The impact of data governance on corporate performance:

The case of a Petroleum Company

A Dissertation presented to the

Department of Information Systems

University of Cape Town

By

Zimasa Ndamase (NDMZIM001)

Supervised by

Professor Michael Kyobe

Date: 29/08/2014

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the Departmental writing guide for citation and referencing. Each contribution to, and quotation in, this Literature Review, from the work(s) of other people has been attributed, and has been cited and referenced.

3. This Literature Review is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature: Date: 15/08/2014

Zimasa Ndamase

NDMZIM001

Table of Contents

Table of figures

Table of Tables

Abstract

While it is acknowledged that data is a valuable corporate asset, many companies fail to exploit it in order to better their performance. Organizations today need to be proactive in their operations and have to make informed business decisions in less time than ever before. This puts pressure on the organisations to better govern the use of data within an organization. Literature has shown that a holistic conceptualization of factors affecting data governance is missing. Also there is limited research on the effects of data governance on firm performance. This study therefore seeks to fill this gap by investigating the factors that affect data governance in organization X which operates in the petroleum industry and also determine the extent to which the quality of data governance influences its corporate performance.

A conceptual model derived from the literature review was used to guide this study. Data was collected from 50 employees in organisation X whose job descriptions are aligned with data management via an intranet web based survey. Quantitative methods were then used to analyse the data. Results of the regression analysis confirmed four out of six research propositions made. Compliance with data policies and regulations, data stewardship and ownership were not found to be significant predictors of data governance. However, data modeling, data integration and data quality are necessary in order to achieve improved data governance. The present study also confirms that poor data governance has a negative impact on corporate performance suggesting that organisation X needs to enhance the quality of data governance in order to realise its full business value and also improved business performance.

Acknowledgements

I wish to extend my sincere gratitude to my supervisor Professor Michael Kyobe for his consistent support and assistance through useful comments and suggestions. Thank you for always making time for the consultations, staying after hours or being late for your next meeting just to make sure I have the right material to work with. I also thank my employer for funding my studies and my bosses for the encouragement to always embark on a journey of growth.

Dedication

I dedicate this dissertation to the Ndamase family, my mum and dad, Mrs Nomhle Beauty Ndamase and Mr Simphiwe Hamilton Ndamase. To my siblings who are my support structure Tabisa, Siviwe, Aluncedo, Asithandile, Asabonga and Athenkosi, who have been there throughout this journey always encouraging me whenever I struggled and encouraged me to dust myself and keep fighting. I dedicate this Masters to Chad Chandalala my number one fan, thank you for always believing in me. To my fellow Masters student who became a wonderful friend and a sister Kudzai Kashora, keep striving for excellence. To my dearest friends Mfundo Nomvungu, Asanda Matolweni, Sinovuyo Msizi, Bulelwa Mpongoma, Lungisile Chayichayi, Vuvu Qokweni, thank you for showing me so much love and support. Last but not least I dedicate this dissertation to my one and only son Endinalo, although this was not an easy journey it was better to sacrifice now. I assure you the rewards will be beneficial to us.

1 Introduction

Organizations today need to be proactive in their operations and have to make informed business decisions in less time than ever before. Organizations face increasing pressure to improve value, accountability, performance, and quality (while reducing risk) to meet the demands of stakeholders, customers, employees, and the government. High quality data is necessary to achieve the organisation's strategic needs and the changing organisational requirements (Newman & Logan, 2006). It is also necessary for business networking (Tellkamp, Angerer, Fleisch, & Corsten, 2004), customer management (Crié & Micheaux, 2006), decision-making and business intelligence (Price & Shanks, 2005), and regulatory compliance (Friedman, 2006).

While it is acknowledged that data is a valuable corporate asset, many companies fail to realize its full business value. This has been attributed to poor data governance (Newman & Logan, 2006; Panian, 2010). An effective data governance program enables the development of formal policies and standards, and ensures oversight over data so that decision-makers may receive accurate and timely information to respond to challenges and opportunities identified above.

Researchers argue that while the concept of data governance is not new to most organizations, there are still many issues relating to lack of effective data governance policies and solutions (Rand secure Data,2013), lack of clarity on the interaction of role and responsibilities in data governance programs, poor design of decision-making structures within the data governance programs and limited information on best practices for the development of governance requirements for IT related systems (Kim et al., 2014; Weber, Otto, & Osterle, 2009).

This study investigates the factors that affect data governance in organization X and also determines the influence of the quality of data governance on the corporate performance of the organisation. This study also identifies which of the factors have the greatest influence on data governance. This study was conducted in a petroleum firm in South Africa. However due to confidentiality agreement the name of firm will not be revealed, it is going to be referred as organisation X henceforth.

1.1 Research Problem

The oil and gas sector is faced with an increasing amount of pressure to report a single version of truth. Data governance processes must therefore be clearly defined, repeatable and auditable, allowing risks to be quantified and mitigated. Literature shows that the quality of corporate data yields to better performance of the firm (Berson & Dubov, 2007b; Cheong & Chang, 2007; Otto, Wende, Schmidt, & Osl, 2007; Sheng, 2003). However the literature discussing the complexity and impact of data governance in the oil and gas industry is limited. Existing literature cover data governance issues in isolation as they tend to focus on single aspects which also results in isolated solution to data governance challenges (Otto, 2011b).

This research aimed to identify the factors influencing data governance in a petroleum firm and the significance of these influencing factors collectively. It is also necessary to determine the extent to which data governance influences the corporate performance. Tallon et al. (2013) stated in their study's recommendation for future research that there is a need of research related to where and how data governance impacts a firm's performance on all aspects and this is very valuable in today's information era.

1.2 Necessity for and Value of Research

In the oil and gas industry, the volume of data is growing rapidly. There are many reasons associated with this rapid growth. One reason being the increase in business activities due to growing competition and costs of mining and processing oil products. This created the need to increase data storage as more fields become cost effective to mine. The ability to store ever-increasing amounts of data introduced the challenge of the organization's ability to manage,
analyze and apply data (Smith & Mckeen, 2008).Organisation X in the past five years participated in joint ventures, merges and acquisitions of assets to expand its business. Inventory data of each party were carefully studied to determine if there are any duplicates and also to identify useful datasets in order to archive or delete unnecessary data. There had also been an improvement in retailing services where petrol stations had evolved into convenience stores with services such as food courts, loyalty cards, car wash and promotions. The improvement was designed to enhance customers' experience, retain customers and build brand loyalty. With this expansion of business activities, there has been more need for data management and also to provide quality data in real-time, clearer communication and collaboration, shared context between production teams and an expanded pool of connected resources.

Recently emerging legal and regulatory components such as Sarbanes-Oxley, ECT Act, and POPI have deeply affected the petroleum industry (Khatri & Brown, 2010; Tallon, Ramirez, & Short, 2013). Due to this, petroleum firms transformed data governance into an essential prerequisite for effective corporate governance. The legal department of Organisation X enforces that data must be properly classified, managed and governed to avoid any legal penalties.

The cost associated with poor data management in oil and gas can reach up to 22% of the annual revenue (Westheimer Energy, n.d.). This cost is related to imprecise or redundant decisions and activities resulting in high costs, unnecessary downtime, suboptimal production rates, non-compliance with regulations and increased safety issues (Industry Series Energy, 2012). This raised a need for the petroleum firms to improve the way they govern and manage their data.

1.3 Research Questions and Objectives

This study was conducted in a petroleum firm. In the oil and gas industry, the volume of data is growing rapidly. There are data challenges that are currently facing petroleum industry due to this rapid growth. The data challenges are: firstly it is hard for someone to get access to the information they need. Secondly, there is a need for systems to better analyse, manage and standardize data for a specific query. Thirdly, there is a growing need for technological capabilities which allow any-time collaboration from anywhere (Adams, 2013). These are also applicable to Organisation X.

Organisation X needs to govern its data efficiently and effectively in order to obtain value and insights to increase profitability as well as achieving a competitive advantage. In order to overcome these challenges these questions need to be answered:

- Which factors influence data governance in organisation X operating in the petroleum sector?
- To what extent does the quality of data governance influence the corporate performance of organization X?

This leads us to the objective of this study:

- To investigate the factors that affect data governance in organization X and also determine the extent to which the quality of data governance influence the corporate performance of organization X.

1.4 Outline of the Dissertation

This document is organized as follows: Chapter 2 presents a review of Literature on data governance and also synthesizes the existing literature. The literature review expands on underlying theories and concepts which are applicable to this research. From the synthesis of the literature, a conceptual model was developed to illustrate the relationships between the identified factors affecting data governance and data governance with corporate performance and creating a concept to properly address research propositions and answer research questions. Chapter 3 presents research design which describes the research paradigm, strategy and methods which are relevant and helpful to get to the objectives of the study. It discusses the research instrument, sampling and data collection and analysis techniques used in this in study. Chapter 4 presents results of the data analysis outlined in the previous chapter. Chapter 5 presents the discussion and Conclusion which interpret and describe the significant of the research findings in relation to the existing literature and draws conclusions. The last section presents recommendations and suggestions for future research and also implications of the findings to practice.

2 Review of Literature

There has been ongoing research for the past three decades on the frameworks or the models that could provide organisations with a set of fully integrated and high quality data (Smith & Mckeen, 2008). It has been data dictionaries in the 1970s to data warehouses in the 1980s to Enterprise Application Integration (EAI) in the 1990s. In the 20^{th} century data governance is the fastest growing research area within the data management research area. Although this area has been intensively researched, there are still huge challenges that organisations are still facing such as poor data quality; the politics of data ownership; synchronization issues; legal and regulatory considerations, which appear to change constantly; the difficulties of agreeing on a single definition of every data item; getting the business to recognize the value of the work involved (Smith & Mckeen, 2008). The previous chapter defined the problem under investigation and outlined the scope and relevance of this research.

This chapter provides a review of literature on data governance and also synthesizes the existing literature. It is organised as follows: *Section 2.1* clarifies the concepts underlying data. *Section 2.2* identifies the concepts underlying corporate, IT governance and data governance and establishes their relationship. The objective is to trace the origins of data governance. *Section 2.3* explains the theoretical foundation which this research is based on; explaining theories which fit or explain the data governance area. *Section 2.4- 2.9* elaborates concepts of data management and identifies relationships between them and how they relate to data governance. *Section 2.10* summarises the literature review with a conceptual model that shows the key elements and relationships that emerged from the literature review. The model acts as a framework for investigating the current research problem. *Section 2.12* presents the research propositions to be evaluated in response to the research question.

2.1 Data

Data is an intangible asset of great value in an organization. It is the key enabler for efficient processes and the real manifestation of the business as it represents an organization's customers, employees, and suppliers; its activities and transactions; and its outcomes and results (Panian, 2010). Data helps in the development of internal capabilities that can be used to absorb and leverage their gains from the utilization of resources (de Abreu Faria, Gastaud Macada, & Kumar, 2013). There has been plenty of research to investigate the demarcation between data and information. The concept of data is delicate and vague and this resulted in a number of competing definitions. According to Cleven & Wortmann (2010), data is defined as raw material such as numbers, symbols or other representations of facts, for information needed for everyday operations and satisfactory decision making. Information is data that has been processed or put into context (Otto et al., 2007; Weber et al., 2009). In the IT field it is common to use data and information as synonyms. We will use the terms in this study interchangeably.

Data is divided into domain data, metadata and reference data. Domain data represents the business domain at hand and is divided into master and transactional data. Master data refers to core business entities a company uses repeatedly across many different business processes and systems, together with their associated metadata, attributes, definitions, roles, connections and taxonomies(Cleven & Wortmann, 2010; Silvola, Jaaskelainen, Kropsu-Vehkapera, & Haapasalo, 2011). Typical master data are parties (customer, prospect, people, vendors, suppliers), places (locations, offices) and things (accounts, assets, policies, product) (Silvola et al., 2011). Master data has the following characteristics:

- Master data objects are independent of other objects.
- Secondly master data usually remain largely unaltered. It is static data.
- Lastly instances of master data classes (e.g. customer data) are quite constant with regard to volume.

Transactional data represents business transactions or events such as sales orders, production requests or invoices etc. Transactional data unlike master data has its data objects changing during its lifecycle and is dependent on master data. Its data volume increases with on-going business activities e.g. Sales orders.

Reference data is described as the representation of an agreed-upon set of values that are used across multiple organizational units or systems to ensure consistent values for attributes of master data, transactional data or metadata e.g. the abbreviation for currency or gender(Cleven & Wortmann, 2010). Although it has the same characteristics as master data it is not limited to domain data but is rather granular and fine.

Metadata relates to domain data and is classified into two separate classes namely operational and informational metadata (Cleven & Wortmann, 2010). Operational metadata enables the design and technical operation of information systems while Informational metadata assists in the understanding and access of domain data and is maintained for end users.

2.2 Governance

Governance is a ubiquitous term in the business, and it has different interpretations depending on the perspective of the user. According to Kooper, Maes, & Lindgreen (2011), governance provides a structure for determining business objectives and monitoring business performance to ensure that objectives are accomplished. Governance ensures that "stakeholder needs, conditions and options are evaluated to determine balanced, agreed-on enterprise objectives to be achieved; setting direction through prioritisation and decision making; and monitoring performance and compliance against agreed-on direction and objectives" (Raval & Dyche, 2012, p1). Governance refers to the approaches that the organization adopts to ensure that strategies are set, monitored, and achieved (Weber et al., 2009). In a nutshell governance empowers the principal to monitor and control the behaviour of an agent (Kooper et al., 2011; Raval & Dyche, 2012).

2.2.1 Corporate and IT Governance

Corporate governance is defined as the set of processes, customs, policies, laws and institutions influencing the way an organisation is administered, controlled and directed. It is largely of interest to the principal stakeholders, that is, board of directors, management and shareholders as it is the discipline which focuses on the proper functioning of management and the goals for which the organisation is governed (Beijer & Kooper, 2010; Kooper et al., 2011).

IT Governance is a sub-discipline of corporate governance which focuses on the governance, risk and performance management of IT systems. It also includes alignment processes, communication tools and decision making structures which ensure that IT sustains and extends an organization's strategies and objectives. It is a well-established discipline and can be viewed as an integral part of corporate governance. Tallon, Ramirez, & Short (2013), define IT Governance as the management, use, and control of physical IT artifacts (hardware, software, networks). IT governance is the organizational capacity exercised by the board, executive management, and IT management to control the formulation and implementation of IT strategy to ensure the merging of business and IT (Kooper et al., 2011). The emergence or the increased interest in IT governance arose after the passing of compliance initiatives i.e. the Sarbanes–Oxley Act in the United States in 2002 and Basel II in the European Union in 2004, but it has existed from a research standpoint as IT infrastructure, IT business value, IT controls, and project management literature for over two decades (Kooper et al., 2011; Tallon et al., 2013).

IT governance is divided into three types of practices namely structural, procedural and relational practises. Structural practice focuses on designating responsibilities to IT executives and relevant IT committees on roles of supervision, directing and planning of IT governance (Ko & Fink, 2010; Tallon et al., 2013). This practice is the most important predictor of whether an organisation will develop value form IT (Ko & Fink, 2010). Procedural practice is for shaping user behaviours through cost control, IT control, IT value analysis and resource allocation. Controls come from policies and practices which are produced through mutual agreements between the principal and an agent. This promotes greater accountability via service-level agreements and IT chargebacks which aids or adds in the main aim of IT governance to promote accountability around IT resource usage and increase of business value through IT (Tallon et al., 2013). Relational practice shapes involvement in IT governance by allowing shared participation of non-IT and IT decision makers through IT knowledge sharing, business-IT partnerships, communications, idea exchange, and conflict resolution (Tallon et al., 2013).

From the literature, it is evident that data may be a minor consideration when an organisation is strategizing on how to govern its physical IT artifacts. Yet data center technologies are seen as an essential aspect of IT governance and storage as a necessary or guiding component of IT infrastructure (Tallon et al., 2013). The focus is more on the physical box and less on the data inside the box. There are implicit restrictions or limitations on IT governance which are inherent or resulting from the concept of IT governance.

These limitations are clearly specified in papers by Beijer & Kooper (2010) and Kooper et al.(2011). The first major inherent limitation is that IT governance is not concerned about how organisations handle information (creation, consume, process and exchange) in order to create business value, and only focuses on purchases and care for physical IT artifacts. Secondly, its main focus is in the control side of the business domain which includes administration, policymaking, responsibility, authorization, reporting, monitoring and audit (Kooper et al., 2011; Tallon et al., 2013). This neglects development issues which are the other half of the business universe and host vital elements such as entrepreneurship, innovation, creativity, improvisation, value creation and experimentation. This creates little room for IT-driven development strategies which could widen the gap between business and IT instead of bridging it. Thirdly, IT governance often suffers from incomplete or half-hearted implementations due to formal or bureaucratic environments which it normally creates and is not favourable for IT professionals (Kooper et al., 2011). Realizing this gap and limitations in IT governance, data governance emerged using some practises which are already in use in IT Governance and new practises which are unique and specific to data. The intelligent and innovative application of information solves business problems and creates customer value at high speed, low cost, and the right scale. (Tallon et al., 2013). Tallon et al. (2013) emphasis that it is not about the box; it is about what is inside the box.

2.2.2 Data Governance

Data needs to be governed in order to address data quality issues and for an organization to be able to quantify and measure its data quality. People and tools shape data and determine where it should go. This implies that data governance (DG) is the governance of the people and technology. There are numerous definitions of data governance because this sphere encompasses many things which are required to ensure data quality. Panian (2010) defined data governance as “the processes, policies, standards, organization, and technologies required to manage and ensure the availability, accessibility, quality, consistency, auditability, and security of data in an organization.”

Cheong & Chang, (2007) state several definitions in his study; data governance is the process by which a company manages the quantity, consistency, usability, security and availability of data. Data governance is the collection of decision rights, processes, standards, policies and technologies required to manage, maintain and exploit information as an enterprise resource. Data governance refers to the organisational bodies, rules, decision rights, and accountabilities of people and information systems as they perform information-related processes and it also sets the rules of engagement that management will follow as the organisation uses data. There is some inconsistency in these definitions as some focus on data while others on information and not all required measures of data quality are mentioned in the definitions. This study will adopt this definition: Data governance is the collection of decision rights, processes, standards, policies and technologies required to manage and ensure the quantity, usability, availability, accessibility, quality, consistency, auditability, and security of data in an organization (Cheong & Chang, 2007; Panian, 2010). In short Data governance defines policies and procedures to ensure proactive and effective data management.

Otto (2011) states the most common business drivers of data governance initiatives are: a) is to ensure compliance; b) enable decision-making; c) improve customer satisfaction; d) increases operational efficiency; e) support business integration. Goals of data governance are to ensure data fulfil business requirements by ensuring that data is reliable, secure, and accessible for decision making. Secondly DG lowers the costs of managing data. Lastly DG protects and treat data as the most valuable business asset ensuring that its value does not diminish through technology or human error, loss of timely access, inappropriate use, or misadventure (Panian, 2010; Tallon et al., 2013). Data governance practises are aligned to IT governance practises which are already in use and tested. IT governance is divided into three types of practices namely structural, procedural and relational practices. Refer to section 2.2.1 for the definition and explanation of these practices.

The next section explains the theoretical foundation which this research is based on; explaining theories which fit or explain the data governance area.

2.3 Theoretical Background

Theoretical foundation for this work is based upon three theories: Resource- based theory (RBV), Dynamic Capabilities theory and Agency Theory. This study looks at data, its governance, how the dynamics within and around the organisations affect its data governance practices and its performance. The agency theory was used to conceptualise governance. Data as an organisation resource was conceptualised using the resource based theory. The Dynamic Capabilities theory was used to reflect the dynamism of the organization. These relationships are represented in the theoretical model below (see Figure 2.1).

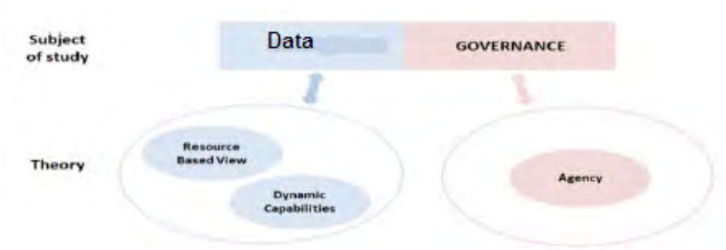

Figure 2.1 : Theoretical model (de Abreu Faria et al., 2013)

Agency theory extends risk sharing literature which arises when there is a difference in attitudes towards the risk within cooperating parties. Agency theory includes the so-called agency problem that occurs when there is a difference in goals and division of labour within the cooperating parties (Bhattacherjee, 2012; Eisenhardt, 1989). The cooperating parties in this theory are the principal who delegates work to the agent. That is the reason this theory is also called principal-agent theory. The goal of Agency theory is to stipulate optimal contracts and conditions which can help to minimise the conflict between the cooperating parties given assumptions about people, organization and information (Bhattacherjee, 2012). Contracts are the essence of the firm, and are made to limit divergences in principal-agent relationships, generating agency costs (Lajara, Carlos, & Maçada, 2013). Eisenhardt, (1989) states two problems which agency theory addresses which occur in agency relationships. The first is the agency problem which arises when the desires or goals of the principal and agent differ, and also it is difficult and expensive for the principal to validate what the agent is actually doing. The second one is a risk sharing problem which arises when cooperating parties have different attitudes towards risk; risk preferences and the actions to take.

In the context of the present study, the Principal is the stakeholders who need accurate information to be available to them when needed in order for them to make informed decisions. The Agent is someone who is responsible for delivering such information to the principal. There should be contracts with clear conditions, and rules in place in order for these parties to work well towards a desired goal of the principal.

The application of this theory is also in an information security setting where management (principal) and employees (agents) have conflicting interests. This conflict arises when responsibility of whether to adhere to or ignore information security policies is delegated to employees (Herath & Rao, 2009). Employees will choose routes based on their interests, either breaking security policies for malicious purposes or choosing to evade security policies for convenience. Monitoring employees' actions related to security policy compliance is hard to accomplish and not practically possible and extremely costly (Herath & Rao, 2009). Agency theory can help in providing a strategy of incentive/disincentive mechanisms which encourages and motivates employees to perform the activities required to protect their organization's information in order to increase level of policy conformance (Bulgurcu, Cavusoglu, & Benbasat, 2010; Herath & Rao, 2009). Key elements of Agency theory such as control, monitoring, risk, rules, alignment and structure were used to describe the governance element.

The Resource- based theory (RBV) is used to determine which resources are critical (or strategic) that the organization needs to focus on. Resources are tangible and intangible assets, that differentiate an organization from its competitors and are difficult to imitate and substitute on which the organizations depends on for long term competitiveness (Rivard, Raymond, & Verreault, 2006). RBV emphasizes that an organization's resources are the essential determinant of competitive advantage and performance (Bridoux, 2004). There are conditions that a resource of an organization must have in order to sustain this competitive advantage. Organizational resources must be heterogeneous, immobile, be valuable, rare, and imperfectly imitable and substitutable in order to be a resource for sustained competitive advantage. Resource heterogeneity is deemed to be a crucial condition for a resource bundle to contribute to a competitive advantage (Bridoux, 2004; Rivard et al., 2006). RBV elements such as heterogeneous resources, data as an asset (resource), quality, information systems (IS) and value (Bhansali, 2013; de Abreu Faria et al., 2013) are considered in the present study.

Dynamic Capabilities theory is considered as an extension of RBV, but distinguished in how competitive advantage is achieved from the resources. Capabilities are what the firm can do with the resources it has. Dynamic Capabilities focuses on the ability of a firm to sense and then seize the opportunities in order to achieve new forms of competitive advantage by reconfiguring organizational resources in a business environment of rapid changes (Augier &

Teece, 2009; de Abreu Faria et al., 2013; Rivard et al., 2006; Wheeler, 2002). Similar to the definition of the authors above, Eisenhardt & Martin (2000) defines dynamic capabilities as the firm's processes that use resources; specifically the processes to integrate, reconfigure, gain and release resources to match and even create market change. These capabilities must be dynamic because an organization must frequently build, adapt, and reconfigure internal and external competences to achieve congruence or balance with the changing business environment because the rate of technological change is rapid (de Abreu Faria et al., 2013; Wheeler, 2002).

Dynamic Capabilities are unique but they are not for long term competitive advantage. They provide more value when they are applied sooner, more astutely, or more fortuitously than rivals (Augier & Teece, 2009). Dynamic Capabilities provides the concepts of flexibility, adaptability, integration and reconfiguration (de Abreu Faria et al., 2013). These are the characteristics which are significantly important to data governance. Organizations often invest a lot in IT in order to sense and respond to rapid changes of the business environment they operate in be it external or internal. Exponential growth in data creation causes rapid changes in data application and importance to the business and it causes major changes in the way organizations operate (de Abreu Faria et al., 2013). The elements of the Dynamic Capabilities theory such as rapid change, skills, context, learning, knowledge and capabilities (Lajara et al., 2013) also capture important characteristics of effective data governance.

2.3.1 Organizational Information Processing Theory

Galbraith (1973) showed that organisations need to process information in order to reduce uncertainty and thereby achieve organizational coordination and control. This theory identifies three important concepts: information processing needs, information processing capability, and the fit between the two to obtain optimal performance (Saunders, Premkumar, & Ramamurthy, 2005). Quality information is crucial in order for an organization to cope with uncertainty and improve their decision making. There are two strategies applied by organizations to cope with uncertainty and increased information needs:

- develop buffers to reduce the effect of uncertainty
- Implement structural mechanisms and information processing capabilities to improve the information flow and thereby reduce uncertainty (Saunders et al., 2005).

2.3.2 Uncertainty & Equivocality

Uncertainty is defined as the difference between the available information and the required information to complete a task, which is the absence of needed information (Cooper & Wolfe, 2005). An organization is made up of sets of groups or departments referred to as subunits/units. The tasks of organizational subunits vary in their degree of uncertainty. Three sources of work related uncertainty, exist namely subunit task characteristics, subunit task environment, and inter-unit task interdependence.

Galbraith's (1973) work was extended by Media richness theory to include equivocality (Cooper & Wolfe, 2005). Equivocality is defined as unclear enquiries due to the presence of multiple and conflicting interpretation (Cooper & Wolfe, 2005). The precise information to execute the task is not clear when equivocality is present. Equivocality results from one of two underlying causes:

- A complex task with cause-effect relationships that are not well understood is characterized by equivocality.
- A task's underlying meaning may not be as well understood, because its compatibility with the organization's history and current direction could be open to question (Cooper & Wolfe, 2005)

Two sources of equivocality are subunits differentiation and task analysability. Subunits differentiation develops as the units have distinct specialties, goals, frames of reference, and jargon. This makes the communication between the subunits complex, ambiguous, and difficult to interpret. When differentiation is high, the equivocality is extensive because it is difficult to process the information due to ambiguity, conflict, and misunderstanding. Analysability is the process where individuals follow objective, systematic procedures in completing a task. When task analysability is low, the equivocality is extensive because task execution relies on judgment and experience, which contribute to multiple interpretations and ambiguity.

Media richness theory argues that organisations process information to reduce both uncertainty and equivocality. “Uncertainty can be reduced by a sufficient amount of information, while equivocality can be reduced by sufficiently rich information” (Goodhue, Wybo, Kirsch, & Ha, 1992)

2.4 Data Stewardship and Ownership

Data in the organization is always shared, integrated and utilized in inter-organizational ways. Data sharing is defined as distribution of data. Either the well-structured or semi structured, among units for further use (Ahmad, Abidin, & Omar, 2011). Data can be easily replicated and shared across a vast distance. Its value does not decline as the usage increases instead it gains more value (Tallon et al., 2013). In order for the effective sharing of this data amongst the units of an organisation and to avoid conflicts, there should be a clear assignment of the right roles to the right decision areas with the right accountability. This is very critical and important due to access levels of data, namely data enclaves, restricted data and public data (Ahmad et al., 2011). Data enclaves are classified as the most restricted data. Restricted data has lesser access priority but its secrecy is imposed by data policies. Lastly public data has no restrictions and is shared with anyone who requires access to the data.

The concept of data stewardship is different from data ownership. Data owners are those individuals or groups in the organisation that have a lawful claim towards data and have control capabilities (obtain, create, have access to and the distribution of data) (Ahmad et al., 2011; Berson & Dubov, 2007a). Data owners often belong to the business and not to the IT department of the organisation. Data ownership is critical and sensitive due to the fear of data manipulation which may lead to negative consequences the organisation.

The main issue which affects ownership in the organisations is data access. Rosenbaum (2010) suggested two approaches to address this namely incentives and treating information as a public good. Departments are reluctant to avail their best resources (data) to other department's projects regardless if the input will give valuable value to the project or whether it is in the interest of the organization as a whole. They need compensation as being the source of that data. Providing an explicit contract that rewards those who create and maintain data, "ownership" will be the best way to provide incentives (Rosenbaum, 2010; Van Alstyne, Brynjolfsson, & Madnick, 1995) This strengthens the recognition of data ownership rights (Rosenbaum, 2010). Choosing best incentives which satisfy both principal and agent will be beneficial because organisations will utilise the available technology to its full potential. The subtle intangible costs of low effort will appear as distorted, missing or unusable data which affects quality of data governance negatively (Van Alstyne et al., 1995).

Alternative approach is to treat data as a public that can be used by different departments of the organisation according to principles of data stewardship. Stewardship entities have broad authority to collect, prepare and support the use data within the organisation (Rosenbaum, 2010). Also to designate certain data uses as being in public interest to be made available to other departments who are able to demonstrate compliance with data stewardship responsibilities (Rosenbaum, 2010).

The role of data stewards is to ensure that the agreed-upon quality metrics are maintained on a continuous basis, and making sure appropriate data quality improvement programs are in place (Berson & Dubov, 2007a; Rosenbaum, 2010). Data stewardship has an ability to help IT departments to effectively improve applicable architectural components to improve data quality. It also ensures that business meta-data is defined, created and effectively maintained across the organisation and measurable data quality goals are identified (Berson & Dubov, 2007a).

In order to effectively manage data sharing, one has to identify the nature of data, its suitable platform, pertinent issues and approaches and how it can be shared among units (Ahmad et al., 2011). The next section focuses on data integration which is essential building block for effective data sharing as it aids in bringing data together from different sources.

2.5 Data integration

Data integration is a process of combining data residing in multiple autonomous and heterogeneous data sources in order to provide the user with a unified view of this data, while the data sources remain unaltered (Lenzerini, Sapienza, Salaria, & Roma, 2002; Souza, Belian, Salgado, & Tedesco, 2008; Ziegler & Dittrich, 2007).
Heterogeneity can be divided into two categories namely structural and semantic heterogeneity (Hristov, 2012). Structural heterogeneity refers to data that is stored in different structures in different sources/systems. Semantic heterogeneity means data has different meaning in different sources/systems (Hristov, 2012; Ziegler & Dittrich, 2007). Goodhue et al. (1992, p. 294) defines data integration as “the standardization of data definitions and structures through the use of a common conceptual schema across a collection of data sources”.

Ziegler & Dittrich (2007), state two main reasons for integration. The integrated view can be used as a single information access point and can aid in facilitating information access and reuse. Secondly, data from different sources gives a more comprehensive basis to satisfy a certain need of information. Data integration is a well-established area in database research. It emerged after database systems were developed around 1980 (Ziegler & Dittrich, 2004, 2007). Earlier integration was easily achieved because they were integrating purely well-structured data. Earlier integration approaches were based on relational or functional data models and were providing a single global schema (tightly coupled solutions). Object-oriented integration approaches were adopted to perform structural homogenization and integration of data in order to address the limitations of earlier integration approaches such as abstraction, classification, and taxonomies (Ziegler & Dittrich, 2007).

Due to the rapid creation of data, the type of data that needs to be integrated shifts from purely well-structured to include also semi and unstructured data over time. This creates more demanding integration because data models will be different and also encounter more problems because data will have heterogeneous semantics. Semantics is described as “people’s interpretation of data and schema items according to their understanding of the world in a certain context” (Ziegler & Dittrich, 2007, p8). There is no complete semantics that can be valid for all users, they depend on context. There is a lot of research which has been done on structural integration to address approaches and issues encountered. More open research challenges in the area for data integration falls in the semantics problems (Ziegler & Dittrich, 2007). Semantic data integration ensures that only data related to the same or closely-related to real-world entities or concepts is structurally merged. Metadata plays a big role in this integration because one depends on it to produce underlying assumptions contained in the data. The context of data needs to be integrated in order to have clear semantics of it. This is difficult to achieve as it is not possible to exhaustively stipulate the intended real-word semantics of all the data and schema elements due to the fact that semantics can be embodied in different areas. They can be embodied in data models, conceptual schemas, the data itself, application programs, and the minds of users (Ziegler & Dittrich, 2007).

Research in this area has led to the design and creation of data integration systems or data mediators (Bennett & Bayrak, 2011). These systems are normally referred to as triple <G, S, M>. G is the global or mediated schema, S is the heterogeneous set of source schemas, and M is the mapping that maps queries between the source and the global schemas (Bennett & Bayrak, 2011). These help the user to query data in order to use and gain the benefits of the data integration. Users pose queries over G and the mapping then asserts connections between the elements in the global schema and the source schemas.

Data systems are designed and implemented using one of two strategies namely GAV (Global as View) or LAV (Local as View) (Bennett & Bayrak, 2011; Hristov, 2012). Global as View is the combination of views exported by heterogeneous, disjoint data sets. Its advantage is that all queries against the global schema if modeled correctly can produce complete answers. GAV has difficulty in coping with dynamically changing data sources. Every time the underlying data changes, the global schema should be written to cater for changes. This becomes a huge problem with large and dynamic number of data sources (Bennett & Bayrak, 2011).

LAV is a materialized view over a global schema. It is designed to cope with the large ever-changing number of data sources. It cannot efficiently answer queries posed to a global schema but rather addresses queries posed against materialized views (Bennett & Bayrak, 2011). Since these strategies have both strengths and weaknesses there has been work in creating a strategy which combines the strengths of both without many disadvantages. Examples of these are BAV (both as view), GLAV (global local as view)and the BGLAV (both global local as view) (Bennett & Bayrak, 2011; Hristov, 2012).

Although there has been extensive research on data integration, real world tools for things such as schema-matching and data federations are far and few between (Bennett & Bayrak, 2011).

The next chapter focuses on data modeling which helps in creating a good conceptual schema that is semantically correct, complete, easy to use in order to easy the laborious work done on data integration.

2.6 Conceptual Data Modeling

Conceptual modeling is defined as the process of gathering requirements and clearly documenting a problem domain by use of conceptual models for the purpose of understanding user's requirements and to aid the primary means of communication between the stakeholders such as owners, service providers, business analysts, developers, and users (Moody, 2005). Conceptual modeling is the cornerstone of many information systems activities not merely to define user requirements, but to also support development, evaluation, reengineering, acquisition, adaptation, standardisation and integration of information systems. Modeling is widely used for database design and management, business process documentation, business process improvement, and software development (Davies, Green, Rosemann, Indulska, & Gallo, 2006).

Research shows that it is cheaper to remove defects discovered during the requirements stage. Removing the same defect costs on average 3.5 times more during design, 50 times more at the implementation stage, and 170 times more after delivery (Moody & Shanks, 2003; Moody, 2005). This shows that the quality of a conceptual model is of concern because it can affect both efficiency (cost, time) and effectiveness (quality of information systems) of IS development. Studies on the impact of requirements errors showed that in practice even if requirements errors are detected after the analysis stage and are not corrected, it is believed that it is often too expensive or politically unaccepted to correct them (Moody, 2005). There are no common/agreed standards or frameworks which guide practitioners in order to evaluate the quality of conceptual models, or what makes a model a good one? This research area is in its infancy stage and there are many proposed frameworks but most of them have not been tested practically (Nelson, Poels, Genero, & Piattini, 2005). The evaluation of quality of conceptual models depends on the agreement of the experts and it is subjective based on experience and common sense.

Conceptual data modeling is just the subset of Conceptual modeling. A conceptual data model is a collection of conceptual tools for describing real world entities to be modeled in the database and the relationships amongst them (Bajaj, 2010; Silberschatz, Korth, & Sudarshan, 1996). "An entity is an object in the world that is distinguishable from all other objects" (Silberschatz et al., 1996, pg 1). According to Batra, Hoffer, & Bostrom (1990), a conceptual data model is an abstraction of real world (organizational) data pertinent to an enterprise. The aim of data modeling is to create a conceptual schema which serves as a communication tool between developers and users. The purpose of a data model is to design a database which performs efficiently, contains correct data and is easy to maintain and extend (Maguire, Worboys, & Hearnshaw, 2006). Also to provide accurate and unambiguous representation of organizational requirements (Young-Gul & Salvalore, 1995). Data modeling has a great impact on the quality of the final system as it is a major determinant of system development costs, system flexibility, integration with other systems and the ability of the system to meet user requirements (Moody & Shanks, 2003).

The major problem which has been researched intensively is to create a good conceptual schema that is semantically correct, complete, easy to use, and comprehensible (Shoval & Shiran, 1997). There are a few guidelines for evaluating quality of data models. Mostly the quality is dependent on the competence of the data modeller. There are two widely known classes for data models which have been used for DBMS development namely logical models and conceptual models.

The three major logical models are hierarchical, network and relational. The relational model has been widely applied in business organizations but this has its limitations. This model has failed to address complex semantics associated with geometric objects and complex unstructured data types such as photos (Liao & Palvia, 2000; Zhang, 2001). Although it has its limitations its strong points gave it the success it achieved. Its advantage is that it is based on a rigorous mathematical foundation, supports high-level query languages and is easy to understand (Silberscatz et al., 1996).

Conceptual models are entity-relationship model, semantic data model and object-oriented model. These have played major roles in research and in practice(Moody & Shanks, 2003; Wand & Weber, 2002; Zhou, Wang, & Xi, 2005). Conceptual models were proposed to address some limitations of logical models (Aguirre-urreta & Marakas, 2008; Bajaj, 2010) . Many of the semantic model concepts are incorporated in the Object Oriented model.

A widely accepted conceptual model is an Entity-Relationship Model (ER-Model) which adopts the view that the real world contains entities and relationships. Entity-Relationship Models represent the meaning of data using three basic concepts i.e. identifiable entities, relationships between entities and their associated attributes. This was later extended to accommodate the notion of categories and it was called the Extended Entity-Relationship Model (EERM) and is a more powerful version of the original one. EERM introduced two additional abstraction constructions namely generalization and aggregation (Bajaj, 2010; Maguire et al., 2006)

2.7 Data quality

Data plays a vital role in the operation of businesses or enterprises in the information age. Data contributes heavily on the wealth and future success of the enterprises as the businesses produce reports, deliver information, monitor performance, make decisions and achieve competitive advantages based on the data collected. Data volume in the organisations is increasing exponential and data generation has dramatically increased due to rapid changes in Information Technologies. Today's business environment is faced with critical issues of managing and improving the quality of data. When data quality is not dealt with seriously it can lead to enormous costs in billions of dollars (Sheng, 2003; Strong, Lee, & Wang, 1997). This shows that firms need to provide more attention to data quality issues, because the business could not last if it does not have high quality data (Azumah & Quarshie, 2012; Sheng, 2002). Although this research area has been intensively researched there is strong evidence that information quality issues have become increasingly prevalent in today's business practices due to little attention or low priority given to data quality areas as it is overshadowed by issues which are deemed to be important or more pressing (Azumah & Quarshie, 2012; Sheng, 2003).

Data Quality is defined in terms of data type and domain, correctness and completeness, uniqueness and referential integrity, consistency across all databases, freshness and timeliness, and business rules conformance (Cheong & Chang, 2007). According to Otto et al. (2007), data quality is defined with two aspects, the dependence of observed quality on the user's needs and "fit for purpose" which is the capability to meet the requirements in a specific situation. Data quality consists of six attributes to determine "fit for purpose". These are called data quality dimensions. These dimensions are accuracy, reliability, timeliness, relevance, completeness, currency, and consistency (Haug, Arlbjørn, Zachariassen, & Schlichter, 2013). Consistency determines if a data unit is specified the same throughout the system that is checking violations of semantic rules defined over data items. Accuracy defines how close a data item is to its true value in terms of meaning and "truthfulness". Completeness is measured according to population checks of completeness of columns of a table containing data. Timeliness describes promptness, freshness and frequency of updates of data (Azumah & Quarshie, 2012).

Data quality dimensions were further grouped into four data quality categories namely: intrinsic, contextual, representational, and accessibility. Completeness and timeliness belongs to contextual and accuracy belongs to intrinsic and consistency to representational data quality categories. Deeper understanding of data quality dimensions helps in effectively addressing problems and issues of data quality (Tayi & Ballou, 1998). Table 2.2 shows DQ Categories and dimensions. Although data quality is important there are regulations and policies that need to be taken into account when taking or performing the steps to improve the quality of the data. The next section explains and describes such regulations and policies.

DQ Category	DQ Dimensions
Intrinsic DQ	Accuracy, Objectivity, Believability, Reputation
Accessibility DQ	Accessibility, Access security
Contextual DQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational DQ	Interpretability, Ease of understanding, Concise representation, Consistent representation

Table 2.1 : **DQ Categories and dimensions**

2.8 Compliance

Companies are required to comply with external regulations and also internal corporate governance policies designed to increase transparency, accountability and to prevent fraudulent activities (Panian, 2010; Cheong & Chang, 2007). Companies must streamline the collection of reporting data to ensure compliance to internal policies (for data security and privacy), to external regulations such as Sarbanes-Oxley (SOX) Act, Control Objective for Information and Related Technologies (CobiT), and standards for data exchange (EDI, HL7, SWIFT, etc.) (Russom, 2008). Cobit is a general accepted framework used by IT auditors to assess SOX compliance(Cheong & Chang, 2007). There are concerns on how to handle data when trying to comply with these regulations. Concerns include sensibility in controlling data access and no clean accurate data for auditors. These concerns can be addressed by Information security policies.

Information security policies are defined as the processes and procedures that the employees should follow in order to protect the confidentiality, authenticity and non-repudiation, integrity and availability of information as it is the valuable asset of the organisation (Vroom, Solms, Technikon, Elizabeth, & Africa, 2004). Auditors use these policies as they are guidelines that dictate the rules and regulations of the organization, which in turn govern the security of information (Vroom et al., 2004). Risks associated with improper information security policy compliance incur huge damages to organisation like corporate liability, loss of credibility, and monetary damage (Bulgurcu et al., 2010; Haug et al., 2013).

Although these policies can be clearly defined and detailed, extensive literature shows that the results are not as desirable, because employees seldom comply with information security procedures (Herath & Rao, 2009) . This creates a major impediment for the organisations because they need to develop strategies for improving their employees' adherence to information security policies. If an organisation can overcome this barrier it can benefit from information security policies. When employees properly adhere to them these are the four outcomes; strategic alignment, value delivery, risk management and performance measurement (Williams, 2001). The next chapter will present the areas of corporate performance which are affected or impacted by data governance.

2.9 Corporate Performance

Organizations are spending lots of money on IT with little to show for it in the output statistics. Although IT investment is associated with superior performance, it is difficult to prove this claim as it is associated with increasing productivity rather than financial impact like return on asset and return on equity. Practitioners should rigorously justify investments in technology and be able to show the gain to get buy in from top management. Corporate performance measurement gives little attention to business process measurement as the focus is strongly on the traditional functional structure (List & Machaczek, 2004). List & Machaczek (2004) emphasises the importance of measurement as measurements are the key. If you cannot measure it, you cannot control it. If you cannot control it, you cannot manage it. If you cannot manage it, you cannot improve it. This clearly shows that measurements are important for control and improvement of the current processes in place.

Most previous research when measuring improved organization performance due to IT investment excessively focuses on financial indicators such as return on investment, return on assets and ratio of expenses to income (Markus & Soh, 1995).Literature on organizational effectiveness showed that organizational performance should not be defined by financial indicators rather it depends on how the organization is viewed. Effective performance of the organization can be measured by the organization's ability to garner scarce resources and effectively turn them into valued outputs. The three main perspectives on organizational performance are: a) Successful goal accomplishment is the appropriate measure of performance if the organization is viewed as rational, goal-seeking entities, successful; b) degree of satisfaction of constituents such as employees and customs if organization is viewed as coalitions of power constituencies; c) last perspective holds organizations to be entities involved in a bargaining relationship with their surroundings, importing various scarce resources to be returned as valued outputs (Markus & Soh, 1995)

Kaplan and Norton (2000) also developed a framework to address the criticism of or to compliment traditional performance measurement named Balance Scorecard (Salterio & Lipe, 2000). Balance Scorecard argued that financial figures are consequences of yesterday's decisions not the indicators of future performance. In order to address this limitation, quality-oriented measurement such as business processes or customer orientation should also be integral part of corporate performance (List & Machaczek, 2004). Balance Scorecard identified four critical key perspectives to be included in measuring corporate performance, namely the financial, customer, internal business process, learning and growth perspectives (Braam & Nijssen, 2004; Salterio & Lipe, 2000). Measures related to customers include customer profitability, customer sales results, and sales from repeat customers. Financial measures are traditional performance measures such as return on assets and net income. Internal business processes measures relate specifically to operational processes (Salterio & Lipe, 2000). The last one is hard to identify because of its subjectivity, Kaplan and Norton suggested learning and growth can be measured by employee motivation and empowerment, employee capabilities and Information systems capabilities (Salterio & Lipe, 2000).

When measuring the extent to which data governance affects corporate performance, Kaplan and Norton (2000) identified that four critical key perspectives can be used. Data governance is associated with increasing productivity rather than financial impact (Panian, 2010; Tallon et al., 2013). Otto (2011) states that the most common business drivers of data governance initiatives are: a) is to ensure compliance; b) enable decision-making; c) improve customer satisfaction; d) increases operational efficiency; e) support business integration.

The next section presents the conceptual model which shows the proposed relationships of independent variables to data governance which has an impact on improving corporate performance.

2.10 Conceptual Model

A conceptual framework logically describes the relationship among the concepts applicable to the problem under investigation (Cavana, 2001). As stated in the previous chapter, the purpose of this study is to gain insight into data governance and synthesize the existing literature on data governance. The main objective of this study is to investigate the factors that affect data governance in organization X and also determine the extent to which the quality of data governance influences the corporate performance of organization X. From the synthesis of literature, approaches regarding the management of data assets were identified. Organisation needs to comply with policies and regulations (Cheong & Chang, 2007; Panian, 2010; Russom, 2008), assigning of appropriate roles such as data owners stewards (Bhansali, 2013; Rosenbaum, 2010), improving or maintaining quality of the data (Haug et al., 2013; Otto et al., 2007), supporting effective data integration (Magnani & Montesi, 2010; Vasista & Al-Sudairy, 2011) and ensuring effective data modeling (Davies et al., 2006; Moody, 2005). These approaches focus on single aspects of data governance and were investigated solely which leads to an isolated solution. The fact that an organisation needs to incorporate these when trying to organise data governance, leads us into our conceptual model in Figure 2. The model shows the proposed independent variables that affect data governance which have an impact on improving corporate performance. The model has seven constructs which were identified from the existing literature; and also shows the relationships between the constructs that were also found in literature. The conceptual model acts as a guideline to organise the measurement, collection and analysis of data.

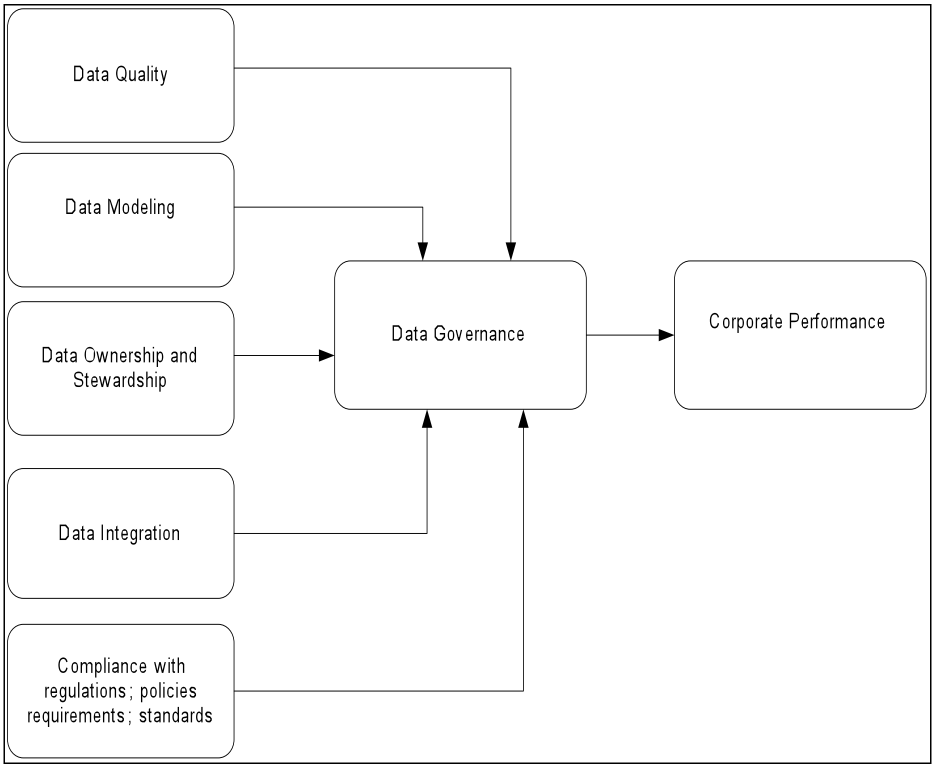

Figure 2.2 : Conceptual Framework

2.11 Definition of Constructs

Definition of data governance which state that; data governance is the collection of decision rights, processes, standards, policies and technologies required to manage and ensure the quantity, usability, availability, accessibility, quality, consistency, auditability, and security of data in an organization (Cheong & Chang, 2007; Panian, 2010) was adopted. After synthesis of literature seven variables namely compliance with data regulations and policies, data ownership, data integration, data modeling, data quality, data governance and corporate performance were identified. The next subsections define these variables.

2.11.1 Data Quality

According to Otto et al. (2007), data quality is defined with two aspects, the dependence of observed quality on the user's needs and "fit for purpose" which is the capability to meet the requirements in a specific situation. Data quality consists of six attributes to determine "fit for purpose" these are not only dimensions but a summary of over 26 dimensions identified (Haug et al., 2013). These dimensions are accuracy, relia0.bility, timeliness, relevance, completeness, currency, and consistency (Haug et al., 2013). Strong, Lee, & Wang (1997) deliver a data quality classification which includes both intrinsic and extrinsic qualities. This data quality classification consists of 15 dimensions classified under four categories. The first one is *Intrinsic* which consists of believability, accuracy, objectivity, and reputation dimensions. The second one is *contextual* and consists of value-added, relevancy, timeliness, completeness, and appropriate amount of data. The third one is *representational* which consists of interpretability, ease of understanding, representational consistency, and concise representation. Lastly is *Accessibility* which consists of accessibility, and access security. Deeper understanding of data quality dimensions helps in effectively addressing problems and issues of data quality (Tayi & Ballou, 1998).

2.11.2 Conceptual Data Modeling

A data model is a collection of conceptual tools for describing the real-world entities to be modelled in the database and the relationships among these entities. Conceptual modeling is defined as the process of gathering requirements and clearly documenting a problem domain by use of conceptual models for the purpose of understanding user requirements and to aid the primary means of communication between the stakeholders such as owners, service providers, business analysts, developers, and users (Moody, 2005). The quality of a data model is crucial in information systems development as it is a major determinant of system development costs, system flexibility, integration with other systems and the ability of the system to meet user requirements (Moody, 2005). A higher quality conceptual model will lead to a higher quality of information system as it impacts both efficiency (time, cost, effort) and effectiveness(quality of results) of IS development (Shanks, 2007).

2.11.3 Data ownership and Stewardship

Who is responsible for data in the organisations? The common answer will be IT. But in reality IT is only responsible for infrastructure and systems. This is not to say that IT has no role in ensuring data quality, but the business must also understand and be held accountable for the quality of their own data. Business should be responsible for data since they know the data best. Data ownership holds the overall responsibility for data that is, definition and management of business metadata and also determines the transformation rules. It assigns ownership of key data elements to data stewards. Data stewardship is an approach addressing data management methods covering acquisition, storage, aggregation, processes and procedures for data release and use (Rosenbaum, 2010). It holds business accountable and responsible to ensure effective control and use of data assets (Tran, Kim, & Hsiao, 2010).

2.11.4 Data Integration

Data integration is a process of providing a user with a unified view of data that resides across multiple and heterogeneous data sources (Magnani & Montesi, 2010; Vasista & Al-Sudairy, 2011; Ziegler & Dittrich, 2004). Data integration frees the user from the knowledge about how data are structured at the sources and how they are to be reconciled in order to answer queries. Data integration is essential in large enterprises that own a multitude of data sources for producing data sets that can develop and improve cooperation among the units of the organisation (Halevy & Ordille, 2006; Vasista & Al-Sudairy, 2011). Ziegler & Dittrich (2004) give two reasons for data integration. Firstly an integrated view is created to facilitate information access and reuse through a single information access point. Secondly, data from different data sources give a comprehensive basis to satisfy a certain information need or query. Data integration can be achieved using one of three approaches namely Application Integration (mediation), database federation and data warehousing (Vasista & Al-Sudairy, 2011). Data integration is one of basic activities used to improve the quality of data as it can both reduce data structural and semantic heterogeneity and redundancy, and increase its availability and degree of completeness (Magnani & Montesi, 2010).

2.11.5 Compliance with regulations; policies requirements and standards

There are risks associated with use of business data, therefore the area of data usage is heavily regulated and must comply with industry standards. Regulatory compliance is a process to ensure that laws and regulations that govern how business is conducted are followed (Bhansali, 2013). Penalties associated with non-compliance are increasing as the need for compliance is growing, therefore organisations must develop policies and procedures that align data as an asset with both the organisation's strategies and also with regulations. Government have stricter regulations over public and private sectors such as Sarbanes-Oxley (SOX) Act, ECT Act 2000 etc. which drive the need to manage risks, ensure privacy and compliance. Privacy is the freedom from unauthorized intrusion; ensuring only authorized personnel have access to certain information. Privacy can be enforced through data security policies. Data security policies are defined as the processes and procedures that the employees should follow in order to protect the confidentiality, authenticity and non-repudiation, integrity and availability of data as it is the valuable asset of the organisation (Vroom et al., 2004; Williams, 2001). These policies should clearly state how such assets are tagged, tracked and monitored (Bhansali, 2013).

2.11.6 Data Governance

Data governance addresses the increasing importance of data in this era where data creation grows exponentially and organisations are starting to realise that data is an enterprise asset, thus fulfilling a gap which is not well addresses by IT governance. Data governance is defined as the "establishment of policies, through formal structures that define rules, procedures and decision- making rights regarding information management, in order to mitigate regulatory and operational risk, reduce costs and optimize the performance of the organization" (de Abreu Faria et al., 2013, p4444). There are three goals of data governance firstly is to ensure data fulfil business requirements. Secondly, to lower the costs of managing data and lastly to protects and treat data as most valuable business asset (Tallon et al., 2013).

2.11.7 Corporate performance

The effective performance of an organization can be measured by the organization's ability to garner scarce resources and effectively turn them into valued outputs. Literature on organizational effectiveness showed that organizational performance should not be defined by financial indicators but rather on how the organization is viewed (Markus & Soh, 1995; Kaplan and Norton, 2000). Kaplan and Norton (2000) state that the performance measurement called balance scorecard identified four critical key perspectives to be included in measuring corporate performance, namely the financial, customer, internal business process, learning and growth perspectives (Braam & Nijssen, 2004; Salterio & Lipe, 2000). The business drivers of data governance initiatives are: a) is to ensure compliance; b) enable decision-making; c) improve customer satisfaction; d) increases operational efficiency; e) support business integration (Otto, 2011a). Therefore balance scorecard deemed to be a good measurement of the impact of data governance on corporate performance.

2.12 Research Propositions

How can an organisation ensure rigorous data governance in order to attain better corporate performance?

Objective:

The objective of this study is to identify and investigate factors that affect data governance in organization X and also determine the influence that quality of data governance has on the corporate performance of the organization.

Propositions:

Proposition 1: Inadequate compliance with data requirements in organisation X will negatively affect quality of data governance.

Proposition 2: Inefficiency of data ownership and stewardship negatively affects good data governance.

Proposition 3: Effectiveness of data integration within the organisation contributes positively to data governance.

Proposition 4: Inadequacy of data modeling has a negative influence on data governance.

Proposition 5: Effectiveness of data quality contributes positively to data governance.

Proposition 6: When the quality of data governance is poor it will impact the corporate performance negatively.

2.13 Chapter Summary

The literature review identified the factors which affect data governance namely compliance with data regulations and policies, data ownership, data integration, data modeling and data quality. Relevant theories which were appropriate to address the problem under investigation were analysed and related to the study. The literature argued that traditional performance measurement will not be appropriate for this study as financial figures are consequences of yesterday's decisions not the indicators of future performance (List & Machaczek, 2004). The Balance Scorecard identified four critical key perspectives to be included in measuring corporate performance namely the financial, customer, internal business process, learning and growth perspectives (Braam & Nijssen, 2004; Salterio & Lipe, 2000). The identified link between theory, data governance and the constructs is shown in Table 2.3.

Identifiable relationships between the constructs were captured through a proposed conceptual model in section 2.10. The conceptual model showed the proposed independent variables that affect data governance which has an impact on improving corporate performance. It incorporated all the factors which affect data governance in order to try to address the gap identified. The identified gap is that literature focused on single aspects of data governance and were investigated solely which to and isolated solution. There is also limited research on the effects of data governance on firm performance. Therefore the objective of this study is to identify and investigate factors that affect data governance in organization X and also determine the influence that quality of data governance has on the corporate performance of the organization.

The next chapter presents the research design. It describes the details of how the study was conducted. It discusses the research philosophy and paradigm, the overall strategy for implementation, and the data collection and analysis techniques adopted for this study.

Table 2.2 : Theories, factors and data management practises

Theory	Theoretical Element	DG Factor Identified	Data management practice
Agency	Control	Accountability Security	• Data ownership and Stewardship • Compliance with regulations; policies requirements; standards
Agency	Monitoring	Monitoring	• Data ownership and Stewardship
Agency	Risk	Compliance Retention	• Compliance with regulations; policies requirements; standards
Agency	Rules	Accessibility Ethics Privacy	• Compliance with regulations; policies requirements; standards • *Data Integration*
Agency	Alignment	Communication Sharing Transparency	• Data ownership and Stewardship • *Data Integration*
Agency	Structure	Formal Structure	• Data ownership and Stewardship
RBV	Heterogeneous Resources	Consumerisation	• Data Integration • Data Modeling
RBV	Data as an asset	Standard	• Data Integration • Data Modeling
RBV	Performance	Mobility	• *Corporate performance*
RBV	Quality	Quality	• *Data Quality*
RBV	Systems	Systems (IS)	• Data Integration • Data Modeling
RBV	Value	Value	• *Corporate performance*
DC	Rapid change, Skills, Learning, Knowledge, Capabilities	Culture	• Data ownership and Stewardship • *Corporate performance*
DC	Context	context	• Data Integration

3 Research Design

This chapter presents the design employed by this research. Research design is the general plan of how the research will go about answering the research questions. It specifies the sources which the data will be collected from and how to collect and analyse this data. Furthermore it discusses ethical issues and some constraints the researcher could encounter. These demonstrate that the researcher has thought through the elements of the particular research design(Saunders, Lewis, & Thornhill, 2011).

The outline of this chapter is as follows sections 3.1 – 3.2 discusses philosophical assumptions which the researcher abides by when conducting the research. Section 3.3 *Research Approach* provides an overview of an approach and reasoning which the researcher adheres to when performing the study. Section 3.4 *Type of Research* describes the nature of the study and also the time frame. This section sets the background for the choice of research strategy and methods. Section 3.5 - 3.10 *Methodology for this Research* describes the research strategy, sampling strategy, data collection and preparation; and 3.11 *Research Ethics* presents the researcher's own ethical duties and how confidentiality issues were solved. Lastly Section 3.12 the summary chapter ties everything together.

3.1 Ontology

Ontology is concerned with the nature of reality that is the study of nature and its exposure to existence. There are two views that clarify ontology namely objectivism and subjectivism. The objectivism ontological stance believes that social entities exist in reality external to social actors concerned with their existence (Saunders et al., 2011). The subjectivism ontological stance states that the perceptions and consequent actions of concerned social actors create the social phenomena therefore reality is social constructed (Cavana, Delahaye, & Sekeran, 2001;Saunders, Lewis, & Thornhill, 2012). The researcher assumes a reality in which one should detach oneself from it in order to obtain the actual truth of the given state of affairs and its operations. Therefore the researcher took the objectivism ontological stance.

3.2 Epistemology

Research is a creation of truth presented as reality (Tien, 2009). Epistemology is related or filters from the ontology the researcher believes in. There are two extremes of epistemology namely positivism (the objectivist approach to social science) and interpretivism (the subjectivist approach to social science).

Interpretivists consider that there is more than one reality which can be accessed or discovered by multiple ways. The knowledge gathered through this perspective is perceived to be subjective interpretations and socially constructed. The aim is to interpret human behaviour rather than to predict the causes and effects. It presents a rich and complex description on understanding motives, meanings, reasons of how people think, react and feel under certain contextually specific situations(i.e. time and context) (Cavana et al., 2001; Hudson & Ozanne, 1988). Rodriguez-ulloa & Paucar-caceres (2005, p.6) states that “life world is an ever changing flux of events and ideas and ‘managing’ means reacting to that flux”.

The Positivism stance was adopted for this research based on ontology which states that objective reality exists. Positivists believe that a single, tangible truth is located 'out there' in the real world and waits merely to be discovered precisely because the facts exist independently of any theories or human observation. Based on the literature review which has been conducted, there are many theories which exist around data governance and firm performance and a conceptual framework was developed. From the conceptual framework developed, there is an assumption of linear causality i.e. these five constructs (compliance, data ownership, data integration, data modeling and data quality) have an effect on data governance which have an effect on corporate performance. This research has pre-specified hypotheses which the research will verify if they are true or false. This involves gathering of facts to confirm or disprove hypothesized relationships among the constructs that have been deduced from propositions (Ghauri, 2005). The acceptance or rejection of the hypotheses helps in explaining or prediction of theory.

3.3 Research Approach

Research approaches are plans and procedures that break down broad assumptions of research into detailed methods of data collection, analysis and interpretation (Creswell, 2013a). There are two states of reasoning which drives the approaches namely deductive and inductive reasoning. Deductive reasoning aims to identify a set of universal laws that can be used to predict general systems of human activity by establishing theoretical positions and moving towards empirical evidence (Cavana et al., 2001). Therefore deductive reasoning is for falsification or verification of theory. Conversely inductive reasoning starts research by collecting data to explore phenomenon and the researcher builds or generates theory (Saunders et al., 2012).

This study follows a deductive reasoning. This is the top down approach where the research wants to confirm some theory on the topic of interest and will be moving from general to particular (Ghauri, 2005). The literature on data governance was synthesised and a conceptual model was developed which serves as the theory to be validated and hypotheses were developed. Hypotheses should be crafted in a way that if the theory is true then certain things should follow in the real world. These will be observed through data collection and analysis. If the hypotheses are confirmed to be correct it implies that our theory is supported. If not the theory needs to be modified and tested again or it will be rejected.

There are three approaches to research namely qualitative, quantitative and mixed methods (Creswell, 2013a). The first two approaches are not discrete but they represent different ends on a continuum. A study tends to be more quantitative than qualitative or vice versa. The less approach results are used to support the findings of the main approach (Creswell, 2013a). Mixed methods combines both qualitative, quantitative methods to conduct the research.

The research has adopted a quantitative approach. This research approach is driven by the ontological and epistemology stance the research has taken. This research adopted a positivist epistemological stance which claims that science involves confirmation and falsification, and that these methods and procedures are to be carried out objectively (Creswell, 2013b; Johnson & Onwuegbuzie, 2009). The quantitative approach maintains that the social science inquiry should be objective, which ties in with the ontological stance. The researcher should be uninvolved with the objects of study, and test or empirically justify their stated hypotheses. It collects data on predetermined instruments that yield statistical data. McMillan & Schumacher (2010) says a quantitative approach is used when we want to find out about the state of something or other or want to explain a certain phenomenon. As we want to find the effect of the data governance on firm performance, this deemed to be the most appropriate approach. They also said quantitative research is especially suited to the testing of hypotheses. Since we are taking the deductive approach as well, hypotheses have been crafted and this research aims to test them.

3.4 Type of Research

The nature of the study can either be exploratory, descriptive or explanatory. An Exploratory study is done when little is known about the situation at hand with the aim to better understand and gain insights about the phenomenon (Cavana et al., 2001; Saunders et al., 2011). Exploratory studies are flexible and adaptable to change as the direction of the research could change as new data and new insights are being revealed. Exploratory studies are widely used in interpretive and critical researches (Ghauri, 2005). They advance knowledge through good theory building (Cavana et al., 2001; Ghauri, 2005).

Descriptive study is done in order to establish and describe factors associated with a phenomenon (Cavana et al., 2001). It is normally used to build profile of factors of interest and provides ideas for further decision-making (Ghauri, 2005). It is suitable for a positivist research but sometimes also an interpretive or critical research.

Explanatory research establishes the causal relationship between variables. It studies the situation or problem in order to explain the nature of a relationship between variables or establish the differences among groups or independence of two or more factors is in a phenomenon (Cavana et al., 2001; Saunders et al., 2011). This is mostly suitable for a positivist research.

This study is an explanatory research, from the conceptual model; the focus is to find the relationship between independent variables and dependent variables. Some causal explanations will be simple while others will be complex and it could be useful to collect qualitative data to explain the reasons why certain conditions hold in this situation (Saunders et al., 2011). Explanatory research attempts to clarify why and how there is a relationship between two or more aspects of a situation or phenomenon. These are the main questions for a case study and guides the research strategy (Rowley, 2000). Answering the why questions involves developing causal explanations. The problem under scrutiny is structured as we are guided by the theory which is a conceptual model in this case.

There are two research time horizons namely cross-sectional and longitudinal time horizons (Saunders et al., 2012). Cross-sectional is the time horizon when a research is taken at a particular time maybe over weeks or months. A Cross-sectional time horizon is normally called a snapshot time horizon. Longitudinal time horizon is when data collection is over a series of times. This could be a number of years where the main focus is to study change and development over time (Saunders et al., 2012).

This study is cross-sectional as this research is for academic purposes and is time constrained (Saunders et al., 2011). Saunders et al. (2011) says cross-sectional studies often employ the survey strategy to explain how factors are related. This holds true for this research as the researcher aims to investigate how these five constructs (compliance, data ownership, data integration, data modeling and data quality) have an effect on data governance which has an effect on corporate performance. This study also uses quantitative methods by gathering numerical data through questionnaires.

3.5 Research Strategy

Yin (1994) provides four types of case study designs namely Single case design, holistic; Single case design, embedded; multiple case design, holistic; multiple case design, embedded. The definition of a Case Study according to Yin (2009) is an empirical inquiry that

- Investigates a contemporary phenomenon in depth and within its real-life context, especially when
- The boundaries between phenomenon and context are not clearly evident.

The case study researcher typically observes the characteristics of an individual unit. The purpose of such observation is to probe deeply and to analyze intensively the multifarious phenomena unit that constitute the life cycle of the unit (Biggam, 2008). This study used a single case design, holistic (single unit of analysis). Selecting units of analysis must be determined by the research purpose, questions, propositions and theoretical context, accessibility and time available (Rowley, 2000). The unit of analysis was the organisation since the purpose of the study is to investigate the impact of data governance on corporate performance. Although the decisions are made by individuals, these individuals are presumed to represent the firm’s decision rather than their personal decisions (Bhattacherjee, 2012). The organisation was treated as a holistic case study and with no interest in logical sub-units within the organisation. A single case is appropriate since the main aim of this study is to test an established theory and also provide a source of new hypotheses. Our literature review shows that there are many established theories around data governance. The aim is to extend the theory by confirming the links between the constructs on the developed conceptual model.

There are critics and limitations in this approach that require to be addressed. Flyvbjerg (2006) in her paper examines five common misunderstandings about case-study research which are categorised as limitations of this approach. One of the criticisms is that one cannot generalize from a single case; therefore, the single-case study cannot contribute to scientific development. Flyvbjerg (2006) argues that it depends on the case in question and how it is chosen. She gave carefully chosen experiments, cases, and experiences which were critical to the development of the physics of Newton, Einstein, and Bohr which were generalized but based on single case studies. In social science too, the strategic choice of a case may greatly add to the generalizability of a case study. If the study could be proved false in the favourable case, then it would most likely be false for intermediate cases. Based on the argument, it was concluded that the criticism that one cannot generalize on the basis of a single case and that the case study cannot contribute to scientific development does not hold. The conclusion is: “One can often generalize on the basis of a single case, and the case study may be central to scientific development via generalization as a supplement or alternative to other methods. But formal generalization is overvalued as a source of scientific development, whereas “the force of example” is underestimated (Flyvbjerg, 2006, p.228).”

The researcher aimed to shed light on what is happening in a particular setting, thereby adding knowledge to a rich picture of data governance. The researcher intended to extend the concept of relatability where other organizations in relating to the situational aspects of the case study may recognize similar issues and problems in their organizations and can learn from the research findings (Biggam, 2008).

3.6 Sampling Strategy

Data gathering contributes to a better understanding of a theoretical framework and it is deemed to a crucial stage of the research (Tongco, 2007). The study used non-probability purposive sampling or judgment sampling; this sampling type chooses its sample based on the qualities possessed by the participants. The Researcher decided what needs to be known guided by the conceptual model and found people who were willing to participate based on their knowledge or experience (Kumar & Phrommathed, 2005). Purposive sampling was appropriate as the research is interested in people who have a rich knowledge of data governance. Data governance incorporates diverse disciplines of data namely data modeling, data quality, data ownership, compliance and data integration. Using purposive sampling ensures that we get participants who have knowledge in all these areas to avoid a situation where participants will answer questions which they have little knowledge or not very familiar about which increases unreliability. Participants were selected based on the knowledge, experience and involvement in data governance in order to receive rich information. These participants included: Application & BI Developers, Data & Application Architects, Solution Specialists, Business Unit Managers, Business Analysts and HR Consultants. Unlike random sampling it is not free from bias. Nevertheless it can provide reliable and robust data and its strength lies in its intention bias (Tongco, 2007). In most papers non-probability sampling methods are associated with qualitative research (Sandelowski, 2000; Tongco, 2007). Kumar and Phrommathed (2005) argue that what determines if non-probability sampling will be used as quantitative or qualitative is the predetermined sample size. In quantitative research it is used to select the predetermined sample. In this research the researcher aimed for 50 or more participants. The population size was 200 that is all staff members who have depth knowledge of data governance. Therefore the response rate was 50/200 which is the quarter of the population.

3.7 Data Collection and Preparation

The data collection technique used affects the results one obtains from the research, therefore one has to ensure rigour and appropriate methods are used (Saunders et al., 2011). There are three methods for conducting research: quantitative, qualitative and mixed methods.
Mixed method uses both quantitative and qualitative methods thereby helping the researcher to understand the research more completely. These methods complement each other, by their strength to overcome limitations of the other (Johnson & Onwuegbuzie, 2009).

As this study adopts a quantitative research paradigm, therefore a quantitative method was appropriate. Data was collected using quantitative method getting quantitative data. Questionnaire was used to gather data and it was hosted within the organisation's intranet. Likert-scaled items, ranging from 1 (Neither agree nor disagree) to 5 (strongly agree) were used in the scale of an observed variable. A few questions required (Yes/No) answers. The first part of the questionnaire collected demographic information about participants (e.g. occupation, department team, level of education, years of experience in IS, etc.). The second part of the questionnaire measured the model constructs, see Appendix B.

3.8 Research Instrument

A Survey helps in gaining the insight of the population by studying a sample of it, and provides a quantitative or numeric description of trends, attitudes or opinions (Creswell, 2012). A survey instrument used in this study was questionnaire see Appendix B. The questions were based on the conceptual model in Figure 2. The first section consisted of demographic questions. The following sections were model construct questions according to the conceptual model.

This instrument was developed for this research with the help of white papers which were investigating adoption of data governance in the organizations. Questionnaires are used mostly for explanatory research as they are not suitable for researches which have large numbers of open-ended questions (Saunders et al., 2012). Explanatory or analytical research examines and explains cause-and-effect relationships between variables(Creswell, 2012; Saunders et al., 2012). A questionnaire was well-suited for the purpose of this study and it required less skill and sensitivity. The questionnaire was hosted on the intranet environment of organization X. It was a self-completed questionnaire and it was sent to the respondents through email as a uniform resource locator (url) link in the body of the email together with the cover letter. This offered great control to be sure that the questionnaire was completed by the intended person because most people manage their email boxes and respond to their own email (Saunders et al., 2012). This improves reliability of the data obtained. A questionnaire is easier to administer, in terms of cost and time, and can be replicated over different groups, times and places for comparison (Creswell, 2012; Saunders et al., 2012). This research was done for purpose of degree with short time frame; the time horizon used was cross-sectional. Data was collected in period of two months. The availability of the respondents was an issue during data collection. The researcher had to constantly remind and pleaded with the respondents to complete the questionnaire. The questionnaire was available 24/7 so that someone can complete it whenever have free time.

The questionnaire consisted of category, rating. Category questions were used on how often questions where a respondent can choose only one category. Rating questions are often used to get opinion data (Saunders et al., 2012). Rating questions are mostly used with a likert-scale rating. Respondents are asked to state how strongly they agree or disagree with the statement. The questionnaire used a five-point rating scale where 1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree and 5=strongly agree. It relies on self-reporting which may result in response bias (Creswell, 2012; Saunders et al., 2012).

3.9 Pilot Study

A pilot study was used to minimise bias and also to test the validity of the research instrument i.e. questionnaire. The researcher pre-tested the questionnaire with 12 people from the purposive sample who have deep knowledge (expert) on data governance and people who work with data. This helped in reviewing the questions for relevance and wording. The main objective of this was to ensure clarity and relevance and also to understand the constructs of the variables to be included in the questionnaire. This aided in refining the questionnaire for ease of response once sent out to the larger population and also made it easier to record the data later (Saunders et al., 2011). Saunders et al. (2011) also suggest that a few questions be included in the pilot questionnaire to gain a better perspective regarding the issues encountered. The sample questions included in the pilot were:

- how long the questionnaire took to complete;
- the clarity of instructions;
- which, if any, questions were unclear or ambiguous;
- which, if any, questions the respondent felt uneasy about answering;
- whether in their opinion there were any major topic omissions;
- whether the layout was clear and attractive;
- Any other comments.

3.10 Data Analysis

Descriptive analysis measures the percentages, measures of central tendency (mean, mode, median) and measures of variability (range, Standard deviation, and variance). Correlation testing will be performed to get a clear view of the relationship of the constructs. Correlation assesses the strength of the relationship between pairs of variables i.e. independent variables. Correlation coefficient ranges from -1 to 1, where a range from -0.7 to 0.7 shows a weak relationship between the variables assuming calculation on probability less than 0.05 which is regarded as statistically significant (Saunders et al., 2011). Linear regression assesses the strength of cause-and-effect relationships between independent variables (compliance, data ownership, data integration, data management practises and information processing capabilities) and dependent variable (data governance) (Ghauri, 2005; Saunders et al., 2011).

Cronbach's Alpha was used for reliability which is the degree to which the measure of a construct is consistent or dependable (Bhattacherjee, 2012). Construct validity was conducted to measure the extent to which a measure effectively represents the underlying construct that it is supposed to measure (Bhattacherjee, 2012). This will be done by conducting confirmatory factor analysis using structural equation modeling on the items of compliance, data ownership, data integration, data management practises, firm performance and information processing capabilities. To check that the model has not suffered from multi-collinearity, the results of Spearman rank, correlation coefficients should not be equal or greater than 0.8 (Field, 2009).

3.11 Access and Research Ethics

The research will be conducted on organization X and this research will deal with sensitive data of the business since the organization relies on data to make informed decisions and strategies. Permission was asked from the CIO and business-related people who are responsible for data governance initiatives. The introductory letter asking for access describes in brief the purpose of the research, what is likely to be involved in participating and also how the organisation can benefit from the study. There are three main organisational concerns, firstly the amount of time or resources that will be involved, secondly sensitivity about the topic and thirdly the confidentiality of the data that would have to be provided and the anonymity of the organisation or individual participants.

The introductory letter provides clear assurances about these aspects in writing giving guarantee that they will be looked into. It also states that the researcher will produce a summary report of the research findings as practitioners also struggle with the same issues as researchers. This will allow the organisation to reflect on the actions they have taken and issues they are experiencing.

3.12 Chapter Summary

This chapter provided an overview of philosophies adopted by the researcher when conducting this study and providing the reasons on why these were chosen over others. The epistemological, ontological and methodological philosophies chosen gave background and shed a light on which methods, research approaches, data collection and analysis techniques will fit in conducting this research. The objective of this study is to investigate the factors that affect data governance in organization X and also determine the extent to which the quality of data governance influences the corporate performance of organization X. Therefore a positivist stance, explanatory and quantitative methods were adopted. Data was collected from 50 respondents who have knowledge, experience and involvement in data governance areas. Table 3.1 outlines the summary of research design for this research. The next chapter presents the results of the tests which were mentioned in section 3.10 to be performed in order to answer our research questions.

Table 3.1: Research methodology Summary

METHODOLOGY	APPROACH
Underlying philosophy	Positivist
Research purpose	Explanatory
Reasoning approach	Deductive approach
Research strategy	Quantitative questionnaire
Data collection techniques	Quantitative

	• Closed Questions
Data Analysis	Quantitative – STATISTICA 10 and Microsoft Excel 2010
Time-frame	Cross –sectional

4 Results

The objective of this research as stated in Chapter 1 is to explore practical and effective data related initiatives which contribute to good data governance in order to create better corporate performance. The researcher formulated research questions; prepositions and a design to collect data from employees who have job descriptions aligned with data initiatives to get an insight of what initiatives affect data governance in order to better corporate performance.

The instrument used gathered both qualitative and quantitative data. The previous chapter clearly stipulates data collection techniques which were used. Qualitative data was gathered in an open ended question which will help in justifying and substantiating the findings of the quantitative data. In quantitative analysis, statistical techniques were used to summarise and describe the data, test for reliability and validity of the instrument and establish the relationships amongst the variables of the conceptual model (Saunders et al., 2012).

This chapter presents the results of data analysis. The outline of the chapter is as follows: section 4.1 presents the response rate analysis. Section 4.2 describes the demographics of the participants. Section 4.3 presents the summary of responses on items of the constructs i.e. compliance, data ownership, data integration, data modeling, data quality, data governance and corporate performance. Section 4.4 presents the results of reliability and validity tests done. Section 4.5 shows *Hypothesis testing* which interprets and discusses the results in relation to research hypotheses and objective.

4.1 Response Rate

Field (2009) illustrated with a graph the sample size needed to achieve a certain level of power as the numbers of predictors vary see figure 3. One can deduce that if one is expecting to find a large effect then a sample size of 80 will always suffice. However, a sample size of 50 is a statistically acceptable number in order to get proper analysis (Saunders et al., 2011). The researcher wanted a larger effect and the study has six predictors. From the graph one can see sample size of 50 is sufficient. And as discussed in section 3.7, the researcher aimed to collect 50 or more responses.

Saunders et al. (2011) state that the likely response rate is 30-50% for intranet responses or within the organization mediated questionnaire. The most time taken to complete the questionnaires is 2-6 weeks from distribution depending on the number of follow ups. The response rate for this study falls within these limits of Saunders et al. (2011); the questionnaire was sent to 170 people and 50 were returned, which is 29.4 % response within four weeks with three follow ups.

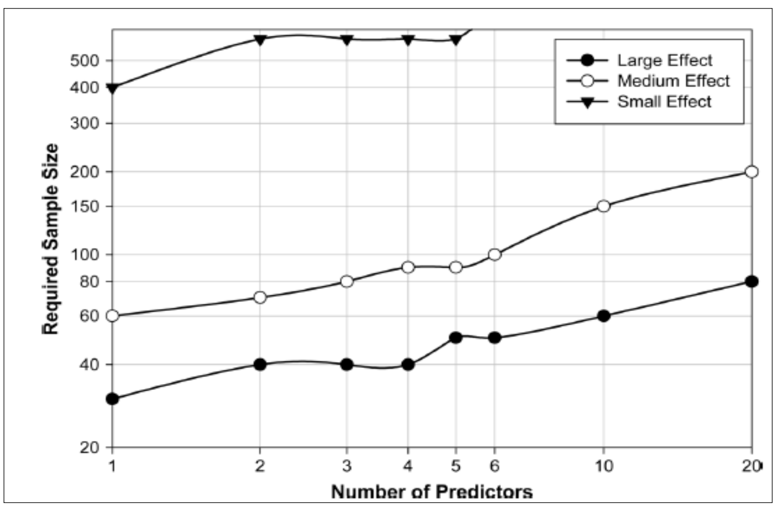

Figure 4.1 : sample size required in regression depending on the number of predictors and the size of expected effect(Field, 2009)

4.2 Demographic Analysis

This section uses descriptive statistics, which enables one to describe and compare variables numerically. This is done through two aspects namely central tendency and dispersion. Central tendency is conducted to give a clear view on which values occur more frequently (mode), values in the middle (median) and average (mean). Dispersion can be described by standard deviation and inter-quartile range. Standard deviation is the extent to which values differ from the mean.

4.2.1 Participants

The participants were coded according to categories If any title had a specialist word in it, it was coded as 2 e.g. Solution specialist, Principal Design Solution Specialist.

Figure 4 below shows that the top three respondents were Specialists with 11 people which is 22% of overall respondents followed by manager with 9 people which is 18% followed by Analyst with 8 people which is 16 %. From descriptive stats it also appears that (Specialist) 2 is the value that occurred most frequently 11 times. From the histogram diagram one can see that data is not normal so the median was used rather than the mean. Median being the mid-point or middle value after the data has been ranked(Saunders et al., 2011). Median is 4 (manager).

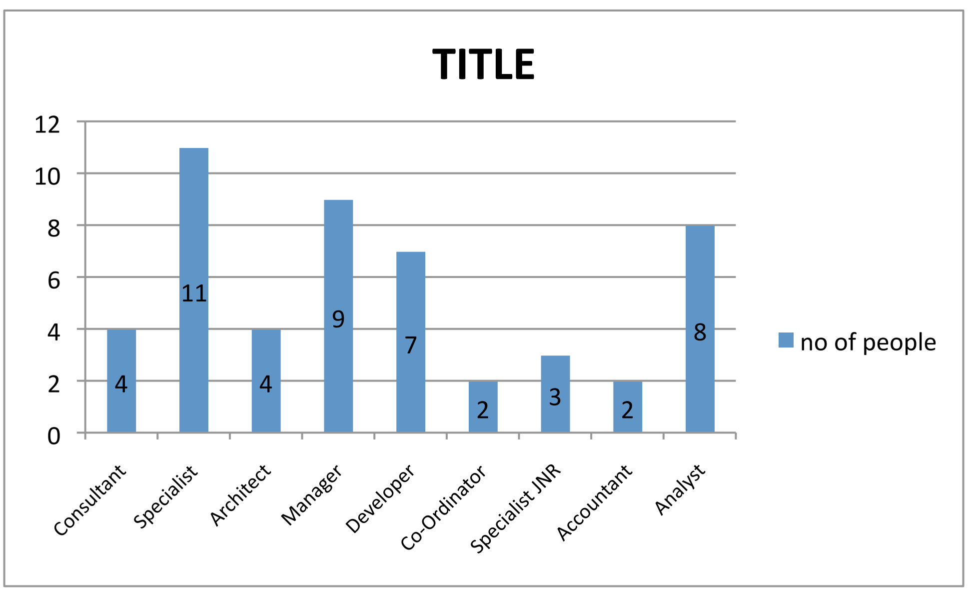

Figure 4.2: Number of people in each title

4.2.2 Level of Education

The level of education which dominated within the sample was diploma or degree category, with 31 people which is 62% of the sample. The Honours and Master's degree categories were the second dominant levels with seven people each or 14 % each.

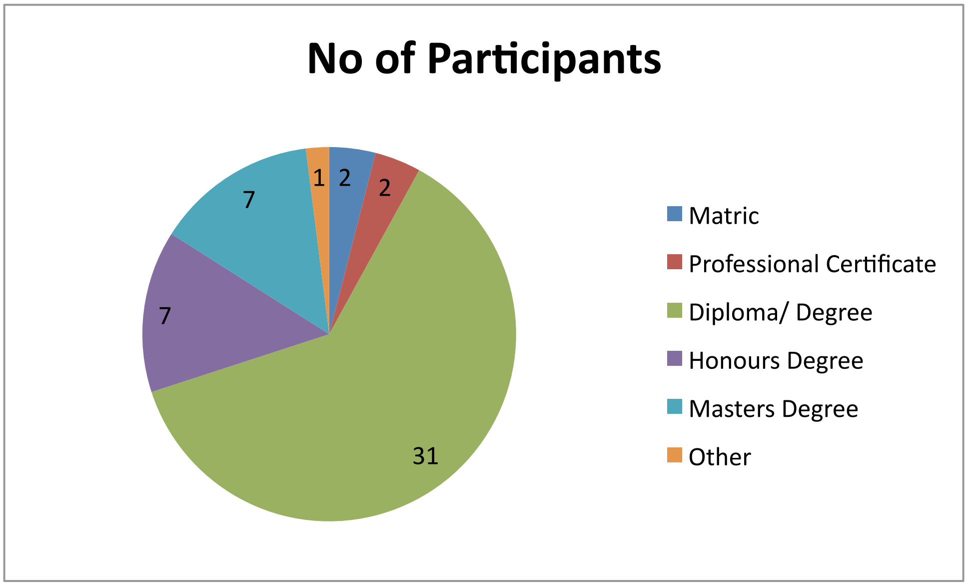

Figure 4.3: Number of people in each level of education.

4.2.3 Years of experience

Figure 6 shows that most of people fall between 6-10 years of work experience, and also it can be seen that a large range of people fall between six or more years of experienced. This could mean that these participants have in-depth knowledge about the subject of data governance.

Figure 4.4: Number of people per years of work experience category

4.3 Constructs

This section presents the results of the items of each construct to see the summary of the respondents' responses. To begin with Table 4.2 summarizes the results of compliance with data policies and regulations. The items which are included in these results are the ones which were not omitted in factor analysis see Table 4.10.

Table 4.1: Data compliance with data policies and regulations item responses

Variable	Description	Valid N	Mean	Std.Dev.
CRL1	**Effectiveness**—information is relevant and pertinent to the processes as well as being delivered in a timely, correct, consistent and usable manner	50	3.66	1.06
CRL3	**Confidentiality**—Protection of sensitive information from unauthorised disclosure	50	3.78	1.13
CRL4	**Integrity**— Accuracy and completeness of information as well as to its validity in accordance with business values and expectations	50	3.58	1.14
CRL5	**Availability**—Information being available when required by the process now and in the future. It also concerns the safeguarding of necessary resources and associated capabilities.	50	3.82	1.06

CRL6	**Compliance**—complying with the laws, regulations and contractual arrangements, to which the process is subject, i.e., externally imposed business criteria as well as internal policies	50	3.68	1.32
CRL7	**Reliability**—Appropriate information for management to operate the entity and exercise its fiduciary and governance responsibilities	50	3.46	1.28

CobiT 4 defines seven control criteria for information to satisfy business objectives. The respondents were asked to indicate to what extent the organisation complies with these controls. The likert-scale with these rating (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree) was used. Table 4.2 shows that the respondents agreed that organisation X complies with five of Cobit 4 control criteria for information(Effectiveness, Confidentiality, Integrity, Availability, Compliance) as they have values above 3.5 so they are close to 4 which is agree on the likert-scale. They disagreed that data is reliable.

Table 4.2: Data ownership and stewardship item responses

Variable	Description	Valid N	Mean	Std.Dev.
DOS2	Data owners contribute to definition and management of business metadata	50	3.32	1.42
DOS3	Data owners determine the transformation rules	50	3.18	1.45
DOS4	Data stewards support the user community regarding data quality	50	3.14	1.43
DOS5-6	Data stewards perform exposure or risk identification and verify the data after load	50	2.74	1.45

Respondents disagree that data owners perform their duties regarding data. Respondents disagree that data owners contribute to definition and management of business metadata. They also disagreed that data owners determine the transformation rules as they are people who know well what intended use of the data is, that is to respond to strategic and operational challenges. Respondents disagreed that data stewards support and help the users regarding data quality. Tran, Kim, & Hsiao (2010) argue that data stewards are nominated as accountable for business responsibilities in order to ensure quality of enterprise data assets. Participants disagree that data stewards perform risk identification and verification of data after load to make sure that everything is according to expected information. Berson & Dubov, 2007b; Bhansali (2013) define the duties of data stewards as to ensure that adequate, agreed upon quality metrics are maintained on a continuous basis and also to ensure that data provided meets all specified requirements. Data validation is critical as it helps in correct data collection, transmission and data derivation processes. It also helps in identifying data outliers and data errors. All these play a vital role in improving data quality (Bhansali, 2013).

Table 4.3: Data Integration item responses

Variable	Description	Valid N	Mean	Std.Dev.
DI2	There is continuous evaluation of existing data integration technology infrastructure and its ability to support data governance practices	50	2.12	1.14
DI3	Data integration lifecycle is followed (Develop and Manage, Access, Discover, Cleanse, Integrate, Deliver, Audit, Monitor and Report)?	50	3.10	1.37

Few items were used to measure data integration construct. DI1 & DI4 were omitted because DI1 was a categorical question (yes or no) and DI4 was ordinal question. These will be used to support the discussion section. Respondents strongly disagree that the organisation continuously evaluates the existing data integration technology infrastructure and its ability to support data governance practices. Hristov (2012) argues that data integration approaches vary very much in their characteristics and it is not easy to replace one approach with another. It is feasible to ask which approach should be used for the specific task not to change whole infrastructure. They also disagree that the data integration lifecycle is followed.

Table 4.4: Data Quality item responses

Variable	Description	Valid N	Mean	Std.Dev.
DQ1	Data is Accurate :data item is close to its true value in terms of meaning and truthfulness	50	3.68	0.87
DQ2	Data is Consistent : data unit is specified the same throughout the organization	50	3.52	0.91
DQ3	Data is Complete: completeness of columns of a table	50	3.26	1.05

	containing data			
DQ4	Timeliness of Data : promptness freshness and frequency of updates of data	50	3.36	1.14
DQ5	The organization has Data Quality tools and plans in place.	50	3.44	1.36
DQ7	How often is data cleaning and monitoring done?	50	2.40	1.47

Four data quality dimensions were used to measure this construct which are accuracy, completeness, consistency and timeliness (Azumah & Quarshie, 2012; Cheong & Chang, 2007). Respondents agreed that the organisation complies with two dimensions, which are data is accurate and consistent. But they disagreed that data is complete and timeliness. They also disagreed that the organization has data quality tools and plans in place. DQ7 had a different Likert-scale which is 1=never; 2=once a year; 3=twice a year; 4=quarterly; 5=monthly. Respondents stated that data cleaning and monitoring done is done once a year.

Table 4.5: Data Modeling item responses

Variable	Description	Valid N	Mean	Std.Dev.
DM2	There is a Data model quality management framework which helps in validating the developed data models.	50	3.02	1.45
DM3	Evaluating the quality of a conceptual data model is critical to the successful development of an information system	50	3.42	1.18
DM4	Data analyst(s) are responsible for developing data models	50	3.12	1.21

In the case of Data modeling (DM), respondents disagreed that data analysts are responsible for developing data models. Data analysts are not the only stakeholders who are responsible for developing data models. Key stakeholders in data modeling processes are Business user(s), Data analyst(s), Data administrator(s) and Application developer(s) as data models acts as communication amongst them(Davies et al., 2006; Moody & Shanks, 2003). Respondents disagreed that evaluating the quality of a conceptual data model is critical to the successful development of an information system. Contrary to this Moody & Shanks (2003); major finding was that the most significant benefits are achieved through improving the process of data modeling rather than through quality assuring the final result. The cost of error detection and correction is 170 times more after delivery rather than error prevention that is removing errors in requirements stages (Moody & Shanks, 2003). Respondents also disagreed that there is a data model quality management framework which helps in validating the developed data models.

Table 4.6: Data Governance item responses

Variable	Description	Valid N	Mean	Std.Dev.
DG1	Organisation performs Month-to-month scorecard/KPIs at business unit-level for accuracy/quality of specific data entities	50	2.62	1.46
DG2	Organisation uses of Data Quality tools i.e. IBM WebSphere QualityStage for data profiling	50	2.70	1.59
DG3	How often are data External audits performed?	50	2.04	1.01

Respondents indicated that the organisation have data External audits performed once a year. Respondents disagree that the organisation uses Data Quality tools. They also disagreed that the organisation performs Month-to-month scorecard/KPIs at business unit-level for accuracy/quality of specific data entities. This could be due to the fact that data owners are not performing their duties regarding data as seen in Table 4.3. As data stewards are the ones who are supposed to do this, they could not because they depend on data owners. This goes in line with the respondents responses on DOS4 and DOS5, as they disagreed that data stewards perform their duties.

Table 4.7: Corporate Performance item responses

Variable	Description	Valid N	Mean	Std.Dev.
P1	Customer-Related Measure : There has been an improvement of customer satisfaction rating as a result of data governance initiatives	50	2.84	1.58
P2	Financial Measure : There has been reduction of costs due to improvement in regulatory compliance/ reduction of regulatory risk	50	2.72	1.62
P3	Internal Business Processes Measure : There has been an improvement in internal business processes e.g. quick responses fewer errors experienced due to data governance initiatives	50	2.92	1.41
P4	Learning and Growth Measure : There has been a reduction of hours of employee training per employee due to consistent usage of data across the Enterprise	50	2.50	1.49

The balance scorecard identified four critical key perspectives to be included in measuring corporate performance namely the financial, customer, internal business process, learning and growth perspectives (Braam & Nijssen, 2004; Salterio & Lipe, 2000). Performance items were around these key perspectives pertaining to data governance initiatives. Respondents disagreed that there has been an improvement of customer satisfaction ratings as a result of data governance initiatives. For the other three perspectives (financial, internal business process, learning and growth) they also disagree that there has been an improvement that was influenced by data governance initiatives. This could be due to the fact that there are still many issues relating to lack of effective data governance policies and solutions (Rand secure Data, 2013).

4.4 Reliability and Validity Testing

Reliability is the degree to which the same results are likely to be produced when measuring the same construct multiple times (Bhattacherjee, 2012). Reliability implies consistency but not accuracy. For qualitative studies this can be affected by the researcher's objectivity which introduces unreliability. Sometimes it can be improved by using quantitative measures e.g. you count the number of the occurrence of things. It can also be improved by using data collection techniques which are less dependent on subjectivity like questionnaires. Bhattacherjee, (2012) identify other sources of unreliability; asking ambiguous or imprecise questions is one source. Piloting the study reduces this because one gets the feedback from the participants and they can raise concerns around not understanding what the question is asking. Some questions can be rectified by using category questions. For example when you ask people what their salary is, it is not clear whether you are referring to monthly, yearly or per hour. If the results are highly divergent and unreliable introducing a salary scale will help (Bhattacherjee, 2012). Lastly, unreliability arises when you ask questions where a respondent has no knowledge of or is not familiar with a question. Respondents will just guess to finish the questionnaire.

There are many ways which reliability can be measured namely Inter-rater reliability, Test-retest reliability, Split-half reliability and internal consistency reliability. For this study internal consistency reliability test was performed.

4.4.1 Internal Consistency Reliability

Internal Consistency Reliability measures consistency between different items of the same construct. Cronbach's alpha test is used to determine reliability of variables. A threshold value of 0.7 is normally used when computing cronbach alpha test (Field, 2009). Although there are some workers *viz.* Moss *et al*. (1998) who have also supported the view that Cronbach's alpha value above 0.6 is generally acceptable. Hair (1998) has also supported the view that in a study with small sample size, low Cronbach's alpha scores such as 0.6 can be taken as the measure of acceptability.

In this study a 0.7 threshold value was used since this is an explanatory research. Field, (2009) states that the value of α depends on the number of items on the construct, therefore as the number of items increases, α will increase. Table 4.9 shows the results of the test. It can be seen that Data Integration, Data Modeling and Data Governance have relatively low cronbach's α of 0.52, 0.62 and respectively. These values are less than 0.7 and this could be the result of the few items in the constructs. The researcher decided to keep these constructs as there is a supporting literature as to why they were included in the conceptual model. Data Quality, Data Owner and Stewards, Compliance and Performance all had high reliabilities meaning they have significantly higher values greater than 0.7.

Table 4.8: Reliability Scores

Construct	no# items	Mean	Avg. Mean	std	Avg. std	Alpha
Data Integration	2	5.22	2.61	2.07	1.04	0.52
Data Quality	7	21.86	3.12	5.55	0.79	0.80

Data Modeling	3	9.56	3.19	2.91	0.97	0.62
Data Owner and Stewards	5	14.74	2.95	5.81	1.16	0.85
Compliance	8	27.08	3.39	6.61	0.83	0.86
Data Governance	3	7.36	2.45	3.21	1.07	0.67
Performance	4	10.98	2.75	5.42	1.35	0.91

4.4.2 Convergent and Discriminant Validity

As part of exploratory data analysis, one has to assess the strength of relationships or associations between pairs of variables. This can be determined by looking at whether two variables covary (Field, 2009). A correlation coefficient is used to measure and quantify the strength of the linear relationship between ranked or numerical variables (Saunders et al., 2011). One needs to be cautious when interpreting correlation coefficient. It only shows the relationship between the variables not causality.

Refer to Appendix C for the results of the Spearman rank correlation coefficient of all the items of all seven constructs, which is R-matrix. The diagonal coefficients are all ones because each variable will correlate perfectly with itself. An R-matrix is used to determine Discriminant validity and Convergent validity simultaneously. Any significant correlation coefficients are shown in red.

Discriminant validity determines if the items of the construct are not measuring other items in different constructs which they are not supposed to measure (Bhattacherjee, 2012). Convergent validity determines the closeness of the items of construct that they were intended to measure (Bhattacherjee, 2012).

Collinearity or multicollinearity is a situation where two or more of the independent variables are highly correlated. This is determined by coefficients which are .8 and higher, it can have a damaging effect on multiple regression (Field, 2009). From the Spearman rank correlation coefficient results, multicollinearity does not exist and there is no correlation at a 0.8 or greater level (see Appendix C).

4.4.3 Factor Analysis

Factor analysis is an alternative statistical method used to demonstrate convergent and discriminant validity. Factor analysis is a technique for identifying clusters of variables (Field, 2009). The main aim is to reduce R-matrix to its underlying dimensions by looking at which variables seem to cluster together in a meaningful way. Field (2009) identifies three main uses of this technique; it is used to understand the structure of a set of variables. Secondly it is used to construct a questionnaire to measure an underlying variable. Lastly it reduces a data set to a more manageable size while retaining as much of the original information as possible.

The existence or presence of groups between subsets of variables indicates that those variables could be measuring aspects of the same underlying dimension. The underlying dimensions are called factors or latent variables (Field, 2009; Saunders et al., 2012). The factor analysis resulted in seven factors thereby confirming the seven constructs of the model. Most of the items that measured each construct loaded on their constructs thereby providing evidence of construct validity. The researcher used 0.5 thresholds. Items of the same construct are expected to have 0.5 or higher value for an adequate Convergent validity and they will be grouped together, whereas for different constructs they should have 0.3 or less value for adequate discriminant validity (Bhattacherjee, 2012; Saunders et al., 2011).

Factor analysis shows which items explain the factor the most, see Table 4.10. These are the items with the highest significant loadings (Hair, et al, 1995). In order to determine the variance explained by the factors, an eigenvalue table was computed see Table 4.11. This helps in determining which factor is statistically important. It seems logical to retain factors with larger values, but what qualifies an eigenvalue to be recognized as large enough to represent meaningful factor? Bhattacherjee (2012) and Field (2009) argue that in order to retain a factor it should have an eigenvalue greater than one. Although there is an argument that one is too strict, there is a suggestion that instead it should look at eigenvalues greater than 0.7 (Field, 2009). The factors satisfied the rigor condition of eigenvalue greater than one, Table 4.11 shows the values of all seven factors. 32 questions were converted into seven factors which explain 71% of the variance in the data.

4.4.3.1 Factor 1 (Corporate Performance)

P1 – There has been an improvement of customer satisfaction rating as a result of data governance initiatives
P2 – There has been reduction of costs due to improvement in regulatory compliance/ reduction of regulatory risk
P3 – There has been an improvement in internal business processes e.g. quick responses fewer errors experienced due to data governance initiatives
P4 – There has been a reduction of hours of employee training per employee due to consistent usage of data across the Enterprise

Corporate performance was the construct underlying factor 1. Factor 1 accounted for 28% of the total variance of the data and this is the highest value meaning it is a most important factor in data governance. It consists of four variables P1, P2, P3 and P4.The factor loading for P1 (0.89) is the highest amongst the four variables and this is the variable which explains the factor the most. The respondents disagreed that there has been improvement in four performance critical key perspectives that was influenced by data governance initiatives. This could be due to the fact that there are still many issues relating to lack of effective data governance policies and solutions (Rand secure Data, 2013).

4.4.3.2 Factor 2 (Compliance with data policies and regulations)

CRL1: Effectiveness—information is relevant and pertinent to the processes as well as being delivered in a timely, correct, consistent and usable manner
CRL2: Efficiency— Delivery of information through the optimal (most productive and economical) use of resources
CRL 3: Confidentiality—Protection of sensitive information from unauthorized disclosure
CRL 4: Integrity— Accuracy and completeness of information as well as to its validity in accordance with business values and expectations
CRL 5: Availability—Information being available when required by the process now and in the future.
CRL 6: Compliance—complying with the laws, regulations and contractual arrangements, to which the process is subject, i.e., externally imposed business criteria as well as internal policies
CRL 7: Reliability—Appropriate information for management to operate the entity and exercise its fiduciary and governance responsibilities
CRL 8: How often is the Organization audited to assess CobiT & ECT Act 2000 Compliance?

The construct that was underlying factor 2 was compliance with data policies and regulations. It accounted for 13.39 % of the total variance. Not all of it variables loaded in this factor. CRL1, CRL3 – CRL7 load well on factor (2) while CRL8 loaded on factor (5) and CRL2 did not load at all and they were excluded. The factor loading for CRL 5 is 0.84 which is the highest amongst compliance variables; this implies that CRL5 explains Factor 2 the most. In Table 5, the mean for CRL 5 is 3.84 and highest amongst other items. This means that the respondents agreed that in organisation X information is available when required by the process now and in the future. They also agreed that data is effective, efficient, kept confidential, have integrity, do comply with regulations, but is not reliable.

4.4.3.3 Factor 3 (Data Ownership and Stewardship)

DOS 2: Data owners contribute to definition and management of business metadata

DOS 3: Data owners determine the transformation rules

DOS 4: Data stewards support the user community regarding data quality

DOS 5: Data stewards perform exposure or risk identification

DOS 6: Data stewards verify the data after load

The construct underlying factor 3 was Data Ownership and Stewardship. Data Ownership and Stewardship variables loaded well on factor 3 except for DOS 6 which loaded on a different factor (6) and it was excluded. This accounted for 8.59 % of the total variance. DOS 2 has the highest loading factor of 0.88 meaning DOS 2 explains Factor 3 the most. Also value for DOS3 = 0.86 is fairly close to DOS2 value and high as well. In Table 4.3, the respondents disagreed that Data owners contribute to definition and management of business metadata and also determine the transformation rules.

The role of data stewards mainly depend on effective and efficient work performed by data owners. Respondents disagree that data stewards perform risk identification and verification of data after load to make sure that everything is according to the expected information. This could be the fact that there are no transformation rules and proper definition and management of business metadata. Literature emphasises that data validation is critical as it helps in correct data collection, transmission and data derivation processes, and plays a vital role in improving data quality (Bhansali, 2013).

4.4.3.4 Factor 4 (Data Quality)

DQ1 Data is Accurate: data item is close to its true value in terms of meaning and truthfulness

DQ2 Data is Consistent: data unit is specified the same throughout the organization

DQ3 Data is Complete: completeness of columns of a table containing data

DQ4 Timeliness of Data: promptness freshness and frequency of updates of data

DQ5 The organization has Data Quality tools and plans in place.

DQ6 How often is data auditing or profiling done?

DQ7 How often is data cleaning and monitoring done?

The construct underlying factor 4 was data quality. This accounted for 6.30 % of total variance. DQ1 – DQ5 & DQ7 loaded well on one factor (4). While DQ6 did not load on any factor and it was dropped. DQ2 had a highest factor loading of 0.86 amongst other variable; therefore this is the variable which explains the factor the most. In Table 4.5, respondents agreed that data is consistent meaning the data unit is specified the same throughout the organization. But they disagreed that there are data quality tools in place.

4.4.3.5 Factor 5 (Data Integration)

DI 2 There is continuous evaluation of existing data integration technology infrastructure and its ability to support data governance practices

DI 3 Data integration lifecycle is followed (Develop and Manage, Access, Discover, Cleanse, Integrate, Deliver, Audit, Monitor and Report).

The construct underlying factor 5 is data integration. This accounted for 5.66 % of the total variance. Both variables have highest factor loadings of DI2 (0.72) and DI3 (0.70) therefore both of them explain the factor the most. In Table 4.4, respondents strongly disagree that the organization continuously evaluates the existing data integration technology infrastructure and its ability to support data governance practices. They also disagree that the data integration lifecycle is followed.

4.4.3.6 Factor 6 (Data Governance)

DG1 Organisation performs Month-to-month scorecard/KPIs at business unit-level for accuracy/quality of specific data entities

DG2 Organisation uses of Data Quality tools i.e. IBM WebSphere Quality Stage for data profiling

DG3 How often are data External audits performed?

Data governance was the construct underlying factor 6. This accounted for 4.69 % of the total variance.DG1 has highest the factor loading of 0.86 meaning this is a variable which explains this factor the most. In Table 4.7, the respondents disagreed that the organisation performs Month-to-month scorecard/KPIs at business unit-level for accuracy/quality of specific data entities. This could be linked to the results found in Table 4.3 that data owners and data stewards are not performing the duties expected of them. They disagreed that the organisation uses Data Quality tools which is aligned to the answer to DQ5 where they disagreed that there are data quality tools in place.

4.4.3.7 Factor 7 (Data Modeling)

DM2 There is a Data model quality management framework which helps in validating the developed data models.

DM3 Evaluating the quality of a conceptual data model is critical to the successful development of an information system

DM4 Data analyst(s) are responsible for developing data models

This accounted for 4.17 % of the total variance which is the least in all seven constructs. This was the underlying construct for factor 7, all the variables loaded well on this factor. DM4 has the highest factor loading of 0.84, but also DM3 has a relatively high value of 0.77 meaning these variables explain this construct the most. In Table 4.6, the respondents disagreed that evaluating the quality of a conceptual data model is critical to the successful development of an information system. However literature emphasises that there are huge costs in doing error detection and correction after delivery rather than removing errors in requirements stages (Moody & Shanks, 2003). The respondents disagreed that data analysts are responsible for developing data models. Although they are responsible they are not solely responsible for data modeling, other key stakeholders in data modeling process are Business user(s), Data administrator(s) and Application developer(s) as data models acts as communication amongst them (Davies et al., 2006; Moody & Shanks, 2003).

Table 4.9: Factor Analysis

Variable	Factor (1)	Factor (2)	Factor (3)	Factor (4)	Factor (5)	Factor (6)	Factor (7)
DOS2			0.88				
DOS3			0.86				
DOS4			0.63				
DOS5			0.65				
CRL1		0.83					
CRL3		0.73					
CRL4		0.73					
CRL6		0.75					
CRL5		0.84					
CRL7		0.51					
DI2					0.72		
DI3					0.70		
DQ1				0.66			
DQ2				0.86			
DQ3				0.74			
DQ4				0.51			
DQ5				0.63			
DQ7				0.50			
DM2							0.51
DM3							0.77
DM4							0.84
P1	0.89						
P2	0.84						
P4	0.79						
P3	0.84						
DG1						0.86	
DG2						0.73	
DG3						0.71	

Table 4.10: Eigenvalues

Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	9.04	28.24	9.04	28.24
2	4.29	13.39	13.32	41.63
3	2.75	8.59	16.07	50.22
4	2.01	6.30	18.08	56.51
5	1.81	5.66	19.90	62.18
6	1.50	4.69	21.40	66.87
7	1.33	4.17	22.73	71.03

4.5 Hypothesis Testing

The primary purpose of the research was to gain insight into the factors that impact data governance and also determine the influence the quality of data governance has on the corporate performance of the organization. The main objective was to synthesize the literature to identify these factors and to propose a model which explains the relationship of the identified factors to data governance refer to section 2.10 Conceptual Model. Five factors were identified from literature and also based on the definition which the research has adopted. Data governance is the collection of decision rights, processes, standards, policies and technologies required to manage and ensure the quantity, usability, availability, accessibility, quality, consistency, auditability, and security of data in an organization (Cheong & Chang, 2007; Panian, 2010). These factors are compliance, data ownership, data integration, data modeling and data quality. In order to address the purpose, research questions were formulated and prepositions were constructed to help answer the questions. The following section presents the findings regarding research question one and its supporting prepositions.

4.5.1 Research question 1: which of the data management practises (factors) have an impact on data governance?

Preposition one, two, three, four and five help in answering the research question one. To test these hypotheses, multiple regression analysis was carried out.

Table 4.11: Impact of data management practises on quality data governance

N=50	b*	Std.Err. of b*	b	Std.Err. of b	t(44)	p-value
Intercept			-0.87	0.64	-1.35	0.18
AVDOS	-0.05	0.12	-0.05	0.11	-0.44	0.67
AVCRL	0.04	0.12	0.04	0.15	0.30	0.77
AVDQ	0.40	0.13	0.53	0.18	3.00	0.004
AVDM	0.25	0.13	0.27	0.14	1.95	0.05
AVDI	0.24	0.13	0.24	0.13	1.81	0.05

Statistic	**Value**
Multiple R	0.70
Multiple R^2	0.49
Adjusted R^2	0.43
F(5,44)	8.49
p	0.00001
Std.Err. of Estimate	0.82

The Multiple R (0.70) is the multiple correlation among the five independent variables and the dependant variable, and R Square (0.49) is the variance in the dependent variable accounted for by the five independent variables. The results also showed that data modeling and data integration are significant at 0.05 levels. The F ratio of 8.49 at 5 and 44 degrees of freedom is statistically significant at the 0.00001 level. Table 4.4 shows that Data Quality, Data Modeling and Data Integration are the factors which significantly impact the quality of data governance. All these predictors have positive b-values indicating positive relationships. So as Data Quality increases, quality of data governance increases; as Data Modeling increases quality of data governance increases; and as Data Integration increases, quality of data governance increases.

Data quality has the highest beta value of 0.53 amongst the three significant independent variables, which shows it is a greater predictor with a greater contribution. Data quality variable is significant at the 0.00001 level (i.e. p<=0.05).

Proposition 1: Inadequate compliance with data requirements in organisation X will negatively affect quality of data governance.

Table 15 shows that the compliance factor was not significantly impacting the quality of data governance. It has a p-value (0.67) which is greater than the statistical significant value of 0.05. It has a beta value of 0.04 which is close to zero. Therefore proposition 2: Inadequate compliance with data requirements in organisation X will negatively affect quality of data governance was not supported.

Proposition 2: Inefficiency of data ownership and stewardship negatively affects quality of data governance.

Table 15 shows that data ownership and stewardship factor was not significantly impacting the quality of data governance. It is not statistically significant at p-value 0.05 as 0.65 is greater than significant value 0.05. As the p-value of 0.65 is not statistically significant, therefore the proposition is not supported. This suggests that changes in quality of data governance are not associated with changes in data ownership and stewardship. The respondents when asked where are data owners residing? They were meant to choose between business unit and corporate IT, 10 out of 50 said corporate IT.

Proposition 3: Effectiveness of data integration within the organisation contributes positively to data governance.

Table 15 shows that the data integration predictor is significantly impacting the quality of data governance. From Appendix C data governance and data integration had a correlation of 0.53, so the value of R^2 will be 0.28. This shows that data integration shares 28 percent of variation in quality of data governance. Beta value of 0.24 is significant at the 0.05 level. Therefore proposition 4 was supported. The positive value of beta weight indicates that if quality of data governance is to be increased, improving data integration is necessary.

Proposition 4: Effectiveness of data quality practices contributes positively to data governance.

Data quality has beta value of 0.53 which is statistically significant at 0.004 level. This is the highest beta value out of all the significant values. It shows that data quality predictor has a greater impact on the model. From Appendix C data governance and data quality had a correlation of 0.62 which shows that these two are highly correlated, so the value of R^2 will be 0.39. Data quality shares 39 percent in total of the variation in quality of data governance. As the beta value is statistically significant therefore proposition 5 is supported. This shows that changes in the quality of data governance are related to changes in the data quality practices. The beta value has a positive value indicates a positive relationship so does data quality increases so as quality of data governance.

Proposition 5: Inadequacy of data modeling has a negative influence on quality of data governance.

From Table 15 the data modeling beta value of 0.27 is statistically significant at 0.05 level. The beta value has a positive value which shows that there is a positive relationship between data modeling and quality of data governance. Appendix C shows that data modeling is highly correlated with quality of data governance with a value of R = 0.49, so R^2 = 0.24. This implies that 24 percent of variation in quality of data governance is shared by data modeling. As beta is statistically significant, that implies that proposition 6 is supported. This shows that changes in the quality of data governance are related to changes in the data modeling.

4.5.2 Research Question 2: What impact does data governance have on the corporate performance of the organization?

Proposition 6: When the quality of data governance is poor it will impact the corporate performance negatively.

Table 4.12: Impact of quality of data governance on corporate performance

N=50	b*	Std.Err.	b	Std.Err.	t(48)	p-value

		of b*		of b		
Intercept			0.74	0.35	2.09	0.04
AVDG	0.69	0.10	0.89	0.13	6.64	0.00

Statistic	**Value**
Multiple R	0.69
Multiple R^2	0.48
Adjusted R^2	0.47
F(1,48)	44.10
p	0.00
Std.Err. of Estimate	1.02

To test this proposition simple regression was carried out. R square (0.48) show that 48 per cent variance in corporate performance is explained by quality of data governance. F-Ratio of 44.10 at 1 and 48 degrees of freedom is statistically significant at 0.00 level that is p-value < = 0.05. Proposition 6 is supported. The beta value has a positive value indicating that there is a positive relationship between the quality of data governance and corporate performance. This shows that changes in the corporate performance are related to changes in the quality of data governance. A linear regression established quality of data governance could statistically and significantly predict corporate performance. The regression equation was: predicted corporate performance = 0.74 +0.89 X (quality of data governance).

4.6 Chapter Summary

This chapter presented the results of data analysis. The sample size was 50 and the response rate was 29 percent. Most of the respondents were managers, specialists and analysts. A large number of the respondents hold degrees or diplomas with vast intensive years of experience. However, the small sample size could limit the reliability of the statistical analysis. The analysis showed that the instrument was reliable although data integration had a low value than the expected 0.6. The number of item in the data integration construct could be the reason for this. The model for this study contains seven factors.

Reliability analysis showed that the factors were internally consistent. Validity analysis showed that there was a high correlation between the constructs but no multicolleneriaty existed. Factor analysis showed that variables of the same construct are grouped under same subsets of variables which indicate that those variables are measuring aspects of the same underlying dimension.

The results also showed that three of the propositions were statistically significant meaning there was a relationships between independent variables and dependent variables. It showed that is there is a relationship between data quality and data governance, data modeling and data governance, data integration and data governance and lastly between data governance and corporate performance. Only two were not statistically significant meaning there was no relationship between compliance with data regulations and data governance, and data ownership and stewardship and data governance. It shows that changes in the quality of data governance are related to changes in the data modeling, data integration or data quality. Also showed changes in the corporate performance are related to changes in the quality of data governance. The results of the tests which confirmed this is shown in figure 7. The next chapter presents an interpretation and discussion of how well the research results support the research propositions.

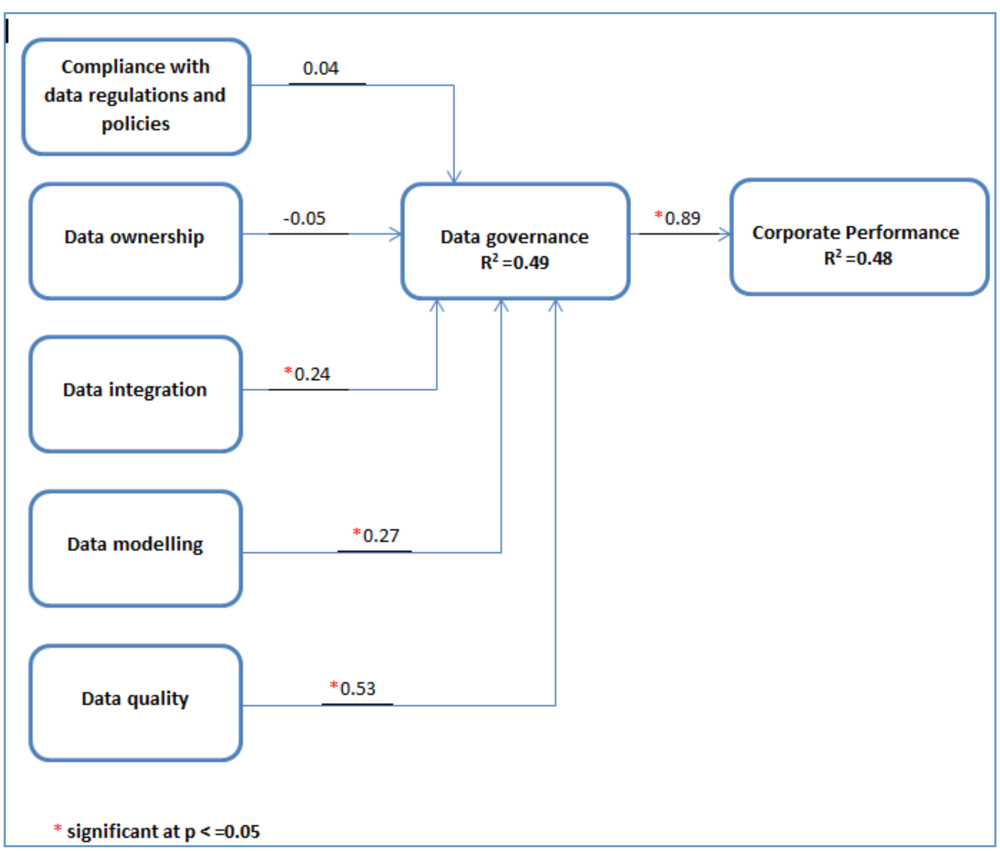

Figure 0.1 : Conceptual Model with test results values

5 Discussion

The previous chapters have clearly defined the research problem, purpose and questions; described the research design and implementation; and presented the results of data analysis. This chapter explains the extent to which the results answer the research questions, establishing if the purpose has been achieved and whether the research solves the problem under investigation. It consists of the following sections: *Discussion section which* interprets the results in relation to the research objective, questions and existing literature. The l*imitations section* outlines the weaknesses in the research and their effect on the findings and how it could be improved.

5.1 Discussion

The primary purpose for this study was to identify and investigate factors that affect data governance in organization X and also determine the influence that the quality of data governance has on the corporate performance of the organization. Literature was synthesised, definition of data governance which state that; data governance is the collection of decision rights, processes, standards, policies and technologies required to manage and ensure the quantity, usability, availability, accessibility, quality, consistency, auditability, and security of data in an organization (Cheong & Chang, 2007; Panian, 2010) was adopted. Seven variables namely compliance with data regulations and policies, data ownership, data integration, data modeling, data quality, data governance and corporate performance were identified. Identifiable relationships were captured through a proposed conceptual model in section 2.10. This section relates the results presented in Chapter 4 to the research objectives, questions and existing literature.

5.1.1 Research question 1: which of the data management practises (factors) have an impact on data governance?

The discussion on research question one consists of discussion on results of preposition one, two, three, four and five as these were formulated to answer this question.

The first proposition states that inadequate compliance with data requirements in organisation X will negatively affect quality of data governance. When compliance with data policies and regulations was tested against data governance, it did not appear to be statistically significant, see Table 15. This implies that compliance with data policies and regulations did not affect quality of data governance. This proposition was not supported because the beta value was not statistically significant; therefore there was no relationship between the two. This is contrary to the finding of a paper by Lajara et al. (2010) which states that data governance is associated with data compliance. Moreover it was found that complying with the data regulations and policies will increase quality of data governance (Bhansali, 2013). With laws such as the Sarbanes-Oxley (SOX) Act, CobiT 4, ECT Act 2000 organisations can ensure that avoidable risks are mitigated; controlled and unavoidable risks are contained in order to improve governance of data (Bhansali, 2013; Cheong & Chang, 2007; Panian, 2010). Data policies and procedures specify how data assets are tagged, tracked and monitored. Enforcing such regulations and compliance with these policies satisfy multiple purposes such as physical security, appropriate authentication, role based access and these contribute to the quality of data governance (Bhansali, 2013). Cheong & Chang (2007) also highlighted the importance of a data governance structure together with policies and procedures for managing data effectively. According to Otto (2011) & Panian (2010) ensuring compliance is the most common business driver for data governance.

A plausible reason for this unsupported proposition could be due to the limited data sample and diversity of the participants. Also it could be due to the fact that Organization X had only recently introduced data governance.

The second proposition states that inefficiency of data ownership and stewardship negatively affects quality of data governance. Panian (2010) argued that in order to achieve the goals of data governance, ownership of the data must be assigned and standards must be clearly defined. He also emphasises, clear definition of roles and assigning specific responsibilities to individuals involved in data governance such as data stewards enforces accountability. Accountability is a data governance factor identified from agency theory see Table 2. Furthermore Otto (2011) showed that in order to balance and match different interests of different stakeholders in data management and also to make binding decisions, data governance council should be formed. Data governance council is formed by data owners and the lead data steward. This shows that there is a link between data ownership and stewardship and data governance.

However the findings as presented in Table 15 show that this proposition was not supported. This implies that there was no relationship between data ownership and stewardship and data governance. Any change in data governance was not related to data ownership and stewardship. According to Table 4.3 and 10 respondents indicated that data ownership and stewardship and data governance were both poor, one would have expected the relationship to be true.

A plausible reason for this unsupported proposition could be due to the fact that there seems to be some uncertainty as to where do data owners and stewards belong either in IT or in the Business. Literature emphasises that data owners resides in the business (Ahmad et al., 2011; Berson & Dubov, 2007a; Rosenbaum, 2010).

The third proposition states that: Effectiveness of data integration within the organisation contributes positively to data governance. In Table 15, it is shown that data integration is a significant predictor of data governance, therefore the proposition was supported. This implies that there is a relationship between data governance and data integration. In Table 4.4 and 10, the respondents indicated that both data governance and data integration were poor. Regression analysis showed that there is a positive relationship between data governance and data integration and the responses confirm that. This is consistent with literature on this topic. For example, Magnani & Montesi (2010), Ziegler and Dittrich (2004) state that data integration promotes data reuse and gives a comprehensive answer to satisfy certain information needs or queries. This is in line with one of the data governance goals which is to ensure data fulfils business requirements.

The fourth proposition states that: Inadequacy of data modeling has a negative influence on data governance. This proposition was supported because data modeling was a significant predictor of data governance. This implies that there is a relationship between data modeling and data governance and regression analysis showed that this is a positive relationship. In Table

4.3 and 10, the respondents indicated that both data governance and data modeling were poor. Therefore high quality of conceptual model will lead to a high quality of information system as it impact both efficiency (time, cost, effort) and effectiveness(quality of results) of IS development (Shanks, 2007). This ties in to one of the business drivers of data governance; that is to increase operational efficiency. This is also linked to one of the data governance goals, which is to lower the costs of managing data.

The fifth proposition states that: Effectiveness of data quality contributes positively to data governance. According to Khatri and Brown (2010) data quality is pivotal to data governance and is one of the decision domains to the information governance framework. The effectiveness and success of any IT initiatives depends on the quality of the data (Cheong & Chang, 2007). Lajara et al. (2013) found that data governance is associated with data quality which supports the research proposition.

Similarly this study showed that data quality was the most significant predictor of quality of data governance therefore this proposition was supported. This shows that there is a strong relationship between data quality practices and data governance. According to Table 4.5 and Table 4.7, respondents indicated that both data quality and data governance were poor. The 'Garbage In, Garbage Out' saying is confirmed in this situation. These are expected results since there is a positive relationship between the independent and dependent variable and it was confirmed by the regression analysis, see Table 15. Poor data quality could be due to the fact that organisations give little attention or low priority to data quality areas as it is overshadowed by issues which are deemed to be important or more pressing (Azumah & Quarshie, 2012; Sheng, 2003).

5.1.2 Research Question 2: What impact does data governance have on the corporate performance of the organization?

The sixth proposition states that: When the quality of data governance is poor it will impact the corporate performance negatively. From Table 16, b-value of 0.89 was statistically significant at 0.00 level this implies that this proposition was supported. It means that there is a relationship between data governance and corporate performance. This is a strong link since data governance explains 48 % of variance in total. In Table 4.7 and 11, the respondents indicated that data governance was poor and they also disagreed that there was any positive change in corporate performance measurements due to data governance initiatives. Tallon et al. (2013) also confirms that there is a link between data governance and firm performance. The energy trading sector interviewees indicated that data governance outcomes reduced risk and improved decision making. In the airline industry it was associated with better decision making in scheduling, market analysis and ticket pricing. Lastly in the automotive industry it was associated with enhanced customer satisfaction and reduced costs (Tallon et al., 2013).

5.2 Limitations

All studies have some limitations. These limitations are related to methodology, some may be related to number and type of participants, others directly related to the survey and others to the procedure followed. This study was also affected by some of these limitations.

The first limitation of this study relates to the selection of participants. Some of the participants did not have depth and broad knowledge of the some areas under scrutiny. In sections where the participants had little knowledge they may have answered the questions by guessing or not answer the question at all. For example, although participants in coordinator role work with data, they may only have in-depth knowledge of compliance of data and may have little or no knowledge in the areas of data modeling and data integration.

Another limitation could have been the sample size. Although the sample size of 50 participants is acceptable but can still be regarded as small. This could have been improved by having an earlier start in data collection as they would be more time to survey additional participants. Also more contact between participants and the researcher would have increased participation. A larger sample would have a wider range of departments and title of participants which could improve the results. Furthermore, a bigger sample would allow for more rigorous testing of the evaluation framework and also enable generalization of the findings.

The last limitation is time horizon which is a cross-sectional approach. This approach can only establish factors of influence and identified relationships, and it cannot confirm causality. Governance is an ongoing process and longitudinal research could be beneficial for this study to address the issue of causality. In order to quantify or clearly identify a value something has on something it should be observed over time. The method of triangulation could have been used to validate findings but due to time constraints this was not possible. Lastly interviews could not be conducted due to time constraints and these could have helped to support the results. The next chapter presents conclusions, recommendations and future research opportunities.

6 Conclusion

The primary purpose for this study was to identify and investigate factors that affect data governance in organization X and also determine the influence that the quality of data governance has on the corporate performance of the organization. The previous chapter clearly discussed the findings of the research. This chapter focuses on what these findings imply and what implication they have in practice and theory. The outline of the chapter is as follows. The *Summary of Findings* draws conclusions from the findings and discusses the contribution of this research to the field of Data governance. The *Recommendations* section presents the recommendations to improve data governance based on the findings and the *Future Research section proposes* areas for further investigation.

6.1 Summary of findings

Data governance as an approach to better govern the use of information within an organization is rapidly gaining popularity. The petroleum industry firms are starting to treat data as a valuable asset and are seeking new ways to exploit data assets in order to expand their businesses and maximise their performances. Many operational domains rely on high quality of corporate data, such as business networking (Tellkamp et al., 2004), customer management (Crié & Micheaux, 2006), decision-making and business intelligence (Price & Shanks, 2005), and regulatory compliance (Friedman, 2006).

While it is acknowledged that data is a valuable corporate asset, the petroleum industry is still facing many issues relating to lack of effective data governance policies and solutions (Rand secure Data, 2013). The cost associated with poor data management in the oil and gas industry is higher than any industries. It can reach up to 22% of the annual revenue (Westheimer Energy, n.d.). This raised a need for the petroleum firms to improve the way they govern and manage their data in order to obtain value and insights to increase profitability as well as achieving a competitive advantage. Organisation X is currently facing these data challenges. Firstly it is hard for someone to get access to the information they need; secondly, there is a need for systems to better analyse, manage and standardize data for a specific query and thirdly, there is a growing need for technological capabilities which allow any-time collaboration from anywhere.

Researchers argue that while the concept of data governance is not new to the oil and gas industry, a holistic conceptualization of data governance is missing. Existing work investigated factors of data governance solely that led to an isolated solution. There is also limited research on the effects of data governance on firm performance (Tallon et al., 2013). This study aimed to address this gap with the investigation of the factors that affect data governance in organization X and also determine the extent to which the quality of data governance influence the corporate performance of organization X.

A case study was used and the data was collected via intranet hosted questionnaires with people whose job titles are aligned with data management. The conceptual model was used as a framework for the study, refer to section 2.10. Data analysis was performed with a statistical tool named STATISTICA 10, to test the hypotheses formed.

This research contributes to the existing theory by building on the dimensions of data governance defined in the literature by identifying some of the data factors which affect data governance. Data factors identified were compliance with data regulations and policies, data ownership, data integration, data modeling and data quality. In addition a data governance framework was developed. The framework incorporates a wide approach to deal with the data explosion challenges. It empirically establishes the relationship among the factors and data governance and also the relationship between data governance and corporate performance. Some factors, such as data modeling and data integration gave unexpected results. According to literature review of this research, they were associated with data quality. However in this study, these factors were associated with data governance.

The results revealed that there is relationship between data model and data governance, data integration and data governance and lastly data quality and data governance. The analysis showed that if there are poor factors (data model, data integration and data quality) they could results in poor data governance. The results also showed that there are positive relationships amongst these relationships which imply that when these independent variables improve there could be improvement on data governance. Therefore petroleum firms should focus on improving these factors in order to address the challenges they are facing. Improvement in data modeling and data integration could lead to clear presentation of information and sharing capabilities. This helps in enhancing the business-intelligence and analytical reporting as the most important information will be present in the most easily consumable way. The collaboration between disciplines could be improved. Data quality was confirmed to be the vital contributing factor to data governance. Core business activities in petroleum industry depend on high quality data. Improving this factor is critical as activities such as monitoring oil reserves and distributions, which help in meeting the current demand, will improve.

Poor data governance will result in poor corporate performance. The results showed that there are positive outcomes regarding this relationship which implies that when the data governance improves there could be an improvement on data governance. This finding confirms that robust data governance is crucial for the profitability and market competiveness in oil and gas producers. No relationship between compliance with data policies and regulations and data governance was found. Furthermore, there was no relationship between data stewardship and ownership and data governance. This might be due to the fact that Organization X had only recently introduced data governance or limited sample size.

6.2 Recommendations

This research showed that compliance with data regulations and policies have no effect on data governance. Recommendations to improve this finding would be that the organisation needs to create more awareness and build expertise in particular in compliance regulations and policies around data. This will create the awareness in employees around risks and penalties of non-compliance. Training will also help to properly show what is expected of the employees so that they do not rebel against policy enforcement. Frequent data audits would help to determine if the organisation is complying or not.

Data ownership and data stewardship was not associated with data governance. This could be rectified by assigning data ownership to the business. There should be a clear definition of duties of data owners in terms of what is expected of them like they should contribute to definition and management of business metadata, determine the transformation rules etc. This is an important exercise as data stewards are guided by these to ensure that the agreed-upon quality metrics are maintained on a continuous basis, and making sure appropriate data quality improvement programs are in place (Berson & Dubov, 2007a; Rosenbaum, 2010).

6.3 Future Research

The analysis showed that there was poor data modeling, poor data integration and poor data quality which results in poor data governance. It may be enlightening to know how these factor practices are implemented or performed. A research strategy which employs both positivist and interpretivist approaches could provide in-depth knowledge in order to determine areas of improvement. A longitudinal research approach could address the issues of causality, and thereby, provide deeper insight into data governance as this study only confirmed relationships. More research in the same or different organisations with bigger sample could provide a better understanding of this framework. For example a study in health and care organisations which have plenty of data governance papers could provide better understanding of this framework.

Bibliography

Adams, Julie. (2013). Oil industry data management—a forward thinking approach? Retrieved from http://www.accenture.com/us-en/blogs/Energy/archive/2013/05/22/oil-industry-data-management.aspx

Aguirre-urreta, M. I., & Marakas, G. M. (2008). Comparing Conceptual Modeling Techniques : A Critical Review of the EER vs . OO Empirical Literature. *ACM SIGMIS Database*, *39*(2), 9–32.

Ahmad, S., Abidin, S. Z. Z., & Omar, N. (2011). Data Sharing in Networked Environments : Organization , Platforms and Issues. In *In Proceedings of the 5th WSEAS international conference on Communications and information technology* (pp. 207–213). Venice, Italy: WSEAS Press.

Augier, M., & Teece, D. J. (2009). Dynamic Capabilities and the Role of Managers in Business Strategy and Economic Performance. *Organization Science*, *20*(2), 410–421. doi:10.1287/orsc.1090.0424

Azumah, K. K., & Quarshie, H. O. (2012). Towards Higher Quality Data : Impact of Perception of Data Quality on IT Investment - Ghana. *Journal of Emerging Trends in Computing and Information Sciences*, *3*(12), 1614–1621.

Bajaj, A. (2010). The Effect of Abstraction of Constructs in Data Models on Modeling Performance : An Exploratory Empirical Study The Effect of Abstraction of Constructs in Data Models on Modeling Performance : An Exploratory Empirical Study. *AMCIS 2010 Proceedings*, *517*(1), 1–13. Retrieved from http://aisel.aisnet.org/amcis2010/517

Batra, D., Hoffer, J. A., & Bostrom, R. P. (1990). Comparing Representations with Relational and EER Models. *Communication of the ACM*, *33*(2), 126–139.

Beijer, P., & Kooper, M. (2010). Information governance: beyond risk and compliance. In *Proceedings of the European Conference on Management, Leadership & Governance*. Poland: Wroclaw.

Bennett, T. a., & Bayrak, C. (2011). Bridging the data integration gap. *ACM SIGSOFT Software Engineering Notes*, *36*(3), 1. doi:10.1145/1968587.1968602

Berson, A., & Dubov, L. (2007a). *Master data management and customer data integration for a global enterprise* (pp. 107–132). New York: McGraw-Hill.

Berson, A., & Dubov, L. (2007b). *Master data management and customer data integration for a global enterprise.* (pp. 111–112). New York: McGraw-Hill.

Bhansali, N. (2013). *Data Governance- Creating Value from Information Assets*. New York, USA: CRC Press.

Bhattacherjee, A. (2012). *Social Science Research: principles, methods, and practices.* South Florida, USA: Global Text Project.

Braam, G. J. M., & Nijssen, E. J. (2004). Performance effects of using the Balanced Scorecard: a note on the Dutch experience. *Long Range Planning*, *37*(4), 335–349. doi:10.1016/j.lrp.2004.04.007

Bridoux, F. (2004). A Resourced-Based Approach to Perfomance and Competition : An Overview of the Connections between Resources and Competition. *Luvain, Belgium Institut et de Gestion, Universite Catholique de Louvain.*, *2*(1), 1–21.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Special Issue Information Security Policy Compliance : An Emperical Study of Rationality -Based Beliefs. *MIS Quarterly*, *34*(3), 523–548.

Cavana, R., Delahaye, B. L., & Sekeran, U. (2001). *Applied business research: Qualitative and quantitative methods*. Australia: John Wiley & Sons.

Cheong, L. K., & Chang, V. (2007). The Need for Data Governance : A Case Study. In *ACIS 2007 Proceedings*. Australia: Toowomba.

Cleven, A., & Wortmann, F. (2010). Uncovering four strategies to approach master data management. In *In System Sciences (HICSS) 43rd Hawaii International Conference*. Kauai: IEEE.

Cooper, R. B., & Wolfe, R. A. (2005). Information Processing Model of Information Technology Adaptation : An diffusion perspective. *ACM*, *36*(1), 30–48.

Creswell, J. W. (2012). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fourth Edi.). Sage.

Creswell, J. W. (2013a). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. *Qualitative Research Journal* (Vol. 6). Sage publications.

Creswell, J. W. (2013b). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. *Sage Publications, Incorporated*. Sage publications.

Crié, D., & Micheaux, A. (2006). From customer data to value: What is lacking in the information chain? Database Marketing & Customer Strategy Management. *Database Marketing & Customer Strategy Management*, *13*(4), 282–299.

Davies, I., Green, P., Rosemann, M., Indulska, M., & Gallo, S. (2006). How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, *58*(3), 358–380. doi:10.1016/j.datak.2005.07.007

De Abreu Faria, F., Gastaud Macada, A. C., & Kumar, K. (2013). Information Governance in the Banking Industry. In *46th Hawaii International Conference on System Sciences* (pp. 4436–4445). Wailea: IEEE Computer Society Press.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57–74.

Eisenhardt, K. M., & Martin, J. a. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*(10-11), 1105–1121.

Field, A. (2009). *Discovering statistics using SPSS* (Third Edit.). Sage publications.

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, *12*(2), 219–245.

Friedman, T. (2006). Gartner Study on Data Quality Shows That IT Still Bears the Burden. In *Gartner Group*. Stamford.

Ghauri, P. N. (2005). *Research methods in business studies: A practical guide*. London: FT Pearson.

Goodhue, D. L., Wybo, M. D., Kirsch, L. J., & Ha, C. (1992). Impact Integration of Data Integrationon the Costs and Benefits of Information Systems. *Management Information Systems Research Center*, *16*(3), 293–311.

Halevy, A., & Ordille, J. (2006). Data Integration : The Teenage Years. In *Proceedings of the 32nd international conference on Very large data bases* (pp. 9–16). VLDB Endowment.

Haug, A., Arlbjørn, J. S., Zachariassen, F., & Schlichter, J. (2013). Master data quality barriers: an empirical investigation. *Industrial Management & Data Systems*, *113*(2), 234–249. doi:10.1108/02635571311303550

Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations : Role of penalties , pressures and perceived effectiveness. *Decision Support Systems*, *47*(2), 154–165. doi:10.1016/j.dss.2009.02.005

Hristov, H. (2012). Choosing Approach for Data Integration. In *Proceedings of the 6th International Conference on nformation Systems & Grid Technologies* (pp. 98–113). Sofia: St Kliment Ohridski University Press.

Hudson, L. A., & Ozanne, J. L. (1988). Alternative ways of seeking knowledge in Consumer Research. *Journal of Consumer Research*, *14*(4), 508–521.

Johnson, R. B., & Onwuegbuzie, A. J. (2009). Mixed Methods Research : A Research Paradigm Whose Time Has Come. *American Educational Research Association*, *33*(7), 14–26.

Khatri, V., & Brown, C. V. (2010). designing data Governance. *Communication of the ACM*, *53*(1), 148–152.

Kim, K. K., Browe, D. K., Logan, H. C., Holm, R., Hack, L., & Ohno-Machado, L. (2014). Data governance requirements for distributed clinical research networks: triangulating perspectives of diverse stakeholders. *Of the American Medical Informatics Association*, *21*(4), 714 – 719.

Ko, D., & Fink, D. (2010). Information technology governance: an evaluation of the theory-practice gap. *Corporate Governance*, *10*(5), 662–674.

Kooper, M. N., Maes, R., & Lindgreen, E. E. O. R. (2011). On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management*, *31*(3), 195–200.

Kumar, S., & Phrommathed, P. (2005). *Research methodology*. *Springer US* (pp. 43–50). New York: Springer US.

Lajara, T. T., Carlos, A., & Maçada, G. (2013). Information Governance Framework : The Defense Manufacturing Case Study. In *Proceedings of the Nineteenth Americas Conference on Information Systems* (pp. 1–10). Illinois, USA: AISeL.

Lenzerini, M., Sapienza, L., Salaria, V., & Roma, I.-. (2002). Data Integration : A Theoretical Perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 233–246). Madison, USA: ACM Press.

Liao, C., & Palvia, P. C. (2000). The impact of data models and task complexity on end-user performance: an experimental investigation. *International Journal of Human-Computer Studies*, *52*(5), 831–845.

List, B., & Machaczek, K. (2004). Towards a Corporate Performance Measurement System. In *Proceedings of the 2004 ACM symposium on Applied computing* (Vol. 2004, Marc, pp. 1344–1350). New York, USA: ACM Press.

Magnani, M., & Montesi, D. (2010). A Survey on Uncertainty Management in Data Integration. *Journal of Data and Information*, *2*(1), 5.

Maguire, D. J., Worboys, M. F., & Hearnshaw, H. M. (2006). Object-Oriented Data Modelling for Spatial Databases. *International Journal of Geographical Information Systems*, *4*(4), 369–383.

Markus, M. L., & Soh, C. (1995). How IT Creates Business Value: A Process Theory Synthesis. In *Proceedings of the International Conference on Information Systems* (pp. 29–42). Amsterdam, The Netherlands: AISeL.

McMillan, J. H., & Schumacher, S. (2010). Research in education. *Pearson Education*.

Moody, D. L. (2005). Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data & Knowledge Engineering*, *55*(3), 243–276. doi:10.1016/j.datak.2004.12.005

Moody, D. L., & Shanks, G. G. (2003). Improving the quality of data models: empirical validation of a quality management framework. *Information Systems*, *28*(6), 619–650. doi:10.1016/S0306-4379(02)00043-1

Nelson, H. J., Poels, G., Genero, M., & Piattini, M. (2005). Quality in conceptual modeling: five examples of the state of the art. *Data & Knowledge Engineering*, *55*(3), 237–242. doi:10.1016/j.datak.2004.12.004

Newman, D., & Logan, D. (2006). Achieving Agility: How Enterprise Information Management Overcomes Information Silos. In *Gartner Research*. Stamford, CT.

Otto, B. (2011a). A Morphology of the Organisation of Data Governance. In *ACIS 2007 Proceedings*. Toowoomba: kluwer academic publishers.

Otto, B. (2011b). Organizing data governance: findings from the telecommunications industry and consequences for large service providers. *Commun AIS*, *29*(1), 45–66.

Otto, B., Wende, K., Schmidt, A., & Osl, P. (2007). Towards a Framework for Corporate Data Quality Management Towards a Framework for Corporate Data Quality Management. In *Proceedings of the 18th Australasian Conference on Information System*. Adelaide, Australia: kluwer academic publishers.

Panian, Z. (2010). Some Practical Experiences in Data Governance. *World Academy of Science, Engineering and Technology*, *62*(1), 939 – 946.

Price, R., & Shanks, G. (2005). A semiotic information quality framework: development and comparative analysis. *Journal of Information Technology*, *20*(2), 88–102. doi:10.1057/palgrave.jit.2000038

Raval, V., & Dyche, G. (2012). Seven Myths of Information Governance. *ISACA Journal*, *4*(1), 1–6.

Rivard, S., Raymond, L., & Verreault, D. (2006). Resource-based view and competitive strategy: An integrated model of the contribution of information technology to firm performance. *The Journal of Strategic Information Systems*, *15*(1), 29–50.

Rodriguez-ulloa, R., & Paucar-caceres, A. (2005). Soft System Dynamics Methodology (SSDM): A Combination Of Soft Systems Methodology (SSM) And System Dynamics (SD). *Systemic Practice and Action Research*, *18*(3), 303–334.

Rosenbaum, S. (2010). Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Services Research*, *45*(2), 1442–55.

Rowley, J. (2000). Using Case Studies in Research. *Management Research News*, *25*(1), 16–27.

Russom, P. (2008). Data Governance strategies: Helping your organization comply , transform , and integrate. *The Data Warehousing Institute*.

Salterio, S. E., & Lipe, M. G. (2000). The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures. *The Accounting Review*, *75*(3), 283–298.

Sandelowski, M. (2000). Focus on Research Methods Combining Qualitative and Quantitative Sampling , Data Collection , and Analysis Techniques in Mixed-Method Studies. *Research in Nursing & Health*, *23*(3), 246–255.

Saunders, C. S., Premkumar, G., & Ramamurthy, K. (2005). Information Processing View of Organizations: An Exploratory Examination of Fit in the Context of Interorganizational Relationships, *22*(1), 257–294.

Saunders, M., Lewis, P., & Thornhill, A. (2011). *Research methods for business students* (fifth edit.). Pearson Education.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (Sixth Edit., pp. 419–467). India: Pearson Education.

Shanks, G. (2007). Conceptual data modelling: an empirical study of expert and novice data modellers. *Australasian Journal of Information Systems*, *4*(2), 63–73.

Sheng, Y. P. (2003). Exploring the mediating and moderating effects of information quality on firms' endeavor on Information Systems. In *Proceedings 8th Internationa Conference on Information Quality* (pp. 344–352). London: Idea Group Publishing.

Shoval, P., & Shiran, S. (1997). Entity-relationship and object-oriented data modeling — An experimental comparison of design quality. *Data & Knowledge Engineering*, *21*(3), 297–315. doi:10.1016/S0169-023X(97)88935-5

Silberschatz, A., Korth, H. F., & Sudarshan, S. (1996). Data models. *ACM Computing Surveys*, *28*(1), 105–108. doi:10.1145/234313.234360

Silvola, R., Jaaskelainen, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2011). Managing one master data – challenges and preconditions. *Industrial Management & Data Systems*, *111*(1), 146–162. doi:10.1108/02635571111099776

Smith, H. A., & Mckeen, J. D. (2008). Developments in Practice XXX : Master Data Management : Salvation Or Snake Oil ? *Com*, *23*(1), 62–72.

Souza, D., Belian, R., Salgado, A. C., & Tedesco, P. (2008). Towards a Context Ontology to Enhance Data Integration Processes. *In ODBIS*, *4*(1), 49–56.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in Context. *Communications of the ACM*, *40*(5), 103–110.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, *30*(3), 141–178. doi:10.2753/MIS0742-1222300306

Tayi, G. K., & Ballou, D. P. (1998). Examining Data. *Communications of the ACM*, *40*(2), 54–57.

Tellkamp, C., Angerer, A., Fleisch, E., & Corsten, D. (2004). From Pallet to Shelf: Improving Data Quality in Retail Supply Chains Using RFID. *Cutter IT Journal*, *17*(9), 19–24.

Tien, C.-C. (2009). What Is the Truth in Market Research ? Being a Truth Teller. *Asian Journal of Management and Humanity Sciences*, *4*(4), 241–258.

Tongco, M. D. C. (2007). Purposive Sampling as a Tool for Informant Selection, *158*, 147–158.

Tran, T., Kim, S., & Hsiao, M. (2010). Data Stewardship and Flow Management for Data Quality Improvement. *D'Atri, Alessandro (Hrsg.)*, *1*(1), 1–13.

Van Alstyne, M., Brynjolfsson, E., & Madnick, S. (1995). Why not one big database? Principles for data ownership. *Decision Support Systems*, *15*(4), 267–284. doi:10.1016/0167-9236(94)00042-4

Vasista, T. G. K., & Al-Sudairy, M. T. (2011). Semantic Data Integration Approaches For E-Governance. *International Journal of Web & Semantic Technology*, *2*(1), 1–12.

Vroom, C., Solms, R. Von, Technikon, P. E., Elizabeth, P., & Africa, S. (2004). Towards information security behavioural compliance. *Computers & Security*, *23*(3), 191–198.

Wand, Y., & Weber, R. (2002). Research commentary: information systems and conceptual modeling—a research agenda. *Information Systems Research*, *13*(4), 363–376.

Weber, K., Otto, B., & Osterle, H. (2009). One Size Does Not Fit All — A Contingency Approach to Data Governance. *Journal of Data and Information Quality (JDIQ)*, *1*(1), 4.

Wheeler, B. C. (2002). NEBIC: A Dynamic Capabilities Theory for Assessing Net-Enablement. *Information Systems Research*, *13*(2), 125–146.

Williams, P. (2001). Information Security Governance. *Information Security Technical Report*, *6*(3), 60–70.

Young-Gul, K., & Salvalore, T. M. (1995). Comparing Data Modeling Formalisms. *Communication of the ACM*, *36*(6), 103–115.

Zhang, Q. (2001). Object-oriented database systems in manufacturing : selection and applications. *Industrial Management & Data Systems*, *101*(3), 97–105.

Zhou, B., Wang, S., & Xi, L. (2005). Data model design for manufacturing execution system. *Journal of Manufacturing Technology Management*, *16*(8), 909–935. doi:10.1108/17410380510627889

Ziegler, P., & Dittrich, K. R. (2004). Three Dedaces Of Data Integration — All Problems Solved ? In *In Building the Information Society* (pp. 3–12). Springer US.

Ziegler, P., & Dittrich, K. R. (2007). Data Integration — Problems , Approaches , and Perspectives. In *In Conceptual Modelling in Information Systems Engineering* (pp. 39–58). Berlin, Heidelberg: Springer.

Appendix A: Cover Letter

Department of Information Systems

Leslie Commerce Building
Engineering Mall, Upper Campus
OR
Private Bag. Rondebosch 7701
Tel: +27 (0) 21 650 4028 Fax: +27 (0) 21 650 2280
Internet: http://www.commerce.uct.ac.za/informationsystems/

21 November 2013

Dear Sir/Madam,

I am currently enrolled in a part-time Masters programme in Information Systems at the University of Cape Town and am required to conduct a research project as part of the deliverables in order to finish the course.

The title of my research is '*The Impact of data Governance on Corporate Performance: The case of a Petroleum Company'* and the purpose of this research is to examine data governance issues in an organization and how these affect a company's performance. This research will be conducted by means of an anonymous web-based survey.

Participation in this research is entirely voluntary and would be greatly appreciated. Participants can withdraw anytime from the survey. The survey will not take more than 15 minutes to complete. This research is being conducted for academic purposes and the results of this research will be submitted as part of my final deliverable. The research findings will also be given to the organization as token of appreciate for giving the researcher permission to conduct this study within the organisation; however, respondents and all of their responses will be treated as confidential and remain anonymous.

This research has been approved by the Commerce Faculty Ethics in Research Committee.

Thank you for your cooperation and time.

Should you have any questions regarding this research, please feel free to contact the researcher.

Sincerely

Signed by candidate

Signature Removed

Zimasa Ndamase

Masters Student

Department of Information Systems

University of Cape Town

Email: zimasa.ndamase@engenoil.com

Signature Removed

Prof. Michael Kyobe

Research Supervisor

Department of Information Systems

University of Cape Town

Email: Michael.kyobe@uct.ac.za

Appendix B: Questionnaire

Demographic Questions

What is your current title?
What function are you part of?
What is your level of education?
How many years of work experience do you have?

Model Constructs

Item	**Data Ownership/ stewardship**
DOS1	Where are data owners or stewards residing? (1= Business unit; 2= corporate IT)
DOS 2	Data owners contribute to definition and management of business metadata (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DOS 3	Data owners determine the transformation rules (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DOS 4	Data stewards support the user community regarding data quality (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DOS 5	Data stewards perform exposure or risk identification (1=never;2=once a year; 3=twice a year;4=quarterly;5=monthly)
DOS 6	Data stewards verify the data after load (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
	Data Integration
DI1	There is data integration technology infrastructure in place. (1=yes;2=no)
DI 2	There is continuous evaluation of existing data integration technology infrastructure and its

	ability to support data governance practices (1=never;2=once a year; 3=twice a year;4=quarterly;5=monthly)
DI 3	Data integration lifecycle is followed (Develop and Manage, Access, Discover, Cleanse, Integrate, Deliver, Audit, Monitor and Report)? (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DI 4	Please prioritise the intergration techniques from 1- 3 in order of use within the organization Number 1 being mostly used and 3 least used Data Warehouse _________ Enterprise application Integration ___________ Business Collaboration Infrastructure ________
	Data Quality
DQ1	Data is Accurate :data item is close to its true value in terms of meaning and truthfulness
DQ2	Data is Consistent : data unit is specified the same throughout the organization
DQ3	Data is Complete: completeness of columns of a table containing data
DQ4	Timeliness of Data : promptness freshness and frequency of updates of data
DQ5	The organization has Data Quality tools and plans in place. (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DQ6	How often is data auditing or profiling done? (1=never;2=once a year; 3=twice a year;4=quarterly;5=monthly)
DQ7	How often is data cleaning and monitoring done? (1=never;2=once a year; 3=twice a year;4=quarterly;5=monthly)
DQ8	Which of these problems caused by poor DQ are often experienced by the organization? Please rank from 1-6 where number 1 is most commonly experienced and number 6 is least experienced. Extra time to reconcile data --------- Delay in deploying a new system ----------- Loss of credibility in a system ---------- Lost revenue ---------------

	Customer dissatisfaction ---------- Compliance problems ------------
Who do you think should be responsible for Data Quality? What often raises an alarm on the Quality of Data e.g. User complaints or data audits? Where the data is normally cleaned from?	
	Data Modeling
DM1	Are there any data modeling processes and standards in place? (1= yes, 2= no)
DM2	There is a Data model quality management framework which helps in validating the developed data models. (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DM3	Evaluating the quality of a conceptual data model is critical to the successful development of an information system 1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DM4	Data analyst(s) are responsible for developing data models (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
	Compliance with Regulation/legislation
CRL1	CobiT 4 defines seven control criteria for information to satisfy business objectives. Please indicate to what extent the organisation complies with these controls. **Effectiveness**—information is relevant and pertinent to the processes as well as being delivered in a timely, correct, consistent and usable manner (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)
CRL2	**Efficiency**— Delivery of information through the optimal (most productive and economical) use of resources (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)
CRL 3	**Confidentiality**—Protection of sensitive information from unauthorised disclosure (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)

CRL 4	**Integrity**— Accuracy and completeness of information as well as to its validity in accordance with business values and expectations (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)
CRL 5	**Availability**—Information being available when required by the process now and in the future. It also concerns the safeguarding of necessary resources and associated capabilities. (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)
CRL 6	**Compliance**—complying with the laws, regulations and contractual arrangements, to which the process is subject, i.e., externally imposed business criteria as well as internal policies (1=neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)
CRL 7	**Reliability**—Appropriate information for management to operate the entity and exercise its fiduciary and governance responsibilities (1=Neither agree nor disagree; 2=strongly disagree; 3=disagree; 4=agree; 5=strongly agree)
CRL 8	Indicate extent to which the organization X comply with ECT Act 2000 (1= poor compliance, 2= satisfactory compliance)
CRL 9	How often is the Organization audited to assess CobiT & ECT Act 2000 Compliance? (1=never;2=once a year; 3=twice a year;4=quuarterly;5=monthly)
	Good Data Governance
DG1	Organisation performs Month-to-month scorecard/KPIs at business unit-level for accuracy/quality of specific data entities (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DG2	Organisation uses of Data Quality tools ie IBM WebSphere QualityStage for data profiling (1=Neither agree nor disagree 2=strongly disagree;3=disagree; 4=agree; 5=strongly agree)
DG3	How often are data External audits performed? (1=never;2=once a year; 3=twice a year;4=quarterly;5=monthly)
	Data Governance – Perfomance
P1	Customer-Related Measure There has been an improvement of customer satisfaction rating as a result of data governance initiatives
P2	Financial Measure There has been reduction of costs due to improvement in regulatory compliance/ reduction

	of regulatory risk
P3	Internal Business Processes Measure There has been an improvement in internal business processes eg quick responses fewer errors experienced due to data governance initiatives
P4	Learning and Growth Measure There has been a reduction of hours of employee training per employee due to consistent usage of data across the Enterprise
A	

Appendix C: R- Matrix

Variable	AVDOS	AVCRL	AVDI	AVDQ	AVDM	AVP	AVDG
AVDOS	1.00						
AVCRL	0.30	1.00					
AVDI	0.26	0.36	1.00				
AVDQ	0.35	0.26	0.51	1.00			
AVDM	0.21	-0.04	0.41	0.37	1.00		
AVP	0.28	0.39	0.49	0.47	0.33	1.00	
AVDG	0.20	0.22	0.53	0.62	0.49	0.68	1.00

Variables	R Squared
DOS	0.04
CRL	0.05
DI	0.28
DQ	0.39
DM	0.24

Appendix D: Ethics Form

UNIVERSITY OF CAPE TOWN

FACULTY OF COMMERCE

Igniting Knowledge and Opportunity

Commerce Faculty Ethics in Research Committee

Updated Ethics Form March 2013

Any individual in the Faculty of Commerce at the University of Cape Town undertaking any research that involves the use of human subjects, or research that may hold ethical consequences for the University of Cape Town, is required to complete this form and obtain approval before conducting research. The completed form should be submitted as an electronic document to departmental Ethics Committee representatives for submission to the Commerce Faculty Ethics in Research Committee. Please also submit electronic copies of your research proposal, informed consent form or other information used to obtain consent, and any questionnaires other material shown to subjects.

1. **PROJECT DETAILS**		
Project title: The Impact of data Governance on Corporate Performance		
Principal Researcher/s: Zimasa Ndamase	Email address(es):	Zimasa.ndamase@engenoil.com
Research Supervisor: Prof. Michael Kyobe	Email address(es):	Michael.kyobe@uct.ac.za

Co-researcher(s):	Email address(es):	

Brief description of the project:

Most companies are beginning to treat data as the most valuable asset as they mainly rely on it to make sound decisions to be competitive, satisfy their customers and increase revenue. They have noticed that data-related problems affect critical areas of running an organization. This creates the need to govern data. Goals of data governance are:

- To ensure that data fulfills business requirements
- To lower the costs of data management and
- To protect and treat data as the most valuable business asset

This study examines data governance issues in an organization X and how these affect a company's performance.

Data collection: (please select)

☐Interviews ☒Questionnaire ☐Experiment ☐Secondary data ☐Observation

☐ Other (please specify): ______

Procedure: (please describe)

Anonymous Questionnaires with open-ended questions will be used to gather data and it will be hosted within the organisation intranet. Open-ended questions have the ability to evoke responses that are unanticipated by the researcher and are rich and explanatory in nature.

Numeric Data collected in the study will be analysed using statistical tools such as SPSS and STATISTICA. Qualitative data collected in open-ended questions will be analysed using thematic analysis.

2. PARTICIPANTS

Characteristics of participants:

Gender: Any

Race / Ethnicity: Any

Age range: Adults (aged 18 and above)

Location: Western Cape

Other:

Race / Ethnicity:

Have you included a "Prefer not to Answer" response category in your questionnaire? (please select)

☐ Yes ☐ No ☒ Not applicable

If you answered 'No' why not?

Affiliations of participants: (please select)

☒ Company employees ☐ UCT staff ☐ General public ☐ UCT Students

☐ Other (please specify): ______

If your sample includes children (aged 18 and below), mentally incompetent persons, or legally restricted groups please explain below why it is necessary to use these particular groups. If subjects are minors or mentally incompetent, please describe how and by whom permission will be granted? If you are including children under the age of 18 and are not getting parental consent, please explain why you believe that their parents would consent if it was possible to contact them.

3. ORGANISATIONAL PERMISSION

If your research is being conducted within a specific organisation, please provide organisational permission or explain how permission will be obtained.

I will ask the permission from the CIO and send a letter to the manager whom I am targeting the respondents from together with CIO approval.

Are you making use of UCT students as respondents for your research? (please select) ☐ Yes ☒ No

If yes, have you contacted Executive Director: Student Affairs for permission? (please select) ☐ Yes ☐ No

Was approval granted? (please select) ☐ Yes ☐ No ☐ Awaiting a response

Are you making use of UCT staff as respondents for your research? (please select) ☐ Yes ☒ No

If yes, have you contacted Executive Director: Human Resources for permission? (please select) ☐ Yes ☐ No

Was approval granted? (please select) ☐ Yes ☐ No ☐ Awaiting a response

Contact Emails: Executive Director: Human Resources (Miriam.Hoosain@uct.ac.za)
Executive Director: Student Affairs (Moonira.Khan@uct.ac.za)

4. INFORMED CONSENT

What type of consent will be obtained from study participants? ☐ written consent ☒ anonymous survey ☐ oral consent (please justify) ☐ other (please specify)
How and where will consent/permission be recorded? The permission letter to conduct the research within the organisation will be signed and stored safely. The participate letter will be part of the survey as the respondents are anonymous no recorded consent will performed.
5. CONFIDENTIALITY OF DATA

What precautions will be taken to safeguard identifiable records of individuals? Please describe specific procedures to be used to provide confidentiality of data by you and others, in both the short and long run. This question also applies if you are using secondary sources of data that is not anonymous.

The participants will take part anonymously on the survey. The organisation will be assured that the information gathered in the study will be confidential and private and also aiming to do the survey on the intranet so the data will be housed in the organisation.

6. RISK TO PARTICIPANTS

Does the proposed research pose any physical, psychological, social, legal, economic, or other risks to study participants you can foresee, both immediate and long range? (please select)

☐ Yes ☒ No

If yes, answer the following questions:

1. Describe in detail the nature and extent of the risk and provide the rationale for the necessity of such risks
2. Outline any alternative approaches that were or will be considered and why alternatives may not be feasible in the study

1.
2.
3.

What authorship agreement have you reached with your co-researchers or supervisor?

☐ This research is not intended for publication

☒ Standard authorship agreement (principal researcher first author, co-researcher(s) and supervisor(s) co-authors)

☐ Customised agreement (please specify below):

I certify that we have read the the UCT Authorship Policy, and Commerce Faculty Authorship Guidelines
☐ **(http://www.commerce.uct.ac.za/Commerce/Information/research.asp)**

I certify that that the material contained herein is truthful and that all co-researchers and supervisors are
☐**aware of the contents thereof.**

I understand that it is my responsibility to conduct research in accordance with the ethical requirements of
☐**UCT.**

Signed by candidate

Signature Removed

Applicant's signature:	
Date:	

CHECKLIST	SELECT
A full copy of a research proposal or a literature review with methodology is attached	☒
Research proposal/ interview schedules / cover letters / questionnaires / forms and other materials used in the study are attached/ consent form	☒
Organisational consent letter / UCT student or staff approval letter	☒
On your cover letter to your questionnaire have you included the following?	NA ☐
1. The following UCT Logo	☒
2. A sentence explaining the aim of the research	☒
3. Sentences of a similar nature to below must be included in the cover letter or consent form:	
This research has been approved by the Commerce Faculty Ethics in Research Committee.	☒
Your participation in this research is voluntary. You can choose to withdraw from the research at any time.	☒

The questionnaire will take approximately X minutes to complete	☐
You will not be requested to supply any identifiable information, ensuring anonymity of your responses.	☒
	OR
Due to the nature of the study you will need to provide the researchers with some form of identifiable information however, all responses will be confidential and used for the purposes of this research only.	☒
Should you have any questions regarding the research please feel free to contact the researcher (insert contact details).	☒
4. Have you scanned in your signature for the last section of the form?	☒

For Ethics committee representative only Recommendation(s): **Signature:** **Date:**
For Ethics committee chairperson only Recommendation: **Signature:** **Date:**