UNIVERSITY OF KWAZULU-NATAL

Factors and perspectives influencing accountability and the sustainability of data quality improvements in higher education

by

Lennard Wood
921362272

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Business Administration

In the
Graduate School of Business & Leadership
College of Law and Management Studies

Supervisor: Alec Bozas

2013

DECLARATION

I, Lennard Robin Wood, declare that:

- The research reported in this thesis, except where otherwise indicated, is my original work.

- This thesis has not been submitted for any degree or examination at any other university.

- This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

- This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

    a) their words have been re-written but the general information attributed to them has been referenced;

    b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

    c) Where I have reproduced a publication of which I am author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

    d) This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed:

Acknowledgments

I wish to express my sincere appreciation and gratitude to the following individuals, without whose assistance, this study would not have been possible:

- My supervisor, Alec Bozas for his encouragement and support.

- Staff at the University of KwaZulu-Natal for giving of their time to share their opinions on data quality.

ABSTRACT

Worldwide, organisational management is increasingly confronted by the need for data quality to make informed decisions. It has been reported that a significant percentage of turnover is lost due to bad data. The higher education sector also requires quality data in order to keep up with the pace of change in this sector. It is against this background that this study investigates the factors and perspectives influencing accountability and the sustainability of data quality improvements at the University of KwaZulu-Natal (UKZN). The study used a questionnaire to elicit responses from approximately 120 information system users (IS-Users) at the University on their perspectives of data quality awareness, quality practices, and the cost, accountability and sustainability of data quality improvement in order to support the implementation of the data quality initiative recently launched at UKZN. The sample was selected from a population of primary information system users. Data collection took place over two months, with a response rate of 50%.

The findings and recommendations of the study revealed different opinions on various issues from the perspective of the three groups of IS-User constituents that were surveyed. The findings include significant and moderate issues relating to the lack of training, skills and leadership; work-around time problems arising from uncertainty with regard to who owns data at the Institution; and the need for stronger leadership and skills in the area of data quality.

The recommendations range from investment in training, to the implementation of performance management to support current data quality activity, service level agreements to enhance data quality from third party suppliers, incentives to reward work that enhances data quality, feedback mechanisms such as metrics or a data quality monitor to report on the condition of data quality in real time and promoting data ownership in order to enhance organisational agility to reduce the work-around or run around time of IS-users.

TABLE OF CONTENTS

CHAPTER TWO – LITERATURE REVIEW

CHAPTER FOUR - DATA ANALYSIS

CHAPTER FIVE - CONCLUSION AND RECOMMENDATIONS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DC : Data Custodian

DMAIC : Define, Manage, Analyse, Improve, Control

DSB : Data Steward (Business)

DST : Data Steward (Technical)

HEI : Higher Education Institution

ICS : Information and Communication Services

IS-User : Information Systems User

MI / II : Management Information

TQM : Total Quality Management

TDQM : Total Data Quality Management

CHAPTER ONE

INTRODUCTION TO THE RESEARCH

1.1 Introduction

In an environment characterised by continuous external and internal change, institutional data quality in higher education institutions, as in ‘organised industry’, has become increasingly difficult to manage. Quality of data is a challenge to organisations across the world as many factors impact the availability of quality data for managerial decision-making. This study examines numerous issues relating to the role of ‘people, processes and technology’ in ensuring data quality. The study gathered empirical data / information on the awareness, practice, accountability, cost and sustainability of data quality improvement in order to support the implementation of a data quality initiative recently launched at the University of KwaZulu-Natal (UKZN). The purpose of the study is to investigate user perceptions of their involvement in data quality activities (this is elaborated on in the Problem Statement in Section 1.7). The results of the study will lay a foundation for data quality improvement.

While there is no single definition of data quality, data appears to be of acceptable quality if it is found to be fit for its intended uses in operations, decision-making and planning (Juran, 1999). It refers to how fitness or appropriateness of data is perceived or assessed in terms of its purpose within a particular context.

The data under discussion relates to higher education institutions, i.e., operational, tactical, strategic data for use in decision-making, funding and planning at the University.

1.2 The Problem of Data Quality

As data is ‘intangible’ in nature, it is a challenge for organisations to manage as a strategic organisational asset. As with other assets, data has to be managed bearing in mind that this process is associated with costs as well as benefits. Worldwide, organisational management is increasingly confronted by the need for data quality to make informed decisions.

High quality data is critical to an organisation’s success, while poor data will impede organisational efficiency (Haug and ArlBjorn, 2011, Zhu, Madnick, Lee and Wang, 2012). As early as 1998, Redman warned that poor data quality can jeopardise the effectiveness of an organisation’s tactics, operations and strategies (Redman, 1998). Poor data quality can cause serious problems in the organisation (Fisher and Kingma, 2001). The impact of data quality and information about data quality on decision-making has been the subject of several studies (Chengular-Smith, Ballou and Pazer, 1999 and Jung, Olfman, Ryan and Park, 2005). Sheng and Mykytyn (2002) assessed the impact of data quality on firm performance. Lee and Strong (2003) investigated whether a certain mode of knowledge, or *knowing-why* affects work performance and whether the knowledge held by different work roles has an effect on work performance.

Madnick and Lee (2009a) postulated that the modern data intensive knowledge and economic environment has increased awareness of data and information quality issues. Wand and Wang (1996) note, that, poor data quality can severely affect the overall effectiveness of an organisation. If this problem is not addressed, the organisation may lose money and operating sub-optimally in terms of efficiency. Data quality also impacts issues relating to corporate and public accountability. Ewell (1989) suggests that satisfyingg stakeholders’ expectations, requires the good use of public money, accountability for resources and supporting the educational goals of a country.

1.3 Data Quality Impact - Value and Costs

The impact of data quality can be appreciated in terms of the business value of data as well as the cost of data quality.

Value:

From a practitioner point of view, O'Neal (2012) recognises that, while data is an intangible asset, the information and knowledge assets of an organisation can represent up to 20% of its value. She demonstrates the theoretical value of information and knowledge within the organisation via the formula VI (theoretical) (value of information) = VOrg x 0.20 where VOrg = Share Price x Number of Shares and VI (theoretical).

Costs:

Poor data quality costs a typical organisation in the industrial sector between 10% and 20% of its revenue (Redman, 2004). Previous studies estimated that 1% to 5% of data found in organisations is of poor quality (Redman, 1998). At the time of Redman's (1998) review, research efforts focused on operation and assurance costs, research and development and the production of data products. More recent 'Data Crunch' reports have underscored the extent of the costs of data quality (Data Crunch Report (Australia) (2011), Data Crunch Report (UK) (2011). It is important for an institution to quantify the impact of its data quality, as well as the cost in order to have a benchmark against which to measure the effectiveness of data quality improvement.

1.4 Change and Complexity

One of the reasons for increasing data complexity in industry in the past two decades is the number of mergers and acquisitions in response to a challenging and increasingly competitive business milieu. The higher education sector that is the focus of this research has not been exempt from these pressures and data quality management has become very important. Data volumes have increased and software and systems have become more complex, putting pressure on the maintenance of data in these systems. Change therefore requires a continuous review of the impact of data quality on the organisation and interventions and programmes need to be launched to counteract its devaluing impact on information. During times of continuous change, data quality improvement efforts are often given lower priority than higher order operational issues (O'Neal, 2011).

This study is situated in a higher education environment. While the 'profit motive' is absent in such an environment, appropriate accountability and reporting structures nevertheless exist with auditing practices that require data to accurately reflect the institution's operations.

1.5 Data and Information:

For the purpose of this study, the terms 'data quality' and 'information quality' are used interchangeably. Strong, Lee and Wang (1997) refer to the difference between data as 'raw facts' and 'information'. This distinction is supported by Zhu, Madnick, Lee and Wang (2012), who also state that '*data quality'* can be used to refer to technical issues around data, while '*information quality'* can be used to refer to non-technical issues. In this study 'data' is a broad term used to cover the concept, 'information'.

1.6

1.6.1 Data Quality Policy: Data Quality Initiative

A data quality initiative was recently approved at UKZN and is in the process of being executed. The scope, stakeholders and role players involved in data quality activity have been formally outlined. The initiative aims to review data in the University's main student administrative system. The initiative came about as a result of concerns expressed by a variety of stakeholders regarding poor data for statutory and internal decision-making purposes coupled with the range and rate of changes experienced by the Institution. The policy has been approved at Executive level, and appropriate documentation has been produced, that details the terms of reference and roles and responsibilities (Data Quality Principles and Guidelines, 2011, Data Quality Terms of Reference, 2011).

The need for a formal approach and formal structures in the area of data quality arose out of the merger of higher education institutions (HEIs) in KwaZulu-Natal, which required the amalgamation of different process and procedures, and software upgrades. Recent changes in UKZN's organisational structure and reporting requirements (for example 'global rankings') underlined the need for a new approach.

UKZN relies on good quality data to achieve its goals as defined in the institutional Strategic Plan. Goal Seven of the Strategic Plan states that the University should "establish and maintain efficient, effective management systems and processes that provide a caring and responsive service to meet internal and external needs in a pragmatic and flexible manner" (UKZN Strategic Plan, 2013, p 19). The University has a strong commitment to data quality and recognises the importance of data for the following objectives:

- to provide efficient and effective services to students, staff and other stake holders;

- to produce accurate management information for effective governance, planning and decision-making;
- to meet the University's statutory obligations to the Department of Higher Education and Training (DoHET);
- to meet external audit standards and requirements.

(Data Quality Principles and Guidelines, UKZN, 2011)

Data quality depends on a data quality strategy that should be the product of appropriate data quality planning and control (Eppler and Helfert, 2007). To give further impetus to the implementation of the policy, it is the researcher's opinion that the policy should be augmented by a *data quality strategy* to guide current data quality activity at operational level in terms of predefined steps.

1.6.2 Data Quality Practice and Problems

Data problems emanate from poor data at various stages of the data information life cycle (or even the student life cycle) and poor operational processes. They manifest at different levels of data gathering and collection and in the way data is extracted and interpreted during various operational phases. Problems also arise as a result of the integration of several operational systems and data transfers, as well as the interpretation of data, particularly more recently with the use of business intelligence applications. It is important to deal with the problem at source and very important to focus on the 'right thing'. According to Gharajedaghi (2006, p 114), "we fail more often not because we fail to solve the problems we face but because we fail to face the right problem".

Institutional data has historically / conventionally been improved in order to support HEIs' statutory submissions to Government. These audit exercises are compulsory and submissions cannot be finalised unless quality controls have been satisfied. While data quality checks for statutory reporting are accepted practice, for the data in information systems 'at large', data correction occurs in an *ad hoc* and reactive manner. Furthermore, a lack of formal accountability structures and feedback mechanisms would compromise the provision of consistent data quality, making it very difficult for the impact of bad data to be reported, particularly to those in authority (Haug & ArlBjorn, 2011).

It is acknowledged that most data problems originate at the data capturing stage (Maydanchick, 2007). Users should be trained and training material should be readily available to support data capture processes as a 'bottom-up' approach. A 'top down' approach should also be followed in terms of having instruments in place to quantify and measure, where possible, the financial impact of data errors. For the 'wider' system, a cycle of programmes could be available to elicit errors where data elements were either incorrect or missing due to violation of the business rules underlying the integrity of these elements.

1.6.3 The Management Information (MI) / Institutional Intelligence (II) Function as a Catalyst for Change

Three groups of information systems staff (referred to in this study as IS-Users) have been recognised as pivotal stakeholders in the success of UKZN's Institutional Data Quality initiative (DQI): Data Owners, the Data Custodian and Data Stewards. The University's Management Information / Institutional Intelligence (MI/II) section has been identified as the Data Custodian. Institutional knowledge *vis a vis* data and systems resides in these three groups, and it is in these groups that quality activity takes place. Therefore, this study sought to establish IS-Users' perspectives of the sustainability of data quality activity.

Data Owners are the institutional functional or business owners of specific areas e.g., Human Resources or Finance. They are referred to in this study as Data Stewards (Business). Another variation of the Data Steward associated with 'IT' relates to the technical assistance provided to the Data Stewards (Business) and

Data Custodian by a Data Steward (Technical). This distinction has been drawn for the purpose of this research study.

1.7 Statement of the Problem

As the world progresses from an industrial to an information economy, the importance of data and information has increased rapidly, supported by technological progress. Data has become the currency of the new economy.

However, the quality of data has been identified as a problem. This has significant consequences in terms of the efficiency and effectiveness of organisations and businesses and has contributed significantly to organisations' operational costs. As data is important in planning and decision-making, organisations confront the challenge of addressing this problem. While many use data cleansing tools to unearth dirty data, technological solutions alone cannot eliminate the causes of poor data quality, as it is as much a business problem as an information technology problem. Research is therefore required to suggest solutions to this dilemma.

This study addresses some of the reasons for poor data quality in a higher education environment to support exist data quality initiatives at the Institution. The study sought to elicit the perspectives of information system users to provide a firm foundation for future data quality improvement and the measurement of data quality improvement.

1.8 Purpose of the Study

The purpose of this research study is to improve data quality at UKZN. Improving the quality of data will promote efficiency, cost savings and customer satisfaction.

- The study seeks to establish the factors that determine data quality in order to design interventions to confront the problem of poor data quality.

- The findings will be used to support the current data quality initiative at the University.

1.9 Objectives of the Study

The objective of this research study is to ascertain data quality stakeholders ('IS Users')'s perspectives of the sustainability of data quality at UKZN. These stakeholders comprise institutional data owners, data custodians and data stewards. The perspectives were obtained *via* a survey focusing on data quality awareness, data quality practices and the cost of data quality, as well as accountability issues and sustainability towards data quality improvement.

1.10 Research Questions

The research study can be encapsulated in the following keywords:

A - Awareness

What is the nature of awareness and communication practices relating to data quality? Are these practices conducive to data quality improvement? Do structures exist to communicate issues relating to data quality (DQ) and the management of data quality? What is the nature of data quality in terms of dimensions and what are the causes and impact?

B – Accountability and Management

What are the perceptions *vis a vis* accountability, the roles involved in accountability for data quality issues, and the role of performance and service levels in promoting data quality? Is the notion of data ownership clear in terms of accountability for data quality?

C - Data Quality Problem Handling / Practice

What are the perceptions *vis a vis* practices or processes that could affect the quality of data?

D – Cost of Data Quality

Is it possible to determine a 'benchmark' cost of addressing data quality issues?

E - Sustainability of Data Quality Improvement

Are there differences in perspectives of the sustainability of data quality among the three groups of data quality stakeholders and if so, are these differences significant? Differences may point to levels of cooperation / synergy that require intervention.

While UKZN's institutional data quality policy has been approved, a 'DMAIC' (Define, Measure, Analyse, Improve and Control) intervention will be proposed to support the implementation of an institutional data quality strategy.

1.11 Motivation for the Research

There is a paucity of research on different perspectives of information system staff *vis a vis* the sustainability of data quality improvement. Earlier studies focused on:

- What data quality means to data consumers (Wang, Strong and Guarascio, 1996);

- The role of data producers (Strong, Lee and Wang, 1997);
- Data quality at British Telecommunications (Tull, 1997).

This research study was inspired by the researcher's involvement in various areas of data quality in the primary role of Information Analyst at UKZN but also *via* secondary involvement as an internal networker and catalyst for change. The reason for undertaking such a research study in the higher education sector is twofold:

- Externally - the changing higher education landscape and commensurate funding mechanisms require decisions based on quality data (a higher premium on accountability for data quality); and
- Internally - the researcher's occupational environment has been subject to continuous change, *inter alia,* relating to mergers, the internal organisational structure and the student system. This led to serious data challenges that raised concerns about data quality among various senior stakeholders that use institutional data, culminating in a formal data quality policy. The researcher utilised this climate of institutional data redress to investigate perspectives that may influence data quality improvements and provide deeper insight into issues that may enhance or impede current initiatives.

1.12 Research Methods

The study adopts a case study approach that incorporates a largely quantitative approach. An analysis of the time devoted to data quality activity will be undertaken to obtain the cost of data quality; this will be used as the basis to evaluate future data quality improvements. Responses from IS-users relating to the sustainability of data quality activity will be assessed in terms of differences in means.

1.13 Beneficiaries of the Research and Outcomes

The beneficiaries of this research study will be the dimensions of 'people, process and technology'. Firstly, the study aims to sensitise 'information workers' to data quality issues and re-examine data quality practices, leading to improved data governance structures to support accountability for data issues. An improved structure will streamline roles and processes, resulting in improved accountability. This will provide clarity and stability and a technology platform to support the monitoring and measurement of data quality, supported by the development of metrics to hold the relevant parties accountable.

1.14 Chapter Outline

Chapter 2 – Literature Review

The literature reviewed in support of the research objectives includes quality and data quality theory; the role of data governance in providing a framework for clarity on the roles and functions of data owners; theory around data ownership and accountability; the costs of data quality; and the sustainability of data quality improvement. It also presents a brief situational overview of UKZN and the context within which the research problem is investigated. The role of management information and data quality at the University is discussed, including its role as an information broker, facilitator and a catalyst for change and data custodian. The roles and activities of the data quality stakeholders are also discussed.

Chapter 3 – Research Methodology and Situational Background

This chapter reviews the research framework relating to the need that inspired this study; the choice and conceptualisation of the research instrument, the process, scope and people involved in data collection and the challenges encountered in the research process.

Chapter 4 – Data Analysis

Chapter 4 analyses the data with reference to the research objectives. It analyses the perspectives of the IS-Users using descriptive statistics and estimates the cost of correcting bad data as well as providing a statistical perspective on the sustainability of data quality improvement from the perspective of the data custodians and data stewards.

Chapter 5 – Conclusions and Recommendations

This chapter summarises the findings ‘from the field’ *vis a vis* the literature, the implications of the cost of data quality, and the implications relating to perspectives of the sustainability of data quality activity. The chapter provides a brief synopsis of data quality improvement and its challenges and presents recommendations arising from the survey results. Recommendations are also made for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of the literature on data quality. It examines the framework within which data quality operates, namely selected building blocks of data governance frameworks and data quality theory relating to awareness and knowledge of data quality dimensions. Dimensions are linked to problems and can be used to diagnose the problems. The causes and impacts of data quality are discussed, as well as the role of data stewardship and accountability. In accordance with the notion of ‘people’ within the ‘people, processes and technology’ theme in the definition of data governance, the notion of ‘stewardship’ is further analysed in terms of responsibilities and job functions. A brief overview of data quality cost is presented as well as the factors that determine the sustainability of data quality improvements. The chapter concludes with a situational overview of data quality at UKZN, in particular the data roles set out in the Data Quality Policy.

2.2 DATA GOVERNANCE

2.2.1 Introduction and Definition

The literature that will be dealt with in this Chapter, notes that data governance provides the broader framework for data quality, data quality management and the roles and commensurate accountability that accompany it. A data governance framework provides the environment for the management of data quality and the execution of data quality initiatives. It sets out the facets of data quality management pertaining to the ‘what’ of decision making and accountability and ‘who’ makes decisions. Data quality participants’ work roles are complex as similar job titles are not always associated with the same functions. The research study focuses on information systems users' perspectives. The research participants are involved in sourcing, capturing, storage, production, maintenance, analysis and reporting of data. The literature that examines the building blocks of data governance, i. e., data quality management and data quality, respectively in terms of citations over the past two decades is shown below in Figures 2.1 and 2.2.

Data Quality Management Research

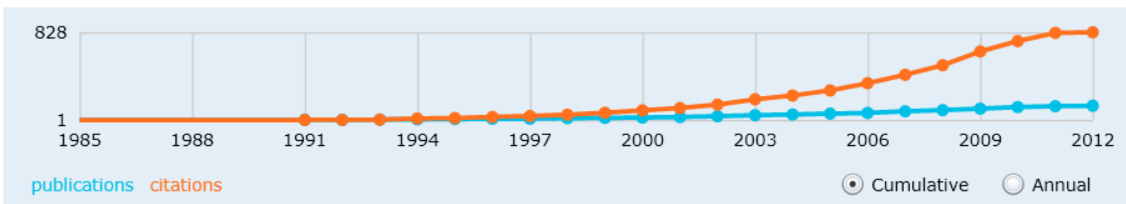


Figure 2.1 - Publications and Citations in Data Quality Management Research

Source: Microsoft Academic Search, http://academic.research.microsoft.com/Keyword/9056/data-quality-management

Data Quality Research

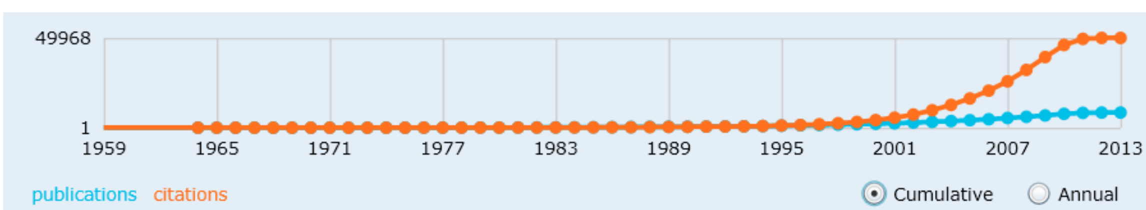


Figure 2.2: Publications and Citations in Data Quality Research

Source: Microsoft Academic Search, http://academic.research.microsoft.com/Keyword/9053/data-quality?query=data%20quality

A comparison of the figures shows that the importance of data quality has increased exponentially over the past 13 years. An understanding and alignment of the goals of business and data governance is important as the objective of a business is to create revenue, decrease costs and increase operational efficiency. The objective of the data governance organisation is to give direction to how data should be captured, collected, transferred and managed (O'Neal, 2011).

2.2.2 Corporate and IT Governance

A discussion of data quality and its position within data governance requires an examination of the position of data governance within corporate governance. Dismute (2009) cites Shleifer and Vishny's (1997, p737) definition of data governance as "the way in which suppliers of finance to corporations assure themselves of getting a return on their investment". He adds that, "the definitions all seem to deal with the direction and performance of a corporation" and involve a number of stakeholders, including directors, senior management and other shareholders (Weill and Ross, 2004, p11). This definition also embraces all the organisation's assets.

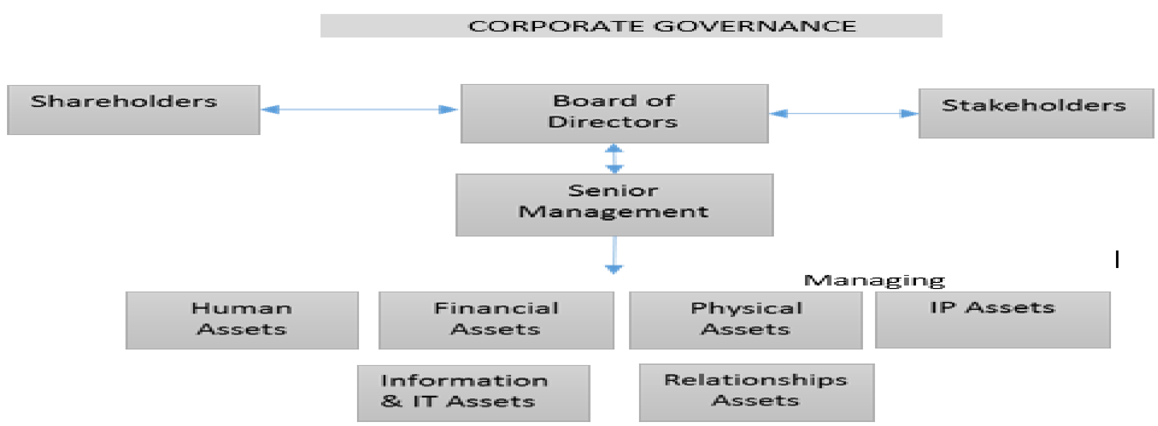


Figure 2.3: Basic corporate governance structure

Source: Dismute, WS (2009) *Data Governance: A Study of the Current State and Emerging Trends*, Master's Thesis, Information Science Department, University of Arkansas, Little Rock.

While governance infrastructure is an amalgamation of technologies, systems, people, policies, practices, and relationships, IT governance is a distinct and subordinate component of corporate governance. Data governance has become a distinct area, as the scope of IT governance requires that it be afforded specific attention. Khatri and Brown (2010) observe that IT Governance creates the context for data governance; this is illustrated in Figure 2.4 below:

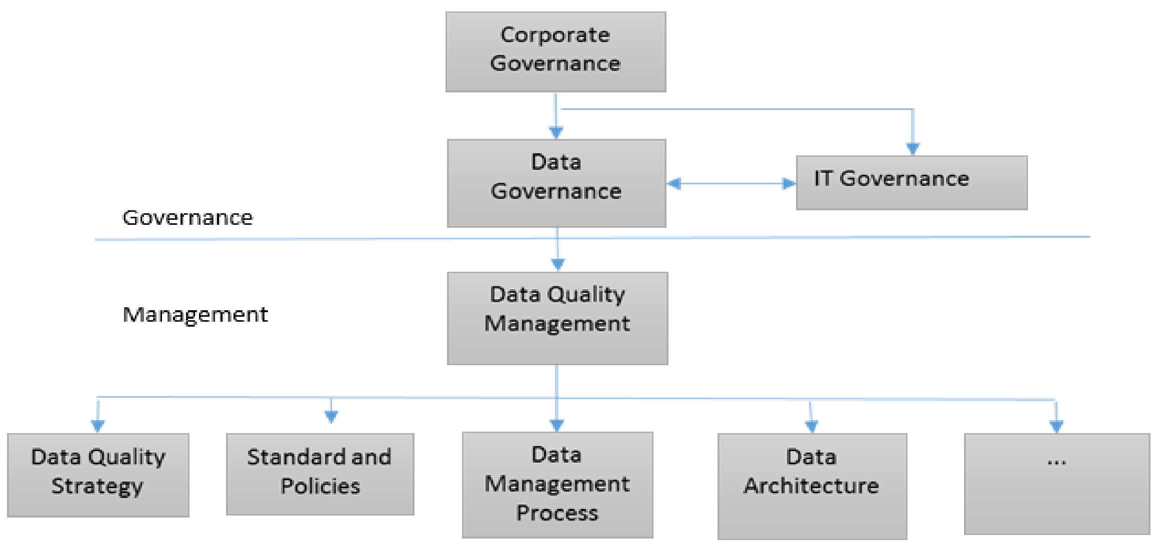


Figure 2.4: Corporate Governance

Source: Khatri, V % Brown CV, (2010), Designing data governance, *Communications of the ACM,* Vol 53, No 1 pp 148-152.

2.2.3 Data Governance

Data governance refers to a system that confers decision making and accountability responsibilities and involves information processes, executed in alignment with models that describe what action can be taken by whom, on the basis of what information, when, under what circumstances, using what methods. It provides a framework for fiduciary obligations for data to be enhanced and protected in order for stakeholders to have access to data of appropriate quality. The success of data governance is determined by the extent to which leadership and management acknowledge and support the notion that it is the responsibility of everybody in the organisation.

Wende and Otto (2011, p6) describe data governance as a “framework that specifies the decision rights and accountabilities to encourage desirable behaviour in the use of data” and sets the ‘rules of engagement’ for management activities relating to data. It is motivated by the need for members of the organisation to exhibit ‘desirable’ behaviour in the use of data and information. In order to inculcate such behaviour, data governance provides for and implements organisation-wide data policies, guidelines and standards in line with the organisation’s mission, strategy, values, norms, and culture. Data governance enables planning at a high, over-arching level and control over data management and coordinates the collaboration between IT and the enterprise. This is underscored by Khatri and Brown (2010) who emphasise the assignment of decision-making rights particularly with regard to an enterprise’s data assets. Decision rights refer to the ‘who’ and ‘what’ in data governance, i.e., who makes the decisions and what processes are involved.

The control aspect of data governance relates to assessment, management, use and improvement in data quality; this is analogous to a ‘control’ imperative in a Define, Measure, Analyse, Improve and Control (‘DMAIC’)/ TQM (Total Quality Management) approach to quality.

Data governance is a framework within which a data asset can be managed (Tyche, 2007). In the case of IT, technology assets are managed. While an organisation’s technological infrastructure provides a platform for data, the data asset itself carries the business value / business information. It is for this reason that efficient and effective management of data is required in order to realise its potential value. Tyche (2007) outlines the criteria that define data as an asset, namely, that the asset has a value that can be quantified, it contributes to the organisation reaching its strategic objectives and the asset requires specialised skills in order for it to be appropriately developed and maintained. As data is converted to information, and knowledge is derived from information, the business value attached to the data increases, producing the intelligence required for sound business decisions.

Organisations around the world are becoming increasingly aware of the importance of data governance (Brunelli, 2012). Brunelli (Ibid.) notes, that a reader survey in 2011 found that 77% of the respondents either had, or were planning to implement, a data governance programme (SearchDataManagement.com 2011 Reader survey). He added that, business users' increasing ownership of data is a major contributor to increased awareness of the importance of data quality. Organizations are also becoming more aware that data problems are business problems; previously, they were regarded simply as IT problems.

2.2.4 Decision Domains

Khatri and Brown (2010) state that, data governance can be expressed in terms of five decision areas or domains: Data Principles, Data Quality, Metadata, Data Access and Data Lifecycle. These domains are described below:

- Data Principles – recognition of data as an asset
- Data Quality – the prerequisites for data quality
- Metadata – the semantics or semiotics underlying data should be appropriately understood and interpreted
- Data Access - the requirements and infrastructure to access the data
- Data Lifecycle – issues involving the production, retention and redundancy of the data

Interrogating the decision domains in terms of the 'what', 'how', 'why', and 'who' of data governance, the 'who' or the human element represents the roles of those that will be involved in data quality, expressed *via* the roles of Data Owner, Subject Data Expert, Data Quality Manager and Data Quality Analyst.

2.2.5 Data Governance - Data Quality

It follows that the link between data governance and data quality can be expressed as the data governance initiatives managed by a number of role players that comprise the team responsible for data according to the five decision domains. At a narrower level within the data quality decision domain, a team consisting of Data Owner, Subject Data Expert, Data Quality Manager and Data Quality Analyst would work together to ensure the accuracy, accessibility, consistency and completeness of data (Khatri and Brown, 2010).

2.2.6 Data Governance – Accountability and Locus of Control

Once a structure is in place to determine how decisions are made *vis a vis* data, who will decide what needs to be done and by when will depend on the concept of control and locus of control. The locus of control is the structure that has primary responsibility for data governance in an organisation. In terms of where it should be positioned, functionally and hierarchically, some authors suggest that it is best located in business departments (Friedman, 2007 in Otto, 2011) while others suggest the organisation's information systems (IS) or information technology (IT) department (Otto, 2011).

2.3 DATA QUALITY - AWARENESS, CHARACTERISTICS, CAUSES AND IMPACT (Objective 1)

Once a data governance environment in place within which a team of data or information workers can operate, it is important to consider the data itself and enhance institutional knowledge of aspects of data quality. It is important that users understand the dimensions of data quality as well as the causes and impacts of such quality relative to the complexities involved in data quality improvement. Data quality can be determined by definitions and the criteria or dimensions that are most often cited when data problems are discussed. It is also important to understand the nature of data quality in order to raise awareness of the need for such quality.

2.3.1 Awareness of Data Quality

As early as 1993, scholars identified data quality awareness as a critical component of a Data Quality 'Jigsaw' matrix (Mare, 1993); this is depicted in Figure 2.5 below. Awareness is seen as a cornerstone of all data quality activities irrespective of it having originated from experience or philosophical thought.

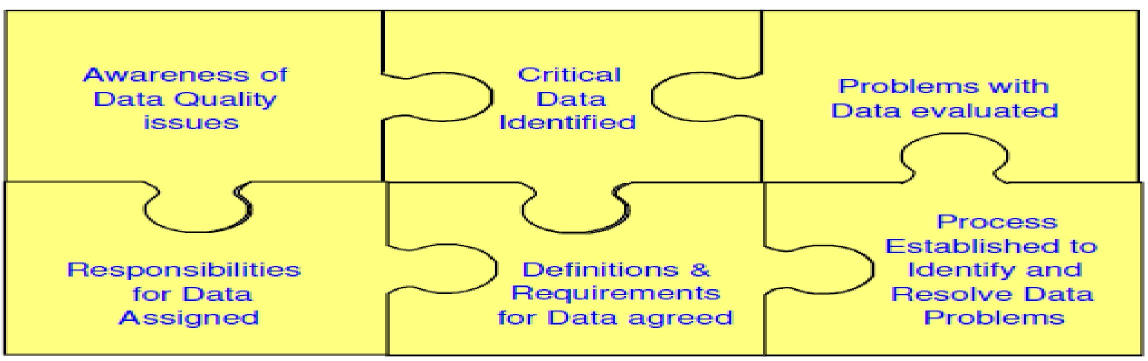


Figure 2.5 - Data Quality 'Jigsaw'

Source: Mare, D (1993), IC Position Paper: Managing Data Quality, Shell Internationale Petroleum Maastschappij B.V., The Hague.

Redman (1998) alerts practitioners to the fact that creating awareness of a problem and its impact is an important step towards resolving the problem.

Data quality dimensions such as those depicted in Figure 2.6 below could be used as feedback mechanisms to sensitise users to unsatisfactory levels of quality as the organisation uses the data. Jones (2012) encourages vigilance and action when dealing with problems, e.g., when errors in spreadsheets, reports or an email are encountered (related to accuracy and completeness of data), users normally fix these themselves. Such errors should ideally be routed to the source department that captures or generates the data.

Table 2.6 below categorises the dimensions into the information quality categories of intrinsic, contextual and representational quality (Wang, 1998).

IQ Category	IQ Dimensions
Intrinsic IQ	Accuracy, Objectivity, Believability, Reputation
Accessibility IQ	Access , Security
Contextual IQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of Data
Representational IQ	Interpretability, Ease of Understanding, Concise representation, Consistent representation

Table 2.6 - Information Quality categories linked to Information / data quality dimensions

Source: Wang, R (1998), A product perspective on total data quality management, *Communications of the ACM,* Vol 41, No 2, pp 58–65.

Pipino, Lee and Wang (2002) suggest that metrics would play a role in the feedback mechanisms, as the process of assessing data quality is a continuous effort that requires awareness of the basic principles underpinning the development of subjective and objective data quality metrics. Madnick and Lee (2009a) suggest that the continuously changing business milieu, regulatory requirements, the increasing varieties of data forms and media, and Internet technologies that dictate how information is generated, stored, manipulated and consumed, heighten quality awareness. This requires constant vigilance in terms of data quality. Madnick and Lee (2009b) add that users' sensitisation to data quality issues has been increasing in light of the importance of data quality in a data-intensive, knowledge-based economy. Lee and Strong (2004) note that, it is important for data workers to possess the 'why knowledge' that provides the foundation for the data production process in order to contribute to the generation of quality data. They add that data collectors' knowledge is more important than that of data custodians. Wang, Storey and Firth (1995) documented employee awareness of issues related to data quality.

Organisations may sometimes be unaware that there are data quality problems. In outlining performance domains, Lwanga Yonke, Walenta and Talburt (2011) state that Information Certified Quality Professionals (ICQPs) can enhance awareness by:

- Formulating a communication strategy that identifies key stakeholders, messages, desired actions and results in order to solicit support for an information quality strategy and governance.

- Critical relationships with senior leaders should be nurtured through regular communication that highlights value and business results; these relationships become a foundation for support and enforcement of the information quality mandate.

- Data quality levels should be monitored and reported on an ongoing basis in order to ensure that data quality levels are maintained. Feedback mechanisms should be put in place to promote awareness.

Grek and Ozga (2008) observe that, regular evaluation and increased regulation, particularly in the education sector, has heightened global demand for more data as well as improved data quality.

Data quality awareness is often created by the practical problems users encounter with operational processes, e.g., data migrations (moving data from one system to another), data mergers (amalgamating data from disparate sources (silos) into a single master file and data consolidation, when organisations ‘purge’ redundant data.

User awareness of the ‘current state of data quality’ could be further facilitated by technologies, for example, using data profiling to provide a ‘helicopter view’ of the condition of data. Data profiling is a relatively new concept that has gained ground as users do not readily have knowledge of the condition of data in ‘real time’. Data is subject to decay from the moment of data entry onwards and is particularly evident in customer information, for example, names and addresses that change continuously. Data loses its value if it is not continually maintained. Knowledge of the current state of data quality would sensitise users to the need for continual maintenance.

Grek and Ozga (Ibid.) acknowledge that while data is a product (or by-product) produced by most organisations, it is not viewed or managed as such. It is important for data to be treated as an asset that has business value and to be aware that costs are incurred if it is not maintained. This thinking lies behind the development of the Six Sigma ‘DMAIC’ concept.

The DMAIC concept refers to a data-driven improvement cycle to improve, optimise and stabilise business processes and designs. The DMAIC improvement cycle is the core process used to drive Six Sigma projects.

2.3.2 Characteristics of Data Quality

This section introduces the various philosophies, principles and dimensions of data quality as well as its causes and impacts in order to understand the complexities of data quality improvement. It is important to understand the foundations of data quality through the definition(s), criteria and dimensions that are most often cited when data problems are discussed.

Du Mars (2008, p1) reports that “missing or incorrect data, duplicate entries, misidentified information, undocumented relationships between data elements” are some of the problems that diminish the value of data on a daily basis in organisations due to errors and mistakes by business users. The compounded effect of these factors means that seemingly insignificant, sporadic data quality problems can ‘snowball’ into big problems for business processes and cause significant losses in both monetary terms as well as employee productivity (Ibid).

Data represents the building blocks of information. While there is no single definition of data quality, data appears to be of an acceptable quality if it is found to be *fit for intended uses* in operations, decision-making and planning (Juran, 1999). The ‘fit for use’ concept of data has been widely accepted in the literature (Strong, Wang and Guarascio, 1996). Among other things, it seems to be related to the context of its use. The earlier literature derided information management as mere data management (Mutch, 1996) in line with the thinking that data quality problems were purely technical nature in that they were restricted to data base elements. On the other hand, information was seen in a holistic dimension as data that was interpreted and reported as meta data and information glossaries and used at strategic levels by information managers and executive management.

Data has generally not been managed or subjected to the same management and quality control measures as other assets in an organisation (English, 2000). Studies relating to the poor quality of information include English (2000) and Huang (1999) in Xu (2009). The Gartner Group estimated that up to 25% of revenues in the financial industry are forfeited due to poor quality data. Dun and Bradstreet reported in 2007 that the cost of poor data quality to the U.S. economy may be $600b annually (Dun and Bradstreet, 2007).

2.3.3 Data Quality Dimensions

Loshin (2011) states that, in order to assess the impact of data quality on a business, it is necessary to classify both data quality expectations and business impact criteria. Data quality dimensions need to be understood in order to determine what the underlying root causes are. While there is no consensus on what the various dimensions should be, those that are cited frequently in the literature are discussed below.

<u>Accuracy</u>: Accuracy is defined in terms of correctness and reliability and also in terms of certified and even audited data (Wang and Strong, 1996). Accuracy of data describes how exact an event describes the real world. Data that has been peer-reviewed (Winningham, 2011) lends itself to higher accuracy. The quality processes involved in the capturing, sampling and storage of data (Ibid) can contribute to improved data accuracy. Winningham (2011) adds that, one of the foundations of research is that the results must be replicable. A researcher needs to come to a similar conclusion working from an identical set of data. Accuracy relates to the differences between an estimated and the true or unknown value. It is may be measured by two main sources of error, namely the sampling error and non-sampling error (Ibid).

<u>Completeness</u>: Completeness of data is measured against the various attribute criteria associated with that data. Hassany, Panahy, Sidi, Affendey, Jabar, Ibrahim and Mustapha (2013) relate completeness to the breadth, depth and scope of the data quality task at hand. Winningham (2011) notes that, while data may be complete it can be inaccurate but still meet the requirements of a stakeholder and this suggests that there are trade-offs between dimensions.

<u>Consistency</u>: Consistency requires agreement between data across the enterprise. Data in *silos* presents problems for reporting. The data attributes may contradict each other in terms of maintenance. While it may be consistent within

a narrow band of data or reporting, it may not be consistent across the organisation. Data may be different if it is reported in various instances. While data can be accurate (representing the real world), it may still be inconsistent. Furthermore, while it may be complete, it might also be inconsistent (Winningham, 2011).

Timeliness: Timeliness refers to how updated the data is in relation to the task it is applied to, or used for (Pipino, Lee and Wang, 2002). This is often evident in organisations' quarterly results that must be reported within a given timeframe (BipM, 2007). While timeliness may be related to user expectations, data may not always be 'timely' as organisations' financial statements are often published after the year-end.

Auditability: "Auditability implies that means that any transaction, report, accounting entry, bank statement etc. can be tracked to its originating transaction" (BiPM, 2007, p1).

Credibility: Credibility refers to the extent to which the source as well as the data content can be relied upon (Khatri and Brown, 2010).

Relevance: Relevance relates to the extent to which statistical information matches clients' real requirements (Wang and Wand, 1996).

Accessibility: Accessibility refers to the ability to easily obtain data from the providing source. It also relates to the ease with which the information can be verified as well as determining whether the format of the available information is appropriate. The importance of this indicator has diminished as data and information that was previously in hard copy format has largely been replaced by on-line information. The cost of deploying information may also have decreased commensurately. Accessibility has become important dimension for both IS professionals and data consumers as technology increasingly enables data and information to be presented in a variety of electronic formats. Accessibility also relates to secure access and a shared understanding of data by various stakeholders (Huang, Lee and Wang, 1999, Strong, Lee and Wang, 1997).

Interpretability: The advent and increase of computing power have made it easier for users to interpret statistical information *via* metadata (Winningham, 2011).

Methodological soundness: Methodological integrity refers to the implementation of international, national or peer-confirmed standards as well as guidelines and practices that are the foundation of statistical outputs. Such standards enable intra-national and international comparison of data and information.

Integrity: If integrity is not built into data, users cannot trust the data or the agency that produced it (Winningham, 2011).

The data quality dimension that deserves mention is the CIA (Confidentiality, Integrity, and Availability) model. The CIA dimension is, however more applicable as guide to policy for information security within an organization (Yan, Olariu & Weigle, 2009). While the dimension of cconfidentiality prevents sensitive information from reaching the wrong people, the dimension of integrity involves maintaining the consistency, accuracy, and trustworthiness of data over its entire life cycle. The dimension of availability in terms of information security relates to the technical or hardware support for the maintenance of data. Of the three CIA dimensions, however integrity is the most relevant for this study. Another approach also is the Parkerian Hexad consisting of the dimensions of confide-entiality, integrity, availability, authenticity, possession, and Utility (Dardick & Secau, 2010).

2.3.4 Other Data Quality Dimension Approaches

Other approaches to data quality dimensions have been advocated (Wang and Strong, 1996) and supported by research. De La Harpe and Marshall (2009) add an intrinsic and contextual as well as a representational approach to the dimensions already discussed.

- **The intrinsic dimension** relates to the quality of data ‘in its own right’ in the sense that believability and reputation require recognition. Intrinsic data quality refers to the dimensions of accuracy, objectivity, believability, reputation, pragmatism, usefulness and usability (Strong, Lee and Wang, 1997).

- **The contextual dimension** relates to data quality ‘within the context of the task at hand’. This dimension places the onus on data consumers to be aware of tasks and their contexts that change over time and to be vigilant in taking these into consideration when conducting research. This dimension relates to the relevance, timeliness, completeness and volume of information data and how much value can be added to it (Strong, Lee and Wang, 1997).

- **The representational dimension** relates to data quality in terms of how it is presented, delivered or disseminated. Representational data quality includes issues relating to how data is formatted and how concisely and consistently it is represented. Data needs to be easy to understand (format) and interpret (meaning). Representational data quality refers to the ease with which data can be interpreted and the ease of understanding, how concisely it can be interpreted and consistently represented (Strong, Lee and Wang, 1997).

The South African Statistical Quality Assessment Framework (SASQAF) draws a distinction between syntactic, semantic and pragmatic data quality dimensions as follows:

- **Syntactic data quality** focuses on the structure of symbols and emphasises the form of data rather than what it means. Syntactic data quality can be related to the dimension of consistency, where data values for particular data elements are represented by a consistent set of symbols, a consistent symbolic representation or a consistent coding taxonomy (Ballou and Pazer, 1995).

- **Semantic data quality** relates to what data means. It focuses on symbols that represent things in the real world. Semantic quality refers to the dimensions of completeness, accuracy, timeliness and currency (Wang, Strong, Guarascio, 1996).

- **Pragmatic data quality** has to do with the use of data, and how useful stakeholders find it in the execution of their work (Wang, Storey & Firth, 1995).

2.3.5 Trade-offs within Dimensions

Scannapieco, Paolo and Batini (2005) report that quality dimensions are not independent of one another, but that a correlation exists among them. While one dimension may be deemed more important than another, the choice of one may negatively affect the others. Establishing trade-offs among dimensions raises complex questions, e.g., a trade-off may need to be made between timeliness and either accuracy, completeness, or consistency. Providing accurate (complete or consistent) data may require time. Conversely, data may be timely, but accuracy may be compromised if it is required at very short notice. Scannapieco, Paolo and Batini (2005) add that consistency and completeness may be traded-off in the sense of either choosing to rely on too little data, but with a higher rate of consistency or having access to more data almost immediately, but with a lower rate of consistency. Opinions differ as to which data quality dimension is the most important. Data accuracy appears to be the most important (Olson, 2013), but relevance could also be regarded as more important; (Brackstone, 2001) states that the other dimensions are meaningless without relevance.

2.3.6 Dependencies within Dimensions – Relationships of Data Quality Dimensions to Data Quality Improvement

In Figure 2.7 below, Hassany et al. (2013) show how data quality dimensions are interdependent using the ACCTI (accuracy, completeness, consistency, timeliness, improvement process) framework.

The first three attributes (accuracy, completeness and consistency) are related to timeliness. Alternatively, all four dimensions can be related to data quality improvement. The authors used the ACCTI framework to model these dimensions and their effect on data quality improvement.

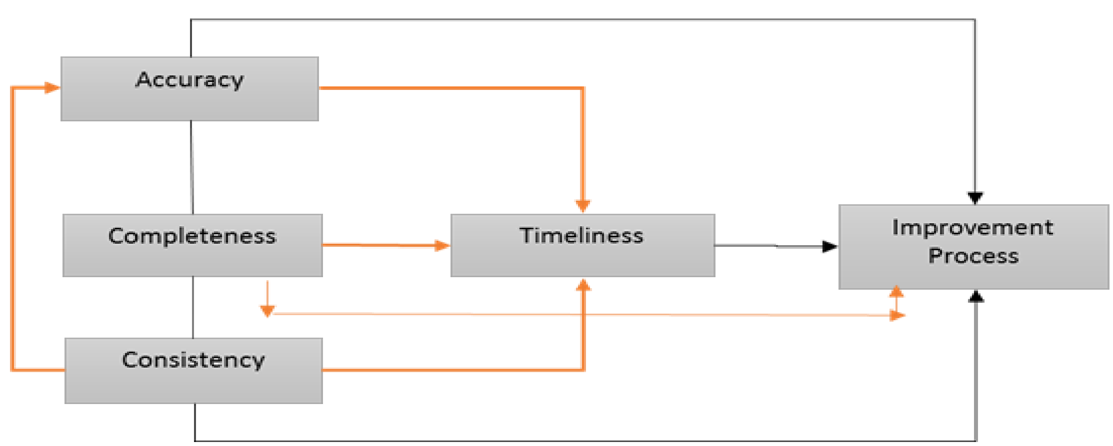


Figure 2.7: ACCTI framework

Source: Hassany, P Panahy, S Sidi, F Affendey, LS Jabar, MA Ibrahim, M Mustapha, A (2013), Discovering Dependencies among Data Quality Dimensions: A Validation of Instrument. *Journal of Applied Sciences*, 13: 95-102).

2.3.7 Causes of Data Quality

To improve data quality it is important to understand the issues and factors that give rise to data quality problems. Haug and ArlBjorn (2011) identify the lack of clearly identified roles and responsibilities, lack of data ownership, inefficient organisational procedures to address data quality, a lack of penalties or incentives and training and education shortcomings as potential barriers to data quality improvement. McKnight (2009) proposes the following ‘real life’ questions in relation to the causes of bad data quality:

- Source - was the data correctly captured or uploaded at the source?
- Process - was the integrity / quality of the information maintained as it was processed through the system (e.g., the student life cycle)?
- Usage - is the data being interpreted correctly?
- Ageing and rate of decay - to what extent is data quality compromised as data ‘age’ in terms of its currency or ‘shelf life’?
- Consistency - can data from disparate systems be reconciled to conform to the way an organisation would ideally want to view the data?

Lee, Pipino, Funk and Wang (2006) add the poor data quality can also be related to a lack of procedures and appropriate technologies. Maydanchik (2007) advocates a process approach to investigate the causes of bad data. This involves the following processes:

- processes that import data from the outside

- data conversions that may not have been done correctly, or importing problems from legacy systems
- system consolidations through mergers and acquisitions and the integration of different systems
- manual data capturing and mistakes *via* various forms and interfaces
- batch updates – the impact of changes in data 'downstream' are hard to determine
- real-time interfaces – while real time interfaces significantly improve IS efficiencies, the capture of data that 'triggers' through to many other systems occurs too rapidly for the data's reliability to be monitored
- processing causes data to decay
- processes that change or transmute data from within the organization, for example, changes in organisational structure (Maydanchik, 2007)

Various organisational processes related to changes to data, the currency of data, upgrades to software and hardware systems, the incorporation of new data, and the automation of processes as well as the loss of organisational skills can contribute to bad data. The causes may relate to deficiencies in the data quality management function itself and manifest through the following stages:

- data processing – are programmes regularly maintained? (poorly maintained programmes with unadjusted business rules will yield incorrect results)
- data cleansing – have data changes been effected to incorporate new technologies?
- data purging – has data intended for deletion or deactivation been purged? (data that is still erroneously current or relevant may cause confusion and slow down processes) (Maydanchik, 2007).

Furthermore, LaValle, Lesser, Shockley, Hopkins and Kruschwitz (2011) note that managerial and cultural factors, that are not always considered when evaluating data quality may impede the use of data to improve an organisation. This is illustrated in Figure 2.8 below.

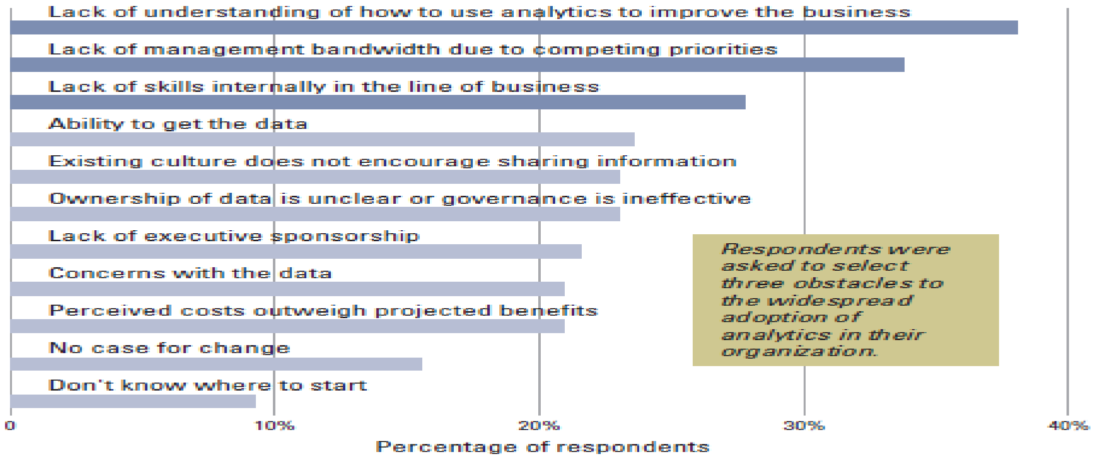


Figure 2.8: Barriers to Information Use

Source**:** LaValle, S Lesser, E Shockley, R Hopkins, MS Kruschwitz, N (2011) - Big Data, Analytics and Lee Y, Strong D (2003), Knowing why about data processes and data quality, *Journal of Management Information Systems,* Vol 20 No3, pp 13–39.

2.3.8 Impact of Data Quality

Slone (2006) has shown that poor data affects organisational performance and that the relationship between information quality and organisational performance can be measured. Marsh (2005) summarises the concerns and findings from several similar surveys:

- 88% of all data integration projects will fail completely or exceed their budgets
- 75% of organisations have calculated and identified the costs arising from bad data
- poor data has caused 33% of organisations to delay or cancel new IT initiatives

- according to Gartner in Marsh (2005), bad data are the number one cause of customer relationships management systems' failure

Loshin (2012) notes that bad data impacts the organisations' financial performance, confidence, productivity and risk or compliance. He proposes a business impact hierarchy where data is made more manageable by managing its impacts in smaller components.

Industry experts including Gartner Group, Price Waterhouse Coopers, and The Data Warehousing Institute note that, while data quality management is in crisis, senior stakeholders have been apathetic in resolving the problems (Marsh, 2005). Madnick and Lee (2009b) concur and add that more research is needed to assess the impact of data quality on individual firms as well as the national economy.

In terms of operational impacts, as early as 1998, Redman predicted that data quality would compromise customer satisfaction and lower employee satisfaction. The impact of poor data quality manifests in delayed decision-making and slow transmission of data from data warehouses. Organisations may also mistrust data sources.

The strategic impact of poor quality data is difficulty in determining and executing strategy. Issues around data ownership can divert attention from organisational objectives, culminating in:

- Increased IT work - scrapping and reworking with more resources devoted to labour and material
- Lost confidence in organisational business intelligence (decisions based on information and knowledge based on faulty data)
- Failure to comply with regulations (may have funding consequences)

- Lost business opportunities and lower profit (opportunity cost of missed ventures)
- Waste (scrapping and reworking)
- Short term remedies more costly (vicious circle; the higher the rate of organisational change, the more maintenance is required but organisations often resort to quick fixes – the momentum of the 'techtonic effect' (Winningham, 2011).

2.3.9 Positioning DQ Dimensions Relative to Data Quality improvement

Shanks and Darke (1998b) proposed a useful framework for understanding data quality in a data warehouse environment that includes both intrinsic and extrinsic data quality dimensions and is based on semiotic theory, the study of the use of signs and symbols to convey knowledge. An important feature of their framework is the separation of data quality goals from the means to achieve them. Other components of the framework include stakeholders, improvement strategies, measures, weightings and ratings. Four types of stakeholders are identified: data producers, data custodians, data consumers, and data managers (Strong, Lee & Wang, 1997, Wang, 1998). This framework is illustrated in Figure 2.9 below and the dimensions are discussed in the following section.

Figure 2.9: Connecting stakeholders to data quality dimensions

Source: Strong, DM, Lee, YW, Wang, RY (1997), Data quality in context, *Communication of the ACM,* Vol 40, Nov 5 , pp 104–108.

2.4 ACCOUNTABILITY AND DATA QUALITY ROLE PLAYERS (Objective 2)

This section discusses the issues pertaining to accountability and data quality role players.

2.4.1 Data Roles

This section focuses on data roles, types of data ownership, ownership models and the roles and perspectives of Data Custodians, Data Stewards and Data Owners in the organisation. This is complex as different functions are associated with similar role names, and roles are blurred, with stewards and custodians involved in similar activities and support for groups rather than individuals.

As noted earlier, the data role players are also associated with the 'locus of accountability for decision making' in the data quality domain. They set 'underlying standards with respect to various dimensions of data quality' as well as mechanisms to show business users how data can be used on a continuous basis; they also prescribe procedures to evaluate the quality of data. Parker, Stofberg, De La Harpe, Venter and Wills (2006) stress the role of people in the progress or 'flow' of data through the organisation. Data only has value when it fulfils purposes that are determined by the data stakeholders. They cite Rothenberg's (1996) observation that the quality of data should be determined and assessed early during the data production stage. Schwolow and Jungfalk (2009) concur and state that, the quality of data should be interrogated in all stages in its lifecycle; indeed, data quality interrogation should occur at source. Those who will be responsible for this process should therefore also be identified early in the lifecycle.

The following approaches have been advocated to data roles or a combination of these roles:

Information supplier - information manufacturer - information consumer approach

Wang (1998) notes that, information suppliers are those who are responsible for creating and collecting the data. Information manufacturers are charged with designing, developing or maintaining the data and information systems infrastructure, while information consumers are the people who apply the intellectual property in their work.

Data producer – Data custodian – Data consumer approach – Strong, Lee & Wang (1997) proposed three roles for those involved in data quality:

- Data producers are people or groups of people that generate the data and who are charged with functions associated with producing the data. They generate data according to specifications and they need to assess whether the data elements are valid and accurate and ensure that the data achieves the purposes for which it is created. They are charged with the scope of the data quality.

- Data custodians provide, marshall and manage the resources required to store, process and secure the data.

- Data consumers use and aggregate the data.

Strong, Lee & Wang (1997) added that the data custodian should have a broader conceptualisation of data quality (which is congruent with the nature of the role as envisaged in UKZN's Data Quality Policy).

Data collector – Data custodian - Data consumer approach - Lee and Strong (2003) maintain that the three major roles in most organisations should be data collectors, data custodians, and data consumers, with data collectors entrusted with initial capture or upload of data, data custodians to store and maintain the data and data consumers using the data to integrate and aggregate, present and interpret (affirmed by Zhu, Madnick, Lee, and Wang, 2012).

Data Manager – Xu, Nord, Nord and Lin (2003) added a fourth data role, data managers, within the data production cycle. They should manage data quality in the system. Mathieu and Khalil (1998) raise the role of process owners and advocate that they should be bear responsibility for the quality of data that the organisation produces. Different data roles might assign different priorities to data quality dimensions (De la Harpe and Roode, 2004).

The research literature emphasises the flow of data and various data roles. The stage of data in the data lifecycle is also important when quality data is required in an organisation. However, the current study focuses on data quality participants that are accountable rather than investigating the quality requirements within each of the lifecycle stages.

Hodkiewicz, Kelly, Sikorska and Gouws (2006) hold that many organisational and behavioural factors influence the quality of data. The people element in the 'people-process-technology' triad has to ensure that 'things happen'. They interrogate the data collection process and propose key performance indicators (KPIs) such as MTTF ('mean time to failure') and the MTTR ('mean time to repair') and the input factors that may affect the MTTF and MTTR at collection stage and also determine 'weak' links in the data collection process, as well as potential remedial initiatives.

2.4.1.1 Data Owners

Various definitions of data owners have been proposed. A data owner has been defined as a "role or group who is empowered to make decisions about how a data entity can be structured, manipulated, or used" http://www.datagovernance.com/glossary_d.html

A data owner is also seen as "the individual responsible for the policy and practice decisions of data. For business data, the individual may be called a business owner of the data" http://www.information-management.com/glossary/d.html

They are also "the individual responsible for the policy and practice decisions of data" http://www.aexis.eu/DataWarehouse-Glossary/
"A data owner is the ultimate accountable person within data governance, although data ownership can also be treated as a shared corporate responsibility" http://www.stibosystems.com/US/Resources/Glossary/B.aspx

Loshin (2001a) states that data ownership is a management issue and that there are complicated issues such as different views of the value of data, issues related to privacy or bureaucratic practices that make the management of data quality difficult. Discussing ownership paradigms, Loshin (Ibid) differentiates between data producers and data consumers but also proposes that whoever produces or generates the data should be the data owner. Some of Loshin's approaches to data ownership are listed below, often from the position of those that have the most interest in the data:

- when all the data that the organisation sourced may be assumed to be produced *within that organisation,* the organisation is the owner,

- where two parties are involved, where one *pays* for the production of the data and the other actually creates the data, the one who pays is the data owner,

- the party that can ‘unlock’ or ‘decode’ the data i.e. the ‘decoder’, will be the owner,

- those in whom the rights are vested such as personal privacy or image copyrights, will be the owner,

- the individual or organisation that buys or licenses data as a purchaser or licenser will be the owner,

- ultimately Loshin (2000) states that everybody in an organisation could be a data owner. If this philosophy was applied throughout the organisation, it would cascade into data quality efficiency at various levels of the organisation.

2.4.1.2 Data Stewards

A data steward may be an individual with data associated responsibilities as stipulated by a data governance or data stewardship programme. Data stewards may be categorised according to various titles such as Data Quality Stewards, Data Definition Stewards and Data Usage Stewards http://www.datagovernance.com/glossary_d.html

Data stewards are the people that implement the data standards and policies that their organisations have adopted. A Data steward may be the person entrusted to maintain a data definition on behalf of the owner of the data definition http://www.damauk.org/glossary.php

Loshin (2000, Ibid) draws a distinction between a data steward and a data custodian. He states that a data steward is responsible for maintaining and distributing the data as directed by a data custodian. A steward should be responsible for procedures that take care of the capture, storage, validation, correction, documentation and production of data from the operational area. Weber, Otto and Osterle (2009) cite Bitterer and Newman (2007) who define data stewards as those accountable for and committed to the 'collaborative business practices' required to manage data as an asset. Weber and Otto (Ibid) also refer to a Chief Steward as a 'master data coordinator' and 'director of data management' or 'data czar' (Dych´e and Levy, 2006). Dych´e and Levy (2006) suggest the need for 'detailed skill profiles' to connect and assign employees who are data workers to data associated roles.

Winningham (2011) states that a data steward should be accountable to implement and enforce data governance policies. Lucas (2011) describes a data steward as a business leader and / or an expert on a particular subject charged with accountability in the following areas:

- the identification of the requirements of a business intelligence system
- laying down the definition of business domain values, and data names in specific subject areas
- ensuring compliance with regulatory requirements and that the organisation adheres to the data policies and standards it has set itself
- analysing and improving data quality

Data stewards typically work with a data custodian, whose responsibility it is to store and move the data http://www.stibosystems.com/US/Resources/Glossary/B.aspx

However, Karel (2007) of Forrester Research recognises that the definition of a data steward has evolved and distinguishes between business and IT stewards. Business stewards should know the subject matter the best. They apply strategic and tactical priorities to the business and the data to support it. This requires a detailed grasp of the requirements of line users. The business steward communicates these needs to the technical data stewards and complies with decisions taken by the Executive member responsible. On the other hand, IT Stewards provide support to the data supply chain front line through all the conversion processes, the infrastructure supporting the data ‘in the middle’ and downstream to a warehouse and business intelligence application. They are also tasked with converting the technical specifications to be interpreted and applied by IT designers and developers.

2.4.1.3 Data Custodians

The literature supports both narrow and wider definitions of the role of data custodians. They also range from the technical to the business side of data quality. Loshin’s (2001b) approach can be considered a *‘narrow’* view in that he considers this role as having the “responsibility of operating systems, data centres, data warehouses, operational databases, and business operations in conformance with the policies and practices prescribed by the data owner” http://www.information-management.com/glossary/d.html

Data custodians are responsible for enforcing data standards. Loshin (2001b) notes that the data custodian is required to provide appropriate administrative data to the organisation in a reliable form that complies with established standards. Data custodians are also entrusted to take appropriate care of the data in the operational system and the definition is *widened* in that they would be directly involved in policy matters. Loshin (2001b, Ibid) acknowledges that the role of the data custodian also accommodates the business side.

Jesionowski (2012) provides for a *wider* interpretation of the role of a data custodian and suggests that they are directly involved in ensuring data quality by:

- collaborating with data stewards to interrogate data issues and inquiries – custodians know the rules for reporting

- resolving data issues, and collaborating on system changes – this is often required for statutory reporting; conversions have to be facilitated between the coding structures of the institution and the coding structures of the body that a submission is made to
- ensuring that data movements are documented and recorded in designated repositories
- assessing the quality of the data produced by the institutional reporting unit, and running data validations and reconciliation processes subsequent to the completion of data capture and corrections movement done by data stewards
- providing source data to support a data repository warehouse – this role is also connected to the coordination and management of service level agreements
- relaying appropriate and relevant issues governing the delivery or quality of source data to the repository team

(Jesionowski's repository team has various parallels with the Data Quality Working Group referred to in this research study).

2.4.2 Connecting Data Roles to Data Quality Dimensions

Giannoccaro, Shanks and Darke (1999) were some of the first scholars to link data stakeholders to the categories and dimensions of data quality (Figure 2.10 below).

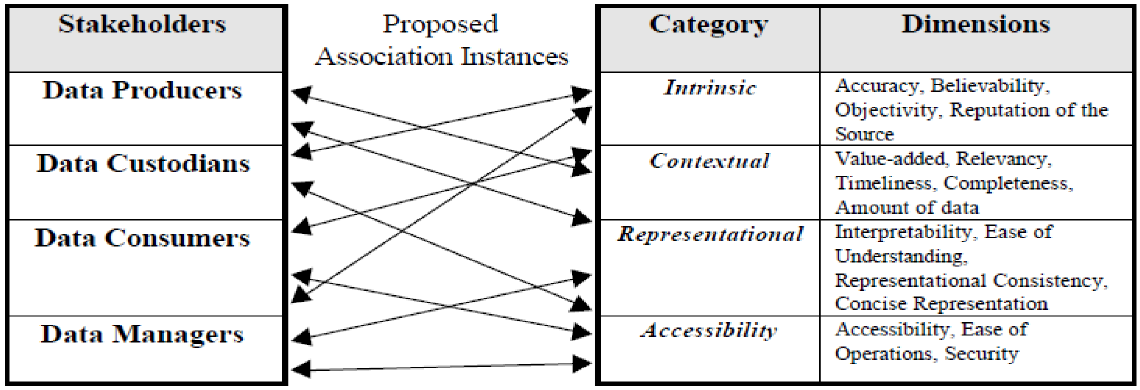


Figure 2.10: Associations between stakeholders and particular data quality dimensions

Source: Giannoccaro A, Shanks G, Darke, P (1999), Stakeholder Perceptions of Data Quality in a Data Warehouse Environment, Proc. 10th Australasian Conference on Information Systems).

Lee and Strong (2003) observe that it is important for data workers to possess the 'why knowledge' underlying the foundations of the data production process in order to contribute to the creation of quality data. They are of the opinion that the knowledge of data collectors is more important than that of data custodians.

2.4.3 Data Quality Board

Weber, Otto & Osterle (2009) advocate for a Data Quality Board that is similar to a business information team (English, 1999), a data governance council or a trustee council (Dych´e and Levy, 2006).

2.4.4 Dimensions of Common Responsibility

Jesionowski (2012) applies a 'team approach to data quality improvement in that data stewards and custodians should work together to apply organisational data management policies and standards. He notes that data stewards and data custodians can facilitate efficiency by collaborating in channeling requests for data access to relevant stewards involved, conceptualizing as well as specifying and developing audits in the areas of data governance or data quality.

2.4.5 RACI Charting in Allocating Roles in Data Quality

The RACI ‘matrix’ refers to a specific framework of data governance activity that consist of a matrix of roles and decision areas, encapsulating four central responsibilities that are usually applied to data governance projects in terms of parties being Responsible, Accountable, Consulted and Informed and defined as follows :

- Responsible – the responsible person is the person who performs an activity or does the work
- Accountable – the accountable person is the person who is ultimately accountable. This is however, not a simple concept. These are the people who identifies those who are responsible for a stated outcome. In many cases, however an accountable and responsible person may be the same individual (Costello, 2012).
- Consulted – the consulted person is the person that needs to both receive feedback and contribute to the activity
- Informed – the informed person indicates the person that needs to know of the decision or action (Grobler and Dlamini, 2010).

The RACI model provides for an opportunity to manage the flow of information relative to the duties and involvement of the stakeholders involved. Wende and Otto (2011) illustrates the position of stakeholders in data quality vis a vis the dimensions of data quality i.e. vis a vis accuracy, timeliness, completeness, comparability. Their model (Figure 2.11 below) is crafted around data quality roles that closely resembles the model implemented at the Institution, i e providing for a Sponsor, a Data Quality Board, a Chief Steward, a Business Data Steward and a Technical Data Steward.

	Exec Sponsor	Data Quality Committee	Chief Steward / (Custodian?)	Business Data Steward / (Owner?)	Technical Data Steward (ICT Resource?)
Plan Data Quality Initiatives	A	R	C	I	I
Establish a Data Quality Review Process	I	A	R	C	C
Define data producing processes		A	R	C	C
Define rules and responsibilities	A	R	C	I	I
Establish policies, procedures and standards for data quality	A	R	R	C	C
Create a business dictionary		A	C	C	R
Define information Systems Support		I	A	C	R
R = Responsible, A = Accountable, C=Consulted, I = Informed					

Figure 2.11 - 'RACI' Chart

Source : Wende, K Otto, B (2011), A contingency approach to Data Governance http://mitiq.mit.edu/iciq/PDF/A%20CONTINGENCY%20APPROACH%20TO%20 DATA%20GOVERNANCE.pdf

There are other approaches as well to analyse roles and activity in terms of data quality eg the COBIT framework (Barnier, 2012).

2.5 DATA QUALITY PRACTICE / SPECIFIC ISSUES

The following are relevant to sound data quality practice.

2.5.1 Currency of Data / Data Decay

In line with the trend of data driven decision making (DDDM) and the pressure for data to be accurate, timeous, current etc, can an institution afford the data decay associated with delays in data correction? In terms of data quality corrections, one of the survey questions (question 17) for this study was: 'In order to minimise the time that data remains incorrect, do you believe that a 'Data Correction Window Period' (e.g. 24 / 48 / 72 hours) should exist for regular data errors to be fixed?' According to the glossary of data compiled by the International

Association for Information and Data Quality (IAIDQ), data quality decay is the extent to which previously accurate data will become not accurate over time because the characteristic about the real world object will change without a corresponding update to the data applied (http://iaidq.org/main/glossary.shtml#D).

McGilvray (2010) refers to data decay as data erosion, describing it as negative changes to data and suggest methods to counteract this phenomenon. Loshin (2010) states that there may be an acceptable level of quality and that knowledge relating to data quality dimensions and data quality measures needs to be applied to determine acceptable thresholds for data quality. He suggests that the threshold for data quality tolerance should be related to the business impact of a data problem. The tolerance levels would be built into a system of data audits that examine and report on data quality in real time. An institution should monitor errors and their impact over a period of time in order to determine acceptable quality levels.

Loshin proposes Service Level Agreements (SLAs) on data quality; his model is illustrated in Figure 2.12 below. While normally associated with issues related to system availability and service turnaround, he is of the opinion that SLAs could play a significant role in data quality improvement.

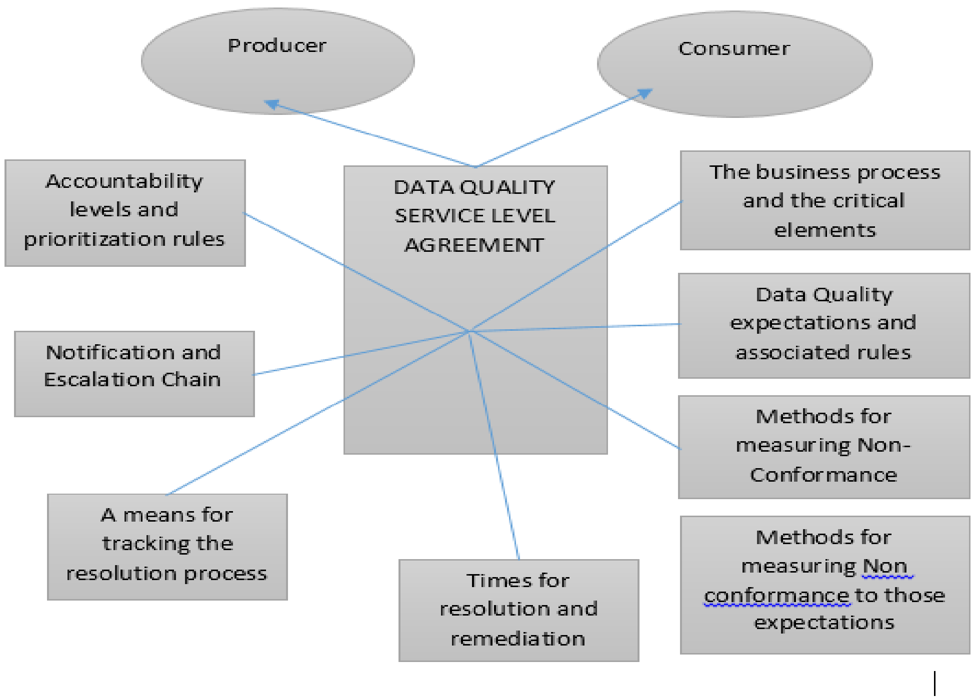


Figure 2.12: Model of a Service Level Agreement within a Data Quality Environment

Source: Loshin, D (2011), The Practitioners' Guide to Data Quality Improvement, Book Online http://books.google.co.za/books?id=B3zd4GCAWeYCandq=SLA#v=onepageandqandf=false

While SLAs may relate to time intervals to correct data, they may also relate to third parties that are responsible for providing data to the systems. Several scholars have noted that it is important to determine data quality problems at the entry points in the data or information lifecycle as data quality may already be compromised before being assimilated into the host system. In this respect, an analogy can be drawn between Michael Porter's (1985) Value Chain Model (Figure 2.13) and Schwolow and Jungfalk's (2009) Information Value Chain Model (Figure 2.14), where Porter's idea is adapted to a data / information environment.

Figure 2.13: The Value Chain Model

Source: Porter, ME (1985), Competitive Advantage, Free Press, New York.

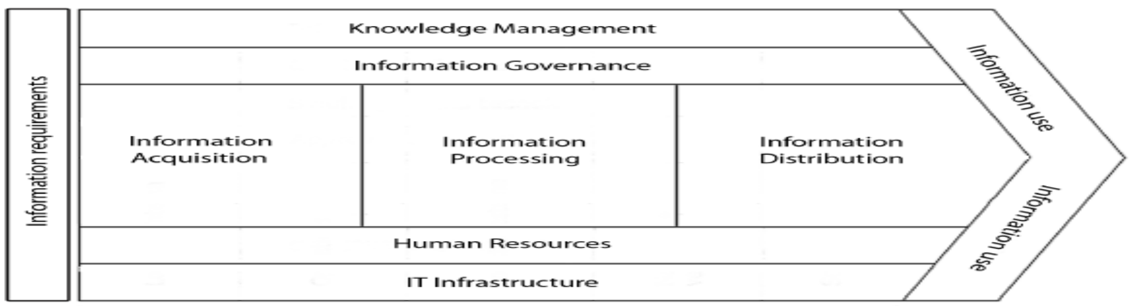


Figure 2.14: The Information Value Chain Model

Source; Schwolow, S and Jungfalk, M (2009), The Information Value Chain: Strategic Information Management for Competitive Advantage, Copenhagen Business School.

http://www.scribd.com/doc/46315789/Information-Value-Chain

While recognising aspects of Michael Porter's value chain in Schwolow and Jungfalk's model, the different phases of information management in the latter model allow for data actors to be involved in data quality improvement activities in each of the three components of information acquisition, information processing and information distribution. As poor data manifest at various stages of the data information life cycle, it may be critical to implement SLAs at Porter's 'inbound logistics' stage and Schwolow and Jungfalk's 'information requirements' stage to address problems at the source.

Loshin (2011) maintains that controls are required to measure if SLAs have been met with appropriate metrics to set off alerts. Dealing with data at the source will underscore that an organisation is focusing on the ‘right thing’.

2.5.2 Performance

In terms of data quality and performance, one of the survey questions (question 18) posed was: ‘In terms of accountability for data and data quality success, do you believe that data quality responsibilities should be included in data owners’ key performance indicators (KPIs)?’ Although the literature has not established a direct link between data quality and performance measurement (it has only been implied *via* SLAs), some studies relate data quality to organisational performance. Anturaniemi (2012) reports that targets and metrics related to data quality are related to organisational performance and that, KPIs as referred to in Figure 2.17 should be used to monitor this performance. Slone’s 2006 study showed that the relationship between data quality and organisational outcomes can be systematically measured and that measuring data quality can be used to predict organisational outcomes. Masayna, Koronios and Gao (2009) distinguish between corporate KPIs (at a strategic level), business KPIs (at a tactical level) and operational KPIs (at an operational level).

Masayna, Koronios, Gao and Gendron (2007) cite KPIs as measures that determine how well business processes are performing in terms of their potential to achieve a particular goal (illustrated in Figure 2.15 below). Masayna (2006) notes that, since KPIs are dependent on data to support KPI metrics, if one relates this to data quality dimensions, it is not practically possible to assure data quality for every dimension (referring to trade-offs). This suggests that one may have to accept satisfactory scores on some of the selected dimensions only. Very little subsequent research has been conducted on the relationship between data quality and KPIs.

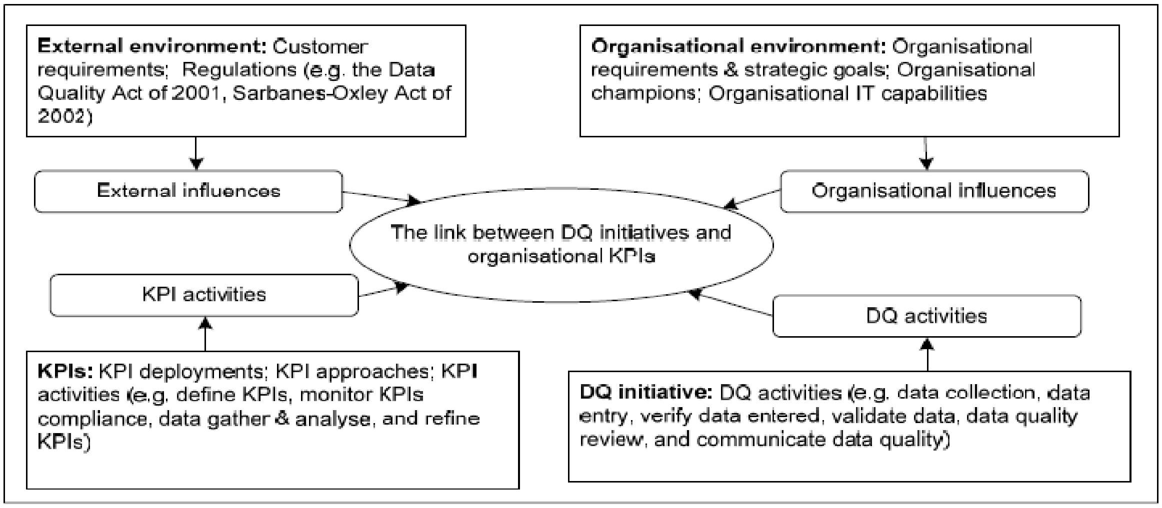


Figure 2.15: A proposed model to determine the links between Data Quality initiatives and organisational KPIs

Source: Masayna, Koronios, Gao, Gendron, (2007), Data Quality and KPIs: A link to be established, paper presented at the Second World Congress on Engineering Asset Management and the Fourth International Conference on Condition Monitoring (WCEAM-CM2007), Harrogate, United Kingdom, 11-14 June.

2.5.3 Quality of Processes

The quality of business processes also impact data quality.

Survey question 15 asked: ‘Do you feel that the operational processes in your areas of work support /underpin work with respect to Data Quality, are robust and of adequate quality ?’

Masayana and Koronios (2011) observe that the efficiency and effectiveness of business processes are two defining criteria of organisational success. Operational processes in users’ areas of work should support or underpin data quality. Business Impact Analysis is one of the processes that guide analysts. It notes that, ‘any potential data-related issues may increase costs, reduce revenues, impact margins or introduce inefficiencies or delays in business activities'. In the CobiT framework, KPIs are defined as measures to gauge the progress of business processes towards a particular goal or objective.

Madnick, Wang, Lee and Zhu (2009) note that as new technologies affect how data is managed and assure its quality, organisations may change processes faster than their systems are updated, with a resultant 'lag' in the system to support the new processes. Heravizadeh, Mendling and Rosemann (2009, p8) explored the quality of business processes. They refer to "the quality of functions, quality of input and output processes, quality of non-human resources, and quality of human resources" from which one may infer that many a data quality antecedent can be traced to a business process.

2.5.4 Access to Reports and Data Quality

According to the Agency for HealthCare Research and Quality (2010), access to well-produced reports can promote a good understanding of the dimensions of data quality. Consumer or user reports contribute to data quality in various ways:

- the dimensions of data quality are better understood
- quality improvement is stimulated among providers as they perceive that performance data can affect their market share or their position in the organisation
- reports that affect providers' or users' 'public image' may be used as a lever to improve the data to safeguard their reputation

(http://www.ahrq.gov/legacy/qual/pubrptguide2.htm)

2.6 DATA QUALITY SUSTAINABILITY

Pipino, Lee and Wang (2002) consider approaches that can be deployed to assess an organisation’s data quality or prospects and sustainability for data quality improvement.

2.6.1 The Comparative Approach, first implemented by Pipino, Lee and Wang (2002) involves the use of data quality surveys and quantifiable data quality metrics. This approach compares the data collected from surveys (perceptions of each class of stakeholder) and the results of the quantitative metrics. The comparisons are used to diagnose and prioritise key areas for improvement. This is referred to as the diagnostic approach due to its diagnostic nature and in order to distinguish it from an alternative comparative approach.

2.6.2 An Alternative Comparative Approach uses aggregated results of data quality surveys to analyse and prioritise key areas of improvement. It includes gap analysis and benchmark analysis. Here the comparisons are not between two techniques. Making use of the survey technique, comparisons are made across stakeholders (collector, custodian and consumer) against an industry standard. This approach falls within the broader category of the AIMQ methodology (Pipino, Lee and Wang, 2002).

Lwanga Yonke, Walenta and Talburt (2011) note that in order to sustain data quality, the Information/Data Quality Professional should assume the role of an internal consultant who will facilitate increasing knowledge and understanding of the data among business customers, and continuously monitor and report on levels of data quality, ensuring that quality is maintained as well utilising it as a feedback mechanism to foster awareness. They should work with project teams in system development lifecycles, promoting information on quality best practices and ensuing that these are included in all information technology development and support processes.

Research on data quality management (involving the ‘hard’ or IT responsibility or involvement systems approach) accelerated from 2000 to 2010. In terms of initiatives to promote a data and information culture or information literacy (question 24) Zu, Fredendall and Robbins (2006) state that (relating to sustainability), organisational culture could be an explanatory variable for the extent that an organisation effectively executes its quality practices. They suggest that organisational culture can be significantly influenced by quality management and different quality practices.

With regard to role of change in sustainability (question 28), and managing data quality in the face of organisational change, Loshin (2010) states that during the process of data quality improvement, policies, processes, and procedures that govern the programme must be able to continue even in the face of change. This applies to situations involving downsizing or re-organisation, outsourcing or IT staff changes.

The role of training in quality (question 26) is so important that it relates to two of the 14 Points of Quality proposed by Edwards Deming (1996), as cited by Masoumeh, Basri and Sadegh (2011).

Turning to the question of how well teams or networks function in terms of sustainability (question 27) Brennan, Bosch, Buchan and Green (2013) state that cross-functional teams is critical to diagnosing process based problems in the area of quality and describe their involvement in developing and testing process improvement as an important element in continuous quality improvement. They also allude to teams overcoming obstacles to cross functional team work such as boundaries in the professional realm and status differences that hinder collaboration (Ibid).

2.6.3 Policy and Strategy – Way Forward

While a data quality policy is a 'directive' or a set of guiding rules intended to ensure that desired practice is implemented, it may need to be complemented by a strategy. A strategy is an operational 'roadmap' where resources are mobilised to achieve an organisation's objectives. According to Dravis (2004), strategy is the implementation of a series of tactical steps, a practical set of procedures relating to what has to be done, the implementation of an action plan to achieve an organisation's goals.

While no formal data governance framework exists at UKZN, sufficient elements appear to have been extracted from the informal application of data governance at the Institution to implement a data quality policy.

2.7 DATA QUALITY COSTS

The impact of poor data quality and the potential of quantifying data quality improvement cannot be fully understood without a brief discussion of costs. JM Juran was the first scholar to address the different forms of waste. Since then, the identification and reduction of waste has become one of the core activities of quality management. Juran (1989) in Dahlgaard and Dahlgaard-Park (2006), referred to this as the cost of poor quality (COPQ).

Dahlgaard and Dahlgaard-Park (Ibid) distinguish between two definitions of data quality costs citing Juran (1951 and 1989). He distinguished between Quality costs – the costs which would disappear if no defects were apparent (1951) and the COPQ – as the sum of all costs that would disappear if there were no quality problems (1989).

The costs associated with data quality issues may include:

- the costs associated with poor data quality, that is process costs caused by data errors and opportunity costs due to lost and missed revenues, as well as indirect costs.

- the costs of assessment and improvement activities, also referred as direct costs.

The costs of poor quality can be classified as follows (English 1999):

- process costs, such as the costs associated with the re-execution of the whole process due to data errors
- opportunity costs due to lost and missed revenues.

O' Riain and Helfert's (2005) proposal to model data quality costs is presented in Figure 2.16 below:

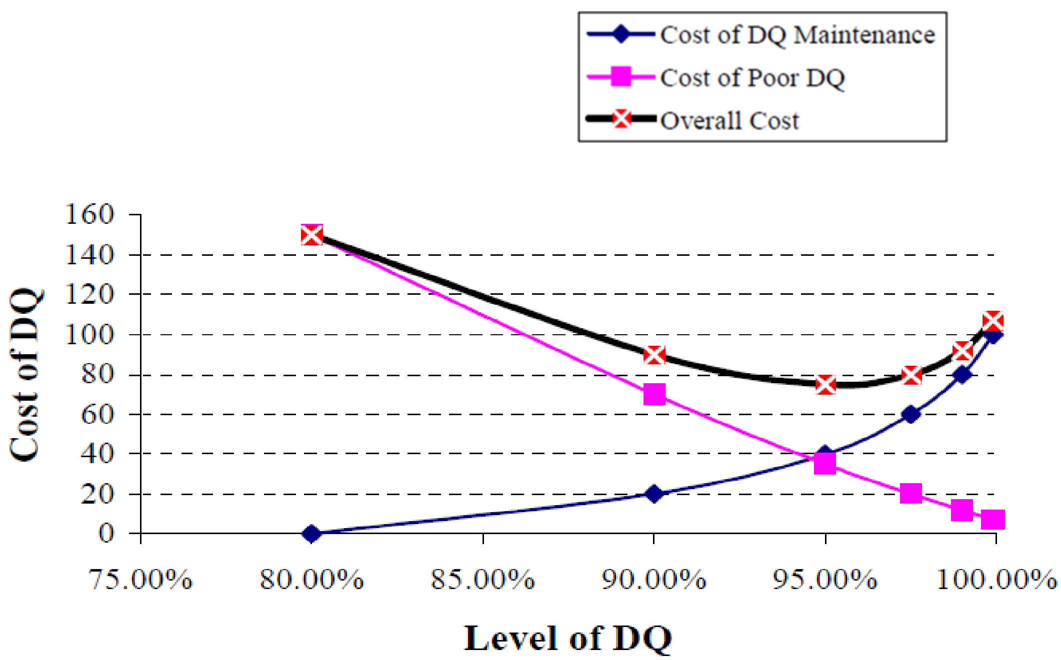


Figure 2.16: Theoretical cost/benefit analysis for investment in DQ assurance in HIS

Source: O' Riain, C and Helfert, M (2005), An Evaluation of Data Quality related problem patterns in health care information systems, School of Computing, Dublin City University, Ireland.

This study will focus only on time ('hours'), namely labour spent on data quality activity. The objective is to obtain an estimate of cost savings through a data quality improvement exercise.

From a practitioner point of view, O'Neal (2012) found that 50% of the time spent by a data quality team is devoted to reconciling data that are very often from disparate systems. The author proposed a calculation of productivity cost, based on the hours spent on reconciliations weighted by the full cost of the staff (Ibid). The cost of accessing data and the cost of project delays can be added. This research study adopts a formula to determine the labour costs in terms of hours spent on data quality activity.

A limitation of the study is that, as a wide variety of factors giving rise to data quality was examined, the scope of the study only allowed for the calculation of labour cost of data quality activity. There are other costs that could also have been incorporated.

2.8 UNIVERSITY OF KWAZULU-NATAL - SITUATIONAL OVERVIEW

The data situation at UKZN is discussed below.

2.8.1 Institutional Data Roles

The positions of those involved in data quality at the Institution are:

- The data stewards (technical) belong to the area Innovation and Development within the Information and Communication Services Division (ICS).
- The data custodians belong to the area of Institutional Intelligence within ICS.
- The data stewards (business) are support staff from the Colleges and Student Academic Affairs who share ownership of the data on the main administrative system (ITS).

The ICS divisional structure consists of sections devoted to User Support, Student Computing, Improvement and Development of Systems, and Networks and Communications as well as Information and Innovation, which includes Institutional Intelligence where the institutional data custodian resides.

2.8.2 Roles and Functions

A list has been compiled of the three information user groups and their functions *vis a vi*s data quality as outlined in the Institutional Data Quality Policy (Data Quality Principles and Guidelines, 2011). For the purpose of the study, the functions of the Data Stewards who are involved in technical issues (DST) have been separated from those involved in the functional business area (DSB).

Data Steward (Technical) (DST):

- Drives data governance initiatives
- Reviews data quality scoreboard
- Reviews / approves validation /business rules
- Designs data cleansing strategy
- Approves user access to systems (in collaboration with DSB)
- Ensures user and system security
- Ensures maintenance / storage / disposal of data
- Manages information system projects

- Configures information system applications

Data Steward (Business) (DSB) / also Referred to as Data Owner or Designate of Data Owner:

- Collaborates with DST to ensure quality
- Designs cleansing strategy with Data Custodian
- Approves user details to access the system
- Advises Data Custodian of data problems
- Cleans data at the root / source
- Advises DC DST of system problems, enabling data quality issues to be addressed at source
- Data owner of functional area

Data Custodian (DC):

- Conceptualises validation / business rules
- Conceptualises / maintains data quality scoreboard
- Assesses data quality / undertakes data profiling
- Assesses data quality impact
- Designs data quality audit system
- Identifies users on the system (in collaboration with DSB)
- Designs cleansing strategy
- Maintenance of metadata / glossary of information
- Data Quality Awareness workshops / campaigns
- Develops Data Quality KPIs
- Designs Data quality scoreboard

2.8.3 Roles and Functions 'Augmented' for the Purposes of this Research Study from the Institutional Data Quality Policy

In terms of the Institutional Data Quality Policy, the Data Custodian is entrusted with coordinating the Data Quality initiative. In support of this function, there has to be synergy between this role and other IS-User roles (below) in order for data quality improvement to be feasible. The 'Principles and Guidelines' (Data Quality Principles and Guidelines, 2011), that are part of the Data Quality Policy define the roles as follows:

- the *Data Owner* is a Senior Manager who is ultimately responsible for the data in a functional area.

- the *Data Stewards* are the source of the data with the best knowledge of the data and the strongest incentive or motivation to take care of it. They will manage the data on behalf of the data owners.

- the *Data Custodian* refers to Management Information (MI) / Institutional Intelligence (II) in ICS that is charged with the overall responsibility of coordinating data quality improvement in all functional areas.

The Custodians are entrusted with:

- Promoting a culture of data quality across the Institution
- Enhancing data quality in the Institution
- Monitoring and measuring functional data
- Implementing systems to identify and correct data errors
- Implementing systems to measure and monitor data quality

The Stewards are entrusted with:

- Managing data on behalf of Data Owners

- Auditing and correcting data
- Identifying data quality issues and reporting them to Data Owners
- Assisting with the conceptualisation and implementation of data audit systems
- Providing regular reports to the Data Owners
- Formulating procedure manuals for the capturing, storing and maintenance of data in the relevant transactional databases

For the purpose of this research, the notion of Data Steward (as suggested by Karel, 2007) has been differentiated into Data Steward Technical (‘DST’) and Data Steward Business (‘DSB’) / also known as the Data Owner. This forms the basis for the differentiated approach that will unfold as the survey data is analysed and the perspectives of each of these roles are obtained.

2.9 CONCLUSION

While the literature review on data quality in this research study may seem very detailed, the issues around accountability and the multidimensionality of the problem in terms of its causes and impacts cannot be overstated. The cost of data quality has become a popular barometer in industry to quantify its impact. The RACI approach was mentioned but can only fully be utilized after further study. As a study of data quality is multi-facetted, other approaches may ‘delve’ deeper than the approach adopted here. One such approach is the D M A I C (Define, Manage, Analyse, Improve, Control) method that is a Sig Sigma approach to data quality improvement. This approach consist of steps that have to be executed in a specific order. One of the main deliverables of a D M A I C approach is a process map at a very high level that may also be used as a basis to craft data improvement at tactical and operational levels.

Chapter Three outlines the framework for this research study.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the research methodology used in this study. It outlines the research approach and design, data collection methodology and mechanisms, sample or population used, issues relating to reliability and validity and ethical considerations.

3.2 Research Approach

A quantitative approach was selected as this approach lends itself to statistical analysis and access to data that can be measured. As quantitative data is expressed in numbers, it is easily expressed in tables and graphs and also facilitates the interpretation of the data (WordPress, 2011). This approach is also relevant to this particular study of data quality, as a benchmark for data quality improvement will be set by calculating the cost of data quality.

3.3 Research Design

Case study has been used as the research design in this study. Yin (1984, p 23) defines the case study research method as an “empirical inquiry that investigates a contemporary phenomenon within its real-life context, when the boundaries between phenomenon and context are not clearly evident and in which multiple sources of evidence are used”. Yin (1989) proposed six types of data collection for case studies, namely, documentation, archival records, interviews, direct observations, observations by participants and physical artefacts.

A case study is appropriate as it contributes to an understanding of a complex issue and can augment experience or add value to what is already known through previous research (Wang, Strong & Guarascio, 1996, Tull, 1997, Wende & Otto, 2007). One of the disadvantages of using a case study is that it may be biased towards verification, i.e., support the tendency to confirm a researcher's own presumptions and allow general theories to develop from the basis of a specific case study.

3.4 Research Process: Academic and Work

This research study was practitioner-based and, as is the case with applied research, occupied "an area situated between academia-led theoretical pursuits and research-informed practice" (Furlong and Oancea, 2005, p1). Furlong and Oancea (Ibid) and Groundwater-Smith and Mockler (2006) note that, in practice-based research, a researcher involved in practitioner inquiry is bound to make use of 'theoretical' as well as 'practical' knowledge to 'move seamlessly between the two'. In line with Tull's (1997) model, the knowledge base for this study rested on two approaches:

- Work-based - as the topic is an industry-wide problem and the researcher is a practitioner, he would have extensive experiential resources to draw on.
- Academic-based - The growth in information technology (IT) has led to a rapid increase in the literature (academic and practitioner-based) in the research area, especially *via* the Internet.

The **work-based approach** produces rich personal information due to the researcher's extensive knowledge of the research area. The researcher's membership of the Institutional Data Quality Working Group (DQWG) provides a close connection to problems in this field of work. With permission from the appropriate institutional authority, the researcher was able to access UKZN's IS-users to collect data.

The **academic approach** contributes to the research study in that the research topic is a well-documented problem that is well-represented in the literature. The broad and multidimensional nature of the data quality problem was illustrated by the bibliographical literature count for this study that was higher than anticipated. A hundred and forty-nine sources were counted and were classified as follows:

- Journal articles - 39%
- Conference / Symposium proceedings - 13%
- Practitioner-based sources - 19%
- Dissertations - 5%
- Other (books, reports etc.) - 24%

The questionnaire design and administration are discussed below.

3.4.1 Questionnaire

A questionnaire was used for data collection as this is a quick method of collecting standardised data. The questionnaire consisted of 33 questions divided into five sections (Appendix 1).

Part 1 - Biographical details and use of systems
Part 2 - Data Quality Awareness (including Data Quality Impact, Causes)
Part 3 - Data Quality Practice
Part 4 – Accountability
Part 5 – Sustainability of Data Quality Improvement

The questionnaire was adapted from the one used in an MBA by Peter Tull (1997) who sourced it from Vaughan Merlin of Omega Point Consulting based on his presentation entitled, “Developing A Data Quality Culture” (1October 1997).

The questionnaire is generic in nature and interrogates users’ perceptions of the use and experience of data quality in University systems, ranging from business (‘soft’) to technical (‘hard’) perspectives. Practitioners have learned that it is not the hard technical issues that stymie an organisation's data quality efforts, but rather the soft, organisational, political and social issues. The structuring of the questionnaire into five parts provided a balance between nominal data describing categories of staff and systems and data quality experience, and ordinal data, enabling opinions to be ranked, and providing interval data to estimate the cost of time spent on data quality activity. Most of the questions could be answered by completing a ‘check’ box, while some allowed for an explanation if the ‘No’ or ‘Other’ option was selected. The questions in the section pertaining to the sustainability of data quality improvements were based on a four point scale.

3.4.2 Questionnaire Distribution - Survey Monkey

The Survey Monkey software was used to make the questionnaire available to users. It lent itself to the easy capture and collection of data for both the interviewees and the researcher. It also supported data collection after the first appeal by tracking and following up on non-responses. The Survey Monkey software stores data in a standard format and makes it available for export to either Excel or SPSS for analysis.

3.4.3 Pilot Study

A pilot study was undertaken was undertaken with 11 users for Part 5 of the questionnaire in July 2012. For practical purposes, the questionnaire items in Part 3 (Practice) and Part 4 (Accountability) were rearranged after the pilot.

The 11 users were chosen in a stratified manner to be representative of the three user groups. The sections were structured for the purpose of analysis as:

Section A - General
Section B - Data Quality Awareness
Section C – Accountability and Management / Practice
Section D - Cost of Data Quality
Section E – Sustainability

3.5 Sampling and Population

Flyvbjerg (2004) notes, that, a representative or random sample might not be the most informative. As one of the precepts of the case study approach is to attempt to collect as much information as possible, a decision was made to target the population of information system users. This comprises all those directly involved in the data quality process in terms of data correction, identifying problems and assessing the impact, as well as the technical users who assist the business users (to correct the problems). Including all the stakeholders who are directly involved in the data quality process means that the numbers in each of the categories are representative.

The survey list included 120 people comprising all the data custodians (8) and data stewards (technical) (22) who work in the Information and Communication Services (ICS) Division, as well as data stewards (business) (90) in the Colleges and Student Academic Affairs sourced from lists of operational workers closely related to data maintenance and correction provided by College and School Managers. The users span the student and staff systems.

3.6 Alternative approaches

Data quality is defined as data that is ‘fit for use by data consumers’. Wang and Strong (1996, p7) define a data quality dimension as a “set of data quality attributes that represent a single aspect or construct of data quality”. Three approaches are used in the literature to study data quality: (1) intuitive, (2) theoretical and (3) empirical (Wang and Strong, Ibid).

An intuitive approach is adopted when the selection of data quality attributes for a particular study is based on the researcher’s experience or intuitive understanding of which attributes are important. One of the key attributes is ‘accuracy’ and in financial literature ‘reliability’ is often used. In information systems, information quality and user satisfaction are two major criteria for evaluating the success of information systems. Wang and Strong (ibid) held that these two dimensions incorporate several data quality attributes, including accuracy, timeliness, precision, reliability, currency, completeness, and relevance.

The theoretical approach is based on data deficiencies occurring during the data production process. During the ‘manufacturing’ of a data element, the process often impacts on its path ‘into the system’ and detracts from its ‘value added’ potential during processing by the system. This research study adopted an empirical approach to enable the voice of consumers or stakeholders to be heard. It was envisaged that this research study could produce the following outcomes:

- Factors found to facilitate the smooth adoption of the University’s data quality policy could enhance the possibility of the policy’s success and sustainability.

- The study's findings could augment the learning experience of the researcher as well as information systems (IS) users, while promoting broad organisational learning at the Institution.

Lee's (2004, p97) study found that "experienced practitioners solve data quality problems by reflecting on and explicating knowledge about contexts embedded in, or missing from, data. Specifically, these individuals investigate how data problems are framed, analysed, and resolved throughout the entire information discourse. Their discourse on contexts of data, therefore, connects otherwise separately managed data processes, that is, collection, storage, and use".

3.7 Summary of Research

Table 3.1: Relationship of the research objectives to the questionnaire, and the literature references

Section	Objectives	Questionnaire	Literature Review
Section A	Awareness	Q 5 to 11, Q 20 to Q 23	Section 2.2.1, 2.2.3,2.2.4, 2.2.7, 2.2.8
Section B	Accountability	Q 17 to Q 19	Section 2.1.6, 2.3.1 to 2.3.3, 2.4.1,2.4.2,2.5.2
Section C	Problem Handling / Management / Practice	Q 13, Q 14, Q15, Q 16	Section 2.4.3, 2.5.3
Section D	Data Quality Costs	Q 14	Section 2.6
Section E	Sustainability	Q 24 to Q 33	Section 2.4.4

3.8 Data Collection Process, Tools and Analysis

The research instrument was circulated during June and July 2013, via the Survey Monkey software and individually followed up to ensure a maximum response (60 responses were received by 14 July 2013), resulting in a response rate of 50%. Since a response rate of 20% is considered the norm, and Baruch & Holtom (2008) cite a mean response rate of 38% for surveys conducted *via* the Web, the response rate for this survey is considered statistically reasonable. A degree of non-response bias was expected as data quality is a sensitive topic; the level of non-response that was encountered may reflect some employees' fear that their views might not be treated in confidence as promised, or that they would reflect negatively on the functional area they represent. Some 20 responses were received after the first round. This was followed up by a second submission *via* the software and by individual e-mails. This resulted in the response rate increasing to 45%. A week later, the response set was closed and finalised at 50%.

The graphic capabilities of the Survey Monkey software and Excel as well as the statistical components of Microsoft Excel and IBM SPSS were used to analyse the data.

A cross tabular analysis of the results was conducted. An analysis of time devoted to data quality activity was undertaken to obtain data quality costs as a benchmark for future data quality improvements. Differences in means were assessed to illustrate different perspectives on the sustainability of data quality activity.

The data to determine the cost of data quality activities in terms of person-hours were obtained though access to institutional records, and from data on the ranks, grades and mid-point salary structures of the individuals that participated in the survey.

3.9 Reliability and Validity

"Validity and reliability are two fundamental elements in the evaluation of a measurement instrument" (Takavol and Dennick, 2011, p 53). Reliability reflects that an instrument measures in a consistent way. The reliability of an instrument is closely associated with its validity. A research instrument needs to be reliable in order to be valid. Cronbach's alpha was developed as a measure of internal consistency. Internal consistency refers to the extent that all the items in a test measure the same concept or construct.

From a total of 60 responses, 45 completed responses were received for Part 5 of the questionnaire. A Cronbach's alpha coefficient of .8 was obtained on the basis of a measure to support the internal consistency of this section. This was an improvement on the coefficient of .61 achieved during the pilot survey that included 11 users.

Non-parametric measures (frequency counts) were used and expressed as integers and percentages of the population. The ordinal data captured (using a four point scale) in Part 5 was summarised as a parametric measure in terms of mean values.

3.10 Ethical Considerations

Diener and Crandall (1978) in Bryman & Bell (2007) note that researchers need to be mindful of ethical considerations, including causing harm to participants, a lack of informed consent, invasion of privacy and the use of deception. Interviewees were assured of the privacy and confidentiality of the opinions expressed. This was particularly important in terms of interviewees not having to fear that their opinions would be divulged at School and College levels. They were informed that their perspectives would be analysed at the 'collective', i.e., institutional level and that demarcation would only be applied at the level of the three different information system user groups.

As users participated in the survey online *via* the use of Survey Monkey, their consent to take part in the study was implicitly assumed.

3.11 Conclusion

What differentiates this research study from other research on data quality is the collection of detailed data from the perspective of the different groups of information systems users at the Institution. Warner and Burke and Argyris in French and Bell (1999) noted that the interactions and inter-relationships between individuals or groups or different groups in a system may 'elevate' the system in terms of its problems. While the interrelationships between groups are not the primary focus of this study, the differences in the perspectives of the information users may point to the need for further research with a strong systems focus.

CHAPTER FOUR

DATA ANALYSIS

4.1 Introduction

This chapter presents 1) an analysis and interpretation of the survey data involving descriptive statistics; 2) the estimation of the cost of correcting bad data; and 3) a statistical perspective on the sustainability of data quality improvement *vis a vis* the data custodians and data stewards. Data from 60 out of a total of 120 IS Users were analysed using the data analysis features of the Survey Monkey software and the SPSS v 13 software. Descriptive statistics were used for sections A to C, while for section D, the cost of data quality activities were calculated in terms of person-hours through access to institutional records and from data on the ranks, grades and mid-point salary structures of the IS Users that participated in the survey.

4.2 BIOGRAPHICAL DETAILS

This section summarises the biographical details of the surveyed sample i.e., level in the organisation, general and specific systems involved in and years of systems experience.

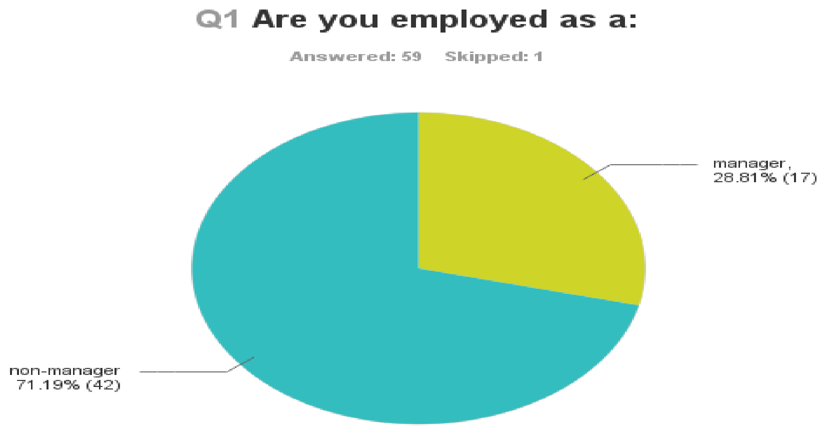


Figure 4.1 - USER PROFILE: MANAGERS AND NON-MANAGERS

Table 4.1 - USER PROFILE: MANAGERS AND NON-MANAGERS

Are you employed as a: * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	manager,	Count	1	14	2	17
		% within IS User	20.0%	31.1%	22.2%	28.8%
	non-manager	Count	4	31	7	42
		% within IS User	80.0%	68.9%	77.8%	71.2%
Total		Count	5	45	9	59
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.1 above shows that 28.8% of the 60 respondents are managers, with one IS User not answering this question. Fourteen of these managers are located in Colleges and Schools and are directly involved in capturing data. This is a significant number, as provides a sound indication of managerial perceptions of the research issues, particularly among Data Stewards (Business).

These individuals oversee the data quality activities in their functional areas by monitoring data quality to ensure that their functional areas are represented at data quality workshops. Fewer managers who are Data Custodians (only one) and Data Stewards (Technical) responded to the survey, as fewer staff are employed in these positions at the Institution.

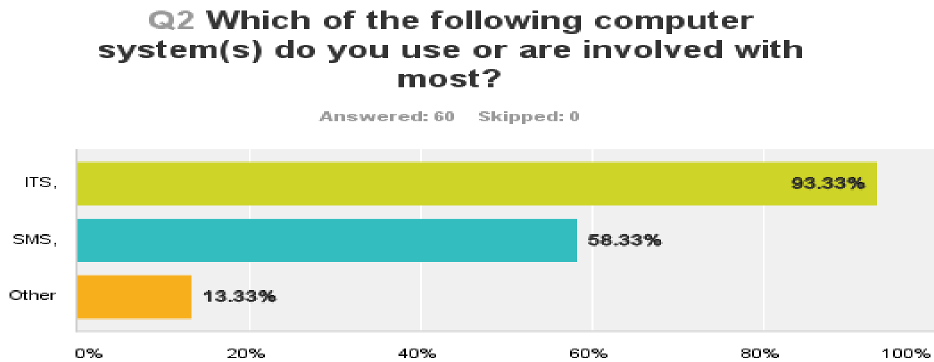


Figure 4.2 - USER PROFILE: USE OF SYSTEMS

Table 4.2 - USER PROFILE: USE OF SYSTEMS

		IS User			
		Data Custodian	Data Steward - Business	Data Steward – Technical	Total
ITS	Count	5	43	8	56
	% within IS User	50.0%	53.8%	47.1%	52.3%
SMS	Count	1	30	4	35
	% within IS User	10.0%	37.5%	23.5%	32.7%
Other	Count	4	7	5	16
	% within IS User	40.0%	8.8%	29.4%	15.0%
Total	Count	10	80	17	107
	% within IS User	100.0%	100.0%	100.0%	100.0%

Many of the survey participants would have had exposure to the ITS system as the main administrative system that was dominant at the institutions that merged to form UKZN (the University of Durban-Westville used the ITS administrative system from 1990 and the University of Natal from 2000). While other very relevant systems such as the SMS (Student Management System) have linkages to ITS; the main data quality activity takes place on ITS.

Table 4.2 shows that Data Custodians only report 50% involvement in ITS (as a transactional system) while other systems (40%), would involve information dissemination systems. As the Data Stewards (Technical) serve a larger constituency and are involved in the maintenance of peripheral systems (to ITS), the 29% response rate is not surprising. Overlaps between the use of ITS and SMS are expected as those with access to ITS will often access SMS as well, due to the SMS to ITS 'refresh' mechanisms.

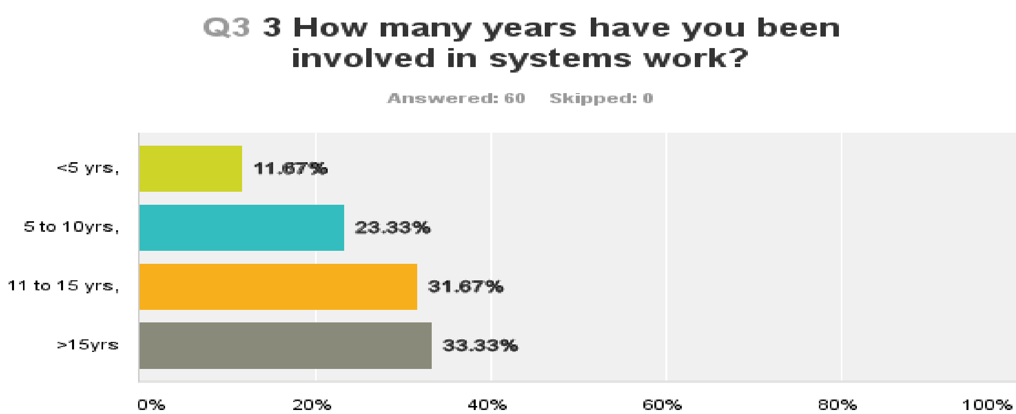


Figure 4.3 - USER PROFILE: YEARS INVOLVEMENT IN SYSTEMS WORK

Table 4.3 - USER PROFILE: YEARS INVOLVEMENT IN SYSTEMS WORK

3 How many years have you been involved in systems work? * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	<5 yrs,	Count	0	6	1	7
		% within IS User	0.0%	13.3%	11.1%	11.7%
	5 to 10yrs,	Count	2	12	0	14
		% within IS User	33.3%	26.7%	0.0%	23.3%
	11 to 15 yrs,	Count	3	15	1	19
		% within IS User	50.0%	33.3%	11.1%	31.7%
	>15yrs	Count	1	12	7	20
		% within IS User	16.7%	26.7%	77.8%	33.3%
Total		Count	6	45	9	60
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.3 indicates that 33.3% of the respondents have 15 or more years' exposure to information systems. Just over two-thirds of the respondents from managerial ranks have experience of 15 years and longer. This is a significant number, as it provides a sound indication of managerial perceptions of the research issues. The respondents in this category are Data Stewards (Technical). This is a reflection of the low turnover of ITS staff at the former Universities of Durban-Westville and Natal as well as UKZN. Mergers and software system upgrades proved beneficial to the Institution in terms of continuity, particularly with regard to the retention of institutional knowledge that is reflected in data quality improvements. Seventy-seven percent of the Data Stewards (Technical) who participated in the survey have more than 15 years' experience, followed by Data Stewards (Business) (26.7%) and Data Custodians (16.7%). Furthermore, 66.7% of the Data Custodians have more than ten years' experience, followed by Data Steward (Business) (60%) and Data Stewards (Technical) (89%).

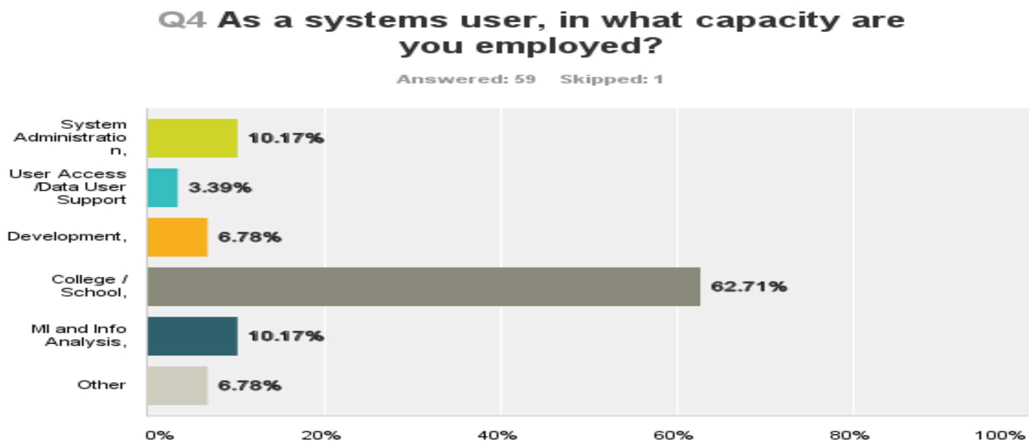


Figure 4.4 - USER PROFILE: CAPACITY IN SYSTEMS WORK

Table 4.4 - USER PROFILE: CAPACITY IN SYSTEMS WORK

As a systems user, in what capacity are you employed? * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	System Administration,	Count	0	4	2	6
		% within IS User	0.0%	9.1%	22.2%	10.2%
	User Access /Data User Support	Count	0	2	0	2
		% within IS User	0.0%	4.5%	0.0%	3.4%
	Development,	Count	0	0	4	4
		% within IS User	0.0%	0.0%	44.4%	6.8%
	College / School,	Count	0	37	0	37
		% within IS User	0.0%	84.1%	0.0%	62.7%
	MI and Info Analysis,	Count	6	0	0	6
		% within IS User	100.0%	0.0%	0.0%	10.2%
	Other	Count	0	1	3	4
		% within IS User	0.0%	2.3%	33.3%	6.8%
Total		Count	6	44	9	59
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.4 shows the categories in which the users are employed. Data Custodians' core function is data quality, i.e. MI and Info Analysis (100%), while 84% of the Data Stewards' (Business) activities are associated with Colleges and Schools.

Data Stewards (Technical) are involved in more diversified areas i.e., system administration (22.2%), system development (44.4%) and other systems (33.3%).

4.3 AWARENESS AND COMMUNICATION

This section summarises the results relating to communication of data quality, the importance of data quality in terms of data quality dimensions, knowledge of the state of data quality and the causes and impact of poor data quality.

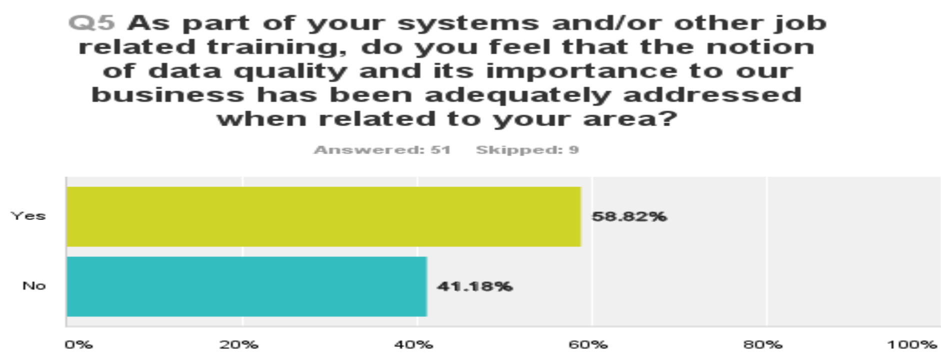


Figure 4.5 - ADEQUACY OF COMMUNICATION ON DATA QUALITY

Table 4.5 - ADEQUACY OF COMMUNICATION ON DATA QUALITY

As part of your systems and/or other job related training, do you feel that the notion of data quality and its importance to our business has been adequately addressed when related to your area? * IS User Crosstabulation

			IS User			Total
			Data Custodian	Data Steward - Business	Data Steward - Technical	
As part of your systems and/or other job related training, do you feel that the notion of data quality and its importance to our business has been adequately addressed when related to your area?	Yes	Count	1	26	3	30
		% within IS User	16.7%	70.3%	37.5%	58.8%
	No	Count	5	11	5	21
		% within IS User	83.3%	29.7%	62.5%	41.2%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

A series of workshops were held over a year with Colleges and School as part of UKZN's Data Quality Awareness campaign. Table 4.5 shows that only 16% of the Data Custodians felt that the notion of data quality had been adequately addressed. This is surprising, given their role of coordinating the data quality process. This result suggests that there are communication problems that needs to be addressed. The high level of positive responses (70.3%) from the Data Stewards (Business) reflects their exposure to the Data Quality Workshops that conveyed and interrogated various aspects of the Data Quality policy. The low percentage for Data Stewards (Technical) (37.5%) may reflect the fact that they were not exposed to the Data Quality Workshops. Management felt that they could not all attend as they were 'on call'.

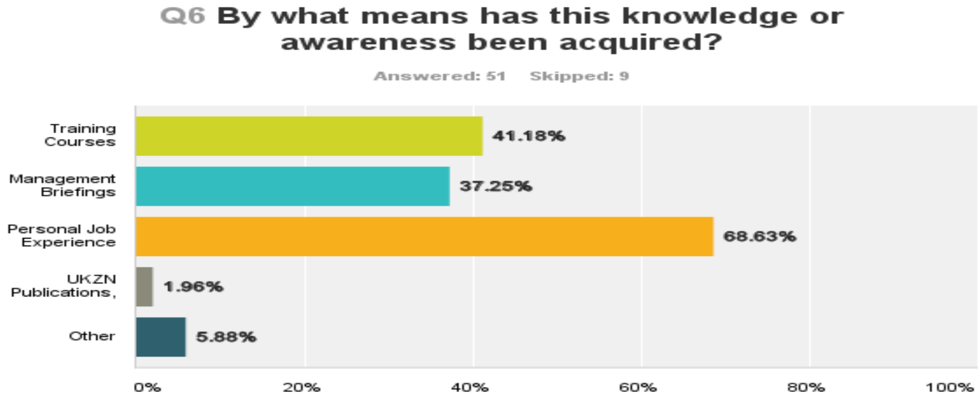


Figure 4.6 - MEANS BY WHICH DATA QUALITY AWARENESS WAS ACQUIRED

Table 4.6 - MEANS BY WHICH DATA QUALITY AWARENESS WAS ACQUIRED

		IS User			
		Data Custodian	Data Steward - Business	Data Steward - Technical	Total
Mgt Briefings	Count	1	15	3	19
	% within IS Staff	14.3%	25.0%	25.0%	24.1%
Training Courses	Count	0	19	2	21
	% within IS Staff	0.0%	31.7%	16.7%	26.6%
Personal Job Experience	Count	5	24	6	35
	% within IS Staff	71.4%	40.0%	50.0%	44.3%
UKZN Publ and Other	Count	1	2	1	4
	% within IS Staff	14.3%	3.3%	8.3%	5.1%
Total	Count	7	60	12	79
	% within IS Staff	100.00%	100.00%	100.00%	100.00%

Table 4.6 above indicates that 69% (34 out of 49 distinct IS Users – the total of 79 includes more than one response per user) reported that they gained awareness of data quality through personal job experience, 43% through training and 37% through management communication. Moreover, 44.4% of all IS Users (duplicated responses) and 71.4% of the Data Custodians stated that they learnt about data quality through job experience.

Only 26% of all respondents stated that they gained awareness of data quality through training. This suggests that there is a need for more training to address data quality problems. It is also possible that the training is ineffective and needs to be improved.

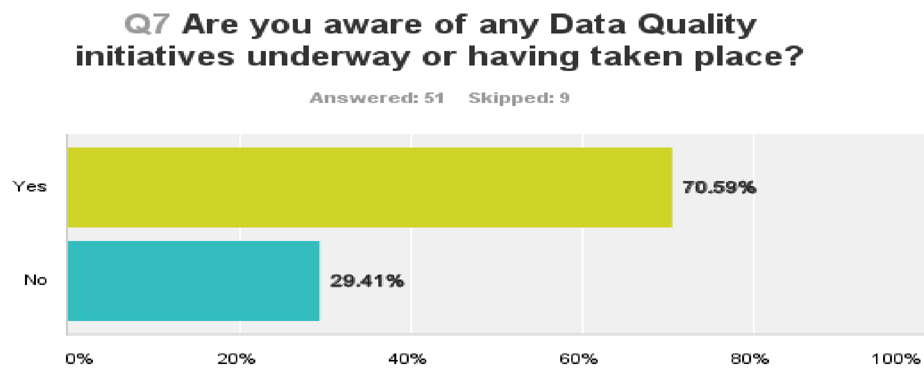


Figure 4.7 - AWARENESS OF DATA QUALITY INITIATIVES TAKING PLACE

Table 4.7 - AWARENESS OF DATA QUALITY INITIATIVES TAKING PLACE

Are you aware of any Data Quality initiatives underway or having taken place? * IS User Crosstabulation

			IS User			Total
			Data Custodian	Data Steward - Business	Data Steward - Technical	
Are you aware of any Data Quality initiatives underway or having taken place?	Yes	Count	5	25	6	36
		% within IS User	83.3%	67.6%	75.0%	70.6%
	No	Count	1	12	2	15
		% within IS User	16.7%	32.4%	25.0%	29.4%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.7 above shows that 83.3% of the Data Custodians stated that they are aware of data quality activity, reflecting the significant role these employees play in terms of 'driving' data quality initiatives. The corresponding statistics for Data Stewards (Business) and Data Stewards (Technical) are 67% and 75%, respectively. It is a fact that not all of the data stewards (business) attended the

workshops. The reason for the non-attendance not clear. The awareness of data quality activity may not have 'trickled down' to everybody. The fact that a reasonable percentage of data stewards (technical) (25%) were not aware, can be understood against the background of their 'overall' non-exposure.

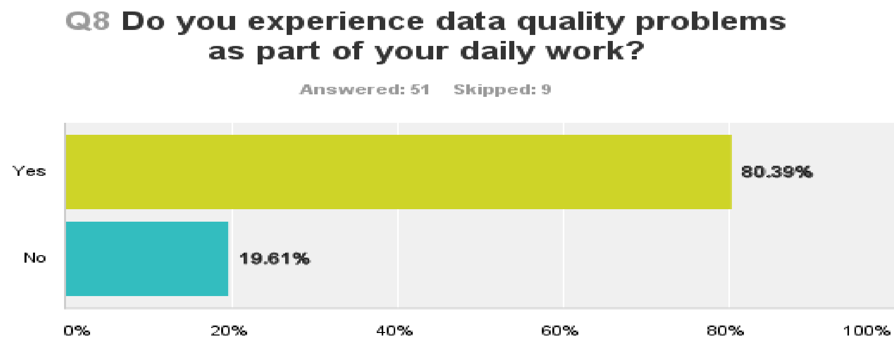


Figure 4.8 - EXTENT OF DATA QUALITY EXPERIENCE

Table 4.8 - EXTENT OF DATA QUALITY EXPERIENCE

Do you experience data quality problems as part of your daily work? * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Yes	Count	6	27	8	41
		% within IS User	100.0%	73.0%	100.0%	80.4%
	No	Count	0	10	0	10
		% within IS User	0.0%	27.0%	0.0%	19.6%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

A significant number of IS Users (80.4%) reported data quality problems as part of their work, which should contribute to a heightened level of awareness.

All the Data Custodians (100%) reported that they experience data quality problems as part of their daily work. This is expected; MI / II as institutional reporter would be well acquainted with the challenge of reporting amidst data quality problems.

Seventy-three percent of the Data Stewards (Business) stated that they experience data quality problems as part of their work. This reflects the extent to which data quality problems surface at the operational, detailed level where Data Stewards (Business) function. The corresponding statistics for Data Stewards (Technical) is 100%. Once again, this is expected. Due to the technical nature of their work, this group of employees finds it difficult to perform procedures on incomplete or inaccurate data.

Amongst the dimensions of data quality below, which are in your opinion the most important to a data user. Please can you rank them (eg 1,4,3,2)

Figure 4.9 - RATING OF DATA QUALITY DIMENSIONS

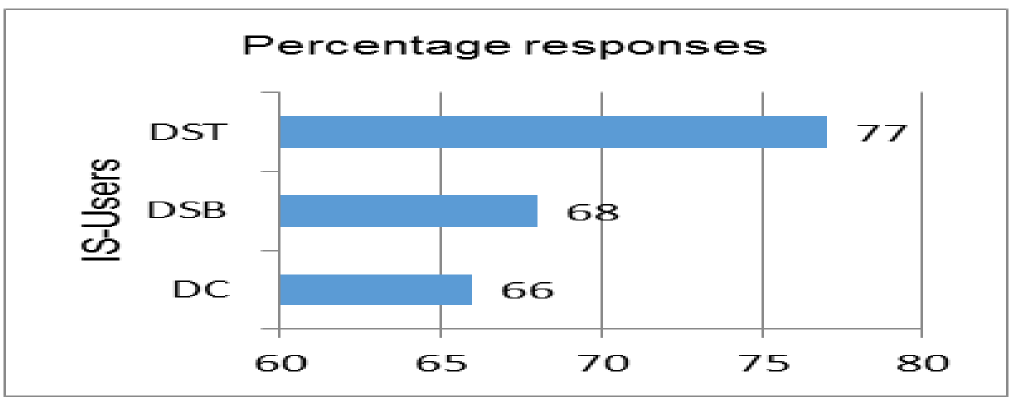


Table 4.9 - RATING OF DATA QUALITY DIMENSIONS

IS-User	DC		DSB		DST
Order of preference	Total	Order of preference	Total	Order of preference	Total
1234	1	1112	1	1234	2
1324	3	1234	4	1324	5
2314	1	1324	11	2431	1
4444	1	1342	1	3421	1
Grand Total	6	1423	5	Grand Total	9
		2314	2		
		4444	2		
		1424	1		
		1	2		
		2134	2		
		2413	1		
		4321	1		
		4332	1		
		4443	1		
		Grand Total	35		

Keys : Accuracy - 1, Completeness - 2, Timeousness - 3, Comparability - 4

Table 4.9 provides an overview of the data dimensions that are most important to an IS User. IS Users were asked to rank their preferences in order, amongst the data quality dimensions of Accuracy (1), Timeousness (2), Completeness (3) and Comparability (4). In order of preference, a coding structure of, e.g., '1234' reveals a preference for Accuracy, followed by Completeness as the most important data dimensions, among the six Data Custodians, 35 Data Stewards (Business) and nine Data Stewards (Technical) who responded. Data Accuracy was placed first by 4 out of 6 DC's (66%), 24 out of 35 DSB's (68%) and 7 out of 9 DST's (77%). There appears to be a significant trade-off between Accuracy and Timeousness amongst the Data Stewards (Business).

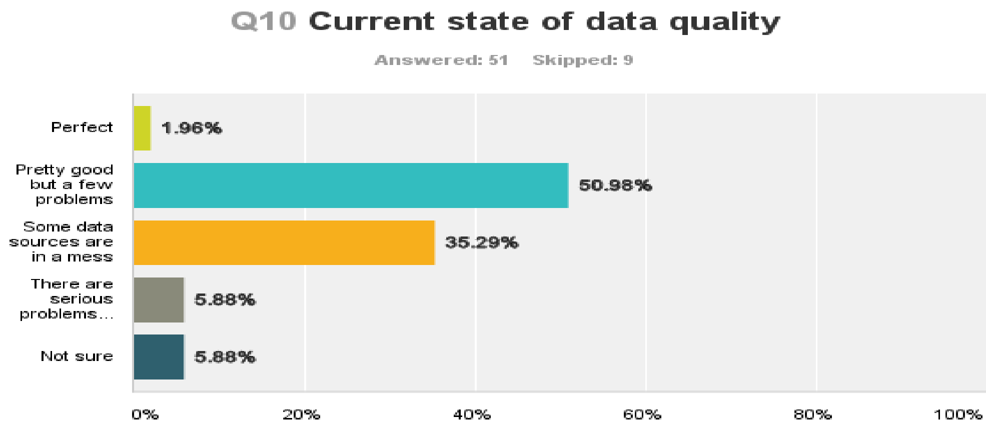


Figure 4.10 - THE CURRENT STATE OF DATA QUALITY

Table 4.10 - THE CURRENT STATE OF DATA QUALITY

Current state of data quality * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Perfect	Count	0	1	0	1
		% within IS User	0.0%	2.7%	0.0%	2.0%
	Pretty good but a few problems	Count	1	23	2	26
		% within IS User	16.7%	62.2%	25.0%	51.0%
	Some data sources are in a mess	Count	2	11	5	18
		% within IS User	33.3%	29.7%	62.5%	35.3%
	There are serious problems overall	Count	2	1	0	3
		% within IS User	33.3%	2.7%	0.0%	5.9%
	Not sure	Count	1	1	1	3
		% within IS User	16.7%	2.7%	12.5%	5.9%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.10 shows that 51 % of the IS Users feel that that the data is good but for a few problems. This statistic is significantly weighted by the Data Stewards' (Business) opinions (62.2%). While 30% of the Data Stewards (Business) deem the data to be 'in a mess', the majority are more optimistic. This may be related to their involvement in a recent (annual) iterative series of cycles of data cleaning; they may thus feel that progress has been made. While this bodes well for their commitment to the data quality initiative, their opinion may be based on the data quality initiatives presented to them thus far. The 'whole picture', e.g.,

the data audits that are still to be designed and enlarging the scope of the audit, are a further process in the data quality initiative. The majority of the Data Stewards (Technical) (62.5%) felt that the data is ‘in a mess’; this is likely influenced by the fact that they have not been always involved in the data cleaning activity undertaken by the Data Custodians and Data Stewards (Business). However, the majority of the Data Custodians either felt that the data is ‘in a mess’ or that there are serious problems overall.

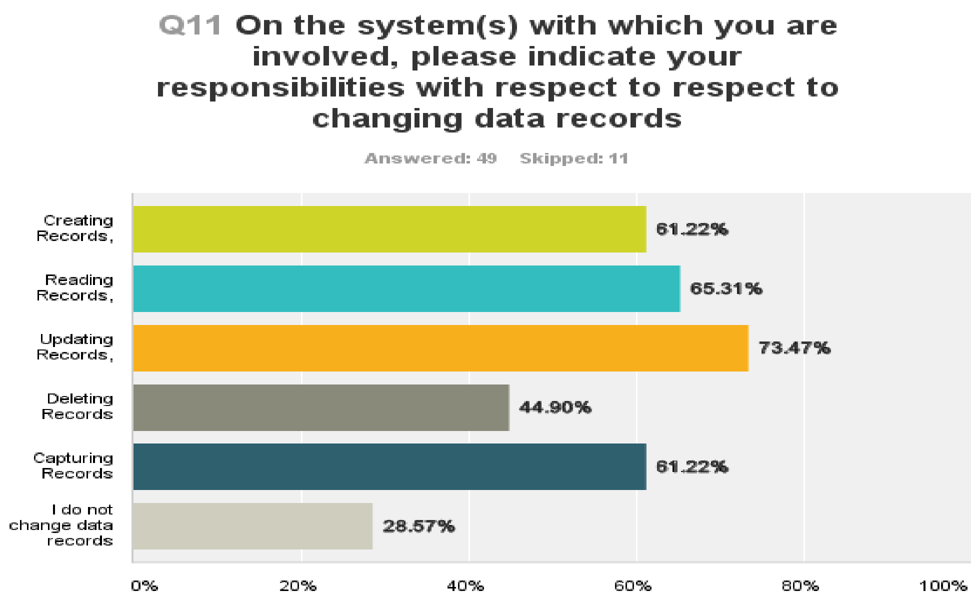


Figure 4.11 - RESPONSIBILITIES IN TERMS OF CHANGES TO RECORDS

Table 4.11 - RESPONSIBILITIES IN TERMS OF CHANGES TO RECORDS

	IS User			
	Data Custodian	Data Steward - Business	Data Steward - Technical	Total
Creating Records	3	24	3	30
Reading Records	2	26	4	32
Updating Records	3	29	4	36
Deleting Records	2	17	3	22
Capturing Records	2	27	1	30
I do not change records	3	7	4	14
	15	130	19	164

Table 4.11 provides an overview of the IS Users' data management activity. The responses illustrate that the Data Stewards (Business) are at the 'front-end' of the system. Data Custodians' activity it is low, as they do not have the authority to create or delete records. This is commissioned by the Data Stewards (Business) as data owners and technical issues are attended to by the Data Stewards (Technical).

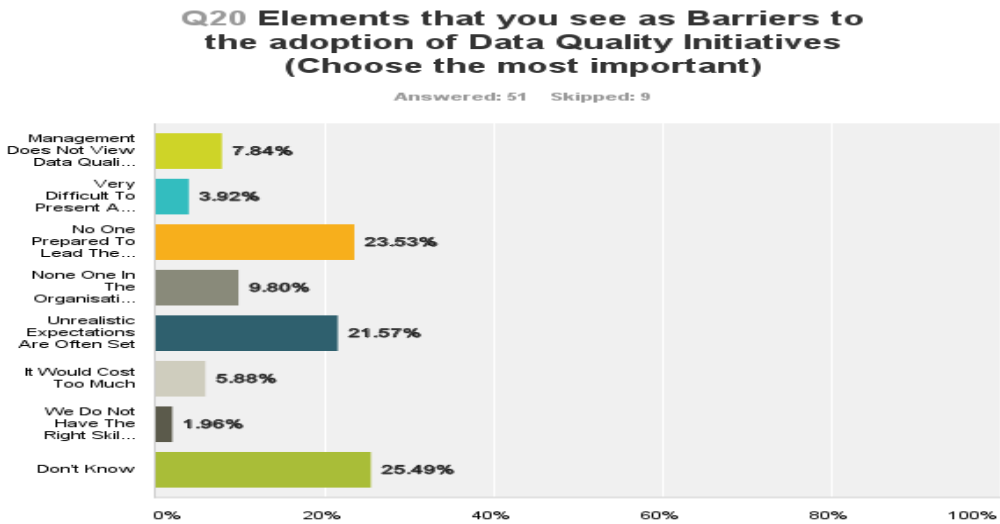


Figure 4.12 – BARRIERS TO THE ADOPTION OF DATA QUALITY INITIATIVES

Table 4.12 – BARRIERS TO THE ADOPTION OF DATA QUALITY INITIATIVES

Elements that you see as Barriers to the adoption of Data Quality Initiatives (Choose the most important) * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Management Does Not View Data Quality As Important	Count	2	1	1	4
		% within IS User	33.3%	2.7%	12.5%	7.8%
	Very Difficult To Present A Business Case	Count	0	1	1	2
		% within IS User	0.0%	2.7%	12.5%	3.9%
	No One Prepared To Lead The Initiative	Count	0	8	4	12
		% within IS User	0.0%	21.6%	50.0%	23.5%
	None One In The Organisation Appears To Care	Count	0	4	1	5
		% within IS User	0.0%	10.8%	12.5%	9.8%
	Unrealistic Expectations Are Often Set	Count	0	11	0	11
		% within IS User	0.0%	29.7%	0.0%	21.6%
	It Would Cost Too Much	Count	1	2	0	3
		% within IS User	16.7%	5.4%	0.0%	5.9%
	We Do Not Have The Right Skill Sets	Count	0	1	0	1
		% within IS User	0.0%	2.7%	0.0%	2.0%
	Don't Know	Count	3	9	1	13
		% within IS User	50.0%	24.3%	12.5%	25.5%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.12 shows, that, 50% of the Data Stewards (Technical) felt that the quality of leadership of data quality initiatives is a significant barrier to the adoption of these initiatives. This may be due to the fact that the Data Stewards (Technical) are not directly represented on the Data Quality Working Group (DQWG); their involvement is only operationalised at DQWG Sub-group level. Some data Stewards (Business) (29.7%) felt that unrealistic expectations are often set. This might be because they have to address data quality issues among other competing priorities.

Data Custodians felt that management at various levels should be more committed to data quality (33.3%). They also stated that the costs of fixing data quality are too high (16%) possibly relating to the costs of fixing software to prevent data capturing errors as well as the cost of developing data quality detection and management software.

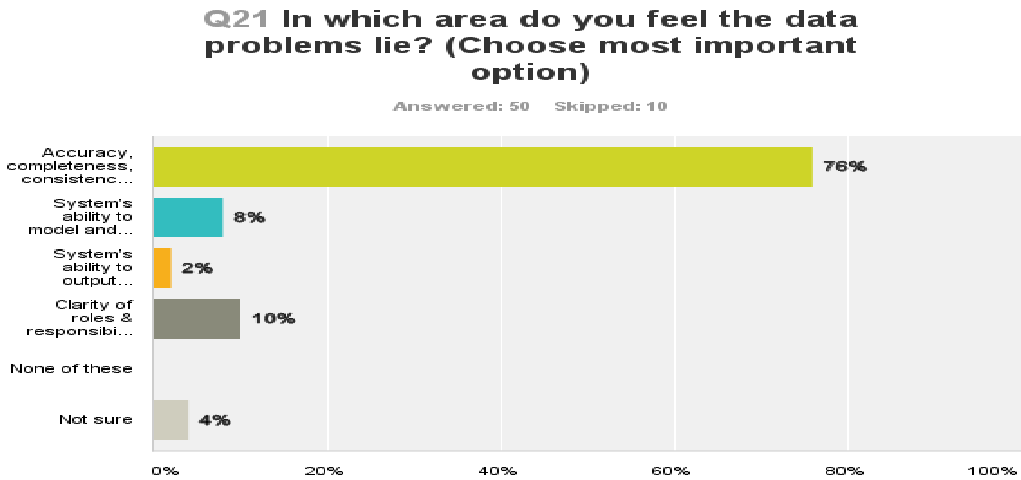


Figure 4.13 - ASSESSMENT OF SOURCE OF PROBLEMS

Table 4.13 - ASSESSMENT OF SOURCE OF PROBLEMS

In which area do you feel the data problems lie? (Choose most important option) * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Accuracy, completeness, consistency, timeliness. (Processes)	Count	4	27	7	38
		% within IS User	66.7%	75.0%	87.5%	76.0%
	System's ability to model and manipulate data representing t	Count	0	4	0	4
		% within IS User	0.0%	11.1%	0.0%	8.0%
	System's ability to output meaningful information	Count	0	1	0	1
		% within IS User	0.0%	2.8%	0.0%	2.0%
	Clarity of roles & responsibilities	Count	1	3	1	5
		% within IS User	16.7%	8.3%	12.5%	10.0%
	Not sure	Count	1	1	0	2
		% within IS User	16.7%	2.8%	0.0%	4.0%
Total		Count	6	36	8	50
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.13 shows that, all the IS Users reported high rates of problems with dimensions of the data (accuracy, completeness, consistency and timeliness) with an overall rate of 76%. Both the Data Custodians (16.7%) and the Data Stewards (Technical) (12.5%) cited 'clarity of roles and responsibilities' as another source of data quality problems. While the Data Stewards (Business)

also felt that data quality as measured by the various dimensions was significant (75%), they cited the system's inability to manipulate data as another reason.

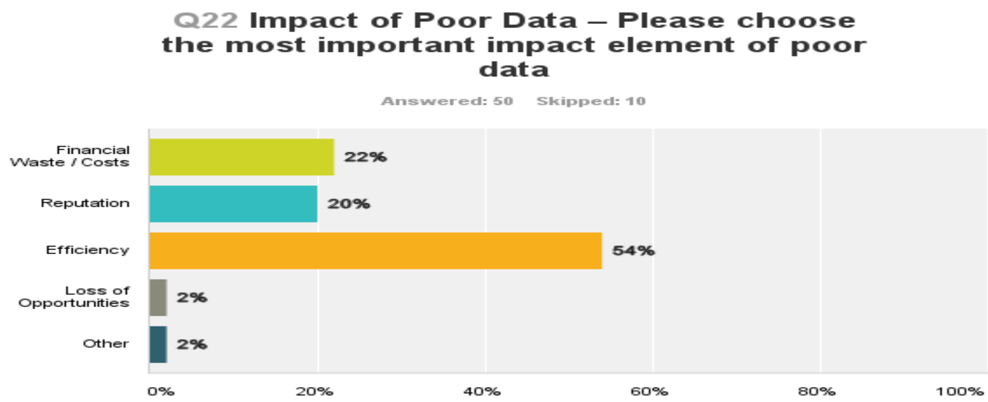


Figure 4.14 – IMPACT OF POOR DATA

Table 4.14 – IMPACT OF POOR DATA

Impact of Poor Data - Please choose the most important impact element of poor data * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Financial Waste / Costs	Count	2	8	1	11
		% within IS User	33.3%	22.2%	12.5%	22.0%
	Reputation	Count	2	8	0	10
		% within IS User	33.3%	22.2%	0.0%	20.0%
	Efficiency	Count	2	18	7	27
		% within IS User	33.3%	50.0%	87.5%	54.0%
	Loss of Opportunities	Count	0	1	0	1
		% within IS User	0.0%	2.8%	0.0%	2.0%
	Other	Count	0	1	0	1
		% within IS User	0.0%	2.8%	0.0%	2.0%
Total		Count	6	36	8	50
		% within IS User	100.0%	100.0%	100.0%	100.0%

The impact of poor data quality (Table 4.14) was most significantly related to financial waste and inefficiency (76%), with 99% of Data Stewards (Technical) and 72% and 66.6% of Data Stewards (Business) and Data Custodians, respectively, citing these factors.

While some Data Custodians (33.3%) and Data Stewards (Business) (22.2%) felt that poor data quality impacts the Institutions' reputation, no Data Stewards (Technical) cited this factor, which may again reflect the 'distance' between the technical function and the business aspects (e.g., reporting the information and the implications thereof) of the organisation.

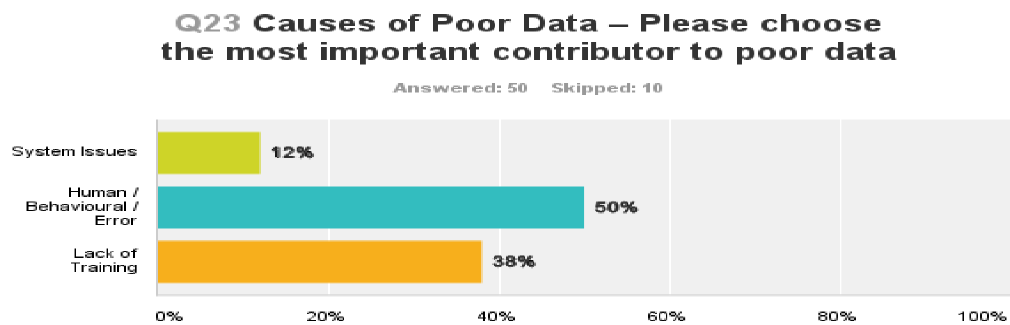


Figure 4.15 – CAUSES OF POOR DATA

Table 4.15 – CAUSES OF POOR DATA

Causes of Poor Data – Please choose the most important contributor to poor data * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	System Issues	Count	0	6	0	6
		% within IS User	0.0%	16.7%	0.0%	12.0%
	Human / Behavioural / Error	Count	5	15	5	25
		% within IS User	83.3%	41.7%	62.5%	50.0%
	Lack of Training	Count	1	15	3	19
		% within IS User	16.7%	41.7%	37.5%	38.0%
Total		Count	6	36	8	50
		% within IS User	100.0%	100.0%	100.0%	100.0%

Commenting on the dimensions that contribute to poor data (Table 4.15), the majority of Data Custodians (83.3%) attributed poor data to human and behavioural issues. While on the surface, this can be related to the capture e.g., inaccurate, incomplete data, it can also relate to processes that require human intervention or that had not been adhered to.

Most of the Data Stewards (Technical) attributed poor data to human and behavioural issues (62.5%). The fact that they did not cite system issues (which is their domain), may indicate that that the quality of the system as a contributor to data quality (or how it ‘dove-tails’ with data quality) is underestimated.

The Data Stewards (Business) felt that both behavioural issues (41%) and training (41%) are significant; some 16% also felt that systems issues cause poor data. More Data Stewards (Business) than Data Custodians and Data Stewards (Technical) cited the need for training; this may be due to the fact that they are operationally closer to the sources of the data than other IS Users.

4.4 ISSUES RELATING TO ACCOUNTABILITY AND DATA QUALITY PRACTICE / MANAGEMENT

4.4.1 ACCOUNTABILITY

This section examines the respondents’ perceptions of how information system users respond in terms of accountability.

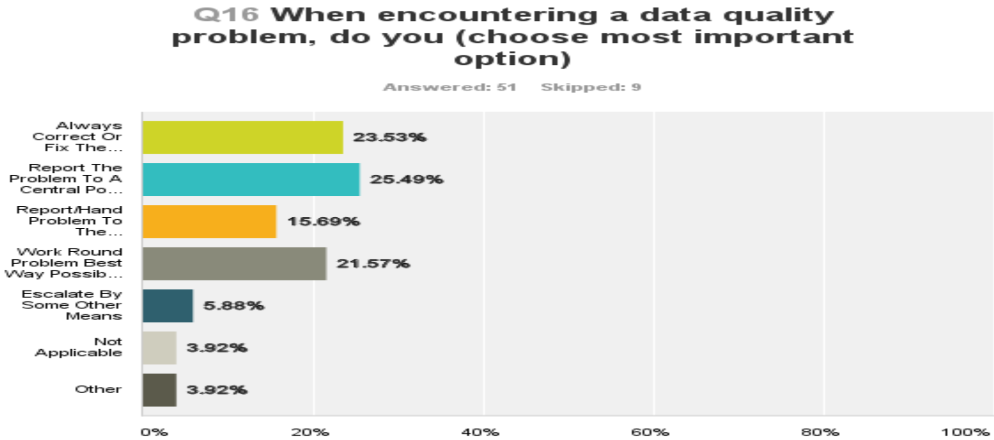


Figure 4.16 - ROUTES TAKEN IN DATA QUALITY ENCOUNTERS

Table 4.16 - ROUTES TAKEN IN DATA QUALITY ENCOUNTERS

When encountering a data quality problem, do you (choose most important option) * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Always Correct Or Fix The Problem Yourself	Count	0	12	0	12
		% within IS User	0.0%	32.4%	0.0%	23.5%
	Report The Problem To A Central Point (Helpdesk / Support Fu	Count	2	9	2	13
		% within IS User	33.3%	24.3%	25.0%	25.5%
	Report/Hand Problem To The Originator	Count	1	5	2	8
		% within IS User	16.7%	13.5%	25.0%	15.7%
	Work Round Problem Best Way Possible To Complete Task	Count	1	8	2	11
		% within IS User	16.7%	21.6%	25.0%	21.6%
	Escalate By Some Other Means	Count	1	1	1	3
		% within IS User	16.7%	2.7%	12.5%	5.9%
	Not Applicable	Count	1	1	0	2
		% within IS User	16.7%	2.7%	0.0%	3.9%
	Other	Count	0	1	1	2
		% within IS User	0.0%	2.7%	12.5%	3.9%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.16 shows responses relating to IS Users' initiative and efficiency in managing data quality problems, as well as their knowledge and perceptions of the 'originators' of the data problem. In terms of the 'critical path' they follow when they encounter a data quality problem, 23.5% of the respondents stated that they generally fix problems themselves, while 25.5% report it to a central point or work around it in the best way possible (21.6%).

The extent of users not attending to data quality issues themselves and reporting or escalating it, is approximately 46%. The Data Custodians reported that they refer problems (66%) or work around the problem (25%). The Data Stewards (Business) responded that they refer the problem (40%) but try to fix or work around it (53%).

Data Stewards (Technical) do not have the authority to change data; therefore they do not correct or fix data themselves. Thus, 62.5% of the Data Stewards (Technical) reported referring the problem elsewhere and 25% attempt to work around the problem.

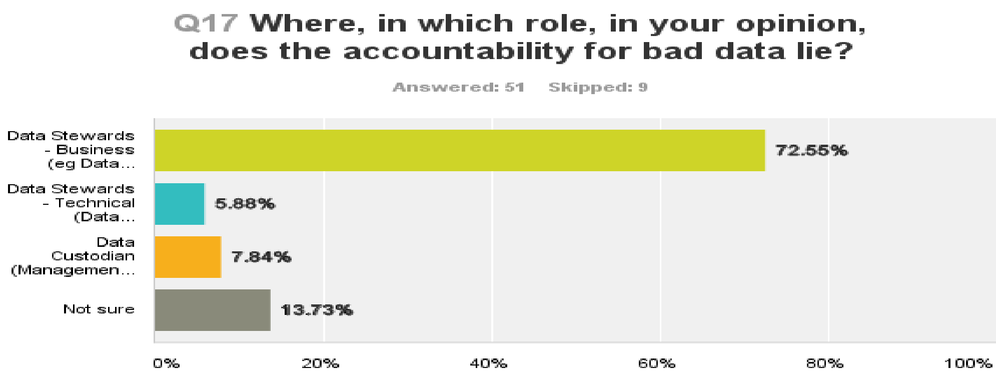


Figure 4.17- ASSESSMENT OF ROLES THAT SHOULD BE ACCOUNTABLE

Table 4.17- ASSESSMENT OF ROLES THAT SHOULD BE ACCOUNTABLE

Where, in which role, in your opinion, does the accountability for bad data lie? * IS User Crosstabulation

		IS User			Total
		Data Custodian	Data Steward - Business	Data Steward - Technical	
Data Stewards - Business (eg Data Owners in the areas of Stu	Count	6	23	8	37
	% within IS User	100.0%	62.2%	100.0%	72.5%
Data Stewards - Technical (Data Stewards in ICS)	Count	0	3	0	3
	% within IS User	0.0%	8.1%	0.0%	5.9%
Data Custodian (Management Information Services in ICS)	Count	0	4	0	4
	% within IS User	0.0%	10.8%	0.0%	7.8%
Not sure	Count	0	7	0	7
	% within IS User	0.0%	18.9%	0.0%	13.7%
Total	Count	6	37	8	51
	% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.17 shows that, the Data Custodians and Data Stewards (Technical) stated that the Data Stewards (Business) should be charged with accountability for bad data. Only 62.2% of the Data Stewards (Business) felt that they themselves should be accountable, and involve the Data Custodians (8.1%) and the Data Stewards (Technical) (10.8%). The fact that the Data Stewards (Business) feel that the Data Custodians and the Data Stewards (Technical) should be accountable 'as a group' may be due to their reliance on the Data Custodians for the data audit systems and for the data problems to be transparent (and measureable) in order for them to maintain and manage data.

Figure 4.18 - INCLUSION OF DATA QUALITY IN PERFORMANCE MANAGEMENT

Table 4.18 - INCLUSION OF DATA QUALITY IN PERFORMANCE MANAGEMENT

In terms of accountability for data and data quality success, do you believe that data quality responsibilities should be included in data owners' performance management agreements (KPA's)? * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Reasonable	Count	6	26	7	39
		% within IS User	100.0%	72.2%	87.5%	78.0%
	Not reasonable	Count	0	6	1	7
		% within IS User	0.0%	16.7%	12.5%	14.0%
	Not sure	Count	0	4	0	4
		% within IS User	0.0%	11.1%	0.0%	8.0%
Total		Count	6	36	8	50
		% within IS User	100.0%	100.0%	100.0%	100.0%

Approximately 78% of all the respondents indicated that it is reasonable for data quality responsibility to be incorporated into data owner's performance management agreements, as expressed in key performance areas (KPAs) (Table 4.18). The corresponding statistics for Data Steward (Business) and Data Stewards (Technical) are 72.2% and 87.5%, respectively. In their role as institutional reporters, the Data Custodians are required to report data at regular intervals. They are pressed for time and cannot be delayed by data quality issues. This explains why all the Data Custodians feel that it is reasonable for data quality responsibility to be included in KPAs. Eleven percent of the Data Stewards (Business) indicated that they were ‘not sure’. This may be related to their reluctance to fully commit to data quality KPAs as other factors may jeopardise their objective of meeting their data correction deadlines.

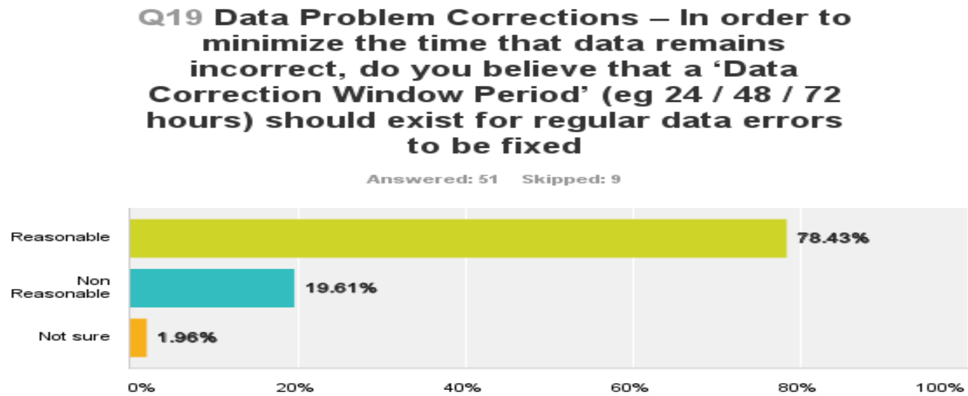


Figure 4.19 - FEASIBILITY OF DATA CORRECTION WINDOW PERIOD

Table 4.19 - FEASIBILITY OF DATA CORRECTION WINDOW PERIOD

Data Problem Corrections – In order to minimize the time that data remains incorrect, do you believe that a ‘Data Correction Window Period’ (eg 24 / 48 / 72 hours) should exist for regular data errors to be fixed * IS User Crosstabulation

		IS User			Total
		Data Custodian	Data Steward - Business	Data Steward - Technical	
Reasonable	Count	4	29	7	40
	% within IS User	66.7%	78.4%	87.5%	78.4%
Non Reasonable	Count	1	8	1	10
	% within IS User	16.7%	21.6%	12.5%	19.6%
Not sure	Count	1	0	0	1
	% within IS User	16.7%	0.0%	0.0%	2.0%
Total	Count	6	37	8	51
	% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.19 shows, that, 78.4% of all the respondents stated that, it is reasonable to impose a ‘Data Correction Window Period’. Taken together with the previous

responses relating to performance management, this suggests that the IS Users' are committed to data quality.

The high level of agreement that a 'Data Correction Window Period' is reasonable amongst the Data Stewards (Business) may reflect the problems they experience with third party data providers. A service level agreement (SLA) to guarantee data quality to the University may have to be considered. SLAs could apply to both third party data providers and internal providers.

In view of their roles as institutional reporters, data quality coordinators, initiators of data quality systems and data quality monitors, the Data Custodians adopted a softer approach to adherence to a window period (66.7%). This may be due to their experience that it is difficult for users to keep within prescribed deadlines, let alone the short window periods that the research question proposed.

4.4.2 - DATA QUALITY PRACTICE / MANAGEMENT

This section examines perceptions of practices or processes that could affect the quality of data.

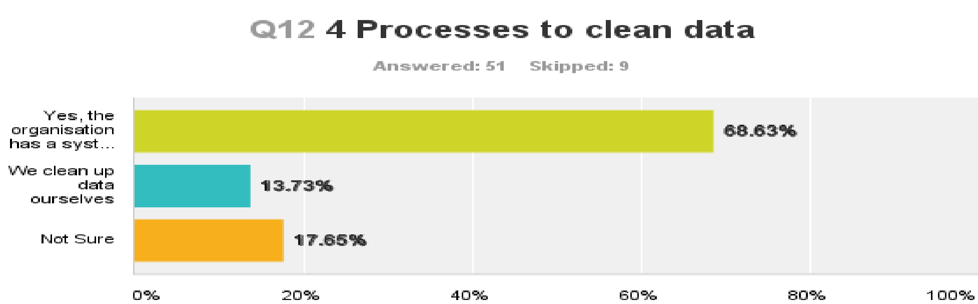


Figure 4.20 – PROCESSES TO CLEAN DATA

Table 4.20 – PROCESSES TO CLEAN DATA

4 Processes to clean data * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Yes, the organisation has a system / division providing data	Count	5	25	5	35
		% within IS User	83.3%	67.6%	62.5%	68.6%
	We clean up data ourselves	Count	0	6	1	7
		% within IS User	0.0%	16.2%	12.5%	13.7%
	Not Sure	Count	1	6	2	9
		% within IS User	16.7%	16.2%	25.0%	17.6%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.20 indicates, that, overall, the respondents reported a high level of awareness of an institutional system for data correction (Data Custodians, 83.3%, Data Stewards (Business), 67.6%, and Data Stewards (Technical), 62.5%).

While stating that they are aware of data cleaning / auditing mechanisms (67%), the Data Stewards (Business) credit themselves with their own efforts to deal with bad data (17%). As the institutional reporters, coordinators of data quality systems and data quality monitors, the Data Custodians reported high involvement (83%) as they conceptualise and implement these systems. The fact that data quality initiatives and formal data correction processes have not been infused at every level may explain the lower awareness of a data cleaning / auditing mechanisms amongst the Data Stewards (Technical) at 62.5%. As with Data Stewards (Business), Data Stewards (Technical) reported their own efforts to correct data at lower percentages (12.5%), while indicating uncertainty (25%) about the totality of the data systems that may exist. They may only be 'part of the picture' in that they may assist the Data Custodians and Data Stewards (Business) with portions of programming towards the data quality system.

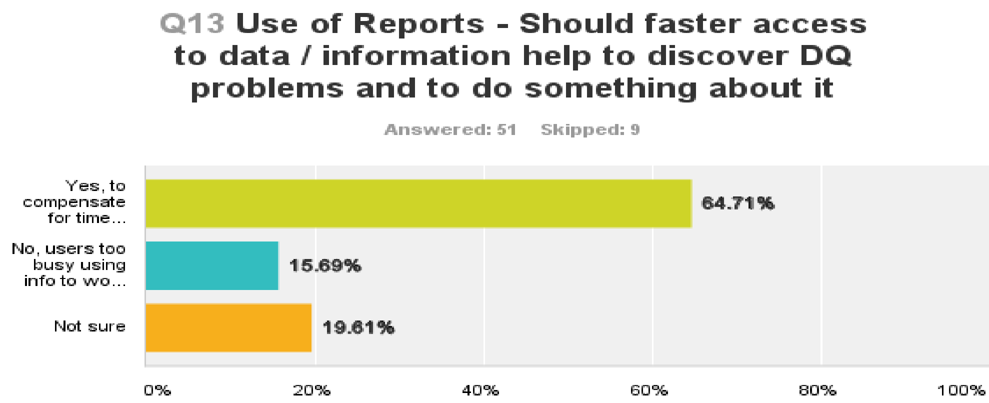


Figure 4.21- ACCESS TO REPORTS AND DATA QUALITY

Table 4.21- ACCESS TO REPORTS AND DATA QUALITY

Use of Reports - Should faster access to data / information help to discover DQ problems and to do something about it * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Yes, to compensate for time saved in accessing info faster,	Count	2	29	2	33
		% within IS User	33.3%	78.4%	25.0%	64.7%
	No, users too busy using info to worry about DQ	Count	2	3	3	8
		% within IS User	33.3%	8.1%	37.5%	15.7%
	Not sure	Count	2	5	3	10
		% within IS User	33.3%	13.5%	37.5%	19.6%
Total		Count	6	37	8	51
		% within IS User	100.0%	100.0%	100.0%	100.0%

Table 4.21 shows, that, while 64.7% of all the IS Users expressed the opinion that faster access or the use of reports may have a positive effect on IS quality improvement, the responses of the Data Stewards (Business) are significantly higher than the other IS Users. This is probably due to the fact that they need reports due to the operational nature of their work. 33.3% of the Data Custodians and 25% of the Data Stewards (Technical) supported the need for reports compared with 78.4% of the Data Stewards (Business). The number of Data Stewards (Technical) who indicated that they were ‘not sure’ (37.5%) may be related to the fact that they are not directly involved in reporting and auditing processes.

It should be bone in mind that IS Users have access to reports designed for them, as well as access to information at a summary level via the II Web Portal.

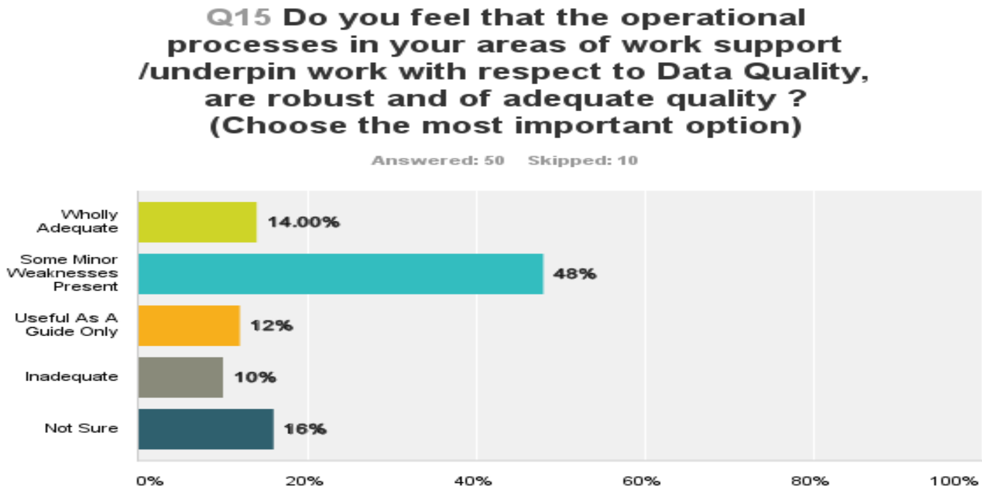


Figure 4.22 - EXTENT OF DATA QUALITY SUPPORTED BY OPERATIONAL PROCESSES

Table 4.22 - EXTENT OF DATA QUALITY SUPPORTED BY OPERATIONAL PROCESSES

Do you feel that the operational processes in your areas of work support /underpin work with respect to Data Quality, are robust and of adequate quality ? (Choose the most important option) * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	Wholly Adequate	Count	0	7	0	7
		% within IS User	0.0%	18.9%	0.0%	14.0%
	Some Minor Weaknesses Present	Count	0	22	2	24
		% within IS User	0.0%	59.5%	25.0%	48.0%
	Useful As A Guide Only	Count	0	3	3	6
		% within IS User	0.0%	8.1%	37.5%	12.0%
	Inadequate	Count	2	1	2	5
		% within IS User	40.0%	2.7%	25.0%	10.0%
	Not Sure	Count	3	4	1	8
		% within IS User	60.0%	10.8%	12.5%	16.0%
Total		Count	5	37	8	50
		% within IS User	100.0%	100.0%	100.0%	100.0%

In terms of the quality of the processes that contribute to IS Users' data quality activity (Table 4.22), Data Steward (Business) expressed a higher level of confidence in the quality of operational or business processes than other IS Users (18.9%) while 60% acknowledged some weaknesses. Data Stewards (Technical) reported a higher level of concern than the Data Custodians or Data Stewards (Business). Between 50% and 60 % were concerned or unsure. The Data Custodians expressed concerns about processes as precursors of data that reflected a lack of insight into the operations (60%). Business process support is very important. All of the users are exposed to batch jobs running at night that are required to update various parts of the system. If this does not occur for various reasons, the quality of the data in the system is compromised and the quality of data in all dependent 'upstream' information reporting mechanisms and operational decisions may be compromised.

4.5 COST OF DATA QUALITY

The results of the survey show that up to 80% of the respondents reported problems with data quality on a daily basis. The response data was converted *via* mean scales of remuneration to calculate and quantify the time that IS Users spend in terms of labour. An analysis of the respondents' time spent on data quality is provided in Table 4.22 below.

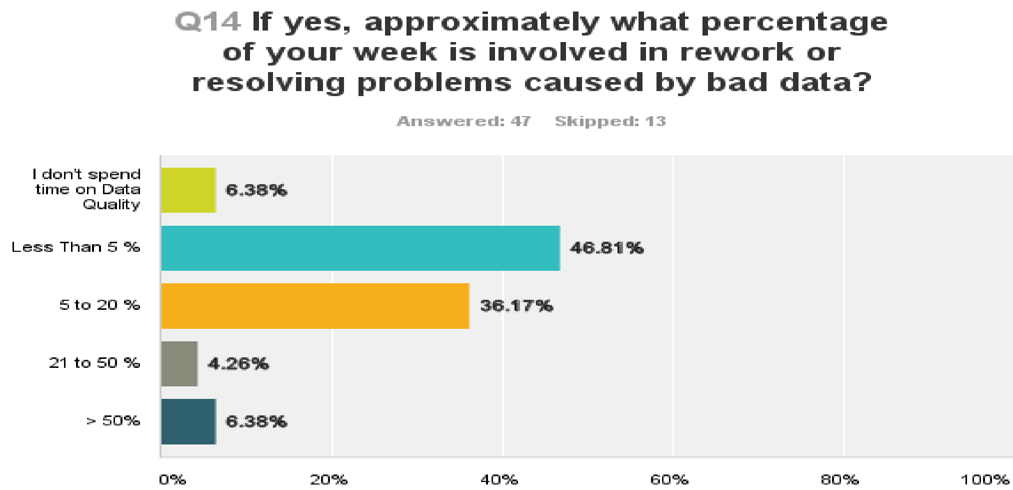


Figure 4.23 - TIME SPENT ON DATA QUALITY

Table 4.23 - TIME SPENT ON DATA QUALITY

If yes, approximately what percentage of your week is involved in rework or resolving problems caused by bad data? * IS User Crosstabulation

			IS User			
			Data Custodian	Data Steward - Business	Data Steward - Technical	Total
	I don't spend time on Data Quality	Count	0	2	1	3
		% within IS User	0.0%	5.4%	16.7%	6.4%
	Less Than 5 %	Count	0	20	2	22
		% within IS User	0.0%	54.1%	33.3%	46.8%
	5 to 20 %	Count	4	10	3	17
		% within IS User	100.0%	27.0%	50.0%	36.2%
	21 to 50 %	Count	0	2	0	2
		% within IS User	0.0%	5.4%	0.0%	4.3%
	> 50%	Count	0	3	0	3
		% within IS User	0.0%	8.1%	0.0%	6.4%
Total		Count	4	37	6	47
		% within IS User	100.0%	100.0%	100.0%	100.0%

There were only four responses from Data Custodians. These responses could be understated as data quality is part of the researcher's official portfolio and he could not be part of the survey. Up to 20% of the Data Custodians' time is devoted to data quality. The time spent by Data Stewards (Business) on data quality varies; 8% of these respondents reported spending more than 50% of their time on data quality. Fifty percent of the Data Stewards (Technical) reported that they spend up to 20% of their time on data quality. The tables below (Tables 4.24a to 4.24c) depict the time spent on data quality by the Data Custodians, Data Stewards (Business) and Data Stewards (Technical). The data are presented in a 'lower range' as well as an 'upper range' scenario.

Table 4.24a - ESTIMATED COST OF DATA QUALITY – Data Custodians

Ordinal representation of time spent	Lower Range	Upper Range	Mid Point Salary Range	Lower - R Per Yr	Upper - R Per Yr	No of Respondents	Total Cost (Lower Range)	Total Cost (Upper Range)
Less than 5%	0.01		R 300 585	R 3 006				
5 to 20%	0.05	0.2	R 300 585	R 15 029	R 60 117	4	R 60 117	R 240 468
21 to 50 %	0.2	0.5	R 300 585	R 60 117	R 150 293			
Over 50 %	0.5	0.6	R 300 585	R 150 293	R 180 351			
						4	R 60 117	R 240 468

Table 4.24b - ESTIMATED COST OF DATA QUALITY – Data Stewards (Business)

Ordinal representation of time spent	Lower Range	Upper Range	Mid Point Salary Range	Lower - R Per Yr	Upper - R Per Yr	No of Respondents	Total Cost (Lower Range)	Total Cost (Upper Range)
Less than 5%	0.01	0.05	R 249 350	R 2 494	R 12 468	20	R 49 870	R 249 350
5 to 20%	0.05	0.2	R 249 350	R 12 468	R 49 870	10	R 124 675	R 498 700
21 to 50 %	0.2	0.5	R 249 350	R 49 870	R 124 675	2	R 99 740	R 249 350
Over 50 %	0.5	0.6	R 249 350	R 124 675	R 149 610	3	R 374 025	R 448 830
						35	R 648 310	R 1 446 230

Table 4.24c - ESTIMATED COST OF DATA QUALITY – Data Stewards (Technical)

Ordinal representation of time spent	Lower Range	Upper Range	Mid Point Salary Range	Lower - R Per Yr	Upper - R Per Yr	No of Respondents	Total Cost (Lower Range)	Total Cost (Upper Range)
Less than 5%	0.01	0.05	R 351 970	R 3 520	R 17 599	2	R 7 039	R 35 197
5 to 20%	0.05	0.2	R 351 970	R 17 599	R 70 394	3	R 52 796	R 211 182
21 to 50 %	0.2	0.5	R 351 970	R 70 394	R 175 985			
Over 50 %	0.5	0.6	R 351 970	R 175 985	R 211 182			
						5	R 59 835	R 246 379

The estimation arrived at was R 728 262 per annum at the lower range and R 1 932 077 per annum at a higher range. If these results are extrapolated to the total IS Users in the survey the labour costs amount to R 1 031 420 per annum at the lower range and R 2 663 623 per annum at the higher range. These costs may be conservative as data quality initiatives at the Institution have not reached their apex. Various areas have yet to be incorporated that will put pressure on current resources for the data quality initiative. In addition, functional areas such as Human Resources and Finance have not been incorporated in terms of errors in the current data quality audit system, and will substantially add to the labour cost that is calculated in this exercise. Furthermore, this exercise does not take the cost of technical solutions and changing business processes to improve data quality into account.

4.6 SUSTAINABILITY OF DATA QUALITY IMPROVEMENTS

Section E of the questionnaire dealt with questions 24 to 33 in the research instrument and sought to gauge IS Users' perceptions of the sustainability of data quality improvements as well as their perceptions on managing the change associated with data quality and how comfortable they are with such change (Tull, 1997). The users were asked to indicate their views on a four point scale with 4 indicators where '1' represented 'high sustainability' or 'Strongly Agree' and a value of '4' that represented 'low sustainability' or 'Strongly Disagree'. Fifty of the 60 IS Users responded. An average mean of 2.25 and a mean of 2.26 were found from the analysis of the 10 questions in this section. A graphical representation is provided below. Detailed descriptive statistics have been attached (Appendix 4).

Question 24 (whether data quality awareness promotes a data and information culture) and Question 29 (whether management views data quality as important) recorded lower levels of optimism, i.e., an average and mean of 1.54 and 1.5 and 1.68 and 2, respectively.

1. **Data quality initiatives / awareness will promote a data and information culture / information literacy (Question 24)**

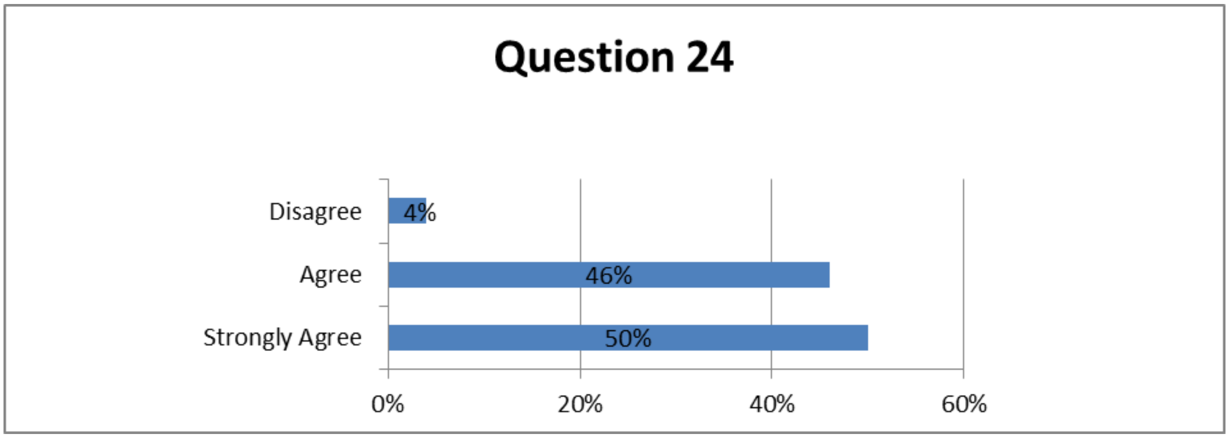


Figure 4.25: Data quality initiatives / awareness will promote a data and information culture / information literacy

Mean	Median	Standard Deviation
1.54	1.5	.08

Fifty users responded to this question. The mean of 1.54 and median of 1.5 suggest that, overall, IS Users are positively inclined towards data quality awareness as a conduit for information literacy that should serve as a catalyst to sensitise users to data quality. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 1.66, 1.55 and 1.375, respectively.

2. **At this institution there are sufficient people to support the data quality initiative with the necessary skills and knowledge to guide the implementation (Question 25)**

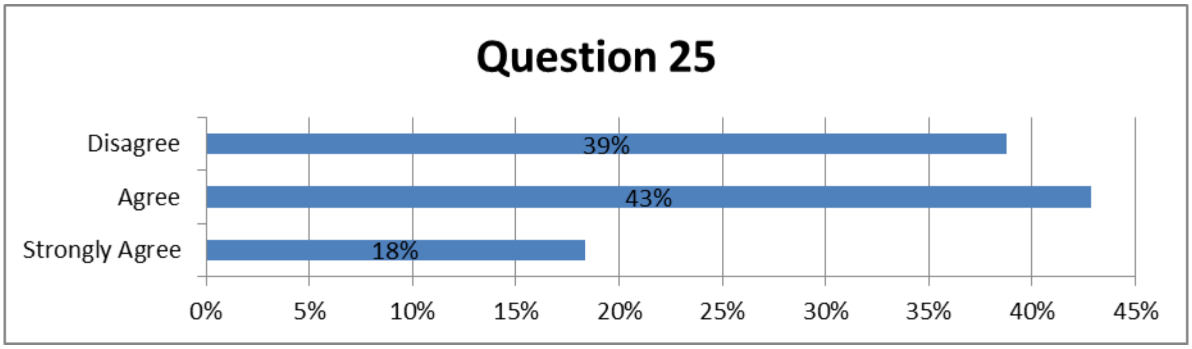


Figure 4.26: Sufficient people to support the data quality initiative with the necessary skills and knowledge to guide the implementation

Mean	Median	Standard Deviation
2.21	2.0	.105

Forty-nine users responded to this question. The mean of 2.21 suggests that IS Users tend to be skeptical about the adequacy of the institutional skills and knowledge base to support data quality improvement. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 3.0, 2.14 and 1.875, respectively. The Data Custodians are the most skeptical.

3. **At this Institution training among data capturers / data owners is adequate to support the attainment of better data quality (Question 26)**

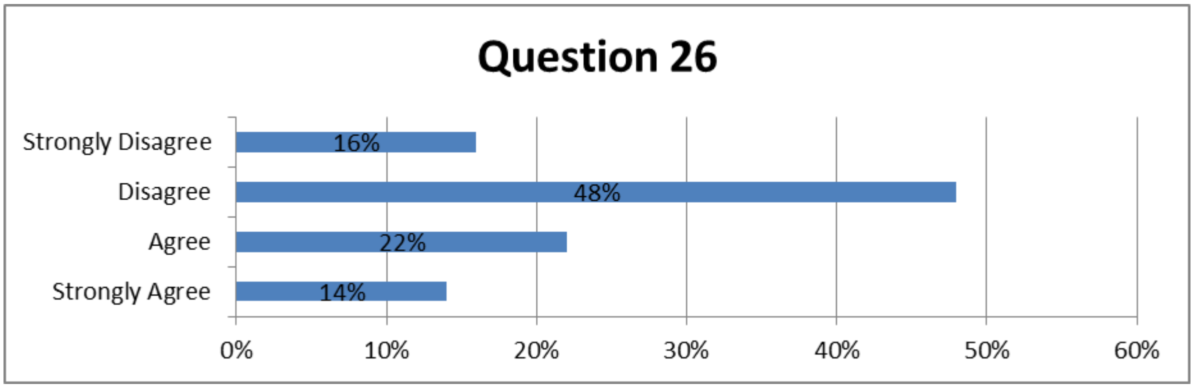


Figure 4.27: Training is adequate to support the attainment of better data quality

Mean	Median	Standard Deviation
2.66	3.0	.129

Fifty users responded to this question. The mean of 2.66 suggests that a significant number of IS Users feel that the levels of training *vis a vis* data quality at the Institution are inadequate and needs more attention. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 3.0, 2.61 and 2.625, respectively. The strongest opinion that training is required was found among the Data Custodians.

4. **The execution of data quality initiatives through dynamic structures (e.g. teams, networks and workgroups) at the Institution is working well (Question 27)**

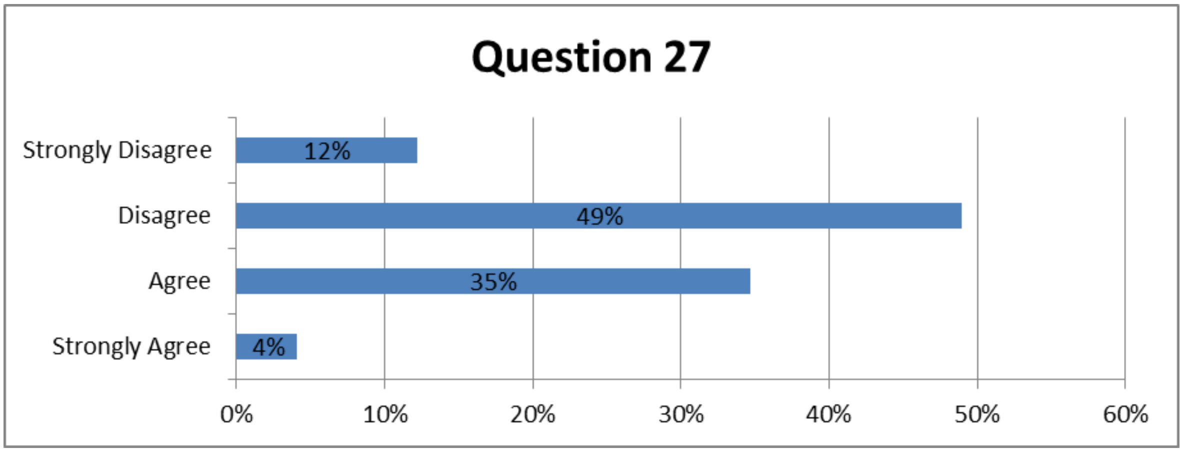


Figure 4.28: Teamwork at the Institution is working well

Mean	Median	Standard Deviation
2.69	3.0	.106

Forty-nine users responded to this question. The mean of 2.69 suggests that the teams and workgroups for data quality projects are not functioning as efficiently as expected. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 3.33, 2.55 and 2.85, respectively.

5. **In this organisation individuals have become comfortable with change and do not seek stability (Question 28)**

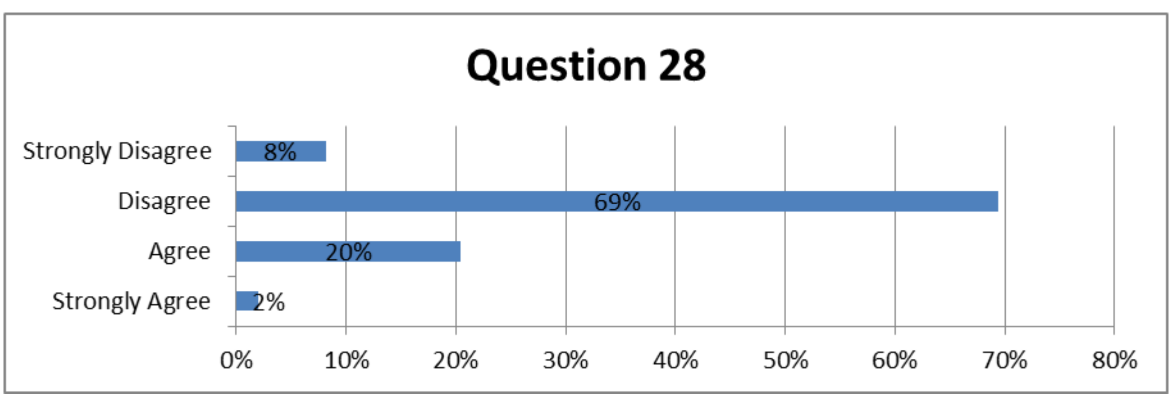


Figure 4.29: Individuals have become comfortable with change and do not seek stability

Mean	Median	Standard Deviation
2.83	3.0	.084

Forty-nine users responded to this question. The mean of 2.83 suggests that IS Users are not comfortable with change and implicitly express a 'wish' for a data environment that is stable. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 2.66, 2.88 and 2.75, respectively.

6. **Management views data quality as important (Question 29)**

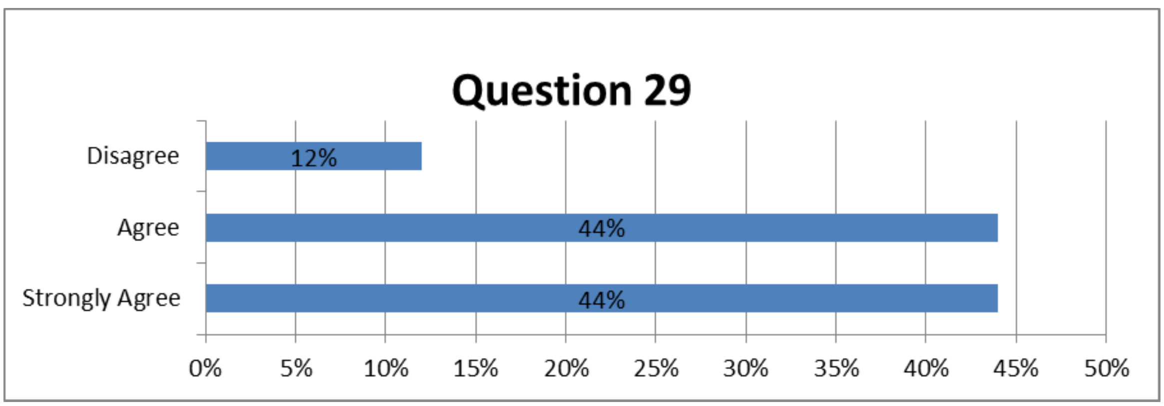


Figure 4.30: Management views data quality as important

Mean	Median	Standard Deviation
1.68	2.0	.096

Fifty users responded to this question. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 2.00, 1.52 and 2.12, respectively.

7. **There are enough people in the Institution to lead a data quality initiative (Question 30)**

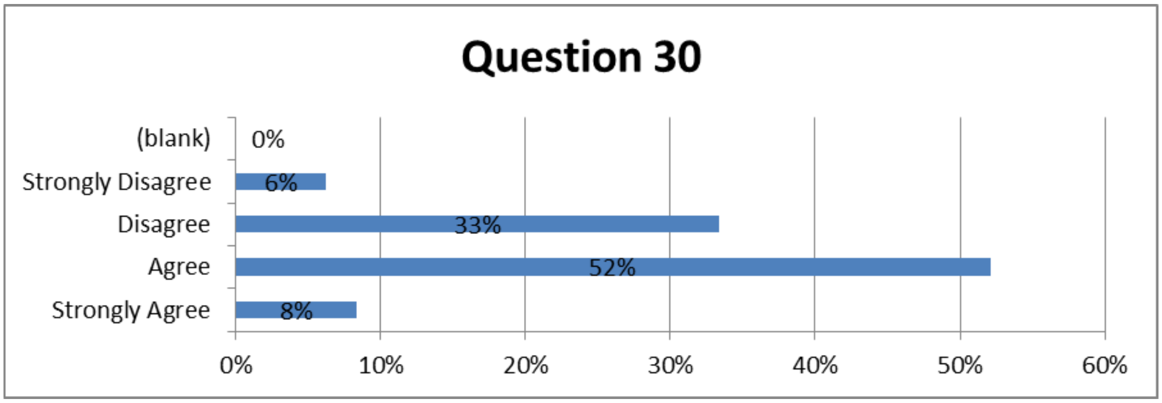


Figure 4.31: There are enough people at the Institution to lead a data quality initiative

Mean	Median	Standard Deviation
2.37	2.0	.108

Forty-eight users responded to this question. The mean of 2.37 suggests that there are not enough people to lead a data quality initiative at the Institution. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 3.2, 2.31 and 2.12, respectively.

8. **There are enough people in the Institution that care about data quality (Question 31)**

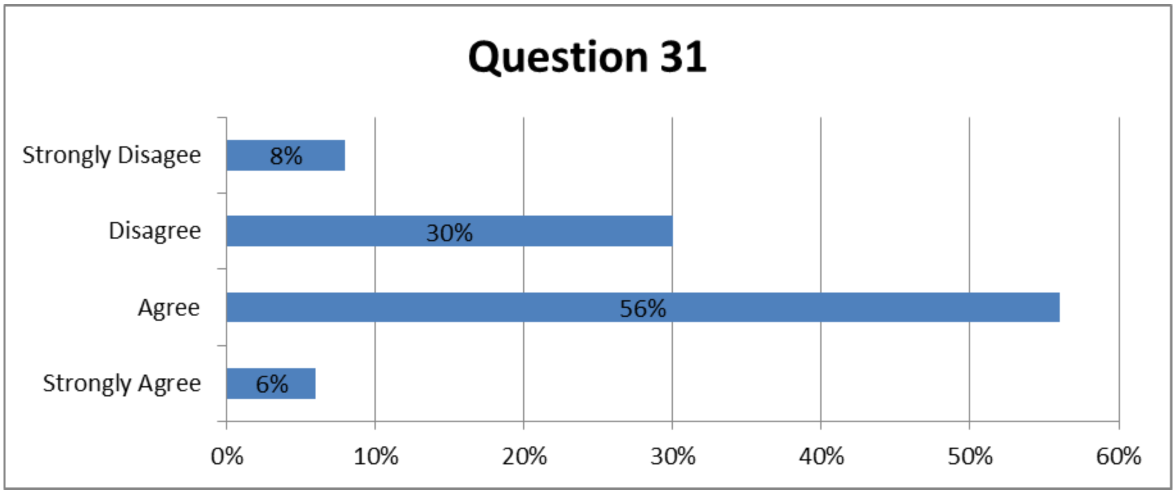


Figure 4.32: There are enough people in the Institution that care about data quality

Mean	Median	Standard Deviation
2.4	2.0	.103

Fifty users responded to this question. The mean of 2.4 suggests that not enough people at the Institution care about data quality. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 2.83, 2.27 and 2.625, respectively.

9. **Expectations about achieving data quality improvement are reasonable (Question 32)**

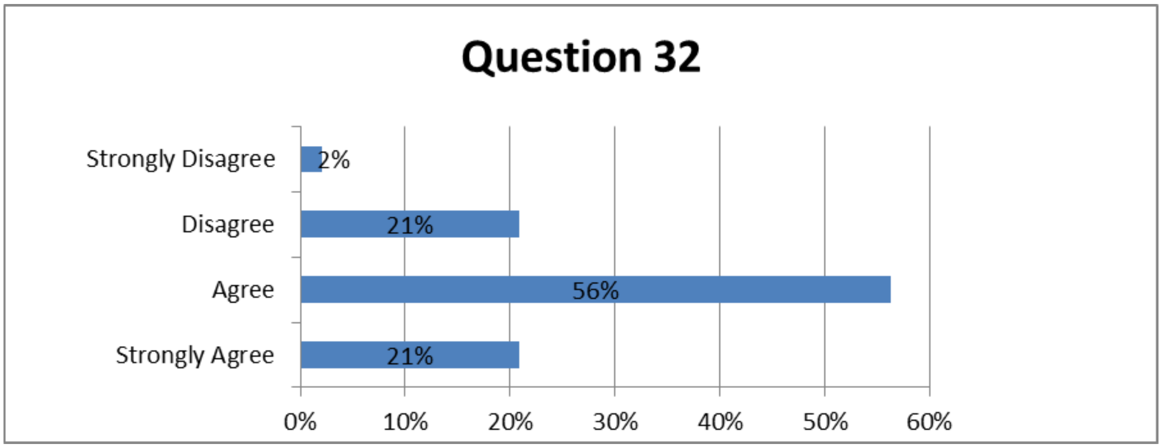


Figure 4.33: Expectations about achieving data quality improvement are reasonable

Mean	Median	Standard Deviation
2.14	2.0	.104

Forty-eight users responded to this question. The mean of 2.14 suggests that IS Users are fairly balanced in their views on whether or not expectations relating to data quality are reasonable. This could be influenced by the fact that, on the one hand, they realise the benefits of improved data quality in their own work areas but, on the other, they may be concerned about the expectations / deliverables in terms of workloads and timelines.

The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 2.4, 2.08 and 2.25, respectively.

10. **Data quality activity / processes will over the longer term be successful and sustainable (Question 33)**

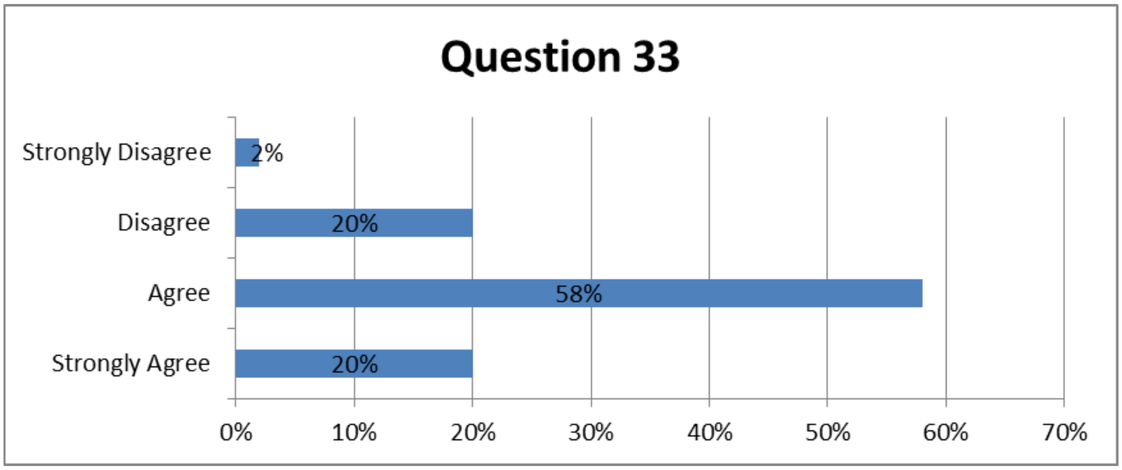


Figure 4.34: Data quality activity / processes will over the longer term be successful and sustainable

Mean	Median	Standard Deviation
2.04	2.0	.098

Fifty users responded to this question. The mean of 2.04 suggests that IS Users are slightly negatively inclined towards the long term sustainability of data quality. The means for the Data Custodians, Data Stewards (Business), and Data Stewards (Technical) are 2.16, 1.97 and 2.25, respectively.

Summary

A summary of the 10 indicators shows that the respondents are of the opinion that there are areas where data quality could be improved, with an overall mean value of 2.26 and a median of 2.35**.**

4.7 CONCLUSION

This chapter provided an analysis and interpretation of the survey data using descriptive statistics pertaining to issues affecting data quality awareness, data quality practice and accountability; an estimation of the cost of correcting bad data; and a statistical perspective on the sustainability of data quality improvement *vis a vis* the data custodians and data stewards. The analysis of the data provided a comprehensive picture of the perspectives of the three groups of IS Users.

To elaborate on accountability and roles, a 'RACI' matrix applied to the data quality environment at the Institution may be provided in Chapter 5 for future research.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 INTRODUCTION

This chapter elaborates on the findings that were discussed in the previous chapter and follow with proposal and conclusions. It includes -

- 5.2 a comparison of the literature vs the fieldwork in terms of the research objectives
- 5.3 suggestions for further research in the form of a proposed 'D M A I C' framework for further research to express the data quality policy through a data quality strategy in terms of concrete actions and steps that may have to take place
- 5.3.1 a proposed RACI diagram to formalise or embody 'Responsibilities and Accountabilities' in support of the data quality initiative at the Institution
- 5.4 a conclusion

5.2 COMPARISON OF THE LITERATURE vs. FIELDWORK

The section that follows draws a comparison between the literature and what was found in the 'field' i.e. via the survey. It provides pointers in each of the research areas from the literature, followed by observations from the 'field' and concluded with Discussion and Recommendations that might include suggestions for action.

5.2.1 OBJECTIVE 1 : AWARENESS

Literature vs Fieldwork

	Literature	Fieldwork
1	The extent to how data quality is incorporated into work areas (question 5) and by what means users obtained knowledge about data quality (question 6) relates to communication strategies may identify key audiences, messages, desired actions and results in order to gain support for the information quality strategy and governance. The extent of the 'big picture' or users' 'knowing-why' of data quality would over time been enhanced by the spread of Quality philosophies eg Total Data Quality and TDQM philosophies.	1. IS-Users indicate that pertaining to their awareness of data quality, that 69% (34 out of 49 distinct IS Users) report that they have become aware of data quality through personal job experience, 43% through training and 37% through management communications.
2	The extent of users' awareness of current data quality activity and the state of data quality (questions 7 and 10) may be related to the extent to which data quality improvements are incentivized and awareness is reinforced via resulting feedback loops to further stimulate data quality improvement.	Up to two thirds of Data Stewards report awareness of data quality activities (question 7). The incidence of Data Custodians and Data Stewards (Technical) reporting that data may be 'in a mess' is higher than that of the Data Stewards (Business).
3	Accuracy of data quality (question 9) as a data quality dimension appears to be more important than other dimensions. Accuracy seems to be followed by Completeness as dimension and there is a recognised	Data Quality accuracy was rated as the most important data quality dimension. The results indicate a variety of other dimensions following accuracy in importance.

	interplay or trade-off between the two.	
4	The main causes of data quality (questions 20 and 23) seem to be lack of clarity of roles, procedures, problems with data at the source and system issues.	In terms of causes of data quality, overall responses to the issue of barriers to the adoption of data quality initiatives indicate that 1) the quality of leadership around data quality initiatives represent a significant barrier to the adoption of a data quality initiative and 2) behavioural issues.
5	The impact of data quality (question 22) seem to be financial, productivity and risk to be compliant.	In terms of impact of data quality, responses to the impact of poor data quality were most significantly related to factors related to financial waste and inefficiency.

Discussion and Recommendation:

1. As data stewards (technical) may not have been involved in the data quality initiative as the other IS-Users, it is important that they become part of the mainstream communication and workshops. There needs to more management briefing, in particular to the Data Custodians and Data stewards (Technical).
2. A higher percentage of Data Custodians and Data Stewards (Technical) than Data Stewards (Business) report that data being ‘in a mess’. The need for a data quality monitor has become an important mechanism for all role players to have access to an objective source to the condition of data quality.
3. While the survey reports accuracy as the most important data quality dimension, other dimensions such as completeness is important. In practical situations, trade-off amongst dimensions are part of convention and acceptable. As the generally accepted definition of data quality is ‘fitness for purpose’, need to be highlighted at data quality briefings.

4. Users responses reflecting that poor data quality cost money being attributed to factors related to financial waste and inefficiency, will in conjunction with the response to time spent on data quality, will represent a strong motivational basis for the current cost of data quality activity to be reduced. Process quality, in addition to data quality needs to be interrogated and root causes identified with in a D M A I C (Define, Manage, Analyse, Improve and Control) exercise in order to address waste and inefficient practice.

5.2.2 OBJECTIVE 2 : ACCOUNTABILITY

Literature vs Fieldwork

	Literature	Fieldwork
1	Relating to accountability and data ownership (question 17), there are considerable differences in the interpretation of the role of a data steward. While early definitions tended to be restricted to associations with an information technology ('IT') approach to data quality improvement, later research extended the involvement of data stewards to include functional, business (or subject) interest in the organisation as well. Considerable differences in the definition of accountability ('everybody's accountable) existed as are there degrees of agreement. Ultimately the data owner of a business or functional subject area is proposed to be accountable for data quality.	The fieldwork is congruent with the literature that accountability for data quality is vested with the data owner. In addressing the issue of data quality accountability, data custodians and data stewards (technical) report that the data stewards (business) should be accountable for poor data.
2	Regarding routing data quality problems (question 16) and 'work around' problems to compensate for 'not knowing', the lack of clarity around accountability leads to a vacuum in leadership a vacuum and result in inefficiencies to get data problems	Relating to which 'critical paths' or 'routing' they may follow in an encounter with a data quality problem, IS-Users report that they generally fix problems themselves (23.5%), or report them to a central point (25.5%) or work around them in the best way possible (21.6%).

	solved.	This may be related to management not informing them what to do in such instances.
3	The application of key performance areas (KPA's) (question 18) as performance criteria measured by the data quality activity of IS-users is recognized in the literature (Loshin, 2011, Anturaniemi, 2012).	In terms of applying performance measures to data quality activity, IS-Users seem to view the application of performance management (KPA's) to data quality improvement to be reasonable. This is affirmed by approximately 78% of all respondents.
4	The use of service level agreements (SLA's) (question 19) as internal and external accountability mechanisms are useful levers in terms of improving the quality of data at the source (Schwolow & Jungfalk, 2009).	In terms of internal and external accountability, IS Users indicate that the application of a service level agreement (SLA) to data provided externally (third parties), as well as a transferred internally as part of the student data lifecycle, is reasonable. In terms of a data correction window period to exist for regular data issues to be resolved, about 78.4% of all respondents indicate that it is 'reasonable.'

Discussion and Recommendation:

1. Following from a very significant response in the research, the idea of a data steward (business) as data owner is well accepted in the Institution. The responsibilities of a data owner may have to work-shopped with the IS-Users as well the tools and mechanisms that are required to execute those functions.

The recognition and importance of 'information' within the traditional structures of 'IT' has been recognised by the creation of the position of a Chief Information Officer ('CIO') in the Institution where a merge of IT (networks, systems development) and management information (the business reporting) has been taking place as one division.

2. Who the data owners are, needs to be formalised through the communication structures in order for people to know how to route a data problem. There are a

significant number of users that ‘work around data’ or ‘try to fix it themselves’. Organisational agility should be vastly improved by improved internal communication structures to facilitate the routing of issues to data owners that on their turn may invoke service level agreements if the problem had been caused by third parties.

3. The inclusion of performance management via the inclusion of data quality in job profiles should be urgently considered for implementation.

4. In terms of data quality issues related to data from third parties that should not be attended to in the Institution, the implementation of service level agreements vis a vis those providers should be urgently considered.

5.2.3 OBJECTIVE 3 : PRACTICE – DATA QUALITY PRACTICES THAT MAY SUPPORT / INHIBIT DATA QUALITY IMPROVEMENT

Literature vs Fieldwork

	Literature	Fieldwork
1	Relating to the use of and faster access to reports and data quality (question 12), the use of management information and data workshops are important and represent an important lever to detect data problems (Braa, Heywood & Sundeep, S (2012). The use of reports is advocated in terms of understand the dimensions of data quality better, provide feedback to users to view and correct data to, inter alia, protect their reputations and see how data may be affecting outcomes that is important to them.	A large number of IS-Users indicate that faster access or the use of management (information) reports may have a positive effect on data quality improvement. While literature underscores reports as a lever vis a vis data quality, there appears to be under-utilisation of management reports at the Institution.
2	Regarding robustness of operational processes and data quality ((question 13), processes underpinning operations need to be robust enough	IS-Users indicate that, in terms of the quality of processes that contributes to data quality, Data Steward (Business) expressed a higher level

	to support data quality (McKnight, 2009).	of confidence in the quality of operational or business processes than other IS-Users while a significant number acknowledge some weaknesses.
3	Data quality correction and monitoring systems should exist with 'levels of tolerance' built into those systems.	IS-Users overall report a high level of awareness of an institutional system for data correction across the spectrum of Data Custodians, Data Stewards (Business) and Data Stewards (Technical).

Discussion and Recommendation:

1. In terms of the use of reports as a lever vis a vis data quality, there appears to be 'under-utilisation' of reports and 'dilution' by other factors. This 'dilution' that have been reported as among others, 'users being too busy' as reported in equal measure by each of the groups of IS-Users. Attempts need to be made to ensure that a platform exists for management reports to be available in real-time.
2. Following from a very significant response that indicate knowledge of systems that exist to clean data, there is still up to a third of the respondents (inclusive of data custodian and data stewards (technical) that appear ignorant. The formalisation of a real time data auditing 'exception' interface and data quality monitor should provide a feedback mechanism to sensitize users and facilitate a slow-down in data decay.

5.2.4 OBJECTIVE 4 - COST OF DATA QUALITY

Literature vs Fieldwork

	Literature	Fieldwork
1	Data quality problems lead to considerable cost to organisations in terms of financial waste and inefficiency (question 14). The	IS- Users at the Institution reported moderate to significant time spent on data quality activity. The costs could very likely be higher if the whole

	literature distinguishes between direct and indirect data quality costs. Data quality is significant organisations in monetary terms. Cost models show that cost of poor data decreases at when levels of data improvement 'peaks' but that cost of data quality maintenance increases.	population was included as well as the costs of functional areas eg Human Resources and Finance that are currently outside of the current scope. Cost of data quality was consequently deemed to be conservative.

Discussion and Recommendation:

While a benchmark has been provided for measuring the (current cycle) of data quality improvement, this exercise, due to its limited scope, provides by necessity only a 'tunnel' view having to measure cost in terms of labour only. While systems need to be in place to measure the reduction in costs that change commensurately with data quality improvements, system input controls as another cost need to be determined and incorporated.

Other costs that that need to be considered but difficult to quantify, may be the opportunity cost of the time devoted to data quality improvement, in other words what activities IS-Users may have spent their time on instead of correcting data. Other costs mentioned by Helfert & Eppler (2007), is related to data re-entry, costs due to increased turnaround due to data quality, costs of acceptance testing and assessment costs.

5.2.5 OBJECTIVE 5 : SUSTAINABILITY of DATA QUALITY IMPROVEMENT

Literature vs Fieldwork

In order to provide for a link to assess whether data quality improvement is sustainable, a few concepts relating to the sustainability of data quality were researched.

	Literature	Fieldwork
1	Data quality practices can play a role in shaping organisational culture facilitating a culture receptive to data quality issues (Ababaneh, 2010, Zu, Fredendal & Robbins, 2006)	IS-Users responded positively that data quality awareness can via a heightened levels of information literacy, affect the organisational culture of the Institution.
2	Team work and in particular cross functional teamwork are important in diagnosing process based quality problems that may affect sustainable data quality improvement (Ababaneh, Ibid)	IS-Users functioning as teams and workgroups on data quality projects are not functioning as efficiently as expected.
3	Training was important for W Edwards Deming to emphasise it as one of his 14 Points of Quality. Data quality assurance program that incorporate training programs should be considered to include quality improvement training (Pipino, Lee & Wang, 2002)	In terms of training, a significant number of IS-Users suggest that the levels of training vis a vis data quality at the Institution are inadequate and need to be addressed.
4	Adequate Skills and knowledge are necessary to ensure continuity in data quality activity.	IS-Users are sceptical towards the adequacy of institutional skills and knowledge base to support data quality improvement.

5	Change induced by internal and external forces can affect sustainability causing stress to maintaining quality as the change puts pressure on consistent application of policies and procedures, implying a trade-off between organisational change and maintaining quality (Loshin, 2000).	In terms of change, IS-Users are positively inclined towards data quality awareness as being a conduit to information literacy that should serve as a catalyst to sensitize users towards data quality.
6	Management support has been cited as a key driver in success of data quality efforts over the long term.	In terms of management supporting data quality, IS-Users responded neutrally to positively to the issue if Management if supporting data quality.
7	Continuous change is a challenge to the sustainability of data quality improvement.	IS-Users are negatively inclined towards being comfortable with change and implicitly express a 'wish' for a data environment that is stable.

Discussion and Recommendation:

Overall the results indicate that the opinions vis a vis sustainability of data quality improvement suggest a slightly skeptical view as illustrated by a mean value of 2.21.

1. The lack of training has surfaced more than once as a problem during the research. Training in the operative student administrative systems should be afforded urgent priority. Optimal staffing in the Institution for the function should be determined and training schedules should be arranged as strong opinions were expressed by all three constituent IS-Users.

2. A user needs analysis should be undertaken to ascertain what data quality skills and at what levels are required. While the data quality skills is non-negotiable at the level of Data Custodians, skills may need to be developed at

the level of the Data Stewards as well. In fulfilling their role as Data Owners some types of investigations may have to be ‘devolved’ and Data Stewards may would need requisite skills to undertake those tasks.

3. Management and leadership to drive data quality appears to a perceived problem. Although the research did not differentiate between top and middle management, it is proposed that there should be representation of management at Data Quality workshops to be continually informed of the data problems.

4. How teams work, perceived synergy, how the three groups of IS-Users that participated in the study work together, should be reconstituted. The Data Stewards (Technical) should have a closer involvement in data quality activity and they should have direct representation in the Data Quality Working Group (DQWG).

5. In terms of continuous change as a challenge to the sustainability of data quality, reporting elements and the data quality monitoring system that should exist, should be subject to regular reviews in the light of continuous change. A ‘Change Control’ item should be on the agenda of the DQWG to manage the effect of change.

5.3 FURTHER RESEARCH

It is the opinion that further research needs to be done in the root causes of data quality. The ‘5 why’ approach referred to in the possible D M A I C - to explore the cause-and-effect relationships underlying a particular problem, particularly underlying the behaviour underlying the actions of the IS-Users and to what interventions can be made to bring about altered behaviour.

5.3.1 RACI DIAGRAM – PROPOSED ACCOUNTABILITIES

The illustration below is a hypothetical representation of the RACI model of Wende and Otto (2011) in terms of data quality activity and roles at the moment at the Institution. As this is hypothetical only and the purpose of this research was not to interrogate this or to determine a 'optimal' model, further research is required to find a RACI model that represents the 'closest fit' of the roles, responsibilities and data role players to the directives in the Data Quality Policy at the Institution.

APPLICATION OF THE 'RACI' TO UKZN

ACTIVITIES / ROLE PLAYERS	Exec Sponsor Registrar	Data Quality Working Group	Data Owner (Business data steward)	Data Custodian	Data Steward (technical)
1. Overall strategic responsibility for data quality and report data quality matters to the Executive	A	R	I	C	I
2. The formulation of a overall data quality strategy, action plans and Identifies and prioritizes data quality initiatives, makes recommendations and motivates for funding	A	A/R	C	C	C
3. The promotion of a culture of data quality across the institution, monitoring and measuring functional data, implementing systems to identify and correct data errors, as well as to measuring and monitoring data quality	A	A	C	A/R	C
4. The correction of data quality in all functional areas, the formulation of procedure manuals for the capturing, storing and maintenance of data in the relevant transactional databases	A	I	A/R	C	I
5. Information Systems Support	I	I	C	A/C	A/R
R = Responsible, A = Accountable, C=Consulted, I = Informed					

Figure 5 : RACI Chart adapted from Wende & Otto (2011)

5.3.2 A PROPOSED ‘D M A I C’ APPROACH

While an Institutional Data Quality policy does exist, and this research has hopefully contributed in terms of benchmarking costs and defining roles and responsibilities, a further ‘structured improvement procedure’ could be proposed in terms of developing a strategy to ‘embody’ the ‘deliverables’ associated with a data quality improvement programme.

The Define-Measure-Analysis-Improve-Control (DMAIC) is one approach within the Six Sigma set of tools representative of a formal, strict process to undertake improvement projects according to Pande et al. (2000) in Zu, Fredendall and Robbins (2006).

A proposed DMAIC approach, to further give form to the current data quality improvement initiative may consist of the following components or building blocks that may assist in expanding the thinking to support the formal basis of a data quality improvement regime.

A) DEFINING

- Implement a University-wide Data Quality Strategy exist that is endorsed by the Executive.
- Develop a Data Dictionary or Glossary exist that identify all production files from source systems and explain information terminology at a higher level.
- Develop Policy and Procedure documents exist as a guide to Data Capture.
- Develop a Data Audit System exist with Exception reports.

B) MEASURING

- Develop systems to measure data or processes to arrive at an assessment of performance.

- Apply data profiling to data quality in terms of determining null or missing values, data values or high or low boundary data.
- Undertaking a gap analysis to determine the ‘tension’ between what information users require and what the system provides.

C) ANALYSING

Undertake a ‘Root Cause Analysis’. By investigating root cause issues at all levels, interrogating “why” (formerly) on various and successive ‘deeper’ levels.

D) IMPROVING

- Improve processes by addressing or eliminating (root causes of) defects.
- Establish or determine roles and accountability in terms of data owners.
- Review business rules that underlie the data that has to be tested.
- Implement business rules into program and distribute to data owners for correction.
  - Implement change controls to ensure change in data elements and systems are incorporated in the data quality systems

E) CONTROLLING / FEEDBACK

- Encapsulate the data, processes, business rules governing that Data Area and its tables/fields into Policy and Procedure guides to serve as a guide to the capturer as well to the analyst.
- Create a platform for the effective dissemination and maintenance of Policy and Procedure documents.
- Incorporate the violation of business rules into an on-line Data Auditing system where recurring problems can be dealt with at data capture entry.
- Create a platform for effective dissemination and maintenance of errors in the Data Auditing System.

- Conceptualise, develop, implement operational reports to serve users general needs as well as serve as feedback on errors that cannot be 'trapped' with a business rule.
- Ensure that, if Peripheral systems exist, that they are 'synchronised' with the reporting infrastructure of the main administrative system.
- Adopt a 'connect and collaborate' attitude in interacting with users in the area of Data Auditing as well as in communicating information terminologies.
- By means of a Data Quality Scorecard, use metrics to monitor that data quality does not fall under a certain threshold or limits.

5.4 CONCLUSION AND RECOMMENDATIONS

This research has assessed perspectives towards the sustainability of data quality improvement at the University. Various shortcomings have been identified and recommendations made as how they can be remedied (sections 5.2.1 to 5.2.5) with the emphasis on;

- training in data quality
- performance management to support current operations
- support teamwork to sustain the current initiatives and use cross-functional teams to diagnose process-based quality problems
- implement service level agreements to enhance data quality from third party suppliers
- adopt incentives to reward work done in the area of data quality
- and develop feedback mechanisms such as data quality monitor to report on the condition of data quality in real time,
- propagate data ownership to enhance organizational agility to reduce '*work around'/ 'run around'* around time of IS-users
- a better understanding of the cost of data quality

In terms of cost of data quality, while the research has provided a benchmark for improvements in terms of cost but has limited the study to direct labour only, there are other data quality costs that need to be studied in order to get a ‘complete picture’ of the cost of data quality.

Over and above the proposals made in section 5.3 in terms of a D M A I C approach to supplement this research, as data quality is directly linked to data and information digested by the Institution, research needs also to be done not only from the perspective of the information system users but from information users themselves. A junction between perspectives and practical experience vis a vis data quality and an information needs analysis in an organisation may contribute significantly to enhancing the quality of the information that organisations need.

LIST OF REFERENCES

Ababaneh, I. (2010), The role of organizational culture on practising quality improvement in Jordanian public hospitals, *Leadership in Health Services*, Vol. 23 Iss: 3, pp.244 – 259

Agency for HealthCare Research and Quality (2010)
http://www.ahrq.gov/legacy/qual/pubrptguide2.htm
Accessed 15 October 2013

Anturaniemi, A (2012), Improving business intelligence data quality in sales insight area, MBA Dissertation, University of Applied Sciences, Haaga-Helia
https://www.theseus.fi/bitstream/handle/10024/49895/ONT_Ari_Anturaniemi_Thesis_public.pdf?sequence=1
Accessed 11 November 2013

Audit Commission (2007),
http://www.wao.gov.uk/assets/englishdocuments/ImprovingInformationToSupportDecisionMaking.pdf
Accessed 5 August 2013

Ballou, D.P. and Pazer, H.L. (1995), Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff, *Information Systems Research*, vol. 6, no. 1, pp. 51-72.

Barnier, B (2012), COBIT 5, An ISACA framework,
http://www.isaca.org/chapters2/New-York-Metropolitan/membership/Documents/2012-04-30%20Spring%20Conference-Meeting/3%20Barnier%20VBA%20COBIT5.pdf
Accessed 15 February 2014

Barone, D., Stella, F., Batini, C. (2010), Dependency discovery in data quality. *Proceedings of the 22nd international conference on Advanced information systems engineering* (CAiSE'10). Barbara Pernici (Ed.). Springer-Verlag, Berlin, Heidelberg, 53-67

Baruch, Y. & Holtom, B. C. (2008). Survey Response Rate Levels and Trends in Organizational Research. *Human Relations 61*(8),pp 1139–1160.

Batini, C., Cappiello, C., Francalanci, A., Maurino, A. (2009), Methodologies for data quality assessment and improvement, ACM Computing Surveys (CSUR) Vol 41 No 3, p 16

BIPM (2007), http://www.executionmih.com/data-quality/accuracy-consistency-audit.php
Accessed 4 May 2013

Braa, J., Heywood, A., Sundeep, S. (2012), Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania, Bulletin of the World Health Organization 2012;90:379-384. doi: 10.2471/BLT.11.099580
http://www.who.int/bulletin/volumes/90/5/11-099580/en/
Accessed 6 September 2013

Brackstone, G. (1999), Managing Data Quality in a Statistical Agency, *Survey Methodology*, Vol 25, pp. 139 - 149

Brennan, S.E., Bosch, M. Buchan, H. and Green, H. (2013), Measuring team factors thought to influence the success of quality improvement in primary care: a systematic review of instruments, *Implement Sci.* Vol 14 , pp. 1-17

Brunelli, M. (2012), Building an effective data governance framework
http://searchdatamanagement.techtarget.com/essentialguide/Building-an-effective-data-governance-framework
http://www.storediq.com/blog/2011/12/05/2012-the-year-for-information-governance/
Accessed 14 March 2013

Bryman, A. and Bell, E.(2007), Business research methods. 2nd ed.
Oxford: Oxford University Press, Canada Symposium

Cappiello, C., Francalanci, C., Pernici, B. (2004), Data quality assessment from the Users perspective, Proceedings from the 2004 international workshop on Information quality in information systems, pp 68 – 73

Carpenter, S. (2008), Studying team shared mental models, International Conference on the Pragmatic Web, Sept 28–30, 2008, Uppsala, Sweden.Copyright 2008 ACM, *Communications of the ACM*, Vol 39, pp 86-95

Chengular-Smith, I., Ballou, D.P. and Pazer, H.L. (1999). The impact of data quality information on decision making: an exploratory analysis, *IEEE Transactions on Knowledge and Data Engineering Vol 11*, No 6, pp 853-865.

Costello, T (2012), RACI—Getting Projects "Unstuck", ITPro, IEEE Computer Society.
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6171964

Accessed 15 February 2014

Dardick, G.S, SECAU, (2010), Cyber Forensics Assurance, Australian Digital Forensics Conferences, Originally published in the Proceedings of the 8th Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia, November 30th 2010
http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1076&context=adf&sei-redir=1&referer=http%3A%2F%2Fscholar.google.co.za%2Fscholar%3Fhl%3Den%26q%3DParker%25E2%2580%2599s%2BHexad%2BModel%2B%26btnG%3D%26as_sdt%3D1%252C5%26as_sdtp%3D#search=%22Parker%E2%80%99s%20Hexad%20Model%22
Accessed 15 February 2014

Dahlgaard, J.J. and Dahlgaard-Park, S.M. (2006), Lean production, Six Sigma quality, TQM and company culture, *The TQM Magazine*, Vol 18 Iss: 3, pp 263 - 281

Data Crunch Report (2011) – India http://www.gs1india.org/upload/CaseStudies/Data_Crunch_Report_2011.pdf
Accessed 13 July 2013

Data Crunch Report (2011) – UK
http://www.gs1uk.org/what-we-do/.../gs1_datacrunch_report.pdf
Accessed 14 July 2013

Data Crunch Report (2011) – Australia
http://www.gs1au.org/assets/documents/news_room/gs1/GS1_Australia_Data_Crunch_Report_23062011.pdf
Accessed 11 July 2013

DeAmicis, F., Barone D., Batini, C. (2006), An analytical framework to analyze dependencies among data quality dimensions. Proceedings of the 11th International Conference on Information Quality, November 10-12, 2006, Cambridge, USA., pp: 369-383.

De La Harpe, R. and Marshall, L. (2009), Decision making in the context of business intelligence and data quality, *SA Journal of Information Management,* Vol 11, No 2

De la Harpe, R. and Roode, D. (2004), An Actor – Network Theory Perspective on Data Quality in Medical Practices, *Studies in Communication Sciences Vol* 4, No 2, pp 69-84.

Dismute, W.S. (2009) *Data Governance: A Study of the Current State and Emerging Trends*, Masters Thesis, Information Science department, University of Arkansas, Little Rock.

Dravis, F. (2004), Data Quality Strategy, a Step by Step Approach, *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04*
http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202004/Papers/DQStrategy.pdf
Accessed 15 March 2013

Du Mars, R. (2008), Internal data quality standards not just a game for business users, http://searchdatamanagement.techtarget.com/feature/Internal-data-quality-standards-not-just-a-game-for-business-users
Accessed 3 April 2013

Dun and Bradstreet (2007), The Cost of Poor Data Quality, Joint research paper of D and B, Canada and Richard Ivey School of Business , University of Western Ontario www.dnb.co.in/Whitepaper.pdf
Accessed 29 April 2013

Dyche, J. and Levy, E. (2006), Customer Data Integration. John Wiley and Sons, New York

ECAR 18th Australasian Conference on Information Systems, 5-7 December 2007, Toowoomba, Editorial ISSN: 2042-6372

EDUCAUSE (1994), The Management of Information Access in Higher Education, *Cause, Cause / Effect*, Vol 17, No 1, Spring 1994

English, L. (1999), Improving Data Warehouse and Business Information Quality, John Wiley and Sons, New York

English, L. (2000) Achieving enterprise wide data quality, 2000, Trillium Software System p.18.

English, L. (2011), The Information Quality Management Maturity Grid, International Association for Information and Data Quality (IAIDQ), *October 2011: Originally published in IDQ Newsletter Vol 7 Issue 4*
http://www.iaidq.com/publications/doc2/english-2011-10.shtml
Accessed 14 April 2013

Eppler, H. and Helfert, M. (2007), a Classification and analysis of Data Quality Costs, *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*

Even, A., Shankaranarayanan, G. (2009), Utility cost perspectives in data quality management, *Journal of Computer Information Systems, Vol* 50, No 2, pp 127-135

Evgeniou, T. and Cartwright, P. (2005), Barriers to Information Management, *European Management Journa*l Vol. 23, No. 3, pp. 293–299

Ewell, P.T. (1989), Institutional characteristics and faculty / administrator perceptions of outcomes - an exploratory analysis, *Research in Higher Education,* Vol 30, pp 13-36

Fisher, C.W. and Kingma, B.R. (2001), Journal of Information Management, Vol 39, Issue No 2, pp 109 - 116

Flyvjberg, B, (2004), Five misunderstandings about case-study research, in Sale,C., Gobo, G., Gudrium, F., Silverman D. (Eds.). Qualitative Research Practice, Sage Publications
http://flyvbjerg.plan.aau.dk/MSFiveMis9.0SageASPUBL.pdf
Accessed 7 February 2014

French, W. L., & Bell, C. H. (1999). *Organization development* (6th ed.). Upper Saddle River, NJ: Prentice-Hall

Furlong, J., Oancea, A. (2005), Assessing Quality in Applied and Practice-based Educational Research
http://www.tlrp.org/capacity/rm/wt/campbell/docs/assessing_quality_shortreport_tcm6-8232.pdf
Accessed 15 June 2013

Galway, L.A. and Hanks, C.H. (1996), Data Quality Problems in Amy Logistics, Classification, Examples, and Solutions, Rand Corporation, Santa Monica

Galway, L.A., Hanks, G. (2011), Classifying Data Quality Problems, IAIDQ's Information and Data Quality Newsletter Vol 7, No 4, pp 1-3

Gharajedaghi, J. (2006), Systems Thinking, Third Edition: Managing Chaos and Complexity: A Platform for Designing Business Architecture, Butterworth-Heinemann, Burlington, MA

Giannoccaro, A., Shanks, G., Darke, P. (1999), Stakeholder Perceptions of Data Quality in a Data Warehouse Environment, Proc. 10, The Australian Conference on Information Systems

Goetz, H. (2011), Harnessing the Power of Feedback Loops, http://www.wired.com
Accessed 21 June 2013

Grek, S. and Ozga, J. (2008), Governing by Numbers? Shaping Education through Data, CES Briefing No 44, University of Edinburgh, http://www.ces.ed.ac.uk/PDF%20Files/Brief044.pdf
Accessed 16 May 2013

Grobler, M.M. and Dlamini, I.Z. (2010), Managing digital evidence – the governance of digital forensics, *Journal of Contemporary Management*, Vol 7, pp 1- 21

Groundwater-Smith, S. and Mockler, N. (2006) Research that counts: practitioner research and the academy, in Counterpoints on the Quality and Impact of Educational Research, *Special Edition of Review of Australian Research in Education*, No 6, pp. 105-117

Hamblin, D.J. and Phoenix, D.A. (2012), A model for managing data assurance in higher education, *Journal of Higher Education Policy and Management*, Vol. 34, Iss. 4, 2012

Harris, J. (2011), The Role of Data Quality Monitoring In Data Governance Aligning Data Quality Metrics with Business Insight, Kalido White paper, 2011
http://www.ocdqblog.com/storage/downloads/The%20Role%20of%20Data%20Quality%20Monitoring%20in%20Data%20Governance.pdf
Accessed 4 May 2013

Hassany, P., Panahy, S., Sidi, F., Affendey, L.S., Jabar, M.A., Ibrahim, M., Mustapha, A. (2013), Discovering Dependencies among Data Quality Dimensions: A Validation of Instrument. *Journal of Applied Sciences,* Vol 13, pp 95-102.

Haug, A. and ArlBjorn, J.S. (2011), Barriers to Master Data Quality, *Journal of Enterprise Information Management* Vol. 24 No. 3, pp. 288-303

Haug, A., Zachariassen, F., Van Liempd, D. (2011), The costs of poor data quality, *Journal of Industrial Engineering and Management*, Vol 4, No 2 pp 168 -193

Helfert, M., Zellner, G., Sousa, C. (2003), Data Quality Problems and Proactive Data Quality Management in Data-Warehouse-Systems

Heravizadeh, M., Mendling, J., Rosemann, M. (2009), Dimensions of Business Process Quality, Business Process Management Workshops, Lecture Notes in Business Information Processing Vol 17, pp 80-91
http://link.springer.com/chapter/10.1007%2F978-3-642-00328-8_8#
Accessed 4 March 2013

Hodkiewicz, M., Kelly, P., Sikorska, J., Gouws, L. (2006), A framework to assess data quality for reliability variables. In: Paper presented at the World Congress of engineering asset management, Gold Coast, Australia, 11–14 July 2006
http://www.ifm.eng.cam.ac.uk/automation/industrial_projects/IQAM/IQMaturity.pdf
Accessed 2 May 2013

Huang, K.T., Lee, Y.W., Wang, R.Y. (1999), Quality Information and Knowledge, *Information Quality*, 1, 3, pp 13/1-13/5.

Jesionowski, T. (2012), Data Steward and Data Custodian Definitions, Data G overnance Blog
http://www.primedataconsulting.com/blogPDC/data-governance/2010/03/03/data-steward-and-data-custodian-definitions
Accessed 24 October 2013

Jones, D. (2012), How to Improve Healthcare Data Quality through Policy and Procedure Management, DataQualityPro,
http://dataqualitypro.com/data-quality-pro-blog/2012/11/13/how-to-improve-healthcare-data-quality-through-policy-and-procedure-management
Accessed 13 November 2013

Jung, W., Olfman, L., Ryan, T., Park, Y. (2005), An experimental study of the effects of contextual data quality and task complexity on decision performance, *Proceedings of IEEE International Conference on Information Reuse and Integration*, August 15-17, pp 149-154.

Juran, J.M. and Godfrey, A.B. (1999), Juran's Quality Handbook. 5 ed, New York: McGraw-Hill

Karel, R. (2007), Data Governance: What Works and What Doesn’t For Information and Knowledge Management Professionals, Forrester Research (Webinar)

Karr, A.F., Sanil, A.P., Banks, D.L. (2006), Data Quality : a Statistical perspective, *Statistical Methodology*, Vol 3, No 137 – 173

Khatri, V. and Brown C.V., (2010), Designing data governance, *Communications of the ACM, Vol* 53, No 1 pp 148-152

Kovac, R., Lee, Y.W., Pipino, L. (1997), Total Data Quality Management: The Case of IRI, IQ, pg 63-79, MIT

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N. (2011), Big data, analytics and the path from insights to value, *MIT Sloan Management Review*, Vol 52 Iss 2, pp 21-32

Lee Y. and Strong, D. (2004), Knowing-why about data processes and data quality, *Journal of Management Information Systems* Vol 20 No3, pp 13–39

Lee, Y.W. (2004), Crafting Rules: Context-Reflective Data Quality Problem Solving, *Journal of Management Information Systems,* Vol. 20, No. 3, pp. 93–119

Lee, Y.W., Pipino, L., Strong, D., Wang, R. (2004), Process Embedded Data Integrity, *Journal of Database Management*, Vol 15, No 1, pp 87-103

Lee, Y.W., Pipino, L.L., Funk, J.D., Wang, R. (2006), Journey to Data Quality, MIT Press

Loshin, D (2000),

Loshin, D. (2001a), BOOK : Enterprise Knowledge Management – The Data Quality Approach, Elsevier Inc Accessed 4 July 2013

Loshin, D. (2001b), Enterprise knowledge management: the data quality approach (on-line) - Overview and Framework for Data and Information Quality Research, *Journal of Data and Information Quality*, Volume 1 Issue 1, June 2009.

Loshin, D. (2010), Managing Data Quality in the Face of Organizational Change, BEyeNetwork
http://www.b-eye-network.com/view/13598
Accessed 13 June 2013

Loshin, D. (2011), The Practitioners Guide to Data Quality Improvement, BOOK ONLINE
http://books.google.co.za/books?id=B3zd4GCAWeYCandq=SLA#v=onepageandqandf=false
Accessed 4 March 2013

Loshin, D. (2012), Evaluating the Business Impacts of Poor Data Quality, Knowledge Integrity Incorporated / Business Intelligence Solutions

Loshin, D. (2012), The Secret to an Effective Data Quality Feedback Loop, Data Quality Pro Expert Journal, http://dataqualitypro.com/data-quality-pro-blog/2012/5/31/the-secret-to-an-effective-data-quality-feedback-loop Accessed 2 February 2013

Lucas, A. (2011), Corporate Data Quality Management in Context (Research in Progess), https://www.repository.utl.pt/bitstream/10400.5/2271/1/ICIQ2010_CDQM.pdf Accessed 26 May 2013

Lwanga Yonke, C., Walenta, C., Talburt, J.R. (2011), The Job of the Information/Data Quality Professional, International Association for Information and Data Quality (IAIDQ) http://www.iaidq.com/publications/doc2/medved-2013-04.shtm Accessed 3 March 2013

Madnick, S., Wang, R.Y. (1992), Introduction to Total Data Quality Management (TDQM) Research Program, TDQM-92-01, Total Data Quality Management Program, MIT Sloan School of Management

Madnick, S.E., Wang, R.Y., Lee, Y.W., Zhu, H, (2009), Overview and Framework for Data and Information Quality Research, ACM Journal of Data and Information Quality, Vol , No 1, pp 1-2

Madnick, S. and Lee, Y. (2009a), Editorial for the Inaugural issue of the ACM Journal of Data and Information Quality, *ACM Journal of Data and Information Quality*, Vol 1, No 1, pp 1 – 6

Madnick, S and Lee, Y (2009b), Where the JDIQ Articles come From: Incubating Research in an Emerging Field, *ACM Journal of Data and Information Quality*, 1, 3, pp 13/1-13/5

Mare, D. (1993), IC Position Paper : Managing Data Quality, Shell Internationale Petroleum Maatschappij B.V, The Hague

Marsh, R. (2005), Drowning in dirty data? It’s time to sink or swim: a four-stage methodology for total data quality management, *Database Marketing and Customer Strategy Management*, Vol 12 No 2, pp. 105-12.

Marsh, J.A., Pane, J.F., Hamilton, L.S. (2006), Making Sense of Data-Driven Decision Making in Education : Evidence from Recent RAND Research, Rand Corporation http://www.rand.org/pubs/occasional_papers/OP170.html

Accessed 7 May 2013

Masayna, V. (2006), A framework for linking data quality efforts to organisational key performance indicators, PhD proposal, University of South Australia.
http://ura.unisa.edu.au/R/?func=dbin-jump-fullandobject_id=60942
Accessed 6 July 2,

Masayna, V., Koronios, A., Gao, J., Gendron, M. (2007), Data Quality and KPIs: A link to be Established, paper presented at the Second World Congress on Engineering Asset Management and the Fourth International Conference on Condition Monitoring (WCEAM-CM2007), Harrogate, United Kingdom, 11-14 June

Masayna, V., Koronios, A., Gao, J. (2009), A Framework for the Development of the Business Case for the Introduction of Data Quality Program Linked to Corporate KPIs and Governance, Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology
http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05361870
Accessed 4 May 2013

Masoumeh, P, Basri, R, Sadegh, BM (2011),*Needs and Necessities for Total Quality Management in Educationedit*. Paper presented at the International conference On Teaching and Learning Education 2011 Universiti Tenaga Nasional, Malaysia, 20-22 June 2011.

Mathieu, R. and Khalil, O. (1998), Data quality in database systems course, *Data Quality Journal Vol* 4, No 1, pp. 1-12

Maydanchik, A. (2007), Data Quality Assessment, Thirteen causes of enterprise data quality problems
http://searchdatamanagement.techtarget.com/feature/Thirteen-causes-of-enterprise-data-quality-problems
Accessed 4 June 2013

McDonough, T. (2012), Higher Education Administrators' Perceptions Of The Academic Quality Improvement Project As Compared To The Program To Evaluate And Advance Quality Within The North Central Association Of Colleges And Schools, PhD Dissertation
https://etd.ohiolink.edu/ap:10:0::NO:10:P10_ACCESSION_NUM:kent1333129303
Accessed 1 August 2013

McGilvray, D. (2006), Data Governance: A Necessity in an Integrated Information World, http://www.information-management.com/ issues/ 20061201/1069951-1.html
Accessed 15 July 2013

McGilvray, D. (2010)(BOOK), Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (on line)MIT Information Quality Industry Symposium, July 15-17, 2009
http://mitiq.mit.edu/IQIS/Documents/CDOIQS_200977/Papers/01_02_T1D.pdf
Accessed 3 February 2013

McKnight, S. (2009), 7 Sources of Poor Data Quality, http://www.melissadata.com/enews/articles/0611/2.htm
Accessed 4 June 2013

Microsoft Academic Search, http://academic.research.microsoft.com/Keyword/9056/data-quality-management
Accessed 12 September 2013

Murphy, G.D. (2009), Improving the quality of manually acquired data: Applying the theory of planned behaviour to data quality, *Reliability Engineering and System Safety,* Vol 94, pp 1881–1886

Mutch, A. (1996) No such thing as …information resource management, *Management Decision,* Vol 34, No 7, pp 58–62

O'Brien,T.(2011),
http://dataquality.computing.dcu.ie/thesis/OBrien_Tony_2011.pdf
DBA dissertation
Accessed 15 May 2013

Olson, J.E. (2013), Data Quality: The Accuracy Dimension http://books.google.co.za/books/about/Data_Quality.html?id=x8ahL57VOtcCandredir_esc=y
Accessed 5 May 2013

O Neal, K. (2011), Coordinating a Successful Data Governance and Master Data Management Strategy Across the Enterprise, IAIDQ Publications
http://iaidq.org/publications/oneal-2013-07.shtml
Accessed 13 May 2013

O'Neal, K. (2012), Practical Formulas for Estimating the value of Quality Data, International Association for information and Data Quality
http://iaidq.org/publications/doc2/oneal-2012-03.shtml
Accessed 17 July 2013

O' Riain, C. and Helfert, M. (2005), An Evaluation of Data Quality related problem patterns in health care information systems, School of Computing Dublin City University, Ireland

Orna, E. (1996), Valuing information - problems and opportunities, in Best DP (Ed), The Fourth Resource : Information and its Management" , Aslib / Gower, Aldershot, pp 18-40

Otto, B. (2011), A Morphology of the organisation of data governance, *ECIS 2011 Proceedings.* Paper 272

Park, V. and Datnow, A. (2009). Co-constructing distributed leadership: District and school connections in data-driven decision-making, *School Leadership and Management* Vol 29, No 5, pp 477-494.

Parker, M., Stofberg, C., De la Harpe, R., Venter, I., Wills, G. (2006) Data Quality : How the flow of data influences data quality in a small to medium medical practice - In, *Community informatics for developing countries: Understanding and organising for a participatory future information society, Cape Town, South Africa, 31 Aug - 02 Sep 2006.*

Porter, M.E. (1979), How Competitive Forces Shape Strategy, *Harvard Business Review*, March/April 1979.

Payyazhi, J., Syed, M., Jamal, H. (2011), Aligning change deployment: a Balanced Scorecard approach, *Measuring Business Excellence*, Vol. 15, No 3 pp. 63 – 85

Pipino, L., Lee, Y., Wang, R. (2002), Data Quality Assessment, *Communications of the ACM* Vol 45, No 4, pp 211–218 *Quality*, 132–141.
Porter, M.E. (1985), Competitive Advantage, Free Press, New York

Pulzello, F. and Bhavsar, S. (2011),The Year for Information Governance,
http://www.storediq.com/blog/2011/12/05/2012-the-year-for-information-governance/
Accessed 26 May 2013

Redman, T.C. (1998), The Impact of Poor Data Quality on the Typical Enterprise, *Communications of the ACM*, Vol. 41, No. 2, pp. 79-82

Redman, T.C. (2001), Data Quality, The Field Guide. Boston: Digital Press. 241

Redman, T.C. (2004), Data: An Unfolding Quality Disaster, *DM Review Magazine*

SASQAF (2010), http://unstats.un.org/unsd/dnss/docs-nqaf/SouthAfrica-SASQAF_Edition_2.pdf

Scannapieco, M., Paolo, M., Batini, C. (2005), Data quality at a glance, *Datenbank-Spektrum* Vol 14, pp 6-14

Schwolow, S and Jungfalk, M (2009), The Information Value Chain : Strategic Information Management for Competitive Advantage, Copenhagen Business School
http://www.scribd.com/doc/46315789/Information-Value-Chain
Accessed 2 February 2013

Shankaranarayan, G., Ziad, M., Wang, R.Y. (2003), Managing Data Quality in Dynamic Decision Environments: An Information Product Approach. *Journal of Database Management*, Vol 14, No 4, pp. 14-32.

Shanks, G. and Darke, P. (1998a) Understanding Data Quality in a Data Warehouse, *The Australian Computer Journal*, Vol 30, pp 122-128.

Sheng, Y. and Mykytyn, P. (2002), Information technology investment and firm performance: A perspective of Data Quality, *IQ*, pp 132-141, MIT

Sidi, F., Panahy, S., Affendey, L.S., Jabar, M.A., Ibrahim, H., Mustapha, A. (2012), Data Quality : A survey of Data Quality Dimensions, International Conference on Information Retrieval and Knowledge Management (CAMP)

Slone, J.P. (2006), Information quality strategy: An empirical investigation of the relationship between information quality improvements and organisational outcomes, Ph.D. dissertation, Capella University
http://web.mit.edu/smadnick/www/IQ%20Dissertations/Skip.Slone.Dissertation.pdf
Accessed 27 May 2013

Sousa, S.D., Nunes, E., Lopes, I.S. (2011), Data Quality Assessment in Performance Measurement, Proceedings of the World Congress on Engineering 2012 Vol III, WCE 2012, July 4 - 6, 2012, London, UK
http://repositorium.sdum.uminho.pt/bitstream/1822/19894/1/WCE2012_pp1530-1535.pdf
Accessed 13 February 2013

Strong, D.M., Lee, Y.W., Wang, R.Y. (1997), Data quality in context, *Communication of the ACM,* Vol 40, Nov 5 , pp 104–108

Sy-Aves, L. (2007), The Effectiveness of Data Provision through Quality Management System: Retooling for the Institutional Monitoring and Evaluation for Quality Assurance in Higher Education (IQUAME), 10th National Convention on Statistics (NCS) EDSA

Tavakol, M and Dennick, R (2011), Making sense of Cronbach's alpha, *International Journal of Medical Education*, Vol2, pp 53-55

Tayi, G.K. and Ballou, D.P. (eds) (1998) , Examining Data Quality, *Communication of the ACM,* Vol. 41, No. 2

Tejay, G., Dhillon, G., Chin A.G. (2006), Data Quality Dimensions for Information Systems Security : A Theoretical Exposition (Invited Paper), *IFIP International Federation for Information Processing,* Volume 193, 2006, pp 21-39

Terenzini, P.T. (1993). On the nature of institutional research and the knowledge and skills it requires. *Research in Higher Education*, Vol 34, No 1, pp 1-10

Tull, P. (1997), *Managing Data Quality within BT - A Feasibility Study* http://www.olney.demon.co.uk/new/MBADIS10.doc, Accessed 14 April 2013

Tyche, J. (2007), A Data Governance Manifesto: Designing and Deploying Sustainable Data Governance http://isites.harvard.edu/fs/docs/icb.topic1167442.files/Day%201/T3_13_WP_JillDyche_DataGovernance_June07.pdf Accessed 14 June 2013

UKZN Strategic Plan, 2013 http://www.ukzn.ac.za/aboutus/strategicplan.pdf

UKZN "Data Quality Principles and Guidelines", 2011

UKZN "Data Quality Terms of Reference", 2011

Wand, Y. and Wang, D. (1996) Anchoring Data Quality Dimensions in Ontological Foundations, *Communications of the ACM*, Vol 39, No 1 pp 96 – 95

Wang, R.Y., Storey V.C., Firth, C.P. (1995), A Framework for Analysis of Data Quality Research, *IEEE transactions on Knowledge and Data Engineering*, Vol 7, Issue 4, pp 623 – 640

Wang, R.Y., Strong, D.M.(1996), Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information System*, Vol. 12, No. 4, pp 5 – 33

Wang, R. (1998), A product perspective on total data quality management, *Communications of the ACM,* Vol 41, No 2, pp 58–65

Wang, R.Y., Lee, Y.W., Pipino, L.L., Strong, D.M. (1998), Manage Your Information as a Product. *Sloan Management Review,* Vol 39 No 4, pp 95-105.

Weber, K., Otto, B., Osterle, H. (2009), One size does not fit all—a contingency approach to data governance, *ACM Journal of Data Inform. Quality* 1, 1, Article 4

Weill, P. and Ross, J. (2004), *How Top Performers Manage IT Decision Rights for Supeior Results.* Cambridge, MA: Harvard Business School Press.

Wende, K. and Otto, B. (2011), A contingency approach to Data Governance http://mitiq.mit.edn u/iciq/PDF/A%20CONTINGENCY%20APPROACH%20TO%20DATA%20 GOVERNANCE.pdf
Accessed 13 July 2013

Winningham, C. (2011), Data Quality Programs in Large Financial Institutions, DAMA-2011, New Jersey

Woodall, P. and Parlikad, A. (2010), Benchmarking information quality performance in asset intensive organisations in the UK In: 5th World Congress on Engineering Asset Management (WCEAM-AGIC 2010), 25-27 October 2010, Brisbane, Australia (Unpublished)

Wordpress (2011), Advantages And Disadvantages Of Quantitative Research. http://picardsflute.wordpress.com/2011/01/12/advantages-and-disadvantages-of-quantitative-research/
Accessed 7 November 2013

Xu, H., Nord, J.H., Brown, H., Nord, G.D. (2002), Data Quality Issues in implementing an ERP environment, *Industrial Management and Data Systems,* Vol 102, No 1

Xu, H., Nord, J., Nord, G., Lin, B. (2003) , Key issues of accounting information quality management : Australian case studies, *Industrial Management and Data Systems,* Vo1 103, No 7, pp 461-470

Xu, H. (2009), Data quality issues for accounting information systems' implementation: Systems, stakeholders, and organizational factors, *Journal of Technology Research,* Vol 1, p 1

Yan, G., Olariu, S., Weigle, M.C. (2009), Providing Location Security in Vehicular Ad-hoc Networks, *IEEE Wireless Communications, Special Issue on On-the-Road Communications*, Vol. 16, No. 6, December 2009, pp. 48-55

Yin, R. (1984), Case study research: Design and methods, Newbury Park, CA: Sage.

Yin, R. (1989). Case study research: Design and methods (Rev. ed.). Newbury Park, CA: Sage Publishing

Zhu, H., Madnick, S.E., Lee, Y.W., Wang, R.Y. (2012), Data and Information Quality Research : Its Evolution and Future (Forthcoming in Computing Handbook Set, Chapman and Hall / CRC)
http://mitiq.mit.edu/Documents/Publications/Papers/2012/Madnick_2012_Data%20and%20Information%20Quality.pdf
Accessed 13 April 2013

Zu, X., Fredendal, L.D., Robbins, T.L. (2006), Organizational Culture and Quality Practices in Six Sigma, The 2006 Annual Meeting of the Academy of Management
http://om.aomonline.org/dyn/award/Best%20paper%202006.pdf
Accessed 12 May 2013

APPENDICES

APPENDIX 1 – Questionnaire

SECTION 1 : General details about you as a computer system(s) user / information worker

1. Are you employed as a:

- ○ manager,
- ○ non-manager

2. Which of the following computer system(s) do you use or are involved with most?

- ☐ ITS,
- ☐ SMS,
- ☐ Other

Other (please specify) ____________________

3. 3 How many years have you been involved in systems work?

- ○ <5 yrs,
- ○ 5 to 10yrs,
- ○ 11 to 15 yrs,
- ○ >15yrs

4. As a systems user, in what capacity are you employed?

- ○ System Administration,
- ○ User Access /Data User Support
- ○ Development,
- ○ College / School,
- ○ MI and Info Analysis,
- ○ Other

Other (please specify) ____________________

SECTION 2 : Data Quality Awareness, Experience and Practice (reflecting experience of the problem)

5. As part of your systems and/or other job related training, do you feel that the notion of data quality and its importance to our business has been adequately addressed when related to your area?

- Yes
- No

Other (please specify)

6. By what means has this knowledge or awareness been acquired?

- Training Courses
- Management Briefings
- Personal Job Experience
- UKZN Publications,
- Other

Other (please specify)

7. Are you aware of any Data Quality initiatives underway or having taken place?

- Yes
- No

8. Do you experience data quality problems as part of your daily work?

- Yes
- No

9. Amongst the dimensions of data quality below, which are in your opinion the most important to a data user. Please rank them (e.g. 1,4,3,2)

Accuracy

Completeness

Timeousness

Comparability

10. Current state of data quality

- Perfect
- Pretty good but a few problems

- Some data sources are in a mess
- There are serious problems overall
- Not sure

11. On the system(s) with which you are involved, please indicate your responsibilities with respect to respect to changing data records

- Creating Records,
- Reading Records,
- Updating Records,
- Deleting Records
- Capturing Records
- I do not change data records

Other (please specify)

12. 4 Processes to clean data

- Yes, the organisation has a system / division providing data quality 'alerts'
- We clean up data ourselves
- Not Sure

Other (please specify)

13. Use of Reports - Should faster access to data / information help to discover DQ problems and to do something about it

- Yes, to compensate for time saved in accessing info faster, users will spend time on DQ
- No, users too busy using info to worry about DQ
- Not sure

14. If yes, approximately what percentage of your week is involved in rework or resolving problems caused by bad data?

- I don't spend time on Data Quality
- Less Than 5 %
- 5 to 20 %
- 21 to 50 %
- > 50%

15. Do you feel that the operational processes in your areas of work support /underpin work with respect to Data Quality, are robust and of adequate quality ? (Choose the most important option)

- Wholly Adequate
- Some Minor Weaknesses Present

Useful As A Guide Only

Inadequate

Not Sure

16. When encountering a data quality problem, do you (choose most important option)

Always Correct Or Fix The Problem Yourself

Report The Problem To A Central Point (Helpdesk / Support Function / Data Quality Desk)

Report/Hand Problem To The Originator

Work Round Problem Best Way Possible To Complete Task

Escalate By Some Other Means

Not Applicable

Other

Other (please specify)

17. Where, in which role, in your opinion, does the accountability for bad data lie?

Data Stewards - Business (eg Data Owners in the areas of Student / HR / Finance)

Data Stewards - Technical (Data Stewards in ICS)

Data Custodian (Management Information Services in ICS)

Not sure

18. In terms of accountability for data and data quality success, do you believe that data quality responsibilities should be included in data owners' performance management agreements (KPA's)?

Reasonable

Not reasonable

Not sure

19. Data Problem Corrections – In order to minimise the time that data remains incorrect, do you believe that a 'Data Correction Window Period' (e.g. 24 / 48 / 72 hours) should exist for regular data errors to be fixed

Reasonable

Non Reasonable

Not sure

20. Elements that you see as Barriers to the adoption of Data Quality Initiatives (Choose the most important)

Management Does Not View Data Quality As Important

- Very Difficult To Present A Business Case
- No One Prepared To Lead The Initiative
- None One In The Organisation Appears To Care
- Unrealistic Expectations Are Often Set
- It Would Cost Too Much
- We Do Not Have The Right Skill Sets
- Don't Know

Prev Next

SECTION 3 : Problem Areas, Impact and Causes related to Data Quality

21. In which area do you feel the data problems lie? (Choose most important option)

- Accuracy, completeness, consistency, timeliness. (Processes)
- System's ability to model and manipulate data representing the real world (including user friendlyness)
- System's ability to output meaningful information
- Clarity of roles and responsibilities
- None of these
- Not sure

22. Impact of Poor Data – Please choose the most important impact element of poor data

- Financial Waste / Costs
- Reputation
- Efficiency
- Loss of Opportunities
- Other

Other (please specify)

23. Causes of Poor Data – Please choose the most important contributor to poor data

- System Issues
- Human / Behavioural / Error
- Lack of Training

Other

Prev Next

SECTION 4: An Organisational Assessment / Sustainability of Data Quality Improvement

This section seeks your overall perception of the business with respect to Data Quality and should not be specifically focused on your current work area.

The following questions/statements reflect key dimensions of data quality initiatives that are predictors of sustainability. Please indicate your views on a scale of than varies from 'Strongly Agree' to 'Strongly Disagree'.

24. Data quality initiatives / awareness will promote a data and information culture / information literacy

	Strongly Agree	Agree	Disagree	Strongly DisAgree
	Strongly Agree	Agree	Disagree	Strongly DisAgree

25. At this institution there are sufficient people to support the data quality initiative with the necessary skills and knowledge to guide the implementation

	Strongly Agree	Agree	Disagree	Strongly Disagree
	Strongly Agree	Agree	Disagree	Strongly Disagree

26. At this Institution training among data capturers / data owners is adequate to support the attainment of better data quality

	Strongly Agree	Agree	Disagree	Strongly Disagree
	Strongly Agree	Agree	Disagree	Strongly Disagree

27. The execution of data quality initiatives through dynamic structures (e.g. teams, networks and workgroups) at the Institution is working well

	Strongly Agree	Agree	Disagree	Strongly Disagree
	Strongly Agree	Agree	Disagree	Strongly Disagree

28. In this organisation individuals have become comfortable with change and do not seek stability

	Strongly Agree	Agree	Disagree	Strongly Disagree
	Strongly Agree	Agree	Disagree	Strongly Disagree

29. Management views data quality as important

Strongly Agree	Agree	Disagree	Strongly Disagree
Strongly Agree	Agree	Disagree	Strongly Disagree

30. There are enough people in the Institution to lead a data quality initiative

Strongly Agree	Agree	Disagree	Strongly Disagree
Strongly Agree	Agree	Disagree	Strongly Disagree

31. There are enough people in the Institution that cares about data quality

Strongly Agree	Agree	Disagree	Strongly Disagree
Strongly Agree	Agree	Disagree	Strongly Disagree

32. Expectations about achieving data quality improvement is reasonable

Strongly Agree	Agree	Disagree	Strongly Disagree
Strongly Agree	Agree	Disagree	Strongly Disagree

33. Data quality activity / processes will over the longer term be successful and sustainable

Strongly Agree	Agree	Disagree	Strongly Disagree
Strongly Agree	Agree	Disagree	Strongly Disagree

Prev Done

APPENDIX 2 – Research Areas by Question

Research Questions

The research can be encapsulated in the form of the following keywords and objectives**.**

A - Awareness

What is the nature of awareness and communication practices around data quality? Are these practices conducive to data quality improvement? Do structures exist to communicate issue around data quality (DQ) and manage data quality?

A.1 Awareness / Structures to communicate DQ problems

Question 5 - has the importance of date quality been incorporated into your work area?

Question 6 - how knowledge of DQ or DQ activity has been obtained (communication structures)

Question 7 - awareness of any DQ activity taking place

Question 8 - the user's experience with DQ problems – large data activity eg: data migration

Question 9 - rate the DQ dimensions

Question 10 - current state of DQ

Question 11 - the user's responsibility in terms of maintaining records (bio with q 1 to 5)

A.2 Awareness of i t o the nature of DQ i e barriers to information use, causes, impact = Questions 20, (cause), 22(impact), 21 and 23 (cause)

B – Accountability and Management

What are the perceptions vis a vis accountability – ROLES - Accountability and consequences

Question 16 - how DQ problems are routed
Question 17 - Where does accountability for DQ lie
Question 18 - Should DQ be incorporated into KPA's (linked performance management)
Question 19 - Minimize data decay – business value of data – should window periods be determined for data correction (linked to SLA's with users)

Perspectives of locii of control and commitment to accountability will be assessed in terms of linking data quality activity to performance management and improved turnover in terms of time spent on data quality improvement (section 2)

C How do are we in terms of Information Quality Problem Handling / Practice –
Processes and Process Quality / Structures to route DQ problems
Question 12 - does Institution have a 'system' to clean up data and so what is it
Question 13 - reports as a FB to assess data – data with metrics – i t o assessment
Question 15 - how does processes support you (robustness, quality, agility)

D – Cost of Data Quality

Can the cost of data quality practices be quantified? Is it significant? In order in inform this question, time (person-hours) and costs will determined that is devoted to data quality improvement in order determine a 'base' from which data quality can be improved and data quality costs can be monitored (Question 14)

E - Sustainability ('means difference')

Are there differences in perspectives toward the sustainability of data quality among the three groups of data quality stakeholders and are the differences significant?. Differences that may be found may point to levels of cooperation / synergy that may require intervention (Questions 24 to 33)

APPENDIX 3 – Cost of Data Quality

COST OF DATA QUALITY

Ordinal representation of time spent	Lower Range	Upper Range	Mid Point Salary Range	Lower - R Per Yr	Upper - R Per Yr	No of Respondents	Total Cost (Lower Range)	Total Cost (Upper Range)	Potential respondents	Lower - R Per Yr	Upper - R Per Yr
Less than 5%	0.01		R 300 585	R 3 006							
5 to 20%	0.05	0.2	R 300 585	R 15 029	R 60 117	4	R 60 117	R 240 468			
21 to 50 %	0.2	0.5	R 300 585	R 60 117	R 150 293						
Over 50 %	0.5	0.6	R 300 585	R 150 293	R 180 351						
						4	R 60 117	R 240 468	6	R 90 176	R 360 702

Ordinal representation of time spent	Lower Range	Upper Range	Mid Point Salary Range	Lower - R Per Yr	Upper - R Per Yr	No of Respondents	Total Cost (Lower Range)	Total Cost (Upper Range)	No of Outstanding Respondents	Total Cost (Lower Range) - Outstanding Respondent	Total Cost (Upper Range) - Outstanding Respondents
Less than 5%	0.01	0.05	R 249 350	R 2 494	R 12 468	20	R 49 870	R 249 350			
5 to 20%	0.05	0.2	R 249 350	R 12 468	R 49 870	10	R 124 675	R 498 700			
21 to 50 %	0.2	0.5	R 249 350	R 49 870	R 124 675	2	R 99 740	R 249 350			
Over 50 %	0.5	0.6	R 249 350	R 124 675	R 149 610	3	R 374 025	R 448 830			
						35	R 648 310	R 1 446 230	45	R 833 541	R 1 859 439

Ordinal representation of time spent	Lower Range	Upper Range	Mid Point Salary Range	Lower - R Per Yr	Upper - R Per Yr	No of Respondents	Total Cost (Lower Range)	Total Cost (Upper Range)			
Less than 5%	0.01	0.05	R 351 970	R 3 520	R 17 599	2	R 7 039	R 35 197			
5 to 20%	0.05	0.2	R 351 970	R 17 599	R 70 394	3	R 52 796	R 211 182			
21 to 50 %	0.2	0.5	R 351 970	R 70 394	R 175 985						
Over 50 %	0.5	0.6	R 351 970	R 175 985	R 211 182						
						5	R 59 835	R 246 379	9	R 107 703	R 443 482
						44	R 768 262	R 1 933 077	60	R 1 031 420	R 2 663 623

Potential responses = 'as if all 60 respondents indicated time spent on data quality

Note : The salary ranges do not constitute any particular level at the Institution. It is calculated as an average of all the different levels and grades that constituted each of the IS-User groups.

APPENDIX 4 – Descriptive Statistics – Section D

Statistics	*Question 25*	*Question 26*	*Question 26*	*Question 27*	*Question 28*	*Question 29*	*Question 30*	*Question 31*	*Question 32*	*Question 33*
Mean	1.54	2.204082	2.66	2.693878	2.836735	1.68	2.375	2.4	2.145833	2.04
Standard Error	0.081866	0.105057	0.129709	0.105962	0.084248	0.096637	0.105794	0.103016	0.103097	0.098809
Median	1.5	2.00	3	3	3	2	2	2	2	2
Mode	1	2.00	3	3	3	2	2	2	2	2
Standard Deviation	0.57888	0.735402	0.917183	0.741734	0.589736	0.683329	0.732963	0.728431	0.714279	0.698687
Sample Variance	0.335102	0.540816	0.841224	0.55017	0.347789	0.466939	0.537234	0.530612	0.510195	0.488163
Kurtosis	-0.67781	-1.05165	-0.54088	-0.22708	1.415376	-0.74352	-0.00577	0.11721	0.610274	0.190338
Skewness	0.496044	-0.34459	-0.41303	-0.07142	-0.5941	0.50657	0.285434	0.528005	0.509331	0.31952
Range	2	2	3	3	3	2	3	3	3	3
Minimum	1	1	1	1	1	1	1	1	1	1
Maximum	3	3	4	4	4	3	4	4	4	4
Sum	77	108	133	132	139	84	114	120	103	102
Count	50	49	50	49	49	50	48	50	48	50

APPENDIX 5 – Gatekeepers Letter

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

9 January 2013

Mr L Wood
Graduate School of Business
College of Law & Management Studies
Westville Campus
UKZN
Email: wood@ukzn.ac.za

Dear Mr Wood

RE: PERMISSION TO CONDUCT RESEARCH

Gatekeeper's permission is hereby granted for you to conduct research at the University of KwaZulu-Natal towards your postgraduate studies, provided Ethical clearance has been obtained. We note the title of your research project is:

"Factors and perspectives influencing accountabilities and the sustainability of data quality improvements in higher education"

It is noted that you will be constituting your sample by handing out questionnaires to select staff members and students in the Department of Information and Communication Services and School of Management, IT and Governance.

Data collected must be treated with due confidentiality and anonymity.

Yours sincerely

Professor JJ Meyerowitz
REGISTRAR

Office of the Registrar
Postal Address: Private Bag X54001, Durban, South Africa
Telephone: +27 (0) 31 260 8005/2206 **Facsimile:** +27 (0) 31 260 7824/2204 **Email:** registrar@ukzn.ac.za
Website: www.ukzn.ac.za

1910 - 2010
100 YEARS OF ACADEMIC EXCELLENCE

Founding Campuses: Edgewood | Howard College | Medical School | Pietermaritzburg | Westville

APPENDIX 6 – Approval

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

4 February 2013

Mr Lennard Robin Wood
Graduate School of Business and Leadership
Westville Campus

Dear Mr Wood

Protocol reference number: HSS/0034/013M
Project title: "Factors and perspectives influencing accountabilities and the sustainability of data quality improvements in higher education"

EXPEDITED APPROVAL

I wish to inform you that your application has been granted Full Approval through an expedited review process.

Any alteration/s to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form, Title of the Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment/modification prior to its implementation. In case you have further queries, please quote the above reference number. Please note: Research data should be securely stored in the school/department for a period of 5 years.

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully

Professor Steven Collings (Chair)

/pm

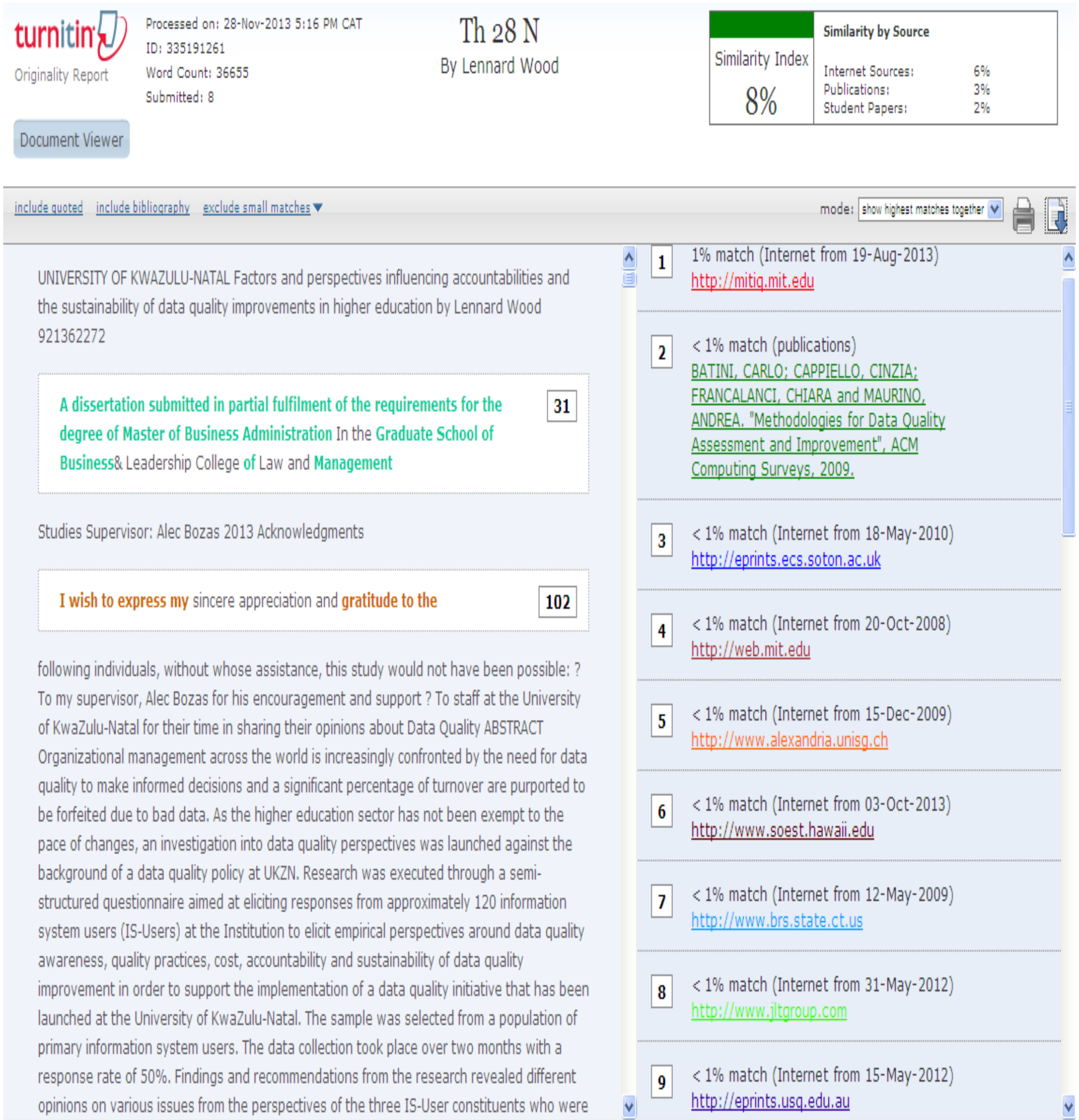
cc Supervisor: Mr Alec Bozas
cc Academic Leader: Dr SA Bodhanya
cc School Admin.: Ms Wendy Clarke

Professor S Collings (Chair)
Humanities & Social Sc Research Ethics Committee
Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban, 4000, South Africa
Telephone: +27 (0)31 260 3587/8350 **Facsimile:** +27 (0)31 260 4609 **Email:** ximbap@ukzn.ac.za / snymanm@ukzn.ac.za

Founding Campuses: Edgewood Howard College Medical School Pietermaritzburg Westville

INSPIRING GREATNESS

APPENDIX 7 – TurnItIn Summary



turnitin Originality Report

Processed on: 28-Nov-2013 5:16 PM CAT
ID: 335191261
Word Count: 36655
Submitted: 8

Th 28 N
By Lennard Wood

Similarity Index: 8%

Similarity by Source
Internet Sources: 6%
Publications: 3%
Student Papers: 2%

Document Viewer

include quoted · include bibliography · exclude small matches · mode: show highest matches together

UNIVERSITY OF KWAZULU-NATAL Factors and perspectives influencing accountabilities and the sustainability of data quality improvements in higher education by Lennard Wood 921362272

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Business Administration In the Graduate School of Business& Leadership College of Law and Management [31]

Studies Supervisor: Alec Bozas 2013 Acknowledgments

I wish to express my sincere appreciation and gratitude to the [102]

following individuals, without whose assistance, this study would not have been possible: ? To my supervisor, Alec Bozas for his encouragement and support ? To staff at the University of KwaZulu-Natal for their time in sharing their opinions about Data Quality ABSTRACT Organizational management across the world is increasingly confronted by the need for data quality to make informed decisions and a significant percentage of turnover are purported to be forfeited due to bad data. As the higher education sector has not been exempt to the pace of changes, an investigation into data quality perspectives was launched against the background of a data quality policy at UKZN. Research was executed through a semi-structured questionnaire aimed at eliciting responses from approximately 120 information system users (IS-Users) at the Institution to elicit empirical perspectives around data quality awareness, quality practices, cost, accountability and sustainability of data quality improvement in order to support the implementation of a data quality initiative that has been launched at the University of KwaZulu-Natal. The sample was selected from a population of primary information system users. The data collection took place over two months with a response rate of 50%. Findings and recommendations from the research revealed different opinions on various issues from the perspectives of the three IS-User constituents who were

1. 1% match (Internet from 19-Aug-2013) http://mitiq.mit.edu
2. < 1% match (publications) BATINI, CARLO; CAPPIELLO, CINZIA; FRANCALANCI, CHIARA and MAURINO, ANDREA. "Methodologies for Data Quality Assessment and Improvement", ACM Computing Surveys, 2009.
3. < 1% match (Internet from 18-May-2010) http://eprints.ecs.soton.ac.uk
4. < 1% match (Internet from 20-Oct-2008) http://web.mit.edu
5. < 1% match (Internet from 15-Dec-2009) http://www.alexandria.unisg.ch
6. < 1% match (Internet from 03-Oct-2013) http://www.soest.hawaii.edu
7. < 1% match (Internet from 12-May-2009) http://www.brs.state.ct.us
8. < 1% match (Internet from 31-May-2012) http://www.jltgroup.com
9. < 1% match (Internet from 15-May-2012) http://eprints.usq.edu.au