after a successful interaction with the business through social media and customers are more likely to purchase a product after they have clicked the like button. Users' social media activity is a source of valuable information. Melder (2019) emphasizes that social media content provides insight about people's interest, who or what they interact with, from which location and their general behaviours. This kind of data carries more risks and complications as the concern about the privacy aspect are growing rapidly.

2.6.6 Documents

A plethora of sources generate large sets of Big Data on a daily basis. Some of this Big Data is in the form of html, csv files, documents, and worksheets unlike the normal transactional data. Processing large amounts or piles of these documents tends to put a strain on the infrastructure and can be extremely expensive. The Big Data technology is very essential when dealing with these documents (Hussain et al., 2019).

Globally, organizations have been intensifying their efforts to acquire increasingly massive amounts of documents and that later creates challenges on the ability to process, preserve and provide access to the data (McAninch, 2015). According to Xu (2013), the main objective when colleting the documents is to extract the much-needed key facts from these documents and later use them as inputs in the decision-making process.

Verma et al. (2016) infer that this kind of data is commonly used in government, for research purposes and general business needs. The normal data only narrates historic actions but documents analysis is able to provide reasons for these actions.

2.6.7 Business Applications

In the Big Data age, we have witnessed that not only the volume of data has grown but also the diversity of its sources. The amount of data sources has since multiplied in a dramatic manner when compared to the pool of traditional

data sources. Most organizations have injected a lot of funds in the efforts to collect, analyse and use Big Data to grow their businesses.

Fitzgerald (2019) states that every time a customer signs up for an application, they are expected to fill in their personal details, location and other relevant prompted information. Some of these applications also record the interests and preferences for a particular customer in order to customize offers or services. Angeles (2019) mentions customer relationship management CRM as a perfect example of a business application that supplies an abundance of actionable data.

Van Rijmenam (2019) states that business applications such as CRM enables the company to interact with existing and potential customers. These interactions involve massive volumes of both structured and unstructured data that later enhance the business insight of each customer in real-time.

Even though preceding generations might be reluctant to embrace the proceeds of investing in Big Data technology, organizations that have leveraged on the rich data generated by the business applications have been able to gain the competitive advantage (McAllister, 2018). The author further states that in the project management space when processed to meaningful information this data can assist project managers in keeping track of their several projects.

2.6.8 Public Web

Nowadays the business environment is evolving in a very rapid manner. This requires organizations to constantly stay updated and aware of the relevant news, trends, developments and topics to effectively run their business. Hakutizwi (2018) contends that social awareness and social media presence is very crucial for any business that seeks to gain a broader understanding of its market. It can also provide the ideal direction to follow and therefore saving valuable time and money.

Herodotou et al. (2011) affirm that making timely and meaningful analysis of Big Data is one of the key success factors for modern businesses. Dhawan and Zanini (2014) concur that the Big Data technology phenomenon has instigated the birth of a fast-developing field of study known as sentiment analysis. Its main objective is to get an understanding of what people are saying or sharing daily in their lives. Businesses collect and store this kind of data in their mission to gain deeper impression of their customers and make improvements where needed.

In various sectors, one of the prevalent methods of collecting Big Data is through gathering of internet search or user behaviour on search engines. Nuti et. al (2014) mention google trends as the leading tool when it to comes to internet search data provision. In the healthcare sector internet search data is

used to draw the sought-after insight of disease outbreaks and patterns. Failure to comply with the rules and regulations governing this type of data may mean that organizations are running the risk of incurring penalties or hefty fines and also reputational damage. To understand the effect Big Data has on traditional data governance, the characteristics unique characteristics and varying sources of Big Data had to be explored.

2.7 Big Data Evolution

Consumer Data, about two decades ago was difficult to obtain as companies had to pay data collection and survey companies for provision. Today, the reverse is the case, as data is in large quantities and companies are now facing several data challenges (Economist Intelligence Unit, 2011). According to Khan et al. (2014) in the age of digital transformation and automation, data is generated and collected at a very fast rate and it is now beyond expectations.

Khan et al. (2014) state that the Big Data era has received an impressive response from businesses, academia and the IT fraternity in general. This growing interest on Big Data is the result of opportunities and competitive advantage that Big Data presents. IDC (2019) states that the exponential growth of the Big Data is expected to hit an annual growth rate of 23.1%. This means that more knowledge and expertise is going to be much needed when critical and data driven decisions are made (Burger, 2019).

With the dawn of the digital age comes an exponential growth in the amount of data that is generated from different sources ranging from social media to online purchase transactions (Ahuja and Moore, 2013). According to Kemp (2020), the number of internet users is growing at a rate of 11 users per second. This growth amounts to an overall total of more than one million users daily. Furthermore, the author states that it is important to note that a great portion of the current growth in data originates from the developing countries that had no internet access in the past few years.

Buttler (2017) claims that Big Data has been a key driver of success in larger corporations for a while, however the small, medium enterprises are now starting to bridge the existing gap and their main attraction being the lower costs associated with managing this type of data. Big Data is tipped to be the future of data warehousing and there is a great confidence that it will decrease the cost of revenue (Minelli, Chambers and Dhiraj 2013).

Ernst & Young (2014) states the current developments in technology have taken data analysis outside of IT backrooms and propelled the mind shift to data-driven decisions for every part of the organization. Even though the improvements in software and hardware are vital for the Big Data era,

technology cannot not be the sole focus. Organizations have to channel their focus to the integration of people, data and processes as well.

Schmarzo (2013) posits that Big Data has mostly transformed business than IT and the author further claim that effective use of Big Data has the potential to augment the timeliness and the quality of the business decisions. Despite all the buzz around the Big Data topic, its crosscutting benefits such as the improvement of decision making are exactly what a lot of organizations have been longing for (Regalado, 2014).

Alongside this disruption, Study International (2019) mentions that the world is currently at an early juncture of the Big Data era. The author further states that data science is radically evolving how business is conducted, how the world communicates, the way people access knowledge and their perception of the world in general. Several ecosystems are being developed, an absolute reconfiguration of the old traditional industries is taking place and the race to attain the best of the data science skills and tools is getting very tense. Businesses that seek to be on the leading edge of the Big Data developments are expected to invest tremendously in attaining the necessary data science skills (DHL, 2013). These skills are very vital to any company that seeks to gain valuable facts or insights from their Big Data as this data needs to be processed, cleaned and critically analysed. Without the necessary data science skills this would be a costly exercise for any organization (Al-htaybat and Albertialhtaybat, 2017).

Forbes (2018) describes the company's ability to exploit Big Data and analytics as matter of life and death, this is a reason why several organizations are investing much of their resources to be part of the leading pack. The growing interest for companies to invest in Big Data can be attributed to benefits such as the ability to accurately predict the near future, monitor trends in real time and acquire substantial insights into the company's business environment and that of competitors (Ernst & Young, 2014).

Effective use and analysis of Big Data enhances the business ability to make very accurate future predictions. Some companies have been making these future predictions for a while predicted future, but the Big Data technology offers them more accurate results and predictions (CGMA, 2013).

Currently data is produced perpetually through our digital interactions and actions. Thus, affording companies the opportunity to collect, analyse and store data with the aim to improve their performance, increase profitability and to solve the key challenges (Kuurila,2016). Smith (2014) identifies the enhanced understanding of customer needs as one of the key Big Data prospects that seem to be appealing to customer obsessed organizations.

Hurwitz, Nugent, Halper and Kaufman (2013) mention that Big Data is one of the factors with the ability to influence the operating model, this may be visible during the initial or mature phase of the Big Data initiative. To fully leverage Big Data, the point of departure should be the identification of the fundamental business challenges, examine components of the business model, identify key sources of data and also understand the structure of the data (Chartered Global Management Accountant, 2013).

Kaisler, Armour, Espinosa and Money (2013) argues that the value gained by an organization from Big Data does not only lie in the ability to process massive data volumes, but it also lies in the capability to analyse various data types (structured or unstructured) to draw business insight. Spencer (2014) states that marketers are now able to make critical decisions based on the large pool of data unlike when they relied on intuition. Mallinger (2015) suggest that intuition has reached a stage of obsolescence, the recent analytic tools will improve the quality of business decisions with their ability to process large data sets within a short space of time.

Organizations stand to gain a number of benefits from the full exploitation of Big Data analytics and implementation of effective Big Data initiatives and programs. These include the transformation of their operations, the delivery of an improved service to their customers and becoming the driver of innovation in the respective markets. Empirical evidence exists to articulate how Big Data can be the determining factor in the battle for competitive edge and increased productivity (Fosso Wamba et al., 2015; Maisel & Cokins, 2014).

It cannot be denied that the Big Data evolution directly impacts traditional data governance as processes, methods and policies that are tailored for Big Data are much needed. This is in line with the objective of this study to determine how the adoption of Big Data affects traditional data governance. Moreover, the evolution presents new data governance challenges which the study seeks to understand in depth.

2.8   Data Governance

Data Governance has no standard or unique definition and hence is it can be defined in various ways. Otto (2011) concurs that there is no single established definition for the term Data Governance. IBM (2007) defines data governance not as a quality control discipline for adding new rigor and discipline to the process of managing, using, improving and protecting organizational information.

Accoding to Kagermann, Osterle and Jordan (2010) the fact that data governance is currently being discussed so intensely is due to a growing number of business requirements which a company's data is expected to meet. Collibri and MapR (2016) adds that data governance comprises the enablement of the appropriate data behaviour in order to minimize the risks posed by ill data behaviour. Data governance is also important in learning the meaning, purpose, and the quality levels of Big Data. Failure to understand the latter can constrain the recognition of data value.

Additionally, Washington (2018) states that data governance involves managing resources, people, technology and processes that are important for leveraging data. According to (Cohn, 2014) data governance regulates the processes and procedures used by the organization to preserve or improve the value of data. The data management process must be well-defined, actively monitored and adherence of these processes must be reviewed constantly. The common denominator to the above definitions is their emphasis on planning, managing people, processes and leading a data driven organizational culture (Tait ,2016).

Building a strong governing program has been identified as one of the key challenges in data value recognition (Marchibroda, 2013). Ndamase (2014) claims that a viable data governance program underpins the formulation of clear data policies and standards while monitoring elements such as quality for the business to receive accurate, timely and consistent data.

There are various approaches to governing Big Data. The components of governance include policies, standards, strategies, plans and strategies. Each of these components adds value to governance and they are key in risk management. Ideally, the governance program seeks to attain strategic goals and that is what makes it distinct from the management function, which seeks to attain operational goals (Liell-Cock, Graham & Hill, 2009).

According to Ernst & Young (2014) effective governance is vital for the success of any Big Data program Ernst & Young (2014) further claims that effective governance comprises of clearly defined procedures, persistent guidance and strategic decision making. The organizational leadership and management are now expected to have much knowledge or experience of Big Data concepts in order to drive the Big Data initiatives of the organization and to also coordinate the activities to convert Big Data into actionable insights (Tabesh, Mousavidin and Hasani, 2019). Manyika et al. (2011) argue that despite Big Data technologies and tools being the enabling factors in business, management and leadership is very imperative in asking the relevant questions, governing the data and exploiting it to make informed decisions.

Deliberations around policy should be prioritized by organizations that aim to exploit Big Data across all sectors of business (Schroeder and Schroeder, 2016). The key challenge encountered by organizations in their attempt to put in place good governance of Big Data is, is that the data falls past organizational lines and relatively concerning third parties. The latter presents severe governance challenges and uncertainties in as far as ownership and protection of the data is concerned. Labadie and Legner (2019) adds that in recent times the public has grown so much interest on data ownership.

This literature is vitally important in answering the main question of this research and also understanding the processes, methods and policies that are involved in governing Big Data.

2.9 Data and Information Management

Data management has been gaining more prominence in the business world as data standards and regulations are rising globally. This prominence may also be attributed to the growing realization of data management benefits and value of data.

Mosley, Brackett, Earley and Henderson (2009) define data management as an extremely broad process that spans over several functional areas. The author provides a deeper definition and defines data management as the process of developing, effective implementation and maintain oversight of data and information policies, plans and processes that regulate, safeguard the value of organizational data.

2.9.1 Good data management

Wilkinson et al. (2016) states that the act of effectively managing data is not an ultimate goal but it is vitally important for uncovering insights, supporting innovation, and ultimately data and knowledge integration. For businesses, the key gain of data management is availability and access to highly valuable information that is attained data that is shared from a variety of IT systems. Another benefit is data quality of a higher degree which is very important for any business that seeks to gain a clear insight of its customers and processes in a competitive terrain (Gordon, 2007).

Steenbeek (2018) suggests that those who are esteemed with the responsibility of driving the organizations data management initiatives should begin by seeking an understanding of the company objectives and where they seek to channel their energies. DAMA International (2017) contends that by implementing a data management program, organizations seek to achieve some of the following goals:

- To gain an insight and support the information needs of the organization and its entire stakeholders.
- Capturing data, ensure that it is securely stored and ensure integrity of the data.
- Drive data quality assurance and improve the quality.
- Ensure that stakeholder data is highly secure, guarantee them privacy and confidentiality.
- To have data stewardship in place, control accessibility and avoid data breach and misuse.
- Successfully leverage data to propel innovation and efficiency.

Effective and efficient management of data is no longer optional. To effectively manage data, an organization needs to intensify its efforts in bolstering data management and analytics competences and augment its expertise on the data management subject (Shamim, Zeng, Shariq & Khan, 2019). In addition, DAMA International (2017) asserts the gains cannot just manifest out of thin air but involves having committed leadership and management with clear objectives and practical plans that are well coordinated.

An appropriate data management architecture is essential for guiding the implementation of a data management program. When thoroughly implemented, an organized approach to data management offers business and technical guidance in various ways (Pflug, 2018). The Enterprise Data Management (EDM) reference architecture by Soares (2014) is illustrated in Figure 5.
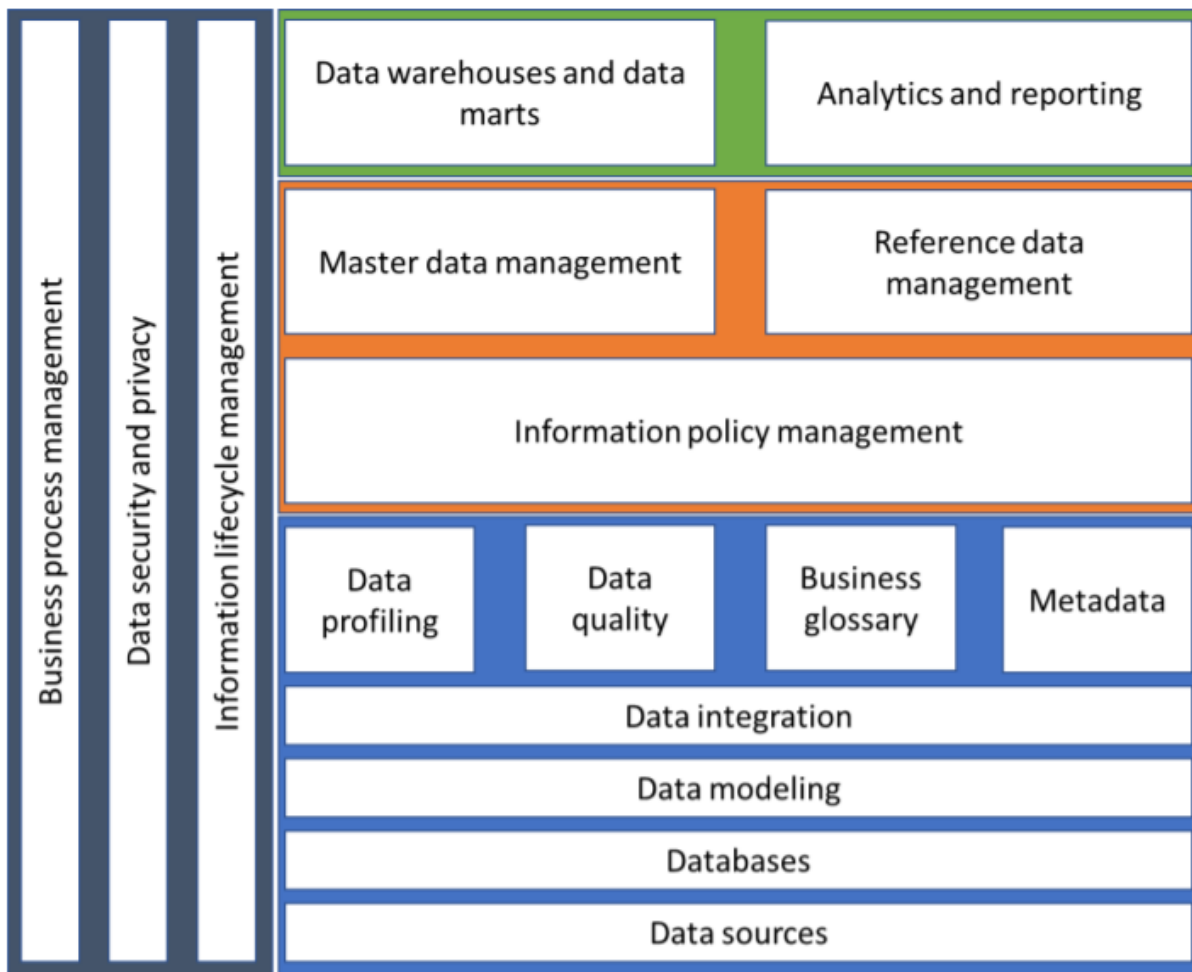
Figure 5: Enterprise Data Management Reference Architecture (Source: Soares, 2014)

Figure 5 depicts an extensive overview of an enterprise data management architecture and it explores all components of concrete data management from the beginning to the end. This involves sources in which the enterprise data is extracted from, the development and execution of data policies and ultimately the continuous efforts to augment the quality of the data assets.

2.9.1.1 Data sources

This is the variety of enterprise data sources originating from internal and external sources and also having different types and formats.

2.9.1.2 Database

The data obtained from the various sources is then stored in a database. Sabesan and Risch (2006) define a database as a collection of data that is managed through a DBMS. It enables the user to access, insert, update and delete data from data using a data management language (DML).

2.9.1.3 Data Modelling

Data modelling is essential for any organization with a goal extracting maximum value from the data it owns. A data model is a clear description of the actual data structure (Zomaya and Sakr, 2017). The primary objective of data modelling is to structure the raw data in order to uncover the data that is most valuable for the stakeholders.

2.9.1.4 Data Integration

According to Ise (2006), data is sourced from multiple sources, is heterogeneous, from various locations and different business areas. This is a common challenge and therefore requires data to be integrated before the business can reap the benefits.

2.9.1.5 Data Profiling

After the data integration process from the data source to the target, the data needs to be assessed for consistency, completeness and validity. Cabral (2021) states that data profiling should be a priority for any organization that aspires to improve the quality of data in order to guide the implementation of informed decisions. This is an extremely critical phase for identifying the raw data issues and the corrective measures that must be taken to resolve this data issues.

2.9.1.6 Data Quality

Having quality data is the prerequisite for analysing, usage and data value assurance. Koziel, Hilber, Westerlund and Shayesteh (2021) define data quality as 'fitness for use'. This means that the data that was collected is evaluated for its ability to serve its intended purpose. Nikiforova (2020) provides a similar definition when defining data quality as the fitness of a certain dataset and its properties for a specific purpose or use case. Haug, Pedersen and Arlbjørn (2009) mention intrinsic, accessibility, and usefulness as the main properties of data quality. However, there is several properties that describe data quality and these vary from one to another.

2.9.1.7 Business Glossary

A business glossary is a repository that offers a set of clear definitions for data related business terms. This is produced to create a common understanding of key business terms, business ideas and how they are related (Tsai, 2020). Paschalidi (2015) concurs that business glossary can be best used to eliminate the possible occurrence of a semantic misunderstanding in larger organizations. Tomson (2019) states that the development of an extensive

business glossary can be a mammoth task and it necessitates a clear business vision, business processes and architecture.

2.9.1.8 Metadata

There is no standard definition of the term "Metadata". Several definitions and explanations do mention similar themes. Metadata is generally defined as information about data stored in any kind of environment, it is relatively used to define various aspects and properties of data such as structure, the elements and the existing relationships among data elements (Kulkarni, 2016).

The author further states that metadata enables the association of similar data assets and this is very important in providing accuracy of search results. Wilkins (2021) affirms that metadata is fundamental for effective management of the data and information assets in any organization.

2.9.1.9 Information Policy Management

Information Management Policy comprises of management and governance of both unstructured and structured information or data that is in custody of the organization in order to execute its primary functions and achieve its objectives (The University of Queensland, 2021). Paananen (2020) adds that Information policy management consists of executing the governance task, managing risk and monitoring compliance of the information.

2.9.1.10   Master Data Management

According to Foote (2021), Master Data is commonly defined as data that embodies the core information of an organization regarding the key assets such as products, people, places, and assets. Master Data Management is therefore a process of implementing and maintaining a well-organized system to coordinated master data as a single source of truth across the enterprise (Papa, 2017).

2.9.1.11   Reference Data

Chen et al. (2013) define reference data as the data that is used specifically for categorization and classification of the data in enterprise applications and databases. Reference data includes the lookup values, list of values, code table data that exists in within the business applications like location codes or currency codes (Informatica,2019). Chisholm (2015) infers that the fundamental function of reference data is translating the data into actionable insights and providing general informational understanding to the users or data stakeholders.

2.9.1.12 Data Warehouse and Data Marts

Similar to a database, a data warehouse is a single storage of data. It houses and manages a collection of data that is sourced from heterogeneous data sources. Bonifati (2001) states that the fundamental objective of a data warehouse is to preserve and analyse historical data in order to support decisions.

The data mart is the simplest version of a data warehouse that focuses specifically on information needs of a business unit or department. Due to its small scope, a data mart sources data from a lesser amount of sources when compared to a data warehouses (Utami, Pratama and Widianto, 2020).

2.9.1.13 Analytics and Reporting

After capturing and storing data, organizations have to extract relevant insight from the data and subsequently convert it into actionable information. Reddy and Reddy (2007) state that computational techniques, tools and algorithms have to be used for the analysis of data before a meaningful understand could be gained. Analytics offer a high degree of customer understanding to businesses and this enhances the organizations. Companies can now customize their products and explore new revenue streams (Porter & Heppelmann, 2014; Loebbecke & Picot, 2015; Kritikos & Barreiros, 2016).

2.9.1.14 Business Process Management

Business Process Management is essential for any data management initiative. According to Skjæveland (2013), Business Process Management is a systematic approach which intends to manage the processes of a business. The activities are designed, monitored and analysed from one end to another with a goal of improving the business process based on data. Hassani and Gahnouchi (2017) state that Business Process Management obtains data from several enterprise applications. The business must exploit data to gain the desired knowledge to improve the processes and business efficiency at large.

2.9.1.15 Data Security and Privacy

Data privacy is a long-standing challenge in the internet environment. Personal data from heterogeneous sources is combined, therefore a lot of details can be revealed about an individual. The act of sharing corporate or personal with a potential of being identified is currently a huge concern (Varga, 2011). Organizations are required to put systems and processes in place in order to preserve its data assets. Investing in top of the range hardware and software is not enough to preserve enterprise data and information. An organization requires systematic policies and strategies to monitor and control data security despite acquiring the best technology (Barbu, 2020).

2.9.1.16   Information Lifecycle Management

Information Lifecycle Management is a holistic approach to overseeing data throughout its lifecycle through the use of policies, strategies and technology to maintain or improve the value of business information (Short, 2006). Information creation is the first phase of information life cycle, follows the information organization, storage, use, and improvement until it is disposed (Dederer, 2015).

2.9.2 The consequences of poor Data Management

Eaker (2016) states that the possibility of having errors and data management challenges does exist, even when data is managed effectively. Nonetheless, the probability increases even more when data is mismanaged. Failure to manage the data asset effectively can have dire consequences and may carry a hefty cost to any business. Having an informal data management plan or system may result to bad decision making and also organizational efficiency (Paterson, 2020). Haug, Zachariassen and Van Liempd (2011) claim that there is plenty of negative repercussions that poor management of data can lead to in an organization and further provides the list the following:

- Lower customer satisfaction

- Increased running costs.

- Inefficient decision-making processes.

- Poor business performance.

- Decreased employee job motivation and satisfaction.

- Higher operational costs.

According to Otto - Hunner (2009), poor data management effects are most likely to hurt the business financially. It is very typical that an organization is heedless of this shortfall until a crisis arises. Instead of proactively managing data, a large portion of organizations opt for reacting, which carries significant costs in the long run. Despite the possible financial loss, organizations who fail to put measures to improve their data management a high likely to commit operational mistakes and ruin its ambition towards gaining the competitive edge. DAMA International (2017) affirms that the necessary media attention may have dire outcomes and possibly lead to reputational damage which can be directly translated to a loss of future opportunities or clients.

This literature played a key role in understanding the components of data management and how they are related to data governance. One of the objectives of this research is to identify data strategies, processes and methods that are employed to govern Big Data and therefore exploring data and information management afforded the researcher with a broad understanding of the components involved in governing Big Data in general.

2.10   Data Governance Challenges

Johnson, Neff and Stuart (2012) assert that Big Data technology is still emerging and it comes with a set of risks. Gerber (2015) concurs that every time an organization adopts a disruptive technology such as Big Data, severe challenges do emerge and they should be treated with the outmost urgency.

2.10.1 Privacy

Organizations use Big Data to guide their decision making, to make projections about people, products and other entities. Given the sensitivity of these sources, privacy and security pose significant risks even if some of the data does not contain personally identifiable information (Franks, 2012).

Kaisler et al. (2013) states that at the current moment, there is a widening gap that is caused by the lack of Big Data specific policies and regulations. Elicitation of unstructured data requires a high degree of governance as it carries severe risks. During the initial phases of collection, this data may not appear to be personal even though the data can potentially be drawn back to the initial source and this may pose a privacy risk (Mayer-Schonberger and Cukier, 2013).

Kulk and Van Loenen (2012) asserts that there is plenty of privacy rights implications around Big Data as some people are not aware of the purpose of collection and use of their data. Using personal data without being granted consent by the subject can potentially mean that security has been compromised, breach of privacy or misuse of data and the result of such an action may have an adverse impact on the business (Kshetri, 2014).

With a good data strategy, the business is promised a great deal of value and benefits. However, data privacy and ownership challenges must be given considerable attention before the data can be processed further. Big Data carries intellectual property fears along, these include issues such as access privileges, sharing business classified information external (Ernst & Young, 2014). Altman, Wood, O'Brien and Gasser (2005) confers that when data activities are carried out effectively, they have the ability to render incredible benefits but on the other side of the coin the expansion in the size of data has the potential in various ways to extend the magnitude of informational harms.

Khan et al. (2014) claims that privacy is one of the major Big Data concerns, due to a growing number of has as some security agencies that use people data for their own gain without any consent or permission. The author further suggests the formulation of policies to deal particularly with user privacy concerns, moreover, the offenders must be brough to book as their action may have detrimental effects.

For customers, privacy is highly important and one of their major concerns is around the preservation of their financial or banking information then follows the privacy of their personal information. Nonetheless customers are so eager to grant access to their personal data where they stand to benefit and obviously with a much-trusted organization (Abraham, Van Kerckhove, Archacki, González & Fanfarillo, 2019). Customers are more comfortable and eager to share their data when there is a substantial amount of trust. Trust is extremely essential for the formation of strong and lasting relationships between the business and its customers.

2.10.2 Quality

The rapid collection and acquisition of Big Data from a variety of sources has undeniably changed the way businesses value data. However, for this value to be fully realized, this data must be factual and be of high quality. Aloysius, Hoehle, Goodarzi and Venkatesh (2016) state that an organization that collects and analyse massive quantities of data are bound to run into various challenges when trying to improve the quality of this data and that has a huge impact on the business decisions.

Data quality challenges tend to hinder the business from making wise business decisions due to issues such as data inconsistency and accuracy. Friedman and Smith (2011) confer that poor data quality has an enormous impact on business productivity and efficiency as it results to bad business decisions. Poor business decisions often lead to poor business performance, therefore raising the possibilities of a financial loss and reputational damage (Loshin, 2013). Wamba and Akter (2015) claim that the revenue loss incurred by an average enterprise due to poor data quality is estimated between the range of 8% and 12%.

Defective data can potentially present a plethora of problems than solutions to the organization, with this form of data the ability to take intelligent decisions is defeated and this often leads to an erroneous understanding of how the organization is performing (Abdullah, Ismail, Sophiayati & Sam, 2015). Eckerson (2002) claims that relying on low quality data is an extremely

precarious decision that any business can take in a market that is rapidly evolving, highly competitive and customer centric.

Achieving the highest level of data quality is one of the challenges and it needs comprehensive assessments to be made. According to TDWI (2019) the quality of data is essential and this applies to all types of data, regardless of where the data is stored and what is the intended use. Organizations are under immense pressure to gain value from their Big Data investment and ensuring the quality of Big Data forms part of these efforts.

Since the Big Data phenomena is not quite an old one, the academia has not yet defined the quality standards and criteria. Within its literature, an abundance of data quality definitions does exist and the commonality between most of these is that they all define good quality Big Data as data that satisfies the initial purpose of the organization (Cai and Zhu, 2015).

Wells (2019) states that the broad range of traditional data quality methods and techniques that have been used for so long to manage the quality of traditional and structured data are not effective in the Big Data environment. Unlike the structured data, several data quality rules are inapplicable when the data is not of structured form and not relational.

According to Toivonen (2014) in the context of Big Data, many quality challenges originate from a lack of data understanding, failure to cope with the cluttered source data and ultimately the ability to deduce insight from the analytical results. Rajpurohit (2017) concurs that a major data quality challenge in Big Data is the various data types such as (unstructured, semi-structured and structured) drawn from a variety of data sources.

Rajpurohit (2017) further argues a single data quality metric cannot be relevant to all the types, thus unique metrics have to be defined for all types of data type. In the Big Data era data emanates from a range of data sources, but the users of this data may not always be its producers. Hence, evaluating the level of data quality becomes a mammoth task for those responsible (Cai and Zhu, 2015). There are three goals of data quality governance. The starting point of data quality improvement is gaining insight on the value of data within the organization, followed by introduction of right levels of control, then comes detecting and prioritizing data issues to resolve (Loshin, 2013).

Accurate and high-quality data is crucial to the success of the business and guarantees that data-driven decisions are mainly based on facts. Furthermore, organizations must also make a decision of where this data fits into the overview organizational value chain in which they are operating to completely uncover and understand the value stored in their data (Enaxisconsulting, 2017). Panchenko (2016) states that it is compulsory for an organization to validate or

confirm the integrity and accuracy of the massive amount of data before it stores it, but this is not being adhered to in some organizations.

Azeroual, Saake and Abuosba (2018) suggest that, if an organization has managed to reach a certain degree of data quality, plans must be made to maintain and further improve that level of quality. As the velocity of Big Data is very high, changes are bound to occur very frequent and this stage of improvement data must be actively monitored. Taleb, Mohamed & Rachida (2018) concurs that the process of measuring and monitoring data quality must take place from the initial to the last phase of data lifecycle.

2.10.3 Security

The rapid increase in volume and the growing demand of data has produced more complications and traditional data governance cannot manage these. The introduction of Big Data was tipped to solve volume related problems. However, security, quality and privacy related challenges manifested (Moreno, Serrano and Fernández-Medina, 2016).

Zaman, Seliaman, Hassan and Marquez (2015) highlight that the volume, velocity, and variety of Big Data have resulted to the rise of security concerns. The author further points out that the rapid growth in internet access and usage, ubiquitous computing, e-commerce, and worldwide expansion of cloud infrastructures offers a wide range of data sources and formats, prodigious volumes of data at a lightning speed. ENISA (2016) concurs that the security factor of Big Data coupled with the privacy challenges and the preservation of its critical infrastructure remains a key concern for many.

The traditional security programs are tailored for smaller data sets with a predefined structure. Even though Big Data is presenting a pool of opportunities in various industries, it comes with a price as the rate of security attacks are constantly growing. With all the improvements in the data space, attackers and hackers are also becoming more advance, omnipresent, extremely sophisticated and malicious (Interaction Design Foundation, 2016).

The contributing factors to the consistent increase in attacks is the rapid evolution of major businesses and the growing focus towards the new ways of working that are more agile and collaborative. Smith, Watson, Baker, and Pokorski II (2007) state that the gains of interconnectivity are accompanied by several risks as the organization's data and IT infrastructure's is more vulnerable to interference due to an array of data related threats. Thakkar (2017) concurs that the rising numbers of both private and public entities rushing to adopt Big Data technology has directly influenced the growth in targeted attacks due to what is at stake.

Large organizations with a global presence, like telecommunications companies, have widespread supply chains that are extremely broad. These critically connected supply chains are potentially prone to attacks and therefore organizations have to shift their focus towards the implementation of security measure and controls to reduce cyberattack vulnerabilities (Acker, 2020).

According to Vargas-Solar, Zechinelli-Martini and Espinosa-Oviedo (2017), Big Data requires security solutions with the ability to monitor all access to critical data. This includes being able to detect any suspicious patterns or activities when matching against the frequent and normal activities. By nature, Big Data consists of numerous instances, layers, and technologies each serving various purposes in sets of Big Data. Thus, providing maximum security for Big Data is regarded as a very complex task than when compared to the traditional data security solutions. Access control is a technique that has been used for a long time in the data space to ensure the security of a system. In the Big Data space, the access control issue comes from the fact that only basic measures of access control are used (Moreno, Serrano and Fernández-Medina, 2016).

2.10.4 Ethics

The rapid growth in stored data about people, their preference and behaviours has stimulated an extensive debate in both the legal and IT fraternity about the ethical and legitimate ways to leverage such data. The most quizzical challenge is the absence of an authority to provide a clear guidance, this necessitates the need to pay critical attention to ethical considerations by data producers, suppliers, and consumers (Hand, 2018).

Gall (2020) states that the preliminary phase towards the consideration of ethics and displaying commitment about ethics is to deliberate about bias. Furthermore, the awareness of the possible bias in Big Data or machine learning and its ethical consequences becomes much critical in the efforts of spotting ethical issues and making relevant improvements to mitigate the risks.

Machine learning algorithms have grown to become a part of our lives, hence it is vitally important to seek a coherent understanding of how these algorithmic decisions are reached. Sweeney (2013) confers that machine learning decision-making systems are gradually used in highly prospective situations and a majority of them affects the society in various ways. The author further states that transparency is extremely imperative when an organization decides to leverage data-driven machine learning, as it is much easier for algorithms to yield biased and discriminatory results.

Indirect discrimination is one of the major concerns facing many data driven organisations. This is mainly because of ambiguous nature coupled with the high possibility of it being overlooked. Barocas and Selbst (2016) assert that in data mining the odd are relatively low for classifiers to be bias and discriminatory towards certain group or class. Workfront (2018) contends that even though businesses are enjoying the prospects of Big Data as they now have massive amounts of knowledge about their customers at their disposal, the biggest fear at the moment is the insufficient understanding of boundaries and not being able to detect when the knowledge becomes detrimental discrimination.

According to Baeza-Yates (2018) current research attempts to address the ethical challenges but lacks a deeper understanding of how to best manage the usage of Big Data driven decision algorithms to guarantee transparency and impartiality. Martin (2015) states that data should be generated, collected, analysed and utilized in a fairest, ethical, lawful, transparent and legitimate manner. The author further states while ensuring maximum confidentiality, ethical and moral conduct, companies must emphasize on compliance with the relevant data privacy and data protection laws.

It is imperative that a critical assessment is undertaken to determine how companies that supply Big D32ata regard privacy and ethics when collecting data. In most instances, consumers supply information only when they are guaranteed privacy, and distributing that information to other external parties may mean that their privacy has been breached (Martin, 2015). Herschel and Mior (2017) states that an ethical assessment can play a critical role in gaining a deeper understanding of the ethical challenges or related concerns and subsequently provides a clear view of the ethical shortcomings in an organization.

2.10.5 Transparency

People have the fundamental right to be well informed about the primary purpose of processing their personal data, this includes the integrity and transparency of the Big Data analytics in the Big Data setting (Debussche, César and Thirot, 2020). Green (2018) confers that Every organization yearns to earn public trust, but that trust can only be received only if the organization is able to substantially demonstrate that their data only sourced ingeniously and subsequently taken through a critical verification process to guarantee a high degree of quality.

According to Bird & Bird LLP (2019), lack of transparency has a substantial potential to massively ruin trust between the organization and its stakeholders, particularly in a Big Data environment. Individuals are indeed not always aware

of the exact nature of the processing activities and of the logic of algorithms and the decision-making process behind Big Data analytics.

Bird & Bird LLP (2019) further state transparency is a major challenge for many organizations while some blatantly refuse to be transparent citing several reasons such as trade secrets protection, protection of critical customer and business information.

As data becomes a competitive advantage and major factor in business growth, gaining customer trust becomes very critical. Organizations that are upfront and transparent about their intended purpose are most likely to be trusted or granted even further access to data by customers (Morey, Forbath & Schoop, 2015).

2.10.6 Compliance

Garcia (2020) defines data compliance as an act of organizing and managing sensitive data with an intention of ensuring that the organization conforms to its own business rules, followed by government and legal regulations. To exploit the data asset profusely, organizations have to ensure that sensitive data is managed comprehensively in order to comply with various laws and rules (Amster, 2016). According to Trint (2019) a compliant organization is the one that implements policies and workflows that clearly stipulate how the organization strives to protect data in accordance with the rules and regulations of its jurisdiction.

The rapid developments in software and data technology, according to (Guamán, Del Alamo & Caiza (2021), has enabled innovative methods for universal collection and sharing of personal data about individuals across different countries, jurisdictions, states and regions. The author adds that this kind of data sharing poses a potential loss of privacy control, as the receiving party may not guarantee comparable degree of protection.

This is supported by Pernot-Leplay (2020) when mentioning that for a while China has been lagging behind its USA and European counterparts and offers a lesser level of protection when it comes to personal data. 128 out of 194 countries had put in place legislation to secure the protection of data and privacy. Trint (2019) suggests that companies must take cognisance of the of the fact that compliance is a universal subject matter and thus their compliance efforts should not only be based on local regulations. The author further states that compliance is mandatory in every corner where the business trades or operates.

The regulatory developments prompt for changes and improvements on the methods employed by organizations to oversee usage and protection of customer data (Acker, 2020). Saqr (2017) concurs that the rapid information

advancement warrants a systematic response from all authorities to seek a broad understanding of the legal, ethical, and cultural regulations. The old solutions are not adequate to resolve emerging compliance challenges. Policies, laws and regulations that are tailored current challenges and beyond are extremely essential.

According to the United Nations Conference on Trade and Development (2020), only 66% of the countries in the world have implemented measures to protect personal data even though there was a promising hike of about 11% between the year 2015 and 2020. The European policymakers have been leading forth when it comes to the adoption of data privacy and protection laws, this has resulted to them being the paragons of compliance (Scott and Cerulus,2018).

Huang (2020) concurs that by implementing the GDPR, some nations have demonstrated how advance they are in terms of regulating and protecting personal data. The author claims that when compared to other existing regulations, GDPR is the most formidable and far-reaching data protection regulation thus far. The adoption of GDPR has afforded citizens the right to request access to the detailed information that is being process about them, therefore imposing improved data management on companies and substantial administrative fines (Burgess, 2020).

In the South African context, the government enacted the Protection of Personal Information Act 4 of 2013 which is currently the world's most recent data privacy law to be endorsed, it contains an ample of similarities with the EU's GDPR. The act seeks to regulate all the activities that comprise of collection, processing and sharing of personal information inside or across the borders (Republic of South Africa, 2013). The Act controls the legitimate processing of personal information. A valid authorisation must be obtained from the data subject before personal information can be shared with third parties (Adams, Adeleke, Anderson, Bawa, Branson, & Christoffels, 2021).

According to (Republic of South Africa, 2013) for non-compliance to certain requirements of POPIA the Information Regulator may enforce penalties that could amount to R10 million or perhaps to imprisonment up to a maximum of 10 years, in some cases both imprisonment and fine. Data privacy compliance is not merely abiding by the laws and regulations but the formulation of policies and processes that aim to safeguard both the data and reputation of the organization. de Bruyn (2014) claims that the risk of reputational damage can have dire consequences on the business and can potentially hinder business growth and customer trust.

2.10.7 Accountability

Accountability is an act of assessing a behaviour of an individual or organization. This includes a responsibility to acknowledge or substantiate for actions or decisions especially if they can be potentially classified as offensive (Unceta, Nin, & Pujol, 2020). At the moment, accountability is one of the central topics in the data protection policy discussions and over the years it has attracted a great deal of attention from different role players in the data protection space.

Weitzner, Abelson, Berners-Lee, Feigenbaum, Hendler, & Sussman (2008) believe that organizations must be transparent about how they use information so that the data subjects can be able to easily identify a possible misuse and therefore be able to hold organizations accountable for being dishonest. According to Voigt and Bussche (2017), a lot still has to be done in ensuring that organizations are held accountable for the sensitive data they collect and process.

The centre for Information Policy Leadership (2018) states that globally, the principle of Accountability has gained an overwhelming support and has formed an integral part of various laws such as the EU General Data Protection Regulation. As featured in the POPIA and GDPR, the accountability principles compel organizations to consistently comply with data protection principles throughout all the stages of data processing and lifecycle (Donnelly, 2020).

The accountability principle is even more critical for data controllers when implementing policies and measures to deal with data minimization or achieve data protection compliance (Alhadeff, Van Alsenoy & Dumortier, 2012). Being accountable means an organization is responsible for compliance and it is also able to prove its degree of compliance. Top management should be held accountable for compliance even though it is executed from the lower level of the hierarchy (European Data Protection Supervisor, 2018). Le Cat (2017) mentions that compliance with the principle of accountability demands a comprehensive governance structure accompanied by cultural and organisational transformation. It is vitally important for all the stakeholders to reach a common understanding of the exact definition of organizational accountability and how it is ought to be implemented within an organizational context (Centre for Information Policy Leadership, 2018).

There is an overwhelming increase of organizations that opt to use algorithms for making informed business decisions and gradually the public has been getting concerned about the use of algorithms for various reasons. Diakopoulos and Friedler (2016) contend that accountability makes algorithmic accountability a responsibility to report and thoroughly explain algorithmic decision making to reduce the likelihood of any social or reputational impact. Even though individuals are actively preserving their own data or information,

accountability transfers the responsibility from the individual to the organization collecting or making use of the data (Abrams, 2009).

2.10.8 Skills

Digitalization and evolution of technology has led to a drastic change on the skills that are demanded in the technology sector. Van Laar, Van Deursen, Van Dijk and De Haan (2017) mention that even though there are several enticing benefits that Big Data offers, what remains as a very critical concern is the scarcity of the necessary digital skills to extensively leverage Big Data. The lack of these critical analytical skills coupled with the lack of skilled managers are hampering a lot of organizations from reaping the benefits of their Big Data investments (Mikalef & Krogstie, 2019).

Big Data is no longer an interest of only the IT department but a broad organizational matter of interest and this has resulted to the increase in its demand of skills and capabilities (European Commission, 2017). The rising popularity of new technology and tools have significantly shifted the landscape. These include the likes of Hadoop, Mongo DB and Apache Spar to mention a few. These tools demand a set of skills that are way unique in nature. Kostis (2016) infers that this has led to organizations digging deep in search of subject matter experts with high-level analytical skills. Gupta and George (2016) contend that companies with a greater commitment to obtain the necessary skills and those who encourage learning are tipped to gain more value from their data.

Davenport and Dyché (2011) report that in many companies the developments in technology and architecture are not accompanied by changes or improvement in skills and organizational structure. The author further states that bigger firms instead attempt to incorporate Big Data on the existing structure rather than establishing a newer structure. Another factor is organizations without a sound data governance strategy in place and one of the reasons may be lack of knowledge of what a strategy entails or lack of the skills needed to execute such action and lack of awareness of what the organization stands to benefit (Baker, 2018).

This study seeks to understand the data governance challenges faced by organizations when governing Big Data and an analysis of literature on these challenges were key to this objective.

2.11 Theoretical Framework

The theoretical framework that guided the research is IBM Information Governance Council Maturity Model as shown in Figure 6. The model was adopted to analyse the Big Data governance dynamics in organizations and it

played a significant role in assessing the organization's Big Data governance capabilities.

Measuring maturity guarantees a critical assessment of the current state of Big Data governance, identifying areas of improvement, and providing proper guidance on how to improve. Huner, Ofner and Otto (2009) affirm that maturity models are key in describing and assessing the maturity of an organization with the ultimate aim of providing the growth plan.

Cooke-Davies (2004) concurs that mature organizations measure plenty of governance elements than immature organizations and they are most likely to witness improvements in results as the maturity levels rise within the organization.

The framework duly guided the researcher in exploring the governance of Big Data using the following core disciplines:

Security and privacy

The growth in data scope creates more security and privacy implications for Big Data and that may result to severe consequences to some organizations. Organizations are expected to formulate policies that guide the use of Big Data while monitoring compliance to the security regulations.

Data quality

In this age of Big Data, data quality management is a very daunting task due to a growing number of data sources and formats. Organizations are expected to define quality standards, measure quality levels and also develop remediation plans to guarantee good quality data for decision-making.

Information lifecycle management

Having a clear understanding of the business requirements is very vital in understand what data to be collected and where is it located. An organization needs a very precise Big Data lifecycle management plan to monitor the data from the phase of collection to the last end. This can be achieved through controls and policies to guide the collection and archiving or disposal of data.
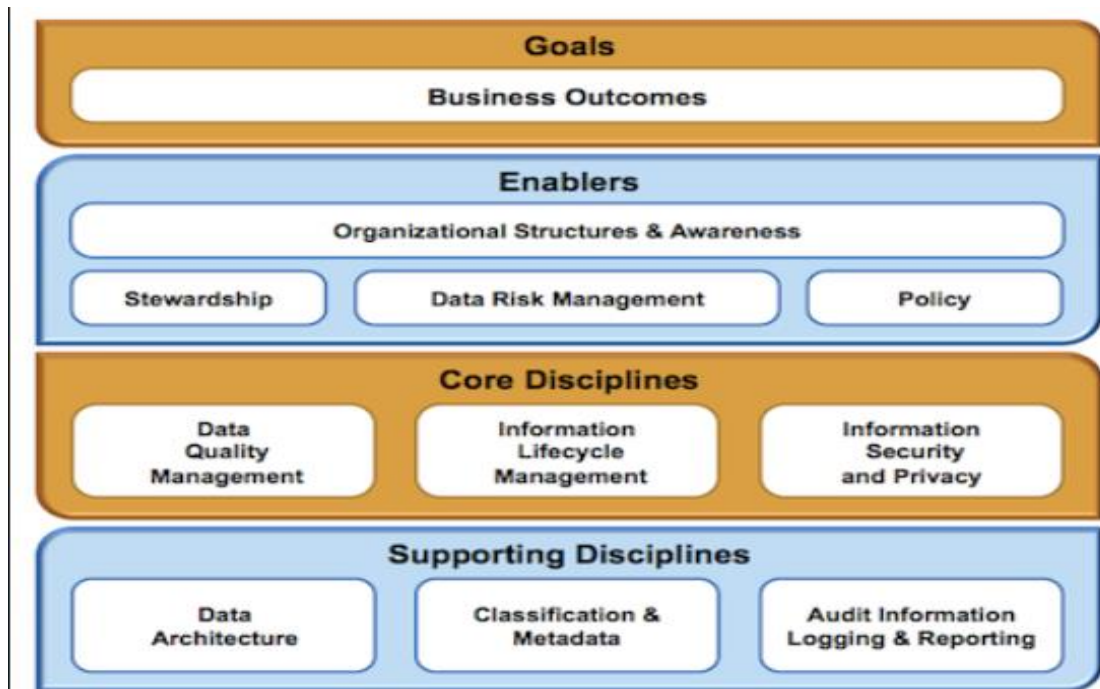
Figure 6: IBM Information Governance Council Maturity Model (Source: IBM 2012)

2.12 Summary of Chapter 2

In this chapter, the researcher reviewed existing literature on Big Data and the governance implications that comes along with it. The researcher performed a review of the literature through various sources. The literature review is presented in sections. Section 2.1 provides a broad overview of governance. Section 2.2 discusses corporate governance, its connotations, as well its possible benefits to an organization. Section 2.3 explores IT governance, analyses its relationship with the business and how it can be best aligned. Section 2.4 deliberates on the value of data and how organizations are leveraging on data as an enterprise asset.

3. CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter will deal with the description of the philosophy that was employed to gather data, followed by the applied approach, the strategy used to conduct this research, the chosen methods for data collection and analysis, the efforts to ensure reliability and validity of the study and ultimately the ethical considerations.

A research methodology provides guidance on how a study should be conducted to achieve the research objectives (Khothari 2005; Schwardt, 2007). Welman, Kruger and Mitchell (2005) concur that a research methodology is a structured manner of conducting and managing a research process through its sequences, procedures and systems. Hart (2005) adds that a research methodology is extremely important for demonstrating how relevant are the chosen data collection techniques and the approaches the researcher has employed in the study. It features the strategy, the action plan and the chosen design which inform the selection and employment of specific methods in order to achieve the research objective (Gray, 2004).

It is important for a researcher to employ very solid theories to support their research methodology, the latter should be coupled with a comprehensive understanding of different methodologies to guarantee an alignment between the researcher's position, research questions and the research objectives (Austin & Sutton, 2014). Creswell (2003) suggests that there are different factors that need to be considered when the researcher decides on a research methodology. It is very critical for the researcher to take note of the research problem, the researcher's experience and research audience for guiding the selection of a methodology.

3.2 Research Philosophy

This research adopted an interpretive paradigm based on the purpose that it seeks to achieve its goal of analysing the data governance implications presented by Big Data. Choosing an interpretive approach as the research framework enabled the researcher to gain a comprehensive understanding of the dynamics and challenges in governing Big Data

When undertaking a research project, philosophical guidance becomes a very key requirement. Lukka (2014) mentions that philosophical orientation is significant in guiding the research action. This section discusses the most appropriate research philosophy for this specific study. According to May (1997) the process of research involves experimental work being done with the data

collection. This outcome can agree, contradict or challenge theories and that stresses the need of different observations.

Each and every research is based on the underlying philosophical paradigm about what defines a valid research and which approaches are deemed appropriate for that particular study, for it to make a contribution to the development of knowledge (Oates, 2006).
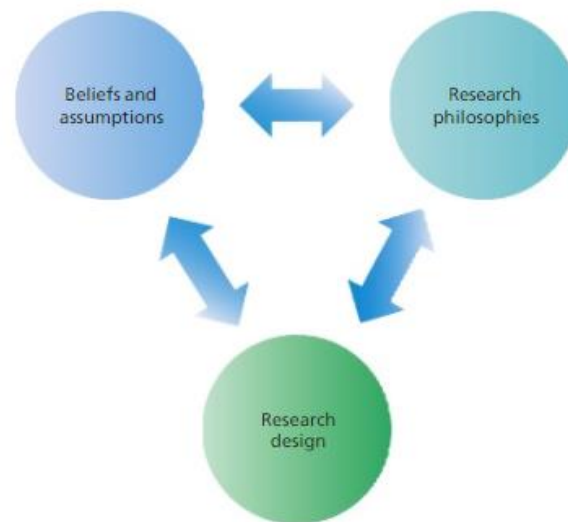
Figure 7: Developing your research philosophy: a reflexive process (Source: Saunders et., al 2015)

Saunders, Lewis, Thornhill and Bristow (2015b) emphasize that a researcher follows certain methods and techniques as shown in Figure 7 to reflect on personal beliefs and assumptions regarding the philosophies available and the research design that was used for the specific research. Saunders et al. (2015b) further conclude it is important to discover the possible clash between the preferred philosophy and the researcher's beliefs and assumptions.

A paradigm is defined as a collection of beliefs or assumptions held by the researcher during the examination of a phenomenon (Guba and Lincoln, 1994; Saunders, Lewis & Thornhill, 2015a). The researcher brings these beliefs into the research process and they are often referred to as 'worldviews' Bhattacherjee (2012) mentions functionalism, interpretivism, radical structuralism and radical humanism as four paradigms of social science research. In various fields of research, interpretivism and positivism remain the two most dominant research paradigms or philosophies (Collis and Hussey, 2014).

3.2.1 Interpretivist research philosophy

Saunders et al. (2015b) claim that the defining factor of the selected research philosophy is the research questions and objectives. Interpretive research aims to interpret how people see the world. Ponelis (2015) concurs that the

interpretive paradigm is driven by a need to acquire an understanding of the world as it is from a subjective point of view and seeks that understating within the same frame of reference of the participant rather than that of an observer.

Interpretivist research philosophy believes that social reality is based on subjective interpretation of actions and that access to reality can only be attained through social constructions such as language, consciousness and shared meanings (Kaplan and Maxwell, 2005). According to Myers (2009) interpretive investigators are more likely to focus on meaning of context.

3.3   Research Method

This study used a qualitative research method to develop a broader understanding of the topic and address the main research question. Qualitative research is suitable for small samples and the results are not measurable or quantifiable. This enables the researcher to broaden the research scope without limitations (Collis and Hussey, 2003).

Maxwell (2012) confers that a qualitative research design possesses a flexible structure and that enables the construction and reconstruction of the design when needed. This therefore means an issue can be analysed in detail and more accurately using qualitative research methods and the participants are afforded the freedom to express what is consistent for them (Flick, 2011).

Maxwel (2012) asserts that this process entails activities such as data collection and data processing, theory construction and transformation, analysing the initial research questions and the detection of possible threats that can potentially flow in parallel and then get to influence each other at a later stage.

It is therefore important that an efficient and appropriate method is prepared before initiating a research operation. Kothari (2004:32) asserts that thoughtlessness in the method of a research project increases the chances of the research exercise being futile. To avoid the latter, the researcher took a reasonable amount of time in the process of identifying the most appropriate data collection method, sampling strategy and data analysis approach to supplement concrete information and support the study in addressing the research problem.

3.4   Research Design

Cooper and Schindler (2011) state that the term "research design" can be defined in several ways. In their definition, they describe it as a structured plan and structure of how an investigation will be conducted to acquire unambiguous solutions to the research questions. Durrheim (2004) defines research design as a strategic framework for action serving as a link between the initial research questions and the research operation and implementation of the strategy.

After the research questions and objectives were defined, a design had to be selected in order to fulfil the research objectives and also answer the defined research questions. Zikmund (2003) referred to the usefulness of categorising research based on function arguing that the nature of the problem would influence the choice of research design.

The study seeks to understand Big Data governance implications and an exploratory case study research was adopted. This research area is less explored and case study research was deemed to be most suitable. Amid all the limitations, it could be established that case studies serve the researcher with more advantages than the disadvantages mentioned in the literature.

Moreover, the case study research design was the most appropriate as the study seeks to take a deep dive into the Big Data governance phenomenon. A case study design allowed the researcher to conduct the study in a natural setting while acquiring valid knowledge about big data governance in various organizations to later develop theories from the practice (Cheong and Chang, 2007). Yin (2014) concurs by defining case study research as a method that enables a scientific in-depth investigation of an existing phenomenon within a defined environment or setting.

Case study research focuses primarily on the formulation of hypothesis which is commonly regarded as theory building (Eisenhardt, 1989, as cited in Teegaravapu et al., 2008). Yin (2006) states that case studies are one of the effective options of conducting research and are relatively used to gain a deeper insight about an individual, process, events and at times a group (Hyland, 2016). This supported the selection of a case study as this study strived to obtain a deeper insight of processes and methods employed by organizations to govern Big Data.

Yin (2014) suggests that a case study can be best used when the main purpose of the study is to provide answers to the "how" and "why" questions and the researcher cannot influence or manipulate the event or behaviour of the participants. In this study the researcher had no influence or control of how the participants behaved and the goal of acquiring answers to the "how" and "why" questions was one of the objectives of this study. Merriam (2009) recommends the use of case study when the study is mainly committed to gaining a deeper understanding of the social processes that occur in a particular context than the relationships that exist between variables.

Yin (2003) classifies case studies as descriptive, exploratory and explanatory. For this study, an exploratory case study was employed to explore and gain a comprehensive understanding of the Big Data governance phenomenon and its implications.

Provided all the various categories, Collis and Hussey (2003) claim that case studies are relatively defined as exploratory research and are commonly employed in under researched areas with minimal theories or insufficient body of knowledge. Flyvbjerg (2011) adds that a large fraction of the available

empirical knowledge in various and disciplines is produces by case study research.

Despite all the positives and advantages of case studies, several authors have unequivocally expressed their criticism and pointed out limitations. Case study research is often criticised for its lack of scientific rigour, requires large numbers to provide meaningful outcomes and it limits the ability to generalize (Yin, 2009). Flyvbjerg (2006) contends that the lack of reliability, and validity is a limitation that is frequently mentioned by the critics of the case study as a scientific method. However, these issues are generally associated with qualitative methods at large (Willis, 2014). To establish trustworthiness for this study, key components such as credibility, transferability, confirmability and dependability were used, and further explanation is given in section 3.8.

3.5   Research Approach

This study employed an inductive orientation. Utilizing an inductive approach, the study commenced with specific observations, gathered evidence to identify patterns, and ultimately drew conclusions (Saunders, Lewis & Thornhill 2019).

The research questions have a direct influence on the choice of a research approach where the adoption of any of the three known approaches does not entirely resolve the research problem, while a combination of research approaches achieve that (Creswell and Plano Clark, 2011). A research approach is an implemented strategy with a goal of collecting relevant data using specific methods. A researcher has an option of choosing between qualitative approach, quantitative approach and mixed (qualitative and quantitative) approach.

It is important to consider other research world interpretations that are within the research methodology prior to choosing a specific approach for a research study (Ihuah and Eaton 2013). Saunders et al. (2009) present two major research approaches, namely deductive and inductive. Teddlie and Tashakkori (2009) add that a researcher can choose to use inductive or deductive in no particular order depending on how the researcher has advanced in studying the phenomenon. For a long while, researchers have been at loggerheads over which approach is the most appropriate approach for conducting research and collecting the relevant data, nonetheless the two approaches are not mutually exclusive and are capable of responding to the same questions using different methods (Soiferman, 2010).

The inductive reasoning proceeds from a specific or set of premises with limited scope towards a generalized conclusion. Substantial scientific research is carried and data is collected and analysed with an intention of understanding the phenomenon to proceed towards theories and generalisations with no intent of validating existing theories (Gray, 2013). While inductive reasoning begins and move from specific towards general statement, deductive reasoning is the

inverse of that, it moves from general principle and ends with the specific case (Williamson, 2002). Creswell and Plano Clark (2007) state that the deductive method of reasoning is often referred to as a 'top down' approach. This approach commences with solid empirical evidence and intends to build theories and test hypotheses. Contrary to deductive reasoning, inductive reasoning is regarded as a 'bottom-up' approach and it aims to search for themes and patterns from observations with an aim to develop a deep understanding and explanations. To develop an understanding and meaning from the interview transcripts, the researcher used an inductive coding in identifying themes and patterns from the data.

3.6   Data Collection Methods

Data was collected through audio recorded semi-structured interviews. The target population was individuals in the Big Data department or division with extensive knowledge of Big Data technology coupled with some knowledge of data governance in general. This means that the title, experience, role and responsibilities played a leading role in selecting participants.

This study adopted a qualitative research paradigm, that means a qualitative method of data collection was used and, in this case, semi-structured interviews was selected. According to Easwaramoorthy and Zarinpoush (2006), interview is an appropriate method for collecting in-depth information on opinions, perceptions, feelings, and experience and they are more effective when the subject of inquiry requires extensive questioning and probing.

3.6.1   Semi-structured interviews

The notion behind the selection of semi-structured interviews as a data collection method was to:

- Allow participants to further elaborate on the topic.
- Understand opinions, experiences and practises of the participants about the topic.
- Have a clear guide and plan of the topics to be covered.
- Afford a discretion to paraphrase and clarify vague responses for positive interactions.
- Gain an in-depth experience of the implications of governing Big Data technologies in various organizations.

When conducting a semi-structured interview, the researcher asks a standard set of questions to the respondents (See Appendix A). Nonetheless, the researcher has the liberty to pose follow-up questions should there be a development of thought-provoking or new line of inquiry throughout the interview process (Young et al., 2018). This kind of interviews can yield rich material that cannot be obtained in any other method, which can support or be

supported by other data from questionnaires and standardised test responses. The interview is easily adaptable and allow more flexibility (Verma & Mallick, 1999).

The aim of the researcher was to develop the interview questions against a pre-existing theoretical proposition as the study seeks to analyse the governance complications that are presented by the introduction of Big Data on data governance. The main objective of the interview questions was to explore and understand the governance measures and process that organizations have in place in order to govern Big Data. The questions also aimed to obtain more insight about the challenges that organizations face when trying to govern Big Data and their approach in dealing with these complications. Those who agreed to participate were given a consent form to complete and they were duly informed about the main objectives of the study.

3.7   Sampling Strategy

Fundamentally, sampling techniques are classified into two types: probability and non-probability (Alvi, 2016). The research used purposive non-probability sampling as the sampling method for data collection. According to Lewis and Sheppard (2006) purposive sampling is a non-random sampling technique that affords the researcher the deliberate choice of participants by identifying certain qualities that a participant possesses. This meant that the researcher had the discretion to define the criteria and characteristics that need to be met in order for an individual to qualify as a participant.

Notwithstanding the researcher's initial decision to use purposive non-probability sampling for this study, the research had characteristics of convenience sampling in some instances. Taherdoost (2016) defines convenience sampling a type of non-probability sampling where research subjects are selected to be part of the sample merely because they are conveniently available to the researchers. The study was conducted during a period where the COVID-19 pandemic was at its peak and that had dire impact on the wellbeing and wellness of some research subjects. This hindered them from participating in this study and that obliged the researcher to act swiftly and recruit available sources with similar attributes.

The choice of a smaller sample was influenced by Turner (2020) when suggesting that sampling from a smaller population is relatively more feasible when compared to a larger population and it saves the researcher ample time and cost. Panneerselvam (2004) also recommends the choice of a sampling strategy has a very huge effect on the budget and timelines of the study, hence the researcher applied much attention and logic before deciding on the sampling plan.

In research, some individuals in the representative group are likely to be more or less credible, bias, or experienced about the phenomenon being studies than

others (Wellington & Szczerbinski, 2007). In this study, the participants had to fulfil the criteria shown in Table 2C to qualify:

Table 2: Sample details

Respondents	No of participants	Criteria
Managing executives	2	**The participant must have extensive knowledge of Big Data coupled with some knowledge of data governance in general.**
Head of Departments	2	**The participant must have extensive knowledge of Big Data and governance**
Big Data Scientists	10	**The participant has extensive knowledge of Big Data and its lifecycle**
Big Data Specialists	2	**The participant must have extensive knowledge of Big Data and its dynamics.**

To confirm whether the participants met the stipulated criteria and would add a valuable insight, their current and past roles, their length of experience, the qualifications they hold and the department or division they work in are some of the elements that were considered. As shown in figure 8 below, a role that dominated the sample was the Big Data scientist role with 10 participants while others had 2 participants each.

3.7.1 Roles of participants

No of Parcitipants by Role

2 (12.5%)
2 (12.5%)
2 (12.5%)
10 (62.5%)

Role
● Big Data Scientist
● Big Data Specialist
● Head of Department
● Managing executive
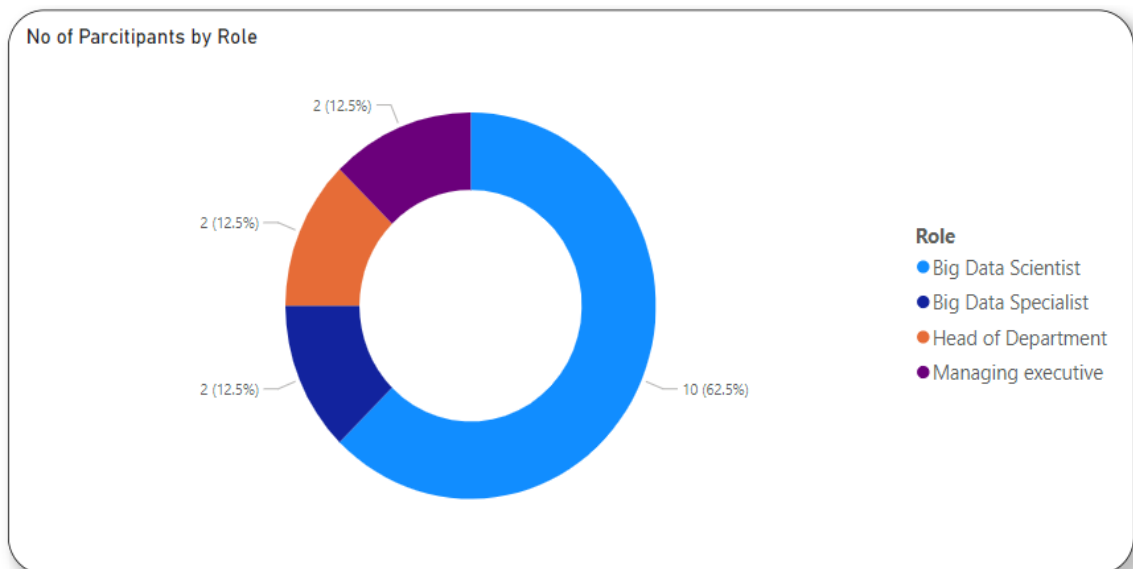
Figure 8: No of Participants by role

Nonetheless, various stakeholders such as legal experts and individuals who are more knowledgeable about the subject of governance were welcome to participate in the study. Their contribution was greatly appreciated, and this is largely more about their involvement in the development of governance policies and frameworks.

3.7.2 Years of experience

Figure 9 indicates that a majority of participants were in the range 16 to 21 years of experience. This denotes that those who participated in the study were significantly aware of the Big Data governance subject and were more knowledgeable about the implications of governing Big Data.
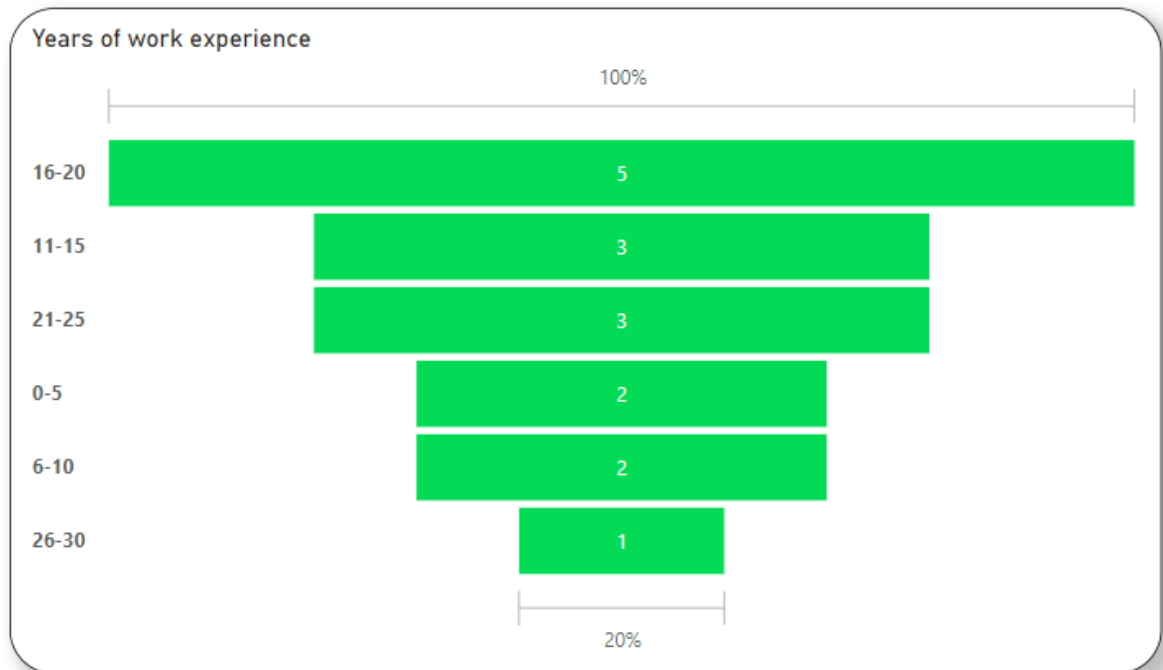
Figure 9: No of participants by work experience

3.8   Case Description

3.8.1 Case 1

This is one of the fastest growing banking institutions in South Africa and it is also one of the authorized financial services and credit providers. The banking products and services offered by the institution vary from savings accounts, cheque accounts, retail banking, personal loans, commercial loans and home loans.

The institution has widened its range of banking and financial products over the years, and it is now offering a comprehensive range of these products to corporate, private, and commercial customers in the country. This institution continues to expand its footprint and networks in all the South African

provinces. The institution has managed to multiply its operations all over the country and currently opening more branches across several provinces.

At the time of the study, the number of employees working in this bank was more than 11000 across all its areas of operation. The organization's commitment towards gaining competitive advantage and improving customer experience has led to the adoption of Big Data to improve various banking features and these include aspects such as targeted advertising and marketing, customer behavioural insight, fraud prevention and detection, risk management, personalized banking solutions and offers.

In the past few years, the organization has been intensifying its efforts towards identification of the critical skills and recruitment of the industry experts in the Big Data space in order to leverage the data at their disposal.

3.8.2 Case 2

This is one of the leading telecommunications and mobile service provider in the African continent with its headquarters located in Johannesburg. Apart from being the provider of both fixed and mobile communications, the company offers a variety of business solutions including insurance and financial services. This organization has achieved a massive amount of growth in the couple of decades and this can be attributed to its diversification of revenue and expansion of markets across the African region.

The company is currently offering its products and services in approximately 7 African countries and partners with several technology and telecommunications companies globally. The organization serves a variety of customers and these include customers in both public and private sector. Its entrance into new markets and innovation product lines and services has prompted the expansion of the workforce and attainment of contemporary skills, technological infrastructure and expertise. At the time of this study, the company had approximately 8000 permanent employees and continues to employ individuals with the required skills to expedite its digital transformation agenda.

3.8.3 Case 3

This company was established in 2006 in the Western Cape, South Africa. Initially the primary purpose of the company was to serve the insurance needs of the low and middle-class population. At the moment the company offers a wide range of insurance products to corporate and individual clients. One of the organization's missions is to expand its operations and accessing even more of the South African that need assistance. This is one of the companies that has been spearheading the designing and development of comprehensive covers and also a forerunner in bringing insurance to township and rural population.

The company's footprint includes offices in the Western Cape, Eastern Cape, Gauteng and KwaZulu Natal provinces of South Africa. At the moment, the company has a workforce of about 120 individuals, and it is devoted to the development of digital skills and boasts a clear digital transformation strategy. In addition, the company is making use of machine learning analytics to grow insurance sales through personalized recommendations on policies.

3.9   The Participants

To address the research questions, a total of 16 participants were interviewed as illustrated in Table 3 below.

Table 3: Participants Codes

Code	Case	Role	Work experience in years
P1_C1	Case 1	Managing executive	20+
P2_C2	Case 2	Managing executive	30+
P3_C2	Case 2	Head of Department	20+
P4_C2	Case 2	Head of Department	25+
P5_C1	Case 1	Big Data Scientist	15+
P6_C2	Case 2	Big Data Scientist	8+
P7_C3	Case3	Big Data Scientist	13+
P8_C3	Case 3	Big Data Scientist	25
P9_C1	Case 1	Big Data Specialist	5+
P10_C1	Case 1	Big Data Specialist	10
P11_C2	Case 2	Big Data Scientist	17+
P12_C2	Case 2	Big Data Scientist	21+
P13_C2	Case 2	Big Data Scientist	11+
P14_C2	Case 2	Big Data Scientist	20+
P15_C3	Case 3	Big Data Scientist	3+
P16_C3	Case 3	Big Data Scientist	21+

3.10 Data Analysis Methods

The data analysis process is very significant for the researcher to acquire extensive understanding of the data collected. Qualitative data analysis is mainly concerned with transforming the qualitative data that was collected and this is performed using various methods of data analysis with an aim of uncovering insights (Gibbs, 2007). This section drills down into the methods that guided the analysis of data for this research.

The study selected qualitative approach of thematic analysis as a method of analysis. Judger (2016) claims that thematic analysis is the most widely used method of analysis in qualitative research studies. According to Alhojailan (2012), thematic analysis is the most appropriate method for a study with an aim of making discoveries using themes. Furthermore, Alhojailan (2012) confers that thematic analysis offers a systematic element to the process of data analysis and it also provides the researcher with advantage of associating the theme's frequency analysis with one of the complete contents.

In efforts of conducting an in-depth analysis of the interview data, the researcher used an inductive coding to code the interview transcripts. According to Braun and Clarke (2012) this type of approach is bottom up and it is determined by the contents of the data. The authors add that themes and coding is generated from the content of the data, essentially the researcher does not commence from a supposition of what codes should be but allows themes to develop during the coding process.

In qualitative research, identifying and discovering themes is one of the most critical tasks, it also carries the potential of being the most confounding task (Ryan & Bernard, 2003). Saldaña (2021) concurs that the task of developing themes is normally as challenging as coding as it demands the researcher to contemplate on the actual meaning of the data at hand. This study obtained themes from the IBM Information Governance Council Maturity Model and subsequently follow the six phases of conducting thematic analysis (see Figure 10) as guided by Braun and Clarke (2006) to pinpoint, examine, and record themes or patterns within the collected data.

Phase 1: Familiarising yourself with the data
Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.

Phase 2: Generating initial codes
Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.

Phase 3: Searching for themes
Collating codes into potential themes, gathering all data relevant to each potential theme.

Phase 4: Reviewing potential themes
Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.

Phase 5: Defining and naming themes
Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.

Phase 6: Producing the report
The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Figure 10: Phases of thematic analysis (Source: Braun and Clarke 2006)

3.11 Reliability and Validity

Realiability and validity improves transparency and minimizes the possibilities of bias in a research study (Singh, 2014). Reliability refers to the degree to which the research method can obtain consistent results across time. Reliability focuses on consistency of the measurement and not its validity Bhattacherjee (2012) states that a measure can be reliable and invalid at the same time.

Reliability and validity are perceived to be different in qualitative studies as compared to quantitative. Lincoln and Guba (1985) suggest the use of key components in establishing trustworthiness such as credibility, transferability, confirmability and dependability instead.

Credibility: In an endeavour to ensure credibility of the study, member check was used. This means participants were granted an opportunity to validate the research outcomes and provide feedback.Lincoln and Guba (1985) regard member check as a key technique for establishing credibility in of research findings.

Dependability: Dependability is the stability and reliability of the research findings over time (Lincoln & Guba, 1985). According to Merriam (1998), ensuring the dependability of a qualitative study involves assessing if there are no faults in the collection of data, interpretation of findings and reporting of the outcomes. Duteous attention was devoted in ensuring that substantial evidence of the research methods is presented precisely, traceable, and well documented (Tobin and Begley, 2004).

Transferability: Transferability refers to the extent to which the findings of a study can be transferred from on setting to another (Shenton, 2004). According to Lincoln and Guba (1985) the onus of describing the research design, participants and the overall methods lies with the researcher and this is very key in enabling the audience to judge whether these findings can be transferable to their preferred setting. In this study, thick and rich description of concepts, the experimental context in which the research was conducted and the methods used for data collection are well articulated and documented for the potential audience.

Conformability: Confirmability is concerned with the neutrality and accuracy of data interpretation (Tobin and Begley, 2004). Ideally, the confirmability criterion should be established after the study has achieved credibility, dependability and transferability (Guba & Lincoln,1989). According to Seale (1999) an audit trail is one of the key methods in establishing confirmability. This means a researcher is able to produce methodological evidence to corroborate the research findings. In pursuit of conformability, an audit of this research has been saved and the collected data has been effectively archived in both manual and electronic forms.

3.12 Ethical considerations

Any research that features human as subjects is bound to encounter several ethical challenges from the beginning to the last phase of the research. These includes concerns that relate to the rights, privacy, dignity and sensitivities of the participants (Cohen, Manion, & Morrison, 2000; Orb, Eisenhauer & Wynaden, 2001). Creswell (2003) confers that the researcher has a duty to guarantee that the rights, values, needs and desires of the subjects are respected. In order to realize the latter, it is therefore critical for the researcher to uphold ethical principles such as anonymity, informed consent, confidentiality and the potential impact or harm on the participants prior to conducting, during

and after the study (Sanjari, Bahramnezhad, Fomani, Shoghi & Cheraghi, 2014).

Below is a detailed description of how the ethical principles have been addressed in this study:

> **Informed Consent:** Prior to their participation, the participants were thoroughly informed about the purpose, objectives and the nature of the study. Furthermore, the researcher clearly stipulated that the data was collected merely for academic purposes and the participants gave consent for interview to be scheduled.

> **Privacy, confidentiality, and anonymity:** The researcher assured the individuals that their identity was preserved at all cost. Essentially, this suggested that any identifiable characteristics of the participants was eliminated to preserve their anonymity and pseudonym was used instead.

> **Harm and risk:** The principle of non-maleficence compels the researchers to ensure that participants are not imposed to any psychological or physical harm when participating in the study (Beauchamp & Childress, 2001;Bertram & Christiansen 2015;66). In this research, substantial amount of attention was directed towards preserving the participants from being imposed to harm or risk of any sort.

> **Honesty and trust:** The researcher ensured that all the ethical guidelines of the Cape Peninsula University of Technology's research ethics committee are followed. The data was collected, analysed and findings were reported in an honest and trustworthy manner without any manipulation.

3.13 Summary of Chapter 3

According to McGhee, Marland and Atkinson (2007) a researcher should develop a decision trail to provide the rationale behind methodological choices and all other decisions that were made during the course of the study. In this chapter the researcher strived to construct an outline to the decision trail in a logical manner by describing the notion behind the design and methodological choices made.

Moreover, the choice of a qualitative data collection technique such as the semi- structured interviews including the sampling strategy and thematic analysis as a data analysis method was clarified. This chapter also consists of the measures that were used to evaluate the quality and credibility of this study including the strategies used by the researcher to establish trustworthiness for every criterion.

Ultimately, ethical considerations in accordance with the ethical principles as mentioned by Cohen et al. (2000) were discussed in depth. The next chapter presents the findings obtained from the interviews with a detailed analysis.

4. CHAPTER FOUR: FINDINGS

4.1 Introduction

In this chapter, the findings obtained from the interviews that were conducted with various participants are communicated. The main purpose of this study was to gain a deeper understanding of the data governance implications that are presented by the introduction of Big Data in various organizations. To achieve the main objective of this study, an examination of the current Big Data governance methods and strategies had to be made. Furthermore, the study also attempted to understand the existing dynamics and the perception of Big Data governance in the organizations.

Subsequent to interviews with the selected participants in various roles and organization, certain themes emerged during the conversations and they encapsulate key empirical findings of this study. These themes are discussed extensively in the following sections and verbatim quotations from the raw data extracted from interviews are used to demonstrate significant findings.

4.2 Theme 1: Regulatory compliance

The discussion below indicates the level of commitment and efforts that organizations are making towards ensuring that compliance with the growing data regulations and legislations is achieved. This theme emerged from the respondent's response to the interview questions that are below:

Research Sub-question
What governance strategies, processes and methods are currently employed by the organizations to govern Big Data?
Interview Question(s)
What measures do you have in place to ensure that compliance with the data regulations and legislations is achieved?
Do you understand privacy regulations and how they affect the use of Big Data?
What data regulation or legislation is the most critical for your role?

Regulatory compliance emerged as a theme that most participants are concerned about. Their concern is clearly justified as failure to comply with the regulations applicable to Big Data comes at a high price for the organization. Nonetheless, the efforts and processes to safeguard compliance are cumbersome and not clearly understood by those involved in daily operations. Big Data has further propelled the increasing need to comply with local and

global regulations and thus adding to the pool of challenges for business, organizations are having difficulties in finding the balance between improving business performance while ensuring maximum compliance to applicable regulations. The risk mitigation systems and processes meant to ensure compliance appears to be extremely sophisticated, these complications are unlikely to be caused by limitations in these solutions but rather by lack of the necessary skills.

P1_C1: *"We are very critical of the privacy and data protection regulations, we do have relevant systems and process in place to proactively oversee and monitor that. It is not as if we have a choice because failure to comply to the government regulations can cost us a lot of money and possibly destroy the business completely."*

A similar sentiment is echoed by P2_C2 below, also emphasizing on how crucial it is for the organization to comply on the data protection regulations:

P2_C2: *"An integral part of my role amongst many is overseeing compliance to all pertinent or applicable regulations. The nature of our business coupled with the magnitude of business compels us to adhere to rules and regulations provided to us by the government. We have seen how global brands and organizations have been dragged down by the media for non-compliance".*

P4_C2: *"We try our outmost best to abide by the Big Data rules and regulations of our country and global ones, we are a company that trades and also makes use of data from other parts of the world."*

Even though most of the participants demonstrated a high degree of commitment towards ensuring that their respective organizations do comply with the relevant regulations, a vast majority of them are much more knowledgeable about the POPI Act than any other existing regulations. This is evident in the quotes below:

P5_C1*: "Several individuals within the business have access to personal data or information on a daily basis, this means we have constantly avoid breaches and promote good conduct. We have achieved that by conduct mandatory POPI training sessions for all our employees through our training portal."*

P1_C1*: "I need to be able to provide all the necessary evidence of POPI compliance to the auditors at any given time. This means that I am responsible for spearheading the development of processes and procedures for complying to the POPI Act. This is mostly vital for our Big Data environment considering the kind of attributes and data we handle."*

The above responses clearly reveal which legislation and regulation the participants are more concerned and cognizant of. A couple of the respondents alluded that they have opted to channel most of their efforts and resources towards ensuring that compliance to the POPI Act is achieved. Due to the nature of their organization being a multinational corporation with a market presence that spreads over several continents, complying to regulatory requirements of various continents becomes very essential. One of the elements cited by these participants, is the similarities that exist between POPI and GDPR, thus being POPIA compliant can be translated to being GDPR compliant.

Judging from the responses, one of the contributing factors towards the prioritization of the GDPR and POPI compliance is the fear of the penalties and loss that could be incurred when found guilty of noncompliance. This is indicated in the extracts below:

P7_C3: *"The fine you can get for breaching GDPR or disseminating personal information is up to 4% of your gross global turnover and this can potentially kill your business as well as its brand reputation in a very bad way."*

P8_C3: *"Our leadership is extremely critical of rules such a POPI, I can attribute this to the dire consequences that come with the failure to comply. If I am correct, exorbitant penalties of up to the value of R10 million can be imposed and one can possibly be imprisoned for approximately 10 years. Any sane individual would try by all means to play safe and remain on the right side of the law."*

Certainly, the existing laws, legislations, rules and regulations have a major impact on many companies and it is even more on multinational companies. ability to meet their data privacy regulatory requirements. Some companies are finding hard to meet all the requirements and this can be attributed to on several factors such as employee awareness, the costs involved and the lack of the relevant tools for driving and monitoring compliance to the rapidly broadening regulatory requirements.

4.3   Theme 2: Policies

Policies are an integral part of data governance and management in any organization or environment. The purpose of policies is to ensure that data is protected and has the required integrity, by applying a formalised set of policies is an indication of how serious an organization is about protecting its data and that of employees and customers.

This theme emerged from the respondent's response to the interview questions that are linked to the research sub-question below:

Research Sub-question
What major effect has the adoption of Big Data had on traditional data governance?
Interview Question(s)
Do you have any set of policies documented specifically for Big Data governance?

The responses provided by the participants reveal how slow or unwilling organizations have been to review and develop Big Data specific policies. This is echoed in the following responses:

P12_C2: *"We do not have policies that are formulated for Big Data as we have not identified a need for that in our environment but who knows? Maybe with time we might be able to see the need."*

P7_C3: *"I feel like the policies we have in place at the moment are very formidable and strong enough, thus we continue using the same policies to govern Big Data."*

P12_C2: *"In my opinion, big or small data is the same and that means the policies and rules that are governing data can be the same because the only distinction between the two is just size."*

Some of the participants have the perception that the existing policies are equally suitable for Big Data. The responses are revealing that most of these organizations do not have Big Data tailored policies in place but continue to rely on existing policies.

It appeared that the organizations have not made any efforts to perform a thorough analysis on how the adoption of Big Data can potentially affect or trigger improvements in the current policies.

4.4 Theme 3: Stewardship

Stewardship is a very crucial element of any data governance program. When executed effectively, the quality and integrity of the data is improved and data as an asset is well preserved in an organization. As some organizations are still in their early phases of Big Data governance, Big Data stewardship remains a new terrain that many organizations are still trying to figure out and others have not yet explored. This theme emerged from the responses by the participants to the interview questions that are linked to the research sub-question below:

Research Sub-question
How do organizations deal with the Big Data governance challenges?
Interview Question(s)
How do you address the stewardship of Big Data?

P1_C1*: "Stewardship is one aspect that we are trying to champion in our governance and management program. We understand the value of having good stewardship in place and we have been missing out on its benefits but that will soon be a thing of the past as we are busy working on the stewardship of our data."*

P6_C2*"The previous management has constantly overlooked data stewardship but the current management as went an extra mile to ensure that the organization receives enough stewardship support, and that process is currently in its initial phase."*

Several factors such as the budget, resources, skills and management support have been mentioned as the main challenges for the absence or perhaps poor data stewardship. This is cited in the quote below:

P14_C2: *"Due to a lack of resources and the necessary knowledge, we have not been able to develop a concrete data stewardship program. Having a couple of individuals or team that is responsible for data stewardship in our business unit would be a great thing but for now it remains part of our discussions."*

P8_C3*: "We have not reached the point of having data stewardship in place even though we have received several recommendations to do so. The challenge has been budget shortages and lack of the needed knowledge."*

The majority of the respondents stated that Big Data stewardship is currently not a priority for their respective organizations due to ambivalent need for this kind of program and some have mentioned the difficulty to quantify its direct benefits to the business. This is cited below:

P6_C2*: "To be quite honest to you, this company has been more concerned about the return of its Big Data investment and everything else has not been the in the list of priorities for our management. Less or no interest is shown on things such as stewardship, the leadership has thrown all its weight in ensuring that we extract the gains from our Big Data investments."*

P15_C3*: "We do have all the willingness and urge to implement important programs such as stewardship, that demands resources and at this point we are understaffed, and we have no other choice but to distribute responsibilities amongst the resource we have."*

Nonetheless, some respondents mentioned that they are slowly attempting to implement stewardship programs even though they are still at early phases while some respondents have unequivocally expressed their willingness to have stewardship in place but several factors have been limiting their efforts.

4.5   Theme 4: Data Quality Management

This theme emerged from the responses by the participants to the interview questions that are linked to the research sub-question below:

Research Sub-question(s)
What major effect has the adoption of Big Data had on traditional data governance?
Interview Question(s)
How has the adoption of Big Data affected how you manage or measure quality?

The respondents felt that managing the quality of Big Data is such a mammoth task in the Big Data environment. This is mainly attributed to a list of factors such as a wide variety of sources, data structure incongruities and time constraints. This is echoed in the quotes below:

P9_C1*: "Our data quality metrics have changed dramatically. We no longer see some of the data quality issues in small data as issues in Big Data. Managing quality has not been a walk in the park in this environment but rather more challenging and complex that it was in small data."*

P13_C2*: "Managing quality of Big Data can be a very strenuous and time-consuming task, but we really cannot complain, instead we channel a lot of our energy and resources in ensuring that our data is of the highest degree. Our Big Data originates from multiple resources and through scripts we transform it to the preferred formats. However, we often face a challenge in normalizing such data thus business involvement becomes a critical factor in helping to determine data quality issues."*

P10_C1*: "By nature, this kind of data is very cluttered, unlike the traditional and some of its sources can be even more messy. Having a standard process of managing quality is almost impossible due to the structural variations of the data. In some instances, data quality metrics or rules are customized for each data set or source, whereas in traditional data rules are standard overboard. In Big Data, the level of acceptable consistency or accuracy for a specific source may not be the same for next one and some of these are entirely dictated by the business use cases or requirements."*

P16_C3*: "The definition of data quality has taken a whole new shape, we now find ourselves sweating over formatting issues and challenges. This is a very challenging aspect of Big Data in my opinion, mainly because it makes our lives difficult when we seek to gain real time and deeper insight on a data set."*

However, one respondent was adamant that there is no need to review or amend the data quality management program in place. This responded believes that Big Data is not special from the traditional data except albeit being larger by volume. This is highlighted in the quote below:

P4_C2*: "Quality is quality in any kind of data. I do not see the need to change how we manage our quality or perhaps any need for fine tuning it at the moment. Big Data is no different barring its huge volume so it deserver similar treatment."*

Achieving a highest degree of quality seems to be relatively difficulty in a Big Data environment. This can be attributed to the variety of Big Data sources which are mostly external and in some cases very messy in structure or format. The variety, velocity and volume of Big Datasets, human intervention alone becomes insufficient for fixing data the data quality issues faced by organizations. At the moment, the tools and technology to manage the quality of Big Data is not yet fully fit to master the Big Data quality complications and as such more measures must be put in place.

4.6   Theme 5: Skills and Roles

This theme emerged from the responses by the participants to the interview questions that are linked to the research sub-question below:

Research Sub-question
What are the perceived Big Data governance challenges encountered by the organizations?
Interview Question(s)
What effect has the adoption of Big Data had on the skills and job roles within your organization?

In this section, the researcher aimed to discover and understand the recent developments in roles and the skills readiness in these organisations. Judging from these findings, a lot of roles have developed and some have transformed in various organizations but a noteworthy gap exists between these new roles and the required skills to execute duties.

The existing skills gap can be attributed to factors such as failure to plan effectively for future staffing requirements and the inability to fine-tune human resource and talent strategies in order to meet the rapidly shifting demand in talent and skills.

One of the major factors is the failure or delay in upskilling the existing workforce as the demand old skills may be obsolete or not less effective over the years.

These sentiments are alluded by respondents below, also detailing their efforts to close the existing gap:

P1_C1*: "We have learnt that a wide range of Big Data and analytics skills are very much desirable for the business to start seeing value in our Big Data investment. The traditional roles and responsibilities have been shifted immensely, some of the responsibilities that were previously key are now automated and this leaves with no other choice but to continuously pursue the needed skills."*

P4_C2*: "To be quite honest, changes have been occurring in a rapid manner and to respond to that we continue to rapidly identify and bridge the divide in order for us to be able to fully exploit our Big Data as an asset. Our HR has been working hand in hand with technical leadership to identify the much needed skills and assisting our staff to acquire them.*

The importance of the role that HR is expected to play in driving this shift cannot be overemphasized as they are needed in planning and preparing for the shift in talent and skills. Building the needed capacity within the business requires business and technical experts to jointly assess the skills at their disposal and how they can be enhanced in order to meet the current and future needs.

P2_C2*: "Over the last decade, we have witnessed several roles within the business transform and new roles and job titles manifesting. For example, the sudden demand for programming languages such as Python has resulted to some of the skills that used to be key in the past decade now regarded as obsolete. We are sitting with a huge shortage of data scientist and we always try by all means possible to upskill our internal employees even though it is a slow process but we are hoping to gain value in a long run."*

P3_C2*: "We have invested a lot of money, time and efforts in trying to get the best out of our Big Data I must say. What gives me a terrible headache has been making sure that our people do have the competencies to support the business in its effort to gain value from its Big Data investment. We try to expedite the process of upskilling by offering in-house training for those in critical roles but we still face the skills gap challenge."*

Organisations have thrown their weight in ensuring that employees are provided with the learning opportunities and the necessary skills are cultivated. The emphasis on personal development and learning is stimulated by the demand of Big Data and analytics skills. Without the critical talent and skills, leveraging Big Data remains a pie in the sky. Some respondents mentioned that their organizations have opted to outsource specific Big Data roles and functions for various reasons. This is highlighted in the quotes below:

P8_C3*: "Not long ago, we were hunting and recruiting software developers and analyst on the market but that has changed with the growing demand for data scientists and engineers. This kind of roles relatively combines an exclusive skills*

set which are very scarce and therefore not easy to find in the market. Due to this shortage of this talent and the costs associated with keeping it, we opted to outsource some of Big Data functions such as analysis while are busy growing and nurturing the talent internally."

A similar view is echoed by respondent P14_C2:

P14_C2*:* "*We have witnessed and experienced a lot of Big Data driven changes and these have led to the development of new job profiles and opportunities such as the highly sought-after data scientists and data engineers. At this stage we cannot we are under pressure gain business value from our Big Data program while operating at a very thin budget hence, after many considerations we realized that outsourcing for a short period of time* will deliver the value."

4.7   Theme 6: Big Data security

Research Sub-question(s)
What are the perceived Big Data governance challenges encountered by the organizations?
What governance strategies, processes and methods are currently employed by the organizations to govern Big Data?
How do organizations deal with the Big Data governance challenges?
Interview Question(s)
Have you made any changes to your data security since the adoption of Big Data?

Most respondents alluded that growing adoption of Big Data is accompanied by potential security risks and concerns. Mitigating the emerging security risks is one of the priorities in their governance programs or organizations. It is indicated protection of Big Data cannot be fully achieved through traditional security models or approaches and that compels organizations to explore Big Data specific security approaches. This is illustrated in the quotes below:

P6_C2*:* "*Yes, we have had to bolster our Identity Access Management efforts and we are not very strict on who should be granted access because Big Data is more sensitive and may contain personal information. We have established an authorization process, we use Claudera for IAM and we just sourced a tool called Microfocus to ensure our data is secure and encrypted. This will help in anonymizing and encrypting our data.*"

P15_C3*: "Due to the highly intensive nature Big Data technologies, as the business we have been compelled to relook at our data security approach, the old approach had gaps and would not work with a data of this magnitude."*

P10_C1*"After much hesitation we identified the potential security risk and decided to develop a Big Data specific security model, mainly because the traditional IT security model lacked the scalability to fully safeguard data from highly untrusted platforms with high volume, variety and velocity."*

The sentiments above clearly spell out the undeniable impact that the characteristics of Big Data such as velocity, variety and volume have on the way governance components such as data security are carried in the Big Data environment as compared to traditional data. The participants indicated that a lot of resistance to enhance or develop new models existed and still exists but some organizations have given in eventually.

P12_C2*: "We have now incorporated Big Data Analytics as a crucial component of our cyber security efforts, we then implemented automated real time detection of potential cyber threats."*

Respondents had different perceptions and some did not see the need for changes in their data security approach but rather opted to augment the traditional approach by adding specific elements:

P14_C2*: "In my honest opinion, the only aspect that lacked from our security program was the real-time monitoring and reporting on vulnerabilities or potential risks."*

P7_C3*: "We have not seen the need to perform a complete overhaul in our data security approach, we have instead strengthened our approach by putting more emphasis on encryption and developing anonymization algorithms."*

Certainly, the findings reveal that different organizations have taken different types of approaches in dealing with the security of Big Data and its demands. Some organizations are yet to see the need for Big Data specific processes, methods, models and tools to bolster the security of their Big Data assets. On the other hand, some organizations feel that their methods of securing Big Data need to be revisited and enhanced in order to close the gaps where they may exist and also fulfil the new security needs that are required to fully safeguard their Big Data and its technologies.

4.8   Theme 7: Metadata Management

Research Sub-question
How do organizations deal with the Big Data governance challenges?
What major effect has the adoption of Big Data had on traditional data governance?
Interview Question(s)
Do you currently manage metadata in your Big Data environment?
Do you understand the benefits of having an effectively managed metadata?

Even though metadata management is supposed to be a key component of any Big Data governance initiative, the responses certainly reveal that some organizations have the awareness of how vital metadata management is and how it could be valuable to them when implemented effectively. However, the responses below indicate that none of these organizations have made it a priority in their quest to achieve good governance of Big Data:

P15_C3*: "I need to state it clear that we have not yet reached the level of maturity to have a solid metadata management program."*

P9_C1*: "We are in the process of developing metadata though it has been a slow process, but we are making strides."*

Maturity of the Big Data program is also considered as one of the factors in the absence of metadata as quoted below:

P6_C2*"Our Big Data program is in its early phase, and we are carefully accelerating our efforts in implementing metadata management. We are currently in the advance stage of sourcing the most suitable solution and tools. Fast tracking this process will allow us the opportunity to implement incrementally."*

P8_C3*: "For our Big Data ecosystem we do not have a solid metadata or data catalogue as yet, we can largely attribute this to challenges such as unstable structure of data and all the changes that take place rapidly in our data due to the nature of our business."*

Additionally, the findings suggest that the respondents have a much better realization of what the organization stands to gain from implementing a solid metadata management solution. These benefits are the main drivers and motivation for the implementation of a metadata management solutions in these organizations as stated in the following quote.

P5_C1*: "Metadata management is very critical, and we have an opportunity to be empowered organizations to fully leverage our data assets. Having organization wide metadata repository makes it data management easier because we will then know what is stored where and its meaning".*

P12_C2*: "A well-managed metadata will provide more insight of our data, and its significance while giving us the much-desired consistency in our data".*

P14_C2*: "Certainly we will be able to save time and resources because everything will be stored in one repository and that makes managing our data much more efficient".*

P10_C1*:" An active metadata or let me rather say metadata that is managed effectively will contribute towards our data discovery, provide an in-depth understanding of existing relationships within the data and these are extremely critical".*

As organizations are rushing to get their Big Data projects up and running, a strategy to manage metadata should be developed and implemented simultaneously. The data volume grows at a very rapid rate and gets disseminated across various business units and therefore a delayed development of a metadata management becomes a mammoth task.

4.9   Summary of Chapter 4

In this chapter, findings on the Big Data governance implications in various organizations are presented. These findings were presented in line with themes which were discussed extensively in chapter.

The findings of this study are summarised for each theme in Table 4 below:

Table 4: Summary of findings

Theme	Findings
Regulatory Compliance	It was clear that, the existing laws, legislations, rules and regulations have kept all organizations on their toes. The increasing urge to meet regulatory requirements can be attributed wholly on the reputational damage and the financial costs of non-compliance. Several organizations are failing to meet all the requirements, and this can be attributed to factors such as employee awareness, the costs involved and the lack of the appropriate tools for championing compliance.
Policies	Some of the participants firmly believe that the current policies are appropriate for Big Data. The findings suggest that majority of organizations do not see the need of Big Data specific policies. It is crystal clear that organizations have not

Theme	Findings
	made any strides to access how the introduction of Big Data can potentially affect the current policies.
Stewardship	The findings reveal that a few organizations have been making efforts to put data stewardship programs in place even though several factors have been restraining their efforts.
Data Quality Management	From the findings, it can be established that achieving highest degree of quality seems to be relatively difficult in a Big Data environment. The variety, velocity and volume of Big Data sets make human intervention to become inadequate for fixing data quality issues.
Skills and Roles	Various organizations have been implementing plans and strategies to upskill their employees. One of the factors influencing the growing emphasis on personal development and learning is the increasing demand of Big Data and analytics skills.
Big Data security	The findings reveal that some organizations have chosen different kinds of strategies in dealing with the security of Big Data and its demands. Some organizations have not seen the need for Big Data specific security processes, methods, models and tools in order to bolster the security of their Big Data assets. Contrary, some organizations are adamant that their methods of securing Big Data need to be revisited and enhanced in order to close the gaps where they may exist and also fulfil the new security needs that are required to fully safeguard their Big Data and its technologies.
Metadata Management	The value of metadata management cannot be disputed. Metadata management is a significant component of any Big Data governance initiative. The findings reveal that organizations are aware of the latter fact and the benefits of an effectively managed metadata. However, the findings indicate that none of these organizations have made it a priority.

The next chapter (Chapter 5) discusses the findings and concludes the study according to its aims and objectives.

5. CHAPTER 5: DISCUSSION AND CONCLUSION

This chapter comprises of a comprehensive discussion of the major findings that were extensively presented in the previous chapter. This chapter also contains the conclusions of the research study, the summary of findings and ultimately the recommendations based on the findings that were discussed in chapter 4. The proposed recommendations serve as contribution towards the pool of knowledge in the space of Big Data governance. The key findings derived from existing literature and interviews with the respondents who were identified as key role players has aided in addressing the research questions.

The discussion is guided by the themes that emerged in Chapter 4 and it is carried out in accordance to the primary questions of this research outlined below:

The main research question:

What data governance implications are presented by the adoption of Big Data?

Sub-Questions

- What governance strategies, processes and methods are currently employed by the organizations to govern Big Data?
- What are the perceived Big Data governance challenges encountered by the organizations?
- How do organizations deal with the Big Data governance challenges?
- What major effects has the adoption of Big Data had on traditional data governance?

Each research sub-question is briefly addressed below:

- **What governance strategies, processes and methods are currently employed by the organizations to govern Big Data?**

Regulatory compliance

The respondents have a considerable understanding about the growing need to comply with the existing data protection laws and regulations, with the majority of the respondents indicating that much of their awareness and focus is on POPI Act. The findings point out that most organizations have heavily invested in educating and training their workforce about POPI and the significance of complying. The respondents also acknowledge that it is vitally important for them to comply as failure to comply might have dire consequences to the business and their jobs.

The literature revealed that it is imperative that organizations look at compliance in a holistic view and also extend their compliance initiatives in order to cater and conform not only to local but to global regulations as well. Compliance is mandatory in every region and location where the organization trades or operates (Trint, 2019). The respondents firmly believe that by achieving POPI Act compliance, they are also in compliance with the European Union's GDPR as they are both meant to regulate all the activities that consist of the collection, processing and sharing of personal information inside or across the borders (Republic of South Africa, 2013).

Policies

Data governance policies are supposed to be at the core of any data governance program. Developing coherent policies helps the organization in ensuring that its data governance program complements the strategic goals of the business.

The respondents were very adamant that the existing policies are suitable to govern Big Data. The findings indicate that no review or amendments have been made by these companies on the current policies that are meant to regulate how data is collected, accessed, processed and stored within the organization. The respondents contend that Big Data specific policy review or development is not one of the items on top of their list of priorities. Provided how rapid access, control and security changes are taking place, any data driven organization or governance program with long term objectives should be carefully planning their policy management while also taking changes in technology and ways of working into consideration.

Stewardship

Having data stewardship in place is important for any organization that seeks to manage and protect its data assets effectively. However, the role of data stewards does not only entail the oversight and protection of data assets but it goes extra length to working with the various units of business to specify the objectives of the data governance program. Sustaining control and oversight of the data assets is the kind of task that needs to be monitored and executed frequently by a dedicated individual or team of data stewards.

The findings reveal that traditional and Big Data stewardship is not given the amount of attention it deserves as one of the most significant components of data governance, thus none of the organizations were found to be having a dedicated and functioning data stewardship program in place. All of the respondents acknowledge how beneficial it would be for the business if an effective data stewardship was in place. The increasing demand to comply with regulations and the need for organizations to leverage its data assets should serve as inspiration to have data stewardship in place.
The findings suggest that the need for a data stewardship program is made to be indiscernible largely by the fact that some duties and processes of data stewards are currently being executed by someone in the organization on a part time or ad hoc basis and therefore giving a false impression that data stewardship is being

performed efficiently. This is supported by Plotkin (2020) who states that even though some organizations do not have an established data stewardship program, they do may have an existing set of processes in place. Discovering, documenting, and formalizing these processes can serve as a good foundation for the implementation of a data stewardship program. Nevertheless, it could also be established that some of these organizations are making strides towards implementing stewardship programs even though it remains a lower priority goal.

Data Quality Management

The findings of this study indicated that managing the quality of Big Data requires added efforts and a different approach to the one used to manage traditional data quality. It is quite evident that the respondents are well aware of the data quality implications in their respective ecosystems of Big Data and organizations are making every effort to attain good quality of Big Data.

In the advent of Big Data, the process of managing quality has changed tremendously, the findings of this study reveal that specific data quality metrics have now been changed to cater for the variety in data types and formats. These metrics are now source and type specific and that clearly means that assessments is not done comprehensively but unique for each data type and format. Other factors mentioned include the high velocity at which the data is generated, it was also alluded that this rapid pace requires the data quality assessment and improvement to be performed in real time because it becomes extremely difficult when delayed or performed at a later stage. Some respondents highlighted that some of their Big Data projects are required to deliver real time results and that leaves them with no other option but to establish data quality metrics that are fitting for real time scenarios.

Skills and Roles

Based on the feedback from the respondents, it is agreed that having the essential skills is critical for the organization to start seeing return from their Big Data investment and certainly that justifies the growing demand for such skills.

The findings suggest that several companies are mindful of the broadening skills gap and they are making strides towards the transformation of their workforce and are heavily investing on the development and expansion of the skills they need in order to reap the benefits of their Big Data initiatives. From the results, it is evident that a number of companies have committed the same error of directing most of their funds towards Big Data technologies than the people involved in generating value out of this technology.

This study also that some organizations are currently exploring different strategies such as automation and outsourcing while their workforce is undergoing the process of reskilling and upskilling. By outsourcing, organizations believe that they

are ought to possess the skill they are lacking while saving substantially on the cost of labour and increase efficiency.

Big Data security

The respondents agree that the adoption of Big Data has challenged them to start revaluating their current approach to data security with an objective of bolstering it or finding alternative approach. This is supported by Tian (2017) when stating that there is an urgent need to enhance the approach to Big Data security and proposed the use of intelligent analytics since the traditional security models are not efficient in the Big Data environment. Cloud security alliance (2013) also suggests that when taking a closer look at the emerging security challenges a very key question that needs to be addressed is what security and privacy strategies would be most capable to resolve existing Big Data security demands.

The findings suggest that a few organizations have already paved the way by deploying solutions such as the Big Data analytics to help in identifying potential attacks, threats and patterns. This study further established that the intensification of the security measures that control access through frameworks such as Identity and Access Management is one of the primacy actions that organization take. Some organizations have not hesitated and quickly explored various tools and solutions to the challenge of anonymizing and encryption.

Albeit the apparent efforts to bolster the Big Data security approach from many organizations, it is evident that there is an existing resistance by some organizations thus they have not made any considerations regarding the improvement of their Big Data security strategies, even though it remains a critical component of Big Data governance. Some respondents seem to understand that it is extremely important that data security tools are assessed regularly in order to validate their ability to secure the organization's data assets from unauthorized users or attackers.

Metadata Management

Metadata management is one of the important components for the effective governance of both big and traditional data. Metadata becomes even more critical when an organization strives to realize and gain value from its data.

Judging from the findings of this study, it is apparent that the adoption of Big Data presents a set of new complications in the process of metadata management. With the increasing challenges, organizations are compelled to start implementing vigorous systems and processes to manage big metadata. An ideal big metadata management with the ability to handle emergent complexity of Big Data sources, increasing volumes and high velocity.

5.1 What are the perceived Big Data governance challenges encountered by the organizations?

Regulatory compliance

The study revealed that Big Data can help organizations improve the operations, product, services and decision making. However, there are growing concerns over the stringent regulations that organizations have to comply with. The literature also supports the notion of organizations having to ensure that sensitive data is managed comprehensively in order to comply with existing laws and rules while exploiting data profusely (Amster, 2016).

The study findings indicated that the growing urge to comply to the data protection regulations is largely driven by the fear of incurring monetary penalties and other severe consequences of non-compliance. According to Wallace and Castro (2018) some organizations have been complaining that the stringent data protection regulations serve as a setback to them when competing with companies from countries or regions with more relaxed regulation compliance requirements.

In the age of Big Data, companies have large volumes of data at their disposal and this stimulates the need to shift towards cloud storage. The cloud servers are relatively located across the United States of America and this on its own is a challenge because the POPI Act intends to regulate the transfer of information outside South African borders. As per information regulator, a legitimate transfer of personal information across the borders of South Africa needs to be accompanied by a guarantee that adequate level of protection will be provided and the principles for processing such data are similar to POPI. Such complications require an in-depth knowledge and understanding of the POPI Act as minimal understanding would mean chances of a data break are elevated.

Policies

The lack of willingness to assess the current policies to identify or project possible implications in the future can be attributed to factors such as lack of resources, knowledge and executive support. Another important factor is the absence of a governance forum or committee that focuses specifically on providing oversight of the organization 's data governance initiatives and identify or manage the potential risks associated with their data assets.

The findings suggest that the Big Data initiatives of these organizations are still in their early phases and therefore the maturity of their policies is most likely to be correlated with the maturity of their Big Data initiatives, if there is an existing commitment and ambition to assess and review these policies. Policy revision is inevitable in this era of rapidly changing data privacy laws and regulations. It is therefore worthwhile for organizations to start thinking ahead and planning for the revision of their data policies and procedures.

The findings also indicate that there is an understanding of the role of policies in the institution and how they should be handled but there are issues regarding people not delivering to deadlines in reviewing policies, external auditors not provided proper information related to change processes and lack of proper management of policies within departments.

Stewardship

Lack of executive support, budget constraints, shortage of staff, scarcity of skills and organizational culture have been cited as some of the major challenges hindering the implementation of an effective data stewardship program. Without the necessary tools to execute stewardship, not much can be achieved by any organization and that requires the executive heads to buy in and support the program financially and otherwise. Unfortunately, without a solid and sound business case for data stewardship data governance business case it is highly unlikely that the executive would be eager to provide the necessary funds and resources.

Data Quality Management

The adoption of Big Data has stimulated the surge in variety of data sources and types while also influencing the noticeable growth in volume of the data in various organizations (Fan, Han and Liu, 2014). It is an undeniable fact that Big Data brings a whole lot of benefits in contrast to traditional data and these are more attractive to a lot of businesses due to the increased prospects of gaining competitive advantage, growing sales and efficiency in production. However, managing the quality of Big Data is indeed a mammoth task and can be very costly to the business when not managed carefully.The respondents do concede that it is a severely challenging task and this is mainly attributed to the known factors such as extreme variation in data types, structure and format.

Skills and Roles

The Big Data phenomenon is relatively a terrain that is still being explored by various organizations globally. The scarcity of the skills and talent in the labour market is considered to be one of the significant challenges that organizations have to encounter when adopting Big Data. This is supported by Sivarajah, Kamal, Irani and Weerakkody (2017), when stating that the scarcity of individuals with the much-needed Big Data analytics skills is one of the critical challenges of Big Data adoption.

There is an apparent gap between the skills that companies have at their exposure and those that are desirable in order to exploit Big Data and gain competitive advantage. The increasing adoption of big technologies continues to transform the set of skills that organizations demand rapidly, making the process of hunting and closing of the skills gap to be an incessant task.

The findings reveal that the workforce is largely disrupted. This is apparent as roles in high demand today were non-existent a few years ago and the traditional roles that were previously sought after by organizations have now vanished (National Academies of Sciences, Engineering, and Medicine, 2017). Nonetheless, over the last decade data scientist roles have extremely grown in demand and that is due to technological advancement and maturity of Big Data programs in various organizations.

Despite the rapid increase in the demand for data scientists, the findings point out that only a few individuals in these organizations have the required skills to execute the necessary duties. It is quite evident that there is a considerable shortage of Big Data related talent and skills despite the growing number of opportunities in this field.

Big Data Security

According to Jagadish et al. (2014) in the era of Big Data confidentiality remains a major concern and a challenging task. It is therefore imperative to follow all the necessary measures to guarantee the security of data in use. Essentially, organizations that collect personal data the responsibility to safeguard and ensure that the data is used appropriately.

The data revolution has proven to massively benefit several organizations through improved operational efficiency, increased agility, and improved customer engagement and experience. This means that companies are now sitting with high volumes of sensitive customer data and information at their disposal. This is sometimes banking or credit information, contact details, geolocation data or perhaps health records.

Notwithstanding all the gains of Big Data, the respondents mentioned that the adoption of Big Data also brings along several security risks and concerns. As organizations are rushing to gain value from Big Data, security attacks also continue to grow aggressively. Each day, the perpetrators continue to proliferate, form global networks, automate their attacks and gather even more expertise.

Metadata Management

The results of this study clearly reveal that metadata management is not held in high esteem in most organizations and this is proven by the absence or lack of the processes and systems to manage metadata especially in Big Data ecosystems. This finding is in agreement with the views echoed by Smith et al. (2014) about the existing Big Data ecosystems lacking a principled approach towards metadata management.

5.2 How do organizations deal with the Big Data governance challenges?

Regulatory compliance

The findings of this study demonstrated that many organizations have a high degree of commitment towards the compliance to the existing regulations. Some organizations have dedicated time towards training their employees on regulations such as the POPI Act and GDPR. In one organization, there is an individual who is assigned to overlook and monitor the use of Big Data and its compliance to the regulations governing it. This clearly demonstrate how eager some organizations are to comply and manage the risk of non-compliance.

Policies

The findings reveal that organizations are adamant that they do not have any policy related challenges when governing their Big Data. There is considerable insistence that the traditional data policies are more than capable to govern Big Data too. None of the organizations have Big Data specific policies in place and there is currently no indication that anyone of them will move to formulate any policies that are tailored for Big Data. It is also apparent that no action or efforts has been made to assess the suitability of the traditional data policies or identify certain areas that could be improved in these policies to govern Big Data and its technologies effectively.

Stewardship

Despite all the mentioned challenges, organizations are bound to gain value from an effective data stewardship program and executives stand to benefit the most. Some organizations are beginning to offer the much-needed support and funds towards the implementation of a data stewardship program. This would assist in making sure that there is consistent data definitions across the organization and individuals who oversee the alignment between the business and its data assets while having a high guarantee of data reliability and quality. With the growing and strict regulatory requirements, it is worth mentioning that an operational and dedicated data stewardship program is essential in ensuring that the risks of owning sensitive data is well managed and people are thoroughly educated and informed about the regulations and the need to comply.

Data Quality Management

Some respondents pointed out that generating data from several sources compels them to develop and consistently run certain scripts to transform the data sets into the preferred format. This was never a case with traditional data but a necessity in achieving their quality objectives in a Big Data environment and thus the involvement of business in the management of Big Data quality is of paramount importance.

Some respondents indicated that in their respective organizations they have not seen a need to find additional dimensions that would be essential in evaluating and monitoring the quality of Big Data. Instead, they opted against transforming and redesigning their approaches to Big Data quality assessment and management by continuing to use the traditional approach, albeit Big Data having different characteristics. It is quite clear that there is a need for some companies to start conducting a comprehensive assessment of their Big Data quality, especially in key business units and areas.

Skills and Roles

Some organizations are making strides towards the implementation of strategies and solutions that are meant to solve the existing Big Data skills. The HR department is collaborating with both the business and technical management in trying to develop plans and solutions to these issues. These entail the rapid development of Big Data and analytics capabilities, driving the incorporation of the Big Data related skillset in their skill development plans, building a culture of learning and encouraging upskilling. One of the organizations has went as far as developing an in-house training programs and courses with an aim to deliver training that is specific to the organization's skills demands.

Big Data Security

The surge in cyber-attacks seems to be proportional to the development of data privacy and security regulations, more states and organizations are trying to restrict access to confidential data to avoid these attacks. Several organizations have started to invest in various tools and solutions with a goal of enhancing their Big Data security.

Cryptography is currently one of the trusted solutions that organizations have implemented in order to ensure that Big Data is secured from one end of its lifecycle to another. The findings of this study also reveal that a few organizations are monitoring the security of their Big Data in real time to detect possible security threats and attacks. Multi-factor authentication is also one of the security methods that have been adopted to ensure that Big Data is strictly accessed by the authorized individuals. Other solutions include Big Data security analytics which are being used in some organizations for tracking and investigation of various system or user activities and behaviour in real time to detect potential changes in user behaviour and attacks.

Metadata Management

Quite surprisingly, the absence of an effective metadata management systems and processes coexists with the understanding of how important is metadata management and the benefits that are ought to be gained by the organization by managing its metadata appropriately. However, some of the Big Data initiatives are still sitting at very low levels of maturity and this becomes a factor when considering the lack of an effective metadata management.

The findings suggest that many organizations regard metadata as a lower priority item and this is apparent as they do not have an effectively managed metadata. On the other hand, a few organizations have been slowly incorporating metadata management in their Big Data strategies and also changing their culture with an ample support from top management. Holom, Rafetseder, Kritzinger and Sehrschön (2020) suggest that a Big Data strategy that is focused on metadata management is extremely important in solving the challenges brought by the adoption of Big Data.

5.3 What major effects has the adoption of Big Data had on traditional data governance?

Regulatory compliance

In the age of Big Data, we have witnessed a rapid increase in the number and complexity of regulations that organizations have to comply to and this spells an increase to the time, resources and costs of ensuring compliance. The findings of this study indicate that the amplified volume, variance in sources and high velocity of data requires real time and constant regulatory compliance monitoring.

One of the main characteristics of Big Data is high volume and this is collected for open-ended purposes and this is against regulations such as the POPI Act. The recent regulations have a huge effect on traditional governance and how the organization collects and uses Big Data. To achieve compliance, organizations have to clearly state how their Big Data is gathered, processed and for what purposed unlike how it was in traditional data.

The mounting pressure to comply is growing in many organizations and it has forced some organization to relook at their governing strategies, especially concerning regulatory compliance. Organizations are now considering methods and solutions in how they can best reach their regulatory obligation while transparent on how they use Big Data and their purpose. A few organizations have started modifying certain roles within the organization and the responsibility to give oversight on compliance is assigned to an individual to actively monitor on a daily basis.

Policies

Having solid policies in place is key to any data related initiative. Without the relevant policies, it is difficult for any organization to monitor and manage the collection, processing, usage and storage of their data. Judging from the findings of this study, organizations are less concerned about the implications of Big Data adoption on their existing policies despite the growing public concerns over misuse, transparency, accountability and trust around Big Data.

The findings of this study also indicate that many organizations have not made any effort to look at how the adoption of Big Data can reshape their policies and that means traditional data policies are still being used to govern Big Data. Examining

existing policies, processes and strategies would be very imperative for organizations before their adoption of Big Data.

Stewardship

Despite all the challenges, the entire organization is bound to gain value from an effective data stewardship program and executives stand to benefit the most. By providing the necessary support and funds towards the implementation of a data stewardship program, they would be able to have consistent data definitions across the organization, stewards to oversee the alignment between the business and its data assets while having a high guarantee of data reliability and quality.

The findings reveal that some organizations have not had a solid data stewardship program even for traditional data even though the data steward duties were performed informally so. With the adoption of Big Data and growing regulatory requirements, most organization are beginning to establish operational and dedicated data stewardship program as they see that it is essential in ensuring that the risks of owning sensitive data is well managed and people are thoroughly educated and informed about the regulations and the need to comply.

Stewardship is most essential component of data governance; this is mainly because of the value it carries when it comes to giving the users of the data insight and context about the enterprise data. The data stewards have a massive role to play in the governance of Big Data and this includes the process of setting up controls to deal with governance and compliance requirements.

Data Quality Management

Authenticating and measuring quality of Big Data is considered to be a mammoth task, this is largely caused by the nature of Big Data and its unique characteristics such as enormous volume, broad variety of sources, and high velocity. At the moment, organizations are looking for different methods and solutions to examine the validity and credibility of the Big Data they collect real time, making it difficult to identify and remove false or even malicious data.

The findings of this study suggest that the adoption of Big Data has shifted how some organization measure or define data quality. Some participants mentioned that they now sitting additional data quality demands that are presented by the variety in sources, high volume and high velocity. In addition, the findings of this study suggest that working with heterogeneous sources has necessitated a data quality management approach that offers an ability to initiate a data quality assessment method that is exclusive to the business use or scenario of the data requirement.

Skills and Roles

The findings of this study suggest that Big Data is rapidly transforming how business utilize and govern data. It is apparent that there a growing skills gap and the demand for Big Data competencies has surpassed the one of traditional data. The adoption of Big Data has had a huge effect on the type of skills and talent that organizations are searching for nowadays. Organizations are now encouraging employees to upskills and acquire Big Data related skills as the traditional data skills are quickly becoming obsolete.

With Big Data adoption comes a change in how work the process of data governance is carried and organizations are searching for professionals with substantial knowledge of cloud technologies and applications in order to facilitate data migration from traditional storage mechanism to the cloud as Big Data is relatively stored in the cloud while managing risk and ensuring compliance.

Big Data Security

In the age of Big Data, some organizations have been forced to remodel and intensify their efforts to secure data. Organizations are now compelled to assess their data security throughout all the stages of its lifecycle, one of the prime reason for the comprehensive assessment is guaranteeing the security of the data and also to preserve it from the rife threats and attacks.

The findings of this study indicate that several organizations are making strides towards improving their security strategies with a goal of protecting Big Data. Organization have rushed to appointed experts to analyse the data monitor and manage any security risk in their Big Data project. This individual is also expected to produce frequent reports on any Big Data security related concerns and also drive development of security controls.

Judging from the findings, it could be established that the protection of Big Data demands Big Data specific traditional security models or approaches. Some organizations have not hesitated to implement process to deal specifically with data authorization and end to end encryption of this data. Contrary to traditional security governance, governing the security of Big Data has compelled organization to introduce trailblazing tools and automated security controls.

Metadata Management

Based on the findings, metadata is often overlooked or neglected in the most phases of Big Data governance and overall initiative. In some instances, the value it carries is well understood but it still remains lesser of a priority and little or no efforts are directed towards its development and maintenance.

It is undeniable that metadata management is an integral part of the comprehensive data governance processes. Therefore, any solid data governance program should feature an effective metadata management. Some organizations

are starting to form small groups of data stewards with a goal of driving the development of metadata in their Big Data initiatives. The findings of this study suggest that effective governance of metadata in the Big Data projects is very important for its success and the organization is likely to gain good return on its Big Data investment.

The growing adoption of Big Data and its related technologies has resulted to organizations collecting and storing exorbitant amounts of disparate data and this requires effective governance of metadata. This rapid increase in the volume, velocity and variance of data has redefined metadata governance immensely and further reshaped the approach meant to govern metadata. The ever growing data regulations have also intensified the importance of metadata in the era of Big Data. Metadata is now vital in classifying each dataset with the relevant regulation that needs to be complied to.

Unlike traditional data, Big Data collected from large heterogeneous sources is relatively unstructured, it is stored in disparate repositories, different data lakes and data warehouses with various structure. With these elements, metadata in the Big Data environment has grown to be more complicated and the traditional mechanisms of governing and managing metadata are no longer fitting.

5.4   The contribution of this study

The outcomes of this study will make an immense contribution towards expanding the body of knowledge in the field of big data governance. This research also outlines the various components of Big Data governance, stating how each of them is vital in effectively governing Big Data which is contextually relevant to organizations that wish to bolster their Big Data governance efforts.

5.5   Research Limitations

Due to the limited time and resources, the study was mainly focused on a case study with a restricted sample size and this serves as a limitation in as far as generalisation of the research results is concerned. The inclusion of several other organizations would have broadened the outcomes of the research but as stated above, the time and resource constraints led to their inclusion.

Big Data governance is an unfamiliar subject to many and that means only a few people are well versed on this matter, thus the selection of a larger sample is severely limited. Lastly, the outbreak of the corona virus constituted limitations as some respondents could not be met physically and that meant only respondents with access to Internet and virtual platforms can be reached at that particular moment.

5.6 Future research

The findings reveal that there is an opportunity to conduct research using a much broader sample with the inclusion of a variety of organizations operating in various sectors of business. Other suggestions for future research include the following:

1. To assess and explore the relationship between the maturity of the Big Data governance vs the maturity of the Big Data program.
2. Measure the impact of poor Big Data governance on the success of a Big Data initiatives.
3. How Big Data skills and training preparedness could aid its good governance

5.7 Recommendations

Below is the series of recommendations of this study, these recommendations are organized based on the findings that were discussed in Chapter 5.

Regulatory Compliance

In the age of Big Data, regulatory compliance has gained substantial amount attention as more laws and regulations are established to control the use of Big Data. Organizations should therefore intensify their efforts in ensuring that more awareness is raised and employees are well informed about these regulations. This may include assigning an individual or a team the responsibility to monitor the organization's compliance to the relevant regulations at a regular time interval. One of the primary responsibilities might as well include introducing systems or processes to identifying potential risks or breach and address them.

Policy

Having an effective a data policy management plan in place is key ensuring that guidelines on usage and handling of the data asset is provided. Policies that were meant to provide in traditional data may not be adequate for Big Data. The study recommends the formulation of Big Data policies taking its characteristics and its dynamics into consideration. The process of developing these policies must ensure maximum participation from all units of business, more especially the departments or individuals with a role to play in the Big Data lifecycle management. Lastly, these policies must be made available to all parties, reviewed frequently and updated to ensure relevancy at all times.

Privacy and Security

Guaranteeing Big Data privacy and security is one of the most significant elements of Big Data governance, it is therefore important to pay a high degree of attentiveness to how data is being accessed, who has access and how it is used. It is highly recommended that a robust security strategy that is tailor made

for Big Data is developed and implemented This will also ensure that risk is mitigated and avoid reputational damage to the organization. Among other recommendations of this study is the augmentation of Big Data security through the use of solutions such multi factor, this is recommended primarily because of its additional layers of protection.

The study also recommends the establishment of a security work group or committee and it must be constituted by representatives from every business area that generate or consumes data.   The committee's sole focus should be to oversee privacy and security concerns, identify threats and vulnerabilities. Furthermore, committee must also conduct continuous analysis of risk and duly implement prevention strategies.

Data quality management

Due to the inherent complexity of Big Data, the traditional approach to data quality management is not fitting and therefore organizations are advised to urgently relook at the approach to Big Data quality management in order to perform certain enhancements and adjustments. When performing these adjustments, one of the most important factors is the revision of the existing data quality dimensions to ascertain if these are still relevant and suitable for large volumes of data with high velocity and variety. Additionally, a lot of efforts should be directed towards the discovery of new dimensions that will be applicable to Big Data.

Achieving a high degree of data quality requires collaboration and equal efforts from all role players, the best processes, technological solutions, tools and the highest level of commitment from the top management. Organizations must display an incredible amount of dedication towards the improvement of Big Data quality by appointing data quality experts to drive the Big Data quality agenda through the deployment of quality control tools, systems and processes.

Skills and Roles

Without a doubt, the introduction of Big Data has changed the job market tremendously in the recent years. The colossal gap keeps widening between the demand and supply of Big Data related talent and skills needed in order to extract value out the enormous data in organizations. This study suggests that organizations should dedicate a greater amount of time and effort towards research and development of the much essential skills. It is also recommended that organizations develop and implement training and development strategies that will drive the process of upskilling, reskilling and attainment of Big Data. This includes investing heavily and developing a culture of learning by availing the necessary tools and training material to stimulate self-development and innovation within the organizations.

Data Stewardship

Having a well-established and effective stewardship program is uncommon for some organizations, those who have a stewardship program in place are sometimes not entirely committed to the program and this can be attributed to factors such as lack of management support or commitment. In the advent of Big Data, data stewardship has grown to be even more indispensable and this study suggest that data driven organizations must put forth the greatest possible effort to garner the support of top management.

In some organizations, some of the data steward duties are executed on a part time basis by individuals who occupy other job roles. This necessitates the establishment of a formal and effective data stewardship program consisting of fully committed individual who are solely responsible for executing data stewardship duties. Despite the suggested formalization of this role, this study recommends the development of a data driven culture where data stewards form an integral part of the data governance structure and be granted the platform to influence critical.

Metadata Management

As Big Data remains a central feature of efficient decision making, the prominence of its quality, security and traceability keeps increasing. Moreover, understanding the importance of an effective data governance program must always be correlated to the understanding of metadata management's importance. The complexity of Big Data makes it essential for organization to start investing on metadata tools in order to get a comprehensive view of their data flows.

It is highly recommended that the implementation of a metadata repository should precede the organization's efforts to manage metadata across all business units. Another recommendation towards effectual management of metadata is the appointment of stewards whose responsibility is to provide oversight of metadata management. Organizations have to take an ample of time to draw their preferred approach to the implementation of a solid metadata strategy and this can be achieved through the assistance of those with advance technical and business acumen.

5.8 Conclusion

The main objective of this study was to analyse the governance implications presented by Big Data adoption on data governance. Moreover, the study aimed to gain a deeper understanding of the data governance dynamics that are presented by the introduction of Big Data.

Responding to the main research question of this study:

"*What data governance implications are presented by the adoption of Big Data?* "

The point of departure was to gain an insight of what literature and research participants at the selected organizations describe as the possible implications of Big Data introduction on data governance. From the existing literature and the empirical findings it was evident that there is an assortment of opinions relating to the effect of Big Data introduction on the process and or the way data is governed.

It is also apparent that adoption of any kind of new technology by an organization is bound to yield a plethora of challenges, thus much research, preparation, planning and analysis should be made prior to the adoption. It is also evident that characteristics such as velocity, volume and variety make Big Data technologies more eccentric and thus governing it requires unique methods and process.

Certainly, dealing with the existence of an inherent resistance to revise the Big Data governance approach cannot be an easy task, therefore it is important for the management to provide unwavering support to the concept of a Big Data specific governance approach.

It is extremely imperative that the process of development and implementation a robust strategy, processes and methods to govern Big Data is driven by the goal to achieve security, quality and regulatory compliance. Big Data governance should be centred around people, strategies, technology, processes, policies and regulations in order for organisations to exploit Big Data and return the investment made on this asset. These have been identified as the critical components of effective management of Big Data and overlooking them could lead to dire consequences for the organization.

REFERENCES

Austin, Z., & Sutton, J. (2014). Qualitative research: Getting started. *The Canadian journal of hospital pharmacy*, 67(6), pp.436.

Abdullah, N., Ismail, S., Sophiayati, S., & Sam, S. (2015). Data Quality in Big Data: A Review. *International Journal of Advances in Soft Computing and Its Applications*, 7(Special Issue 3), pp.16-27.

Abhyankar, P.P. and Powar, S.A. (2016). Big Data: Trends and Challenges. *Asian Journal for Convergence In Technology (AJCT) ISSN -2350-1146*, [online] 2. Available at: http://asianssr.org/index.php/ajct/article/view/168/202 [Accessed 5 Nov. 2018].

Abraham, M., Van Kerckhove, J., Archacki, R., González, J.E. & Fanfarillo, S. (2019). The Next Level of Personalization in Retail. [online] BCG. Available at: https://www.bcg.com/publications/2019/next-level-personalization-retail.aspx [Accessed 8 Jan. 2020].

Abrams, M. (2009). Data Protection Accountability: The Essential Elements. A Document for Discussion. Accountability: A Compendium for Stakeholders, pp 8-9.

ACCA & IMA. (2013). Digital Darwinism: Thriving in the Face of Technology Change. Available at: http://www.accaglobal.co.uk/content/dam/acca/global/PDF-technical/other-PDFs/Five-mins-on-Digital-Darwinism.pdf

Acker, R. (2020). *Big Data, Big Risks*. [online] Infosecurity Magazine. Available at: https://www.infosecurity-magazine.com/blogs/big-data-big-risks/.

Adams, R., Adeleke, F., Anderson, D., Bawa, A., Branson, N., & Christoffels, A. (2021). POPIA Code of Conduct for Research. *South African Journal of Science*, 117(5-6), pp.1-12.

Agata, T., Miyata, Y., Ishita, E., Ikeuchi, A. & Ueda, S. (2014). Life span of web pages: A survey of 10 million pages collected in 2001, IEEE/ACM Joint Conference on Digital Libraries, pp. 463-464.

Ajah, I. & Nweke, H. (2019). Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data and Cognitive Computing*, 3(2), pp.32.

Alhadeff, J., Van Alsenoy, B., & Dumortier, J. (2012). The accountability principle in data protection regulation: origin, development and future directions. *In Managing privacy through accountability* (pp. 49-82). Palgrave Macmillan, London.

Alhojailan, M. I. (2012). Thematic analysis: A critical review of its process and evaluation. Zagreb, Croatia: WEI International European Conference Proceedings.

Ali, A., Qadir, J., ur Rasool, R., Sathiaseelan, A., Zwitter, A. & Crowcroft, J. (2016). Big Data for development: Applications and techniques. *Big Data Analytics*, 1(1). doi:10.1186/s41044-016-0002-4.

Al Mubarak, M. & Hamdan, A. (2016) 'The impact of corporate governance on market capitalization: evidence from Bahrain Bourse'*, Corporate Ownership and Control Journal*, 13(3), pp.120–129.

Aloysius, J.A., Hoehle, H., Goodarzi, S. & Venkatesh, V. (2016). Big Data initiatives in retail environments: Linking service process perceptions to shopping outcomes. Annals of Opera-tions Research, pp.1-27.

Altman, M., Wood, A., O'Brien, D. R., & Gasser, U. (2018), Practical approaches to Big Data privacy over time, International Data Privacy Law.

Amster, A. (2016*). Compliance Challenges and Big Data* - DATAVERSITY. Available at: https://www.dataversity.net/compliance-challenges-big-data/ [Accessed 26 September .2019].

Anderson, C. (2010). Presenting and evaluating qualitative research. *American Journal of Pharmaceutical Education*,74(8), pp.141.

Angeles, S. (2019). Big Data vs. CRM: How Can They Help Small Businesses? [online] *Business News Daily.* Available at: https://www.businessnewsdaily.com/6053-big-data-vs-crm.html [Accessed 5 Jan. 2020].

Austin, C., & Kusumoto, F. (2016). The application of Big Data in medicine: current implications and future directions. *Journal of Interventional Cardiac Electrophysiology*, 45, pp.1–9.

Azeroual, O., Saake G. & Abuosba, M. (2018). Investigations of concept development to improve data quality in research information systems, In Proceedings of the 30th GIWorkshop on Foundations of Databases, 2126, pp. 29-34, Wuppertal, Germany.

Baeza-Yates, R. (2018). Bias on the web. Commun. ACM 61, pp.54–61.

Baker, S. (2018). What Causes the Data Governance Gap? Available at: https://www.linkedin.com/pulse/what-causes-data-governance-gap-scott-baker [Accessed 7 March. 2021].

Barbu, I. (2020). Cyber Security: The Vulnerabilities of future information systems. Ph.D. University "Politenhica" of Bucharest.

Barocas, S., Selbst, A.D. (2016). Big Data's disparate impact. California Law,104(3). pp.671–732.

Beauchamp, T.L. and Childress, J.F. (2001) Principles of biomedical ethics. 5th Edition, Oxford University Press, Oxford, 59.

Beedgen, C. (2012). Log Data is Big Data | Sumo Logic. [online] Sumo Logic. Available at: https://www.sumologic.com/blog/log-data-is-big-data/ [Accessed 14 Jan. 2020].

Bertram, C, & Christiansen, I. (2015). Understanding research: An introduction to reading research. Pretoria: Van Schaik Publishers.

Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta., A, & Paraboschi, S .(2001). Designing Data Marts for Data Warehouses. ACM Transactions on Software Engineering Methodology.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2). pp.69-14.

Braun, V. & Clarke, V. (2012). *Thematic analysis.* (3rd ed.). American Psychological Association. https://doi.org/10.1037/13620-004

Broeders, D., Schrijvers, E., van der Sloot, B., van Brakel, R., de Hoog, J. & Hirsch Ballin, E. (2017). Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. *Computer Law & Security Review*, 33(3). pp.309-323.

Brynjolfsson, E., Hitt, L. M., Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance?

Bhattacherjee, A. (2012) *Social Science Research: Principles, Methods, and Practices.* Available at: https://scholarcommons.usf.edu/oa_textbooks/3/ [Accessed 5 May. 2021].

Buttler, P. (2017). *10 Challenges to Big Data Security and Privacy - Dataconomy.* [online] Dataconomy. Available at: http://dataconomy.com/2017/07/10-challenges-big-data-security-privacy/ [Accessed 5 Jun. 2018].

Burger, L. (2019). *Why Big Data Means Big Opportunities.* [online] Simplilearn.com. Available at: https://www.simplilearn.com/why-big-data-means-big-opportunities-article [Accessed 6 Jan. 2020].

Burgess, M. (2020). *What is GDPR? The summary guide to GDPR compliance in the UK.* Available at: https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018 [Accessed 6 February .2021].
Cabral, A. F. G. (2021). Data Profiling in Cloud Migration: Data Quality Measures while Migrating Data from a Data Warehouse to the Google Cloud Platform (Doctoral dissertation).

Cai, L. & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14, pp.2.

Centre for Information Policy Leadership. (2018). The Central Role of Organisational Accountability in Data Protection [Discussion Paper] (1st ed.). Available at:
https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_account ability_paper_1_-_the_case_for_accountability_-_how_it_enables_effective_data_protection_and_trust_in_the_digital_society.pdf
CGMA. (2013). From insight to impact. Unlocking the opportunities in Big Data.

Chen, H., Chiang, R.H. & Storey, V.C. (2012). "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly (36:4), pp. 1163-1170.

Chen, W.-J. (2013). *A practical guide to managing reference data with IBM InfoSphere Master Data Management reference data management hub.* Poughkeepsie, NY: IBM Corp., International Technical Support Organization.

Cheong, L. K. & Chang, V. (2007). The Need for Data Governance: A Case Study. In Proceedings of the 18th Australasian Conference on Information Systems, Toowoomba (Australia).

Chessell, M., Scheepers, F., Nguyen, N., van Kessel, R., & van der Starre, R. (2014). Governing and managing Big Data for Analytics and Decision Makers. IBM Redguides for Business Leaders

Chisholm, M., (2015). *The Foundations of Successful Reference Data Management.* [online] Topquadrant. Available at:
https://www.topquadrant.com/docs/whitepapers/TopBraid_ReferenceDataManag ementWhitepaper-3-18-15.pdf [Accessed 18 March 2021].

CloudMoyo (2019). Sources of Big Data: Where Does It Come From?. [online] CloudMoyo. Available at: https://www.cloudmoyo.com/blog/data-architecture/what-is-big-data-and-where-it-comes-from/ [Accessed 15 Nov. 2019].

Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education.* 5th ed, Routledge Falmer, London.

Collibri & Mapr. (2016). The Need for Big Data Governance | MapR. [online] Available at: https://mapr.com/resources/need-big-data-governance/ [Accessed 15 Jun. 2018].

Collis, J, & Hussey, R. (2003). Business research: A practical guide for undergraduate & postgraduate students .2nd ed. Basingstoke, Hampshire UK: Palgrave Macmillan.

Collis, J., & Hussey, R. (2014). *Business Research, a Practical Guide for Undergraduate and Postgraduate Students* .4th ed. London: Palgrave Macmillan.

Conrad, C. & Serlin, R.C. (2011). *The SAGE handbook for research in education: pursuing ideas as the keystone of exemplary inquiry*. Los Angeles; London: Sage.

Cloud Security Alliance. (2013). Expanded Top Ten Security and Privacy Challenges.

Cooke-Davies, T. J. (2004). Measurement of organizational maturity. In D. P. Slevin, D. I. Cleland & J. K. Pinto (Eds.), Innovations: Project management research (pp. 211–228). Project Management Institute.

Cooper, D. & Schindler, P. (2008). *Business research methods*. 10th ed. New York, McGraw-Hill/Irwin.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed method approaches* .2nd ed. California: Sage Publications.

Creswell, J. W. (2007). *Qualitative inquiry & research design: choosing among five approaches.* Thousand Oaks, CA: Sage Publications.

Creswell, J. W, & Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research. London*: Sage Publications.

Creswell, J. W, & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research.* 2nd ed. London: Sage Publications Ltd.

Cukier, K, & Mayer-Schönberger, V. (2013). The rise of Big Data. Foreign Affairs (May/June): pp.28–40.

DAMA International. (2017). *DAMA-DMBOK: Data Management Body of Knowledge.* 2nd ed. Technics Publications.

Datafloq (2019). Data Veracity: A New Key to Big Data. [online] Datafloq.com. Available at: https://datafloq.com/read/data-veracity-new-key-big-data/6595 [Accessed 14 Nov. 2019].

Davenport, T. H, & Dyché, J. (2011). Big Data in big companies: executive summary. White paper, SAS Institute Inc.

de Bruyn, M. (2014). The protection of personal information (POPI) Act - Impact on South Africa, *International Business and Economics Research Journal (Iber)* .13 (6), pp. 1315-1340

Debussche, J., César, J. & Thirot, M. (2020). *Big Data & Issues & Opportunities: Transparency, Consent, Control and Personal Data Ownership.* [online] Bird & Bird. Available at: https://www.twobirds.com/en/news/articles/2019/global/big-data-and-issues-and-opportunities-transparency-consent-control-and-personal-data-ownership [Accessed 1 February 2020].

Dederer, M.G. (2015). '8 steps to effective information lifecycle management', *Information Management Journal*, 49(1), pp.32.

De Haes, S. and Van Grembergen, W. (2004). IT governance and its mechanisms, *Information Systems Control Journal* (1), pp. 27–33

De Mauro, A., Greco, M. & Grimaldi, M. (2015). What is Big Data? A Consensual Definition and a Review of Key Research Topics, AIP Conference Proceedings, vol.1644, pp.91-102

Deloitte (2019). *Developing an effective governance operating model A guide for financial services boards and management teams*. Available at: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/dttl-fsi-US-FSI-Developinganeffectivegovernance-031913.pdf [Accessed 29 Oct. 2019].

Denzin, N.K, & Lincoln, Y.S. (Eds.) (2005). *The Sage Handbook of Qualitative Research*. 3rd ed. Thousand Oaks, CA: Sage.

Dhawan, V. & Zanini, N. (2014). Big Data and social media analytics. Research Matters: A Cambridge Assessment publication, 18.

DHL. (2013*). Big Data in Logistics-A DHL perspective on how to move beyond the hype*. [online] Available at: https://www.dhl.com/content/dam/downloads/g0/about_us/innovation/CSI_Studie_BIG_DATA.pdf [Accessed 27 Jan. 2020].

Diakopoulos, N, & Friedler, S. (2016). *How to Hold Algorithms Accountable.* Available at: https://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/ [Accessed 4 April. 2021].

Dial, M. & Storkey, C. (2017). *Futureproof: How To Get Your Business Ready For The Next Disruption.* 1st ed. Pearson UK: Pearson Education Limited, pp.27.

Dolma, S. (2010). The central Role of the Unit of Analysis Concept in Research Design. *Journal of the School of Business Administration, Istanbul University*, 39(1), pp.169-174.

Donnelly, D. L. (2020). Privacy by (re) design: a comparative study of the protection of personal information in the mobile applications ecosystem under United States, European Union and South African law (Doctoral dissertation).

Durrheim, K. (2004). Research design. In M. Terre Blanche, & K. Durrheim (Eds.), Research in practice: Applied methods for the social sciences (pp. 29-53). Cape Town: University of Cape Town.

Easwaramoorthy, M. & Zarinpoush, F. (2006). Interviewing for research. In: CANADA, I. (ed.). Toronto.

Eaker, C. (2016). What Could Possibly Go Wrong? The Impact of Poor Data Management.

Eaton, C, et al. (2012). Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data. vol.1. New York, The McGraw-Hill Companies.

Eckerson, W. W. (2002) Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data: The Data Warehouse Institute: pp.1-33.

Egan, P. (2011). Effecting Data Quality Through Data Governance: A Case Study in the Financial Services Industry. Master. Regis University.

Eisenhardt, K.M. (1989). Building Theories from Case Study Research, The *Academy of Management Review*,14(4). pp. 532-550.

Elgendy, N. & Elragal, A. (2014). Big Data analytics: A literature review paper, Springer, pp.214-227.

Enaxisconsulting. (2017). *Governing and Managing Data in the Big Data Era – Enaxis Consulting*. [online] Available at: http://www.enaxisconsulting.com/governing-and-managing-data-in-the-big-data-era-blog/ [Accessed 6 Jun. 2018].

ENISA, (2020). *Big Data Threat Landscape And Good Practice Guide*. [online] Heraklion. Available at: https://www.enisa.europa.eu/publications/bigdata-threat-landscape [Accessed 22 November 2019].

Enofe, M.O. (2017). Data Management In An Operational Context A Study At Volvo Group Trucks Operations. Master thesis. Lund University.

Ernst & Young. (2014). *Big Data Changing the way businesses compete and operate Insights on governance, risk and compliance* [ebook], pp.11-14. Available at: http://www.ey.com/...changing_the_way_businesses_operate/.../EY-Insights-on-GRC-Big-d... [Accessed 11 Jun. 2018].

European Commission, (2017). *Digital skills coalition*. [online] Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/policies/digital-skills-coalition [Accessed 5 January 2021].

Fan, J., Han, F., Liu, H. (2014). Challenges of Big Data analysis National Science Review, 1 (2), pp. 293-314

Fatt, Q. and Ramadas, A. (2018). The Usefulness and Challenges of Big Data in Healthcare. [online] Healthcare-communications.imedpub.com. Available at: http://healthcare-communications.imedpub.com/the-usefulness-and-challenges-of-big-data-in-healthcare.pdf [Accessed 25 Oct. 2018].

Flyvbjerg, B. (2011). Case Study. In Norman K. Denzin and Yvonna S. Lincoln, (Eds.), The Sage Handbook of Qualitative Research (pp. 301-316). Thousand Oaks, CA: Sage.

Flick, U. (2006). An introduction to qualitative research, Thousand Oaks, CA: Sage.

Flick, U. (2011). Introducing research methodology: A beginner's guide to doing a research project. London: SAGE Publications Ltd.

Foote, K.D. (2021). A Brief History of Master Data. [online] DATAVERSITY. Available at: https://www.dataversity.net/a-brief-history-of-master-data/ [Accessed 10 March 2020].

Forrest, C. (2019). Ten examples of IoT and Big Data working well together | ZDNet. [online] ZDNet. Available at: https://www.zdnet.com/article/ten-examples-of-iot-and-big-data-working-well-together/ [Accessed 20 Nov. 2019].

Funk, T. (2013). Advanced Social Media Marketing: How to Lead, Launch, and Manage a Successful Social Media Program.New York: Apress.

Gall, R. (2020). Machine Learning Ethics: What You Need To Know And What You Can Do | Packt Hub. [online] Packt Hub. Available at: https://hub.packtpub.com/machine-learning-ethics-what-you-need-to-know-and-what-you-can-do/ [Accessed 5 January 2020].

Gandomi, A. and Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. International Journal of Information Management, [online] 35(2), pp.137-144. Available at: https://www.sciencedirect.com/science/article/pii/S0268401214001066 [Accessed 27 Oct. 2018].

Garcia, T. (2020). What is Data Compliance? Available https://reciprocitylabs.com/resources/what-is-data-compliance/ [Accessed 11 February.2021]

Gibbs, G. (2007). Analyzing qualitative data. Sage Publications

GlueReply. (2014). The Valuation of Data as an Asset: a Consumption-Based Approach. Reply, 12 June 2015, www.reply.com/Documents/13903_img_The-valuation-of-data-as-an-asset.pdf . [Accessed 05 Jun. 2018].

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. The Qualitative Report, 8, 597-606.

Gompers P, Ishi JL, Metrick A (2003) Corporate governance and equity prices. Q J Econ 118(1). pp.107–155.

Goundar, S. (2013). Research methodology and research method. Available at: http://www.researchgate.net/publication/333015026 [Accessed 18 February 2020].

Gray, D. E. (2004). *Doing Research in the Real World.* London SAGE Publications.

Gray, D. E. (2013). *Doing research in the real world.* London SAGE Publications.

Green, J. (2018). *The Importance of Transparency in Big Data.* Xmode. Available at: https://medium.com/@xmodesocial/the-importance-of-transparency-in-big-data-ec4eeae09165 [Accessed 7 April 2020].

GSMA (2018). *Assessing Regulatory Requirements of Privacy Management for Members Offering IoT Services Using Personal Data.* [online] Available at: https://www.gsma.com/iot/wp-content/uploads/2018/11/GSMA_Assessing-regulatory-requirements-of-privacy-management-for-members-offering-IoT-services-using-personal-data.pdf [Accessed 7 April 2020].

Guamán, D. S., Del Alamo, J. M., & Caiza, J. C. (2021). GDPR Compliance Assessment for Cross-Border Personal Data Transfers in Android Apps. IEEE Access, 9, 15961-15982.

Guba, E. G, & Lincoln, Y. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). *Big data and predictive analytics for supply chain and organizational performance.* Journal of Business Research, 70, 308-317.

Denzin, N.K. & Lincoln, Y.S. (1994). *Handbook of qualitative research.* Thousand Oaks: Sage Publications, pp.105–117.

Hagen, C., Evans, H., Thota, B., Wall, D., Seshadri, A. & Khan, K. (2014). IT's Challenge: Bringing Structure to the Unstructured World of Big Data - Paper - A.T.

Hakutizwi, B. (2018). *5 Ways Google Trends Can Improve Your Business.* [online] Business2community. Available at: https://www.business2community.com/seo/5-ways-google-trends-can-improve-your-business-02118759 [Accessed 18 Oct. 2019].

Halwani, M. (2018). The Evolution of Big Data and Modern Business Intelligence Applications. pH.D. University of North Texas.

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? Big Data 6: 176–90.

Harris, J., Hofer, M., Blanchard, M., Grandgenett, N., Schmidt, D., van Olphen, M. & Young, C. (2010). "Grounded" Technology Integration: Instructional Planning Using Curriculum-Based Activity Type Taxonomies. Journal of Technology and Teacher Education, 18(4), 573-605. Chesapeake, VA: SITE.

Harris, J. (2015). *Can Big Data be governed?* [online] The Data Roundtable. Available at: https://blogs.sas.com/content/datamanagement/2015/07/30/can-big-data-governed/ [Accessed 7 Jun. 2018].

Harwell, M. R. (2011). Research Design in Qualitative/Quantitative/ Mixed Methods. Opportunities and Challenges in Designing and Conducting Inquiry, University of Minesotta, 147–182. http://doi.org/10.4135/9781412961288.n380

Hassani, A, & Gahnouchi, S. A. (2017). A framework for business process data management based on Big Data approach. Procedia computer science, 121, 740-747.

Haug, A., Pedersen, A. and Arlbjørn, J.S. (2009), A classification model of ERP system data quality, Industrial Management & Data Systems, 109 (8), pp. 1053-68.

Haug, A., Zachariassen, F., & Van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management* (JIEM), 4(2), pp. 168-193.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B. & Sethupathy, G. (2016). "The Age of Analytics: Competing In a Data-Driven World". McKinsey Global Institute.

Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B., & Babu, S. (2011). *Starfish: A Self-tuning System for Big Data Analytics.* In CIDR, pp. 261–272.

Herschel, R. & Miori, V. M. (2017). Ethics & Big Data, Technology in Society, 49: 31– 36.

Hoch, J., Parycek, P., & Schöllhammer, R. (2016). Big Data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), pp.147–169.

Horodyski, J. (2015). The metadata maturity model. *Journal of Digital Media Management*, 3(4), pp.349-354.

Holom, R. M., Rafetseder, K., Kritzinger, S., & Sehrschön, H. (2020). Metadata management in a Big Data infrastructure. *Procedia Manufacturing*, 42, pp.375-382.

Huang, M. (2020). Privacy and Data Protection in the Digital Era: The Global v. National Scope. In UNTAR Conference. Indonesia. Retrieved from

https://www.researchgate.net/publication/339310131_Privacy_and_Data_Protecti
on_in_the_Digital_Era_The_Global_v_National_Scope

Hull, G. (2015) Successful Failure: What Foucault Can Teach Us about Privacy
Self-Management in a World of Facebook and Big Data, *Ethics and Information
Technology.*

Huner, K.M., Ofner, M. & Otto, B. (2009) *Towards a Maturity Model for Corporate
Data Quality, Proceedings of the 2009 ACM Symposion on Applied Computing*,
Waikiki Beach, Honolulu, Hawaii, USA.

Hurwitz, J., Nugent A., Halper, F. & Kaufman, M. (2013). *Big Data for Dummies,
John Wiley & Sons.* Available at: http://www.wiley.com/.

Hussain, S., Hussain, M., Afzal, M., Hussain, J., Bang, J., Seung, H. & Lee, S.
(2019). Semantic preservation of standardized healthcare documents in Big Data.
International Journal of Medical Informatics, 129, pp.133-145.

IBM. (2007). *The IBM Data Governance Council Maturity Model: Building a
Roadmap for Effective Data Governance, IBM Corporation.*

Information Resources Management Association. (2016). *Big Data: concepts,
methodologies, tools, and applications. Hershey, PA*: Information Science
Reference

IDC. (2019). *IDC Forecasts Revenues for Big Data and Business Analytics
Solutions Will Reach $189.1 Billion This Year with Double-Digit Annual Growth
Through 2022.*

Ienca M, Ferretti A, Hurst S, et al. (2018) Considerations for ethics review of Big
Data health research: *A scoping review. PLoS One.*13(10).

Ihuah, P.W. & Eaton, D. (2013). *A framework for the sustainable management of
social (public) housing estates in Nigeria: a pilot study.* [online] usir.salford.ac.uk.
Available at: http://usir.salford.ac.uk/id/eprint/30993/ [Accessed 18 Dec. 2021].

Informatica. (2019). *How to Manage Reference Data.* [White Paper] Available at:
https://www.informatica.com/content/dam/informatica-com/en/collateral/white-
paper/how-to-manage-reference-data_white-paper_3641en.pdf [Accessed 18
February 2020].

Ise, O. A. (2016). Integration and Analysis of Unstructured Data for Decision
Making: Text Analytics Approach. *International Journal of Open Information
Technologies*, 4(10).

IT Governance Institute. (2006). *Information Security Governance: Guidance for
Boards of Directors and Executive Management* (2nd ed.). IT Governance
Institute.

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

Jiang, F. (2017). Data Analytics Helps Business Decision Making. Master's thesis,

Johnson, R. B., Onwuegbuzie, A., & Turner, L. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, pp.112-133.

Judger, N. (2016). *The thematic analysis of interview data: An approach used to examine the influence of the market on curricular provision in Mongolian higher education institutions*. Hillary place papers, 3rd edition, University of Leeds.

Kabir, S. M. (2016*). Basic Guidelines for Research: An Introductory Approach for All Disciplines*.1st ed. Bangladesh: Book Zone.

Kagermann, H., Osterle, H., & Jordan, J. M. (2010). IT-driven business models: Global case studies in transformation. Hoboken: Wiley.

Kaisler, S., Armour, F., Espinosa, J. & Money, W. (2013*). Big Data: Issues and Challenges Moving Forward. 46th Hawaii International Conference on System Sciences.*

Kaplan, B., & Maxwell, J. A. (2005) Qualitative research methods for evaluating computer information systems, In J. G. Anderson & C. E. Aydin (Eds.), *Evaluating the organizational impact of healthcare information systems* (2nd ed.), Springer, New York, NY, pp.30–55.

Kapow (2019). *The Big 9 Big Data Sources [*Infographic]. [image] Available at: https://www.getelastic.com/big-data-infographic [Accessed 13 Oct. 2019].

Kearney. (2013*). Big Data and the Creative Destruction of Today's Business Models*, [Online]. Available: http://www.atkearney.com/paper//asset_publisher/dVxv4Hz2h8bS/content/id/5152447 [Accessed 27 Oct. 2018].

Kemp, S. (2020). *Digital trends 2020: Every single stat you need to know about the internet.* [online] Podium | The Next Web. Available at: https://thenextweb.com/podium/2020/01/30/digital-trends-2020-every-single-stat-you-need-to-know-about-the-internet/ [Accessed 15 Feb. 2020].

Khan, N. et al. (2014) Big Data: survey, technologies, opportunities, and challenges. *Scientific World Journal*, 2014.

Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A. A., & Salehian, S. (2018). The 10 Vs, issues and challenges of big data. In Proceedings of the 2018 international conference on big data and education (pp. 52-56).

Koziel, S., Hilber, P., Westerlund, P., & Shayesteh, E. (2021). Investments in data quality: Evaluating impacts of faulty data on asset management in power systems. Applied Energy, 281, 116057.

Kostjens, P. (2016). Anomaly Detection in Application Log Data. Master of Science in Computing Science. Utrecht University.

Kostis, A. (2016). The role of Big Data in the evolution of Platform based Ecosystems: A case study of an emerging platform-based ecosystem in the software engineering industry.

Kothari, C.R. (2004). Research Methodology: Methods and Techniques. 2nd Ed. New Age International Publishers, New Delhi.

Kouanou A.T., Tchiotsop, D., Kengne, R., Tansaa, Z.D., Adele, N.M., & Tchinda, R. (2018) An optimal Big Data workflow for biomedical image analysis. Inf Med Unlocked 11, pp.68–74

Kritikos, K, & Barreiros, J. (2016). Business model innovation to explore data analytics value; A case study of Caterpillar and Ericsson.

Kshetri, N. (2014). Big Data's impact on privacy, security and consumer welfare. Telecommunications Policy. 38(11), pp.1134–1145.

Kulk, S. & van Loenen, B. (2012) Brave new open data world. International Journal of Spatial Data Infrastructures Research 7: pp.196–206.

Kulkarni, A. (2016). A Study on Metadata Management and Quality Evaluation in Big Data Management. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 4(VII).

Kumar, S. & Singh, M. (2019). Big Data Analytics for Healthcare Industry. Impact, Applications, and Tools.Big Data Mining and Analytics, 2(1), pp.48-57.

Labadie, C, & Legner, C. (2019). Understanding Data Protection Regulations from a Data Management Perspective: A Capability-Based Approach to EUGDPR.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety, META Group Research Note.

Laudon, K.C & Laudon, J.P. (2014). Management information systems: managing the digital firm. 13th ed. Boston: Pearson.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S, & Kruschwitz, N. (2011). Big Data, Analytics and the Path from Insights to Value. [online] MIT Sloan Management Review. Available at: https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/.

Lee, I. (2017). Big Data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), pp.293–303. doi:10.1016/j.bushor.2017.01.004.

Leedy, P. & Ormrod, J. (2001). *Practical Research: Planning and Design*. 7th ed. Upper Saddle River: Prentice-Hall.

Lewis, J.L. & Sheppard S.R.J. (2006). Culture and communication: can landscape visualization improve forest management consultation with indigenous communities? *Landscape and Urban Planning* 77, pp.291–313.

Li, G., He, J. & Fu, Y. (2008) A group-based intrusion detection scheme in wireless sensor networks, *Computer Communications*, 31(18), pp.4324–4332.

Liell-Cock, S., Graham, J. & Hill, P. (2009). *IT Governance aligned to KING III.*

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, Calif: SAGE.

Lippell, H. (2016). Big Data in the media and entertainment sectors. In *New Horizons for a Data-Driven Economy* (pp. 245-259). Springer, Cham.

Loebbecke, C. & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and Big Data analytics: A research agenda. *The Journal of Strategic Information Systems*, 24(3), pp.149–157. doi:10.1016/j.jsis.2015.08.002

Loshin, D. (2013). *Big Data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph.* Amsterdam: Elsevier/Morgan Kaufmann.

Lukka, K. (2014). Exploring the possibilities for causal explanation in interpretive research. *Accounting, Organizations and Society*,39(7), pp.559–566.

Lukoianova, T. & Rubin, V.L. (2014) Veracity roadmap: Is Big Data objective, truthful and credible? *Advances in Classification Research Online* 24(1), pp.4–15.

Luna, D., Mayan, J.C., García, M.J., Almerares, A.A, & Househ, M. (2014) *Challenges and potential solutions for Big Data implementations in developing countries*. Yearb Med Inform.9:36–41.

Melder, B. (2019). *Understand, Predict, and React: How Big Data Social Media Marketing Transforms Your Campaigns.* [online] Socialbakers.com. Available at: https://www.socialbakers.com/blog/big-data-social-media-marketing [Accessed 29 Jun. 2019].

McAfee, A. & Brynjolfsson, E. (2012) Big Data The Management Revolution. Harvard Business Review.

McAninch, G. (2015). Big Data and Big Challenges for Archives. [online] bloggERS!. Available at: https://saaers.wordpress.com/2015/03/05/big-data-and-big-challenges-for-archives/ [Accessed 5 Feb. 2020].

McGhee, G., Marland, G.R., & Atkinson, J. (2007). Grounded theory research: Literature reviewing and reflexivity. *Journal of Advanced Nursing*, 60(3), pp.334–342

Madden, S. (2012). From Databases to Big Data. *IEEE Internet Computing*, 16(3), pp.4–6.

Maisel, L, & Cokins, G. (2014). Why analytics will be the next competitive edge. *Journal of Corporate Accounting & Finance*, 25(5), pp.63-72.

Marchibroda, J. (2013). *A Policy Forum on the Use of Big Data in Health Care.*

Mathes, C.A. (2016). Big Data Has Unique Needs for Information Governance and Data 73 Quality. *Journal of Management Science and Business Intelligence*, 1(1), pp. 12–20.

Mathur, N. & Purohit, R. (2017). Issues and Challenges in Convergence of Big Data, Cloud and Data Science. *International Journal of Computer Applications,* 160(9), pp.7-12.

Maxwell, J. A. (2012). *Qualitative Research Design: An Interactive Approach*. 3rd Ed. Thousand Oaks, CA: Sage Publications.

May, T. (1997). *Social research: issues, methods and process*. 2nd ed. Trowbridge: Redwood Books. Available at: http://www.mheducation.co.uk/openup/chapters/0335206123.pdf

Martin, K.E. (2015). Ethical Issues in the Big Data Industry. *MIS Quarterly Executive* 14, pp.67–85.

Martinez-Uribe, L. & Fern´andez, P. (2015). Data services: a strategic function of 21st century libraries. *Profesional de la informaci´on*, 24 (2), pp.93-199.

Merriam, S.B. (1998). Qualitative Research and Case Study Applications in Education. Jossey-Bass Publishers, San Francisco.

Merriam, S. (2009). Qualitative research: A guide to design and implementation. Jossey-Bass Publishers, San Francisco.

Mikalef, P., Krogstie, J. (2019). Investigating the data science skill gap: an empirical analysis. In 2019 IEEE Global Engineering Education Conference (EDUCON) (pp. 1275-1284). IEEE.

Minelli, M., Chambers, M., Dhiraj, A., (2013). Business Analytics. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, 99-125.

MIT Technology Review. (2015). The Rise of Data Capital.

Moreno, J., Serrano, M. & Fernández-Medina, E. (2016). Main Issues in Big Data Security. *Future Internet*, 8(3), pp.44.

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. *Harvard Business Review*, 93(5), pp.96-105.

Mosley, M., Brackett, M. H., Earley, S., & Henderson, D. (2010). DAMA guide to the data management body of knowledge. Technics Publications.

Motadata. (2019). *Data Logging Guide: Types, Sources and Formats - Motadata*. [online] Available at: https://www.motadata.com/blog/data-logging-guide-types-sources-and-formats/ [Accessed 14 Jun. 2019].

Myers, M. D. (2009*). Qualitative research in business and management*. London: Sage Publications.

Myers, P. (2012). Information governance-control content chaos. *KM World*, 21 (8), 9.

Neirotti, P & Paolucci, E. (2007) Assessing the strategic value of information technology: An analysis on the insurance sector. *Information & Management* 44(6), pp.568–582.

National Academies of Sciences, Engineering, and Medicine. (2017*). Information technology and the US Workforce: Where are we and where do we go from here*? Washington, DC: The National Academies Press.

National Computing Centre (2005). *IT Governance Developing a successful governance strategy A Best Practice guide for decision makers in IT*. (2019). 1st ed. [ebook] Manchester: The National Computing Centre, [Accessed 11 Oct. 2019].

Nikiforova, A. (2020). Definition and Evaluation of Data Quality: User-Oriented Data Object-Driven Approach to Data Quality Assessment. *Baltic Journal of Modern Computing*, 8(3), pp.391-432.

Nolan, R. & McFarlan, F.W. (2005) Information technology and the board of directors, *Harvard Business Review,* 83(10), pp.96-106.

Nuti, S., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R., Chen, S. & Murugiah, K. (2014). The Use of Google Trends in Health Care Research: A Systematic Review. PLoS ONE, 9(10), p.e109583.

Oates, B. (2006). *Researching information systems and computing.* London, Sage Publications.

OECD. (2004). OECD Principles of Corporate Governance, Paris: OECD Publications Service

Oghojafor, B. E. A., Olayemi, O. O., Okonji, P. S., & Okolie, J. U. (2010). Poor Corporate Governance and its Consequences on the Nigerian Banking Sector.

Onwuegbuzie, A.J., Johnson, R.B. & Collins (2009). Call for mixed analysis: A philosophical framework for combining qualitative and quantitative approaches. *International Journal of Multiple Research Approaches*, 3 (2), pp.114-139.

Oracle (2013*). Information Management and Big Data: A Reference Architecture. An Oracle White Paper.* [online] Available at: https://www.oracle.com/technetwork/topics/entarch/articles/info-mgmt-big-data-ref-arch-1902853.pdf [Accessed 8 Oct. 2019].

Orb, A., Eisenhauer, L., & Wynaden, D. (2001). Ethics in qualitative research. *Journal of nursing scholarship*, 33(1), pp.93-96.

Paananen, J. P. (2020). A New Data Governance Model for the Bank of Finland.

Pae, J. & Choi, T., H. (2011) Corporate Governance, Commitment to Business Ethics, and Firm Valuation: Evidence from the Korean Stock Market, *Journal of Business Ethics*, 100, pp. 323– 348.

Panchenko, A. (2016) Nine Main Challenges in Big Data Security.

Paterson, K. (2020). *What Is Data Management And Why It Is Vital | Blue-Pencil.* [online] Blue-Pencil. Available at: https://www.blue-pencil.ca/what-is-data-management-and-why-it-is-important [Accessed 2 February 2021].

Paschalidi, C. (2015). Data Governance: A conceptual framework in order to prevent your Data Lake from becoming a Data Swamp. (Masters). Luleå University of Technology.

Papa, A. (2017). Master Data Management in Global Enterprise. Masters. University of Tuku.

Peco, M. (2019). *The Imperative for Reimagining Data Governance | Transforming Data with Intelligence.* [online] Transforming Data with Intelligence. Available at: https://tdwi.org/Articles/2018/10/26/BIZ-ALL-Reimagining-Data-Governance.aspx?Page=2 [Accessed 29 May 2019].

Pernot-Leplay, E. (2020). Data Privacy Law in China: Comparison with the EU and US Approaches.

Pflug, M. (2018). *Without data management, forget AI and machine learning in health care.* [online] Government Data Connection. Available at: https://blogs.sas.com/content/statelocalgov/2018/11/19/data-management-health-care [Accessed 4 January 2021].

Plotkin, D. (2020). Data stewardship: An actionable guide to effective data management and data governance. Academic Press.

Ponelis, S. (2015). Using interpretive qualitative case studies for exploratory research in doctoral studies: a case of information systems research in small and medium enterprises. *International Journal of Doctoral Studies*, 10, pp.535-550.

Porter, M. & Heppelmann, J. E., (2014). How smart, connected products are transforming competition. Harvard Business Review, pp. 66-88.

Raschke, R., Krishen, A.S. & Kachroo, P. (2014). Understanding the components of information privacy threats for location-based services. *Journal of Information Systems*, 28(1), pp.227–242.

Reddy, K., & Reddy, S. (2007). Olap Reporting Application Using Office Web Components (Doctoral dissertation, University of Akron).

Regalado, A. (2014). Data and Decision Making. MIT Technology Review, 117(2),pp.61-63.

Republic of South Africa. 2013. Act No. 4 of 2013: Protection of Personal Information Act.Vol. 581 No. 37067. Cape Town: Government Gazette.

Rhodes, R. A. W. (2007). *Understanding Governance: Ten Years On, Organization Studies*, 28(8), pp.1243-1264

Ricci, A. & Kalyva, F. (2015). Unlocking the full potential of Big Data: A change management approach. Master's Programme in Management. Lund University.

Rowley, J. (2000) Is Higher Education Ready for Knowledge Management? *The International Journal of Educational Management*, 14 (7), pp.325-333.

Ryan, G.W. & Russell Bernard, H. (2003) Techniques to Identify Themes. *Field Methods*, 15, pp.85-109.

Sabesan, M., & Risch, T. (2006). Querying mediated Web services. 8th International Information Technology Conference, IITC 2006. Colombo, Sri Lanka.

Saldaña, J. (2021). *The coding manual for qualitative researchers*. sage.

Sanjari, M., Bahramnezhad, F., Fomani, F. K., Shoghi, M., & Cheraghi, M. A. (2014). Ethical challenges of researchers in qualitative studies: The necessity to develop a specific guideline. *Journal of Medical Ethics and History of Medicine*, 7 (14).

Saqr, M. (2017). Big Data and the emerging ethical challenges. *International journal of health sciences,* 11(4), 1.

Satish, L. & Yusof, N. (2017). A Review: Big Data Analytics for enhanced Customer Experiences with Crowd Sourcing. *Procedia Computer Science*, 116, pp.274-283.

Saunders, M., Lewis, P. and Thornhill, A. (2015a). *Research Methods for Business Students*. 6th Ed. London, UK: Pearson.

Saunders, M., Lewis, P., Thornhill, A. & Bristow, A. (2015b). Understanding Research Philosophy and Approaches to Theory Development. In M. Saunders, P. Lewis, & A. Thornhill, Research Methods for Business Students (pp. 119-150). Harlow: Pearson.

Saunders, M., Lewis, P., Thornhill, A. & Bristow, A. (2019*) Research Methods for Business Students*, 5th ed., Harlow: Pearson Education.

Schmarzo, B. (2013). *Big Data: Understanding How Data Powers Big Business*. Indianapolis: John Wiley & Sons Inc.

Schroeder, R. & Schroeder, R. (2016). *Big Data business models: Challenges and opportunities*. pp.2–10.

Scott, M., & Cerulus, L. (2018). *Europe's new data protection rules export privacy standards worldwide*. [Accessed 18 March. 2018].

Seale, C. (1999). Quality in qualitative research. *Qualitative Inquiry*, 5(4), pp.465-478.

Shamim, S., Zeng, J., Shariq, S. & Khan, Z. (2019). Role of Big Data management in enhancing Big Data decision-making capability and quality among Chinese firms: *A dynamic capabilities view. Information & Management*, 56(6), pp.103135.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, pp.63-75.

Short, J. E. (2006). *Information lifecycle management: an analysis of end user perspectives*. San Diego, pp2-37.

Sileyew, K.J. (2019). *Research Design and Methodology.* [online] Available at: https://www.intechopen.com/books/cyberspace/research-design-and-methodology [Accessed 28 March 2021].

Singh, A. S. (2014). Conducting Case Study Research in Non-Profit Organisations. *Qualitative Market Research: An International Journal*, 17, pp.77–84.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, pp.263-286.

Skjæveland, B. (2013). Analyzing business process management. Master's thesis. University of Stavanger, Norway.

Smeda, J. (2015). Benefits, business considerations and risks of Big Data. Master's thesis. University of Stellenbosch, South Africa.

Soares, S. (2012). A Framework That Focuses on the Data in Big Data Governance. *IBM Data Management.*

Soares, S. (2014), Data Governance Tools: Evaluation Criteria, Big Data Governance, and Alignment with Enterprise Data Management, MC Press, Ketchum.

Soiferman, L. K. (2010). Compare and Contrast Inductive and Deductive Research Approaches. Online Submission.

Smith, G. E., Watson, K. J., Baker, W. H. & Pokorski II, J. A. (2007). A Critical Balance: Collaboration and Security in the IT-enabled Supply Chain. *International Journal of Production Research* 45(11) pp.2595–2613.

Smith, K., Seligman, L., Rosenthal, A., Kurcz, C., Greer, M., Macheret, C., ... & Eckstein, A. (2014, June). " Big Metadata" The Need for Principled Metadata Management in Big Data Ecosystems. In Proceedings of Workshop on Data analytics in the Cloud (pp. 1-4).

Spur Corporation. (2017). Integrated report 2017: IT Governance. [online] Spur Corporation, p.1. Available at: https://www.spurcorporation.com/financials/ar/2018/downloads/it-governance-and-risk-committee.pdf [Accessed 24 Oct. 2019].

Stander, J. (2015). *The modern asset: Big Data and Information Valuation.* [online] Stellenbosch University. Available at: https://scholar.sun.ac.za/handle/10019.1/97824 [Accessed 13 Jun. 2018].

Stake, R. E. (2003). Case studies. In N. K., Denzin, & Y. S., Lincoln (Eds.). Strategies of Qualitative Inquiry, London: Sage Publications. pp. 134–164.

Steenbeek, I. (2018). *The Data Management Cookbook: A pocket guide for implementation of data management.* 1st ed. Data Crossroads, pp.10.

Study International. (2019). *Data scientists will lead the digital age - Study International.* [online] Available at: https://www.studyinternational.com/news/data-scientists-will-lead-the-digital-age/ [Accessed 15 Dec. 2019].

Schwandt, T. A. (2007). The SAGE Dictionary of Qualitative Inquiry. 3 ed. SAGE Publishing

Sweeney, L. (2013). Discrimination in online ad delivery. Communications of the ACM 56(5),44–54.

Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electronic Journal*, 5(2), pp.18–27.

Tait, A. (2016). 5 Big Data Security Challenges | Learning Tree Blog. [online] Learning Tree Blog. Available at: https://blog.learningtree.com/five-big-data-security-challenges/ [Accessed 5 Jun. 2018].

Taleb, I., Serhani, M. and Dssouli, R. (2018) ''Big Data quality: A survey'' in 2018 IEEE International Congress on Big Data (BigData Congress), pp. 166–173, IEEE.

TDWI. (2019). Data Quality In The Age Of Big Data. [online] Available at: https://tdwi.org/articles/2019/04/19/diq-all-data-quality-in-the-age-of-big-data.aspx [Accessed 16 January 2020].

Tabesh, P., Mousavidin, E. & Hasani, S. (2019). Implementing Big Data strategies: *A managerial perspective. Business Horizons,* 62(3), pp.347-358.

Teddlie, C. & Tashakkori, A. (2009). *Foundations of Mixed Methods Research*. Thousand Oaks, CA: Sage.

Teegavarapu, S., Summers, J. D., & Mocko, G. M. (2008). Case study method for design research: A justification. In International design engineering technical conferences and computers and information in engineering conference (Vol. 43284, pp. 495-503).

Teradata. (2019) *5 Big Benefits of Data and Analytics for Positive Business Outcomes*. [ONLINE] Available at: https://www.teradata.com/Blogs/5-Big-Benefits-of-Data-and-Analytics-for-Positive-Business-Outcomes. [Accessed 28 October 2019].

Tesfaye, B. (2017). What is the influence of Big Data and analytics on Management Control Systems? Masters. Radbound University.

Thanh, N. C. & Thanh, T. T. (2015). The Interconnection Between Interpretivist Paradigm and Qualitative methods in Education. *American Journal of Educational Science*, 1(2).pp.24-27.

The University of Queensland. (2021). 6.40.01 Information Management Policy - Policies and Procedures Library - The University of Queensland, Australia. [online] Ppl.app.uq.edu.au. Available at: https://ppl.app.uq.edu.au/content/6.40.01-information-management-policy [Accessed 17 May 2021].

Tian, Y. (2017). Towards the Development of Best Data Security for Big Data. *Communications and Network,* 9, pp.291-301.

Tiglias, J. (2014). Impossible to Ignore: The Importance of IT Governance. [online] Corporate Compliance Insights. Available at:

https://www.corporatecomplianceinsights.com/impossible-to-ignore-the-importance-of-it-governance/ [Accessed 14 Oct. 2019].

Tobin, G. A., & Begley, C. M. (2004). Methodological rigour within a qualitative framework. *Journal of Advanced Nursing*, 48, pp.388–396.

Toivonen, M. (2014). Big Data Quality Challenges in The Context of Business Analytics. Masters Thesis. University of Helsinki.

Tomson, S. (2019). Business glossary creation, management and integration with data models on the example of a financial company. (Masters). Tallinn University of Technology.

Trint. (2019). *What is data compliance, and why is it so important?* [online] Available at:https://trint.com/resources/7n1weo10/what-is-data-compliance-and-why-is-it-so-important [Accessed 23 May. 2020]

Tsai, P. (2020). *Data Governance 101: How to Build a Business Glossary | BigID.* [online] Available at :https://bigid.com/blog/data-governance-101-how-to-build-a-business-glossary/ [Accessed 17 Jan, 2021]

Turner, D. P. (2020). *Sampling Methods in Research Design. Headache*, 60(1), pp.8-12.

Tykheev, D. (2018). Big Data in marketing. Undergraduate. Saimaa University of Applied Sciences.

Unceta, I., Nin, J., & Pujol, O. (2020). Risk mitigation in algorithmic accountability: *The role of machine learning copies. Plos one*, 15(11), e0241286.

UNCTAD. (2020). Data Protection and Privacy Legislation Worldwide | UNCTAD. [online] Available at: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide [Accessed 24 Jan. 2021]

Utami, A., Pratama, B. & Widianto, S. (2020). Data mart design in bkpp bandung using from enterprise models to dimensional models method. JITK (*Jurnal Ilmu Pengetahuan dan Teknologi Komputer),* 5(2), pp.279-284.

Varga, M. (2011). Challenges of Data Management in Always-On Enterprise Information Systems. In Enterprise Information Systems: Concepts, Methodologies, Tools and Applications (pp. 1695-1714). IGI Global.

Van Laar, E, Van Deursen, A, Van Dijk, J, De Haan, J. (2017). The relation between 21st-century skills and digital skills: *A systematic literature review. Computers in Human Behavior*, 72, pp.577–588.

van Rijmenam, M. (2019). *How Big Data Turns CRM into Something Truly Valuable.* [online] Medium. Available at: https://medium.com/dataseries/how-big-data-turns-crm-into-something-truly-valuable-c44f55482557 [Accessed 13 Sep. 2019].

Valenti, M. A., Luce, R., & Mayfield, C. (2011). The effects of firm performance on corporate governance. *Management Research Review*, 34(3), pp.266-283.

Varah, S. (2014). Video: The Next Frontier for Big Data. [online] WIRED. Available at: https://www.wired.com/insights/2014/09/video-big-data [Accessed 21 Jun. 2019].

Vargas-Solar G, Zechinelli-Martini J, Espinosa-Oviedo J. (2017). Big Data management: what to keep from the past to face future challenges? *Data Science and Engineering*, 2(4), pp.328-345.

Venter, F. & Stein, A. (2012). *Images & videos: really Big Data - Analytics Magazine*. [online] Analytics Magazine. Available at: http://analytics-magazine.org/images-a-videos-really-big-data/ [Accessed 24 Aug. 2019].

Verma, G. & Mallick, K. (1999). Researching Education: Perspective and Techniques.(London- Falmer Press).

Verma, J., Agrawal, S., Patel, B. and Patel, A. (2016). Big Data Analytics: Challenges and Applications for Text, Audio, Video, and social media Data. *International Journal on Soft Computing, Artificial Intelligence and Applications,* 5(1), pp.41-51.

Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR):* A Practical Guide, 1st Ed., Cham: Springer International Publishing, 10, 3152676.

Vossen, G. (2013). Big Data as the New Enabler in Business and other Intelligence. *In Vietnam Journal of Computer Science.* Available at: http://link.springer.com/article/10.1007/s40595-013-0001- 6/fulltext.html [Accessed 4 Aug. 2019].

Wallace, N., & Castro, D. (2018). The impact of the EU's new data protection regulation on AI. Center for Data Innovation, 27.

Wamba, S.F, Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'Big Data' can make big impact: findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234–246.

Wamba, S.F. & Akter, S. (2015). 'Big Data analytics for supply chain management: a literature review and research agenda', Workshop on Enterprise and Organizational Modeling and Simulation, 61–72.

Washington, E. (2018). *Why Data Governance is Crucial for Big Data Environments*. [online] ReadITQuik. Available at: https://www.readitquik.com/articles/data/why-data-governance-is-crucial-for-big-data-environments/ [Accessed 4 Jun. 2018].

Weitzner, D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., & Sussman, G. J. (2008). *Information accountability. Communications of the ACM*, 51(6), pp.82-87.

Wellington, J., & Szczerbinski, M. (2007). *Research methods for the social sciences*. A&C Black.

Wells, D. (2019). *A Data Quality Framework For Big Data*. [online] Eckerson Group. Available at: https://www.eckerson.com/articles/a-data-quality-framework-for-big-data [Accessed 4 October 2019].

Wilkins, J. (2021). *What is Metadata and Why is it Important?*. [online] Info.aiim.org. Available at: https://info.aiim.org/aiim-blog/what-is-metadata-definition-and-value [Accessed 16 March 2021].

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), pp.1-9.

Williams, G. (2019). *Applied qualitative research design*. Scientific e-Resources.

Williamson, K. (2002). Research methods for students, academics and professionals. 2nd ed. Information Chandos Publishing.

Workfront. (2019). Data Discrimination: The Dark Side Of Big Data. [online] Available at: https://www.workfront.com/blog/data-discrimination-the-dark-side-of-big-data [Accessed 30 September 2020].

World Economic Forum. (2014). *The global Information technology report 2014 -* Rewards and risks of Big Data.

World Economic Forum. (2016). *What do we mean by 'governance'?* [online] Available at:https://www.weforum.org/agenda/2016/02/what-is-governance-and-why-does-it-matter [Accessed 11 Oct. 2019].

Wu, S. (2015). *Big Data Processing with Hadoop.* Bachelors's thesis. Turku University of Applied Sciences.

Su, X. (2013). *Introduction to Big Data*. [online] Ntnu.no. Available at: https://www.ntnu.no/iie/fag/big/lessons/lesson2.pdf [Accessed 5 Apr. 2019].

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks,CA: Sage.

Yin, R. K. (2006). *Case study methods*. In J. L.Green, G.Camilli, & P. B.Elmore (Eds.),Handbook of complementary methods in education research (3rd ed., pp. 111–122).Washington, DC: American Educational Research Association.

Yin, R. K. (2014). *Case Study Research: Design and Methods,* (5th Ed.) Los Angeles: Sage.

Young, J.C., Rose, D.C., Mumby, H.S., Benitez-Capistros, F., Derrick, C.J., Finch, T., Garcia, C., Home, C., Marwaha, E., Morgans, C., Parkinson, S., Shah, J., Wilson, K.A. and Mukherjee, N. (2018). A methodological guide to using and reporting on interviews in conservation science research. *Methods in Ecology and Evolution*, [online] 9(1), pp.10–19. doi:10.1111/2041-210x.12828.

Youssef, T. (2010) Corporate Governance an overview – Around the Globe, Grant Thornton, Egypt.

Zaman, N., Seliaman, M., Hassan, M. & Marquez, F. (2015). Handbook Of Research on Trends and Future Directions In Big Data And Web Intelligence. Hershey, PA: Information Science Reference.

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of Industry 4.0: *A review. Engineering*, 3(5), pp.616–630.

Zicari, R.V. (2012). *3 Big Data: Challenges and Opportunities*. [online] Available at: http://www.odbms.org/wp-content/uploads/2013/07/Big-Data.Zicari.pdf.
Zikmund, W.G. (2003). *Business Research Methodology (7th ed.).* United States: Southwestern.

Zomaya, A. Y.& Sakr, S. (2017). *Handbook of Big Data Technologies*. Springer.