



HAL
open science

Data Governance: A Challenge for Merged and Collaborating Institutions in Developing Countries

Thandi Charmaine Mlangeni, Ephias Ruhode

► **To cite this version:**

Thandi Charmaine Mlangeni, Ephias Ruhode. Data Governance: A Challenge for Merged and Collaborating Institutions in Developing Countries. 14th International Conference on Social Implications of Computers in Developing Countries (ICT4D), May 2017, Yogyakarta, Indonesia. pp.242-253, 10.1007/978-3-319-59111-7_21 . hal-01650060

HAL Id: hal-01650060

<https://inria.hal.science/hal-01650060>

Submitted on 28 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Data Governance: A challenge for merged and collaborating institutions in developing countries

Thandi C Mlangeni¹ and Ephias Ruhode²

¹Information Technology Department, Cape Peninsula University of Technology, South Africa

Email: charmza22@gmail.com

²Research, Innovations and Partnerships Department, Cape Peninsula University of Technology, South Africa

Email: RuhodeE@cput.ac.za

Abstract. Organisations now invest in ICT solutions to drive business activities and to provide the agility sought within changing environments. Owing to many reasons including inadequate financial resources, organisations in developing countries are characterised by mergers of two or more institutions. It means therefore that disparate systems with different data management schemes are merged or made to collaborate making access to quality data almost impossible. In turn, a level of inefficiency finds its way with potential to generate inaccurate, missing, misinterpreted and poorly defined information. This research is motivated by the need to investigate data governance challenges in institutions within developing countries that are characterised by complex dynamics rooted in merged and collaborating environments. The study has been empirically scoped to explore data governance challenges in a large university of technology in the Western Cape Region of South Africa as a developing country. The challenges with regards to ICT and data governance are equally applicable in higher education institutions as they do in business organisations. Higher education institutions have a growing ICT infrastructure used in everyday activities and online functionality, making them prone to data problems. Challenges related to data management in universities are a lot more pronounced in universities which were established through the merging of independent institutions and also those that exchange data through collaborations. Thematic analysis has been employed within the theoretical lens of two models, contingency model (Wende & Otto, 2007) and the data governance decision domain model (Khatri & Brown, 2010). Analysis of data through the two models led to the development of a data governance framework applicable to the case under study and deemed to apply to any organisation in the same context. Challenges related to data principles, data access, data quality, data integration, metadata, data lifecycle, and design parameters emerged as the main findings from the study. Since the institution under study was established through a merger of independent institutions, the findings were deemed to be applicable to many other institutions where mergers and collaborations characterise their environment.

Keywords: Data Governance, Data Quality Management, Higher Education, Higher Education Institution, ICT, Cape Peninsula University of Technology, South Africa

1 Introduction

Organisations, both private and public, are attempting to construct a paradigm for data quality management in the backdrop of ubiquity of information and data. Godfrey et al. (1997) portray information as an asset with economic value. This infers that looking after data has the potential to bring efficiency in the running of the organisation. To this effect, many studies have identified data governance as a discipline that can address data quality issues (Korhonen, et al., 2013). The pervasive use of IT in organisations mandates IT governance as a corporate imperative (IoD, 2009:14). IT governance, under which data governance falls, is defined by Van Grembergen (2004:1) as:

“... an integral part of corporate governance and consists of the leadership, organisational structures and processes that ensure that the organisation’s IT sustains and extends the organisation’s strategy and objectives”

This paper recognises that organisations in many sectors in developing countries face challenges in implementing data governance principles. An organisation in South Africa is examined to empirically identify the data governance challenges from a merger environment. Owing to many reasons including inadequate financial resources, organisations, especially public institutions in developing countries are characterised by mergers of two or three institutions. Merging of organisations, as has been witnessed in the South African higher education landscape since 2003, results in data governance challenges. Just like business organisations, universities are concerned about brand perceptions, business processes and human presence in the IT ecosystems (people, process and technology). Data repeatedly used across various business processes in universities mostly originate from these entities: students, classes, faculty, campus, facilities, location and employees. This data is often dispersed among units, departments or divisions (Drucker, 2005:102) and, therefore, a level of inefficiency finds its way with potential to generate inaccurate, missing, misinterpreted and poorly defined information (Redman, 2005:1).

This paper seeks to investigate the challenging obstacles faced by merged and collaborating institutions in obtaining clean, reliable, relevant data from their IT systems and how such electronic systems can be managed. These challenges are framed within the broader data governance paradigm. Under this background, the two critical questions which then emerge are:

What are the data governance challenges faced by organisations as a result of mergers and collaborations?

Which data governance framework can be adopted for merged and collaborating institutions?

The paper is organised as follows: the next section presents the literature review followed by the methodology adopted in the study. Data collection and analysis

methods are explained in the methodology which is followed by a section on presentation of findings and implications to policy and practice before the paper ends with a summary and conclusion.

2 Literature Review

The literature review section presents the data governance concept followed by a presentation of the theories which underpinned the study.

2.1 Data governance concept

Organisations seek to break down the silos of data that result in poor quality of information which, in turn, lead to organisational costs, risks and wrong decisions (Korhonen et al., 2013:11). Most of them realise that their strategic initiatives depend on the quality of data and their ability to manage fast-growing volume of information. Bryant (2014) asserts that this can be achieved through data governance. Data governance can be defined as an organisational approach to data management that formalises a set of policies and procedures to encompass the full life cycle of data (Korhonen et al., 2013:11). It transforms an organisation's data, its management technology, who owns it and how it should be used (Russom, 2008:4). According to Russom (2012), a well-designed data governance programme should consist of both business and IT people. They must ensure information strategy and business strategy are aligned with the organisation's overall mission and strategy (Korhonen et al., 2013:14).

Most organisations deal with data quality problems emerging from both systematic and structural perspective. In seeking for solutions they develop new systems to replace old ones and, as a result, neglect to address the issue inherited from the old systems (Lee et al., 2006). Some authors agree that if organisations have poor data quality that is inappropriately integrated, business operations will continue to be afflicted with data deficiencies that will make it hard to use data (Fisher, 2009; Lee et al., 2006). According to Olson (2003), poor data management is said to cost many organisations some billions of dollars each year and a large portion of that cost is due to data quality inaccuracies. Redman (2001:45) suggests that 10% of organisations revenue is impacted by poor data quality. Both authors recognise the impact data quality can have on the organisation's profit. According to Redman (2008), data quality issues experienced by most organisations include the following:

- People cannot find the data they need
- Incorrect data
- Poor data definition
- Data privacy/data security
- Data inconsistency across sources
- Too much data and
- Organisational confusion.

It is the contention of researchers in this study, that data quality issues that are experienced by organisations in developing countries are at a far wider scale than organisations in developed countries. Reasons for this lie in limited resources that range from financial to skills. An empirical case in a developing country was therefore selected to identify the data governance challenges.

2.2 Theoretical lens

Two models have been identified and considered relevant in this study, namely the contingency model (Wende & Otto (2007) and the data governance decision domain model by Khatri and Brown (2010). Firstly, the relevance of the contingency theory (Wende & Otto (2007) lies in its design parameters which form the basis of a data governance framework. The two design parameters are organisational placement of decision-making authority and coordination of decision-making style. According to *ibid*, these two design parameters affect the configuration of data governance model as their value influences the assignment of responsibilities.

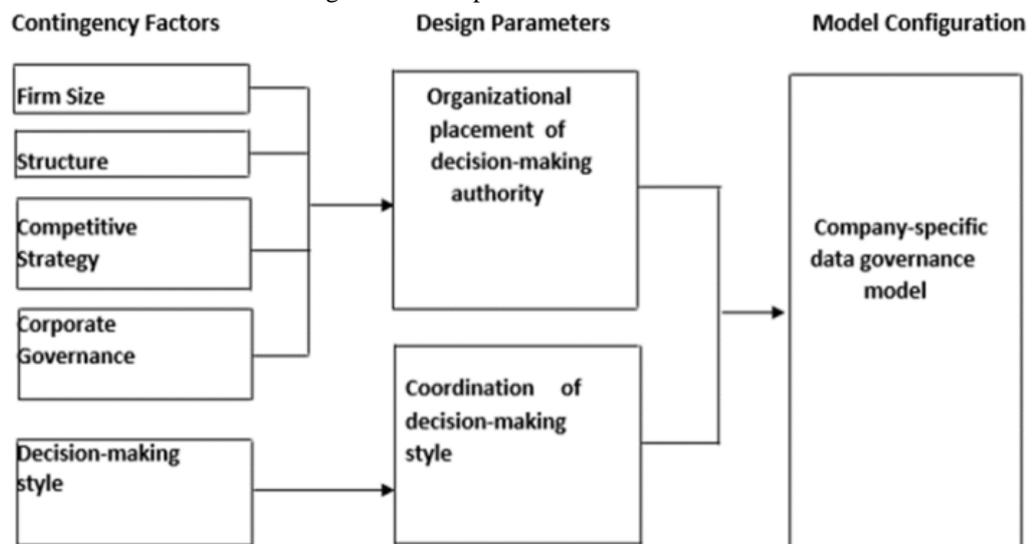


Figure 1: Contingency model for Data Governance (Source: Wende and Otto, 2007)

Secondly, Khatri and Brown (2010) used data governance decision domains to structure their data governance decision domains. They identified the following data governance components:

- Data Principles – establish the linkage with the business, by describing the business uses of data and ensuring data is treated as an enterprise wide asset.
- Data Quality – involves ensuring accuracy and integrity data that is always available for an enterprise.
- Metadata – describes what the data is about and provides a mechanism for a concise and consistent description of the representation of data.

- Data Access/Data authorisation – involves data security and explaining variety of ways in which a dataset can be accessed.
- Data life cycle – involves understanding how data is used, and how long it must be retained to minimise the total cost of storing over its life cycle

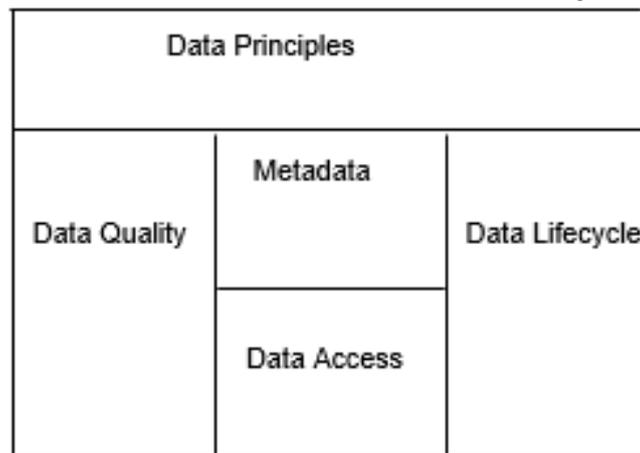


Figure 2: Data governance decision domains (Source: Khatri & Brown, 2010)

The data governance domain model was coalesced with the design parameters of the contingency model and used as one framework to guide data collection from the empirical case. These components from the data governance decision domains model need to be assigned to roles that will be accountable for them (from the design parameters of the contingency model), also referred as locus of accountability.

3 Methodology

The study was conducted at the Cape Peninsula University of Technology in South Africa. Considering the case rationale provided in the case description section, the researchers contend that findings from CPUT will be applicable to all other organisations which exist as a result of mergers and collaborations in developing countries. In the next section, the empirical case is presented, as well as the context of mergers within which the case is located.

3.1 Case description

In South Africa, the first decade of the 21st century was characterised by a massive transformation of the higher education landscape. Technikons were collapsed and merged with each other and in some cases with already existing universities. This transformation of higher education occurred “within the context of a formidable overall challenge of pursuing economic development (including restructuring economic relations to address inequitable historical patterns of ownership, wealth and income distribution), social equity and the extension and deepening of democracy simultane-

ously” (Moses, 2014). Post-1994 (the democratic dispensation), the new South African government sought to redress a myriad of economic and social challenges which had been created by the apartheid government. The South African society under apartheid government had been characterised by social, political and economic discrimination and inequalities of class, race, gender, institutional and spatial nature (Badat, 2010). Like all other sectors of the economy, the higher education landscape underwent a major restructuring and reconfiguration in the first decade of the 21st century. Merging of institutions of higher learning became a common phenomenon. After the new dispensation that saw mergers of institutions in higher education in South Africa, universities are divided into three broad categories, namely:

- Universities of Technology that focus on vocationally-oriented education;
- Comprehensive universities that offer a combination of academic and vocational diplomas and degrees;
- Traditional universities offering theoretically oriented university degrees.

The development of UoTs was regarded as the core in mergers between two or multiple technikons where one institution was always being historically disadvantaged and the other historically advantaged (UoT, 2008).

Within the above landscape, the Cape Peninsula University of Technology (CPUT) was selected as a case under study. CPUT was formed on the 1st of January 2005 with the merger of the previously Peninsula and Cape Technikons and started operating as a new merged institution on the 1st of February 2006. The university is currently located in six campuses in the Western Cape Province of South Africa. The campuses are dispersed within Western Cape of which the farthest is around seventy kilometres from the central campus. Both former institutions had their own institutionalised systems which through the merger, were compelled to integrate and bring together both their IT departments and systems. As in all other mergers in the country, the integration process has been very slow such that Bellville campus, previously the main campus of the Peninsula Technikon, and the Cape Town campus, previously the main campus of the Cape Technikon, are still regarded as two different infrastructures. The computing services of the university are divided into three domains – Computer and Telecommunications Services (CTS), Management Information Systems (MIS) and E-learning. The CTS department is responsible for IT infrastructure, network, facilities, desktop support, printing and helpdesk support. MIS is responsible for institutional data which uses the Integrated Tertiary Software (ITS) integrator to capture both staff and student data. E-learning manages the university’s learning management system (LMS).

The rationale for selecting CPUT as a case in this study is two-fold. Firstly, CPUT is one of the biggest merger universities with over 33,000 thousand students. Secondly, many merger related challenges are still experienced by CPUT. The challenges are both social and technological. Social challenges which are evident at CPUT relate to cultural and racial diversity as the two merged institutions had been defined by race. Technological challenges, which are the focus of this study range from difficulties to retrieve, manipulate and analyse aggregate data for metrics and planning, difficulties to manage unstructured data, and general data integration challenges. These data management challenges arise from institutions not thoroughly dealing with data con-

tent, records management, quality, stewardship, governance and research data management (Albrecht & Pirani, 2009:3).

3.2 Data Collection

This study followed a deductive approach, which is theory-driven, using both the contingency theory's design parameters (Wende & Otto, 2007) and the data governance decision domains (Khatri & Brown, 2010). Both models were used as guideline to test empirical data with the aim of analysing the impact of data governance at CPUT. A questionnaire and interviews were used to collect data. The questionnaire was structured using the data governance decision domains (Khatri & Brown, 2010) and interviews followed the contingency theory's design parameters (Wende & Otto, 2007). Data collected from the questionnaire allowed the researcher to determine and understand the impact of data governance in higher institutions of higher learning, in this case, CPUT as an institution. The interviews allowed the participants to express their opinions on experiences and challenges related to data, and their understanding of the issues related to data responsibilities and decision making by the business users.

This research employed purposive non-probability sampling to select the sample mainly from business users and IT technical personnel. Business people in this context included executive-level board members (for example, Vice Chancellors, Deputy Vice Chancellors and others). The following explain selected participants or unit of analysis for this study. Participants chosen to represent executive-level members included:

- Deputy Vice Chancellor of knowledge and information services: this individual was selected because he oversee the whole ICT function in the entire institution.
- Registrar: this individual was selected because he is liable for student data and he is a custodian for the institution policies.

Participants chosen to represent ICT technical personnel included:

- IT manager: this individual is responsible for the integration of systems,
- IT Risk and Compliance officer: is responsible for the development and compliant of policies in the IT department, and lastly and
- IT coordinator: this individual focus more on IT projects involving student data. These are the participants that were used as unit of analysis.

3.3 Data Analysis

Thematic analysis is used to analyse both questionnaire and interviews data. It is used to analyse classifications and present themes that are related to data and further illustrates data in great detail, while dealing with diverse subjects via interpretations (Alhojailan, 2012; Boyatzis, 1998). It also provides description and understanding of answers through discovering patterns and developing themes. Themes come from both data itself (an inductive approach) and from the investigator's prior theoretical understanding of the phenomenon under study (Ruhode, 2016). In this case, themes from the questionnaire emerged from the components within the Data governance

decision domains (Khatri & Brown, 2010). Burnard, Gill, Stewart, Treasure and Chadwick (2008) highlight that in deductive thematic analysis, a predetermined framework is used to analyse data. (ibid) contend that this approach is useful when one has specific research questions that already identify the main themes. Interview themes emerged from the data itself and the actual data was used to derive the structure of analysis.

4 Findings

As presented in the data analysis section, themes were derived from the data governance decision domains (Khatri & Brown, 2010). An important phenomenon which evolved from the data is the emergence of one more theme – data integration. The findings per theme and related data challenges are discussed in the subsections which follow.

Data Principles. The responses from the questionnaire revealed that the institution recognises data as an asset and it is considered to have value at both strategic and operational levels in relation to analytics which, in turn, can help the institution with decision-making. The results also reveal that even though there is process in place, most business users do not take ownership of their data because they think IT people are responsible for it. From the interviews, it was identified that this challenge is exacerbated by poor communication among data users and technical people, a problem which mainly results from social aspects of merging culturally diverse groups.

Data Quality. Khatri and Brown (2010:150) state that Data Quality involves ensuring accuracy and integrity of data that is always available for an enterprise. While the data quality committee has been established, the existence of too many systems whose origin is the different technikons makes it difficult for the team to work coherently. The university has, since the merger, been producing erroneous student examination results and this has cost the university huge amounts of money.

Metadata. Metadata describes what the data is about and provides a mechanism for a concise and consistent description of the representation of data (Khatri & Brown, 2010:150). The findings suggest that there are mechanisms that provide clear description of data representation and authorised users have access to it. The findings also indicate that data is documented and passed on to new employees through documentation. However, the interface between the ITS and the LMS is not as effective as anticipated. The schematic structure of data varies in these two systems and this has posed challenges in data integration.

Data Access. According to Khatri and Brown (2010:151), Data Access involves data security and specifying access requirements of data. The findings reveal that the institution does recognise the importance of data security and access. CPUT has established an IT risk and compliance office that is responsible for data security and developing policies that guard data security and data access. The challenge to data access lies in users who rarely adhere to policies that focus on data security.

Data Lifecycle. Data Lifecycle involves understanding how data is used and how long it must be retained to minimise the total cost of storing over its life cycle (Khatri & Brown, 2010:151). The findings reveal that electronic data is stored for longer than

required and there are no policies that focus on how long data can be used, retained and archived. The findings also show that the lifecycle of paper-based data is actively and properly managed in the institution by an established Records and Archive department, which determines the use of data, how long it should be retained and its archival value.

Data Integration. The MIS is not properly integrated with the sub-systems, which leads to data flow challenges that cause unsynchronised data which, in turn, affects data quality. Managing data that come in different forms is quite problematic and incorrect data capturing and the way data is received in the institution is an issue that contributes to data integration challenges. The institution is not entirely aware that data integration challenges caused by the lack of properly integrating the main system with sub-systems or the merger of the previous institutions are the reasons it is currently experiencing data quality and data management challenges.

Design Parameters. The study analysed data using the contingency theoretical framework and a theme was also generated from the framework. It was found that the institution is using both centralised and decentralised models for decision-making, meaning some decisions are made by the IT department in terms of the infrastructure and other decisions related to data are made by business. The findings also show that the decision-making structure incorporates both hierarchical and cooperative models, which mean there are instances where people coordinate and work together to ensure that the university is sustainable and cases where direct control is used where subordinates report to their superiors.

5 The emergent model for data governance in organisations

The contingency theory's design parameters (Wende & Otto, 2007) and the Data governance decision domains model (Khatri & Brown, 2010) were coalesced to construct the emergent model for data governance (Figure 3). The concept and constructs of the emergent model for data governance were discussed in the preceding section on findings. This new model forms a baseline for institutions to developing their own data governance strategies.

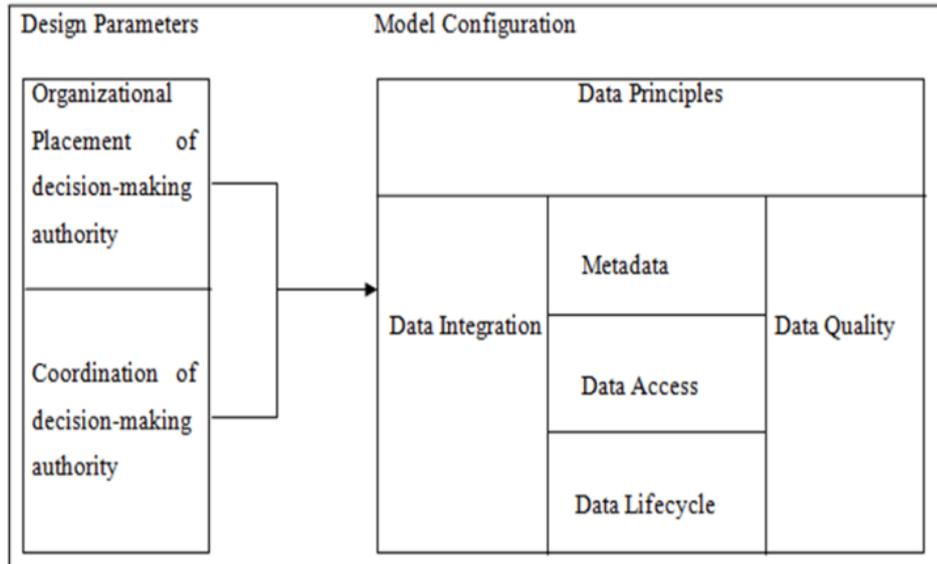


Figure 3: Emergent Data Governance model

The Emergent Data Governance model places attention on design parameters and model configuration. Decision-making authority in CPUT uses a hybrid approach where some decisions are made by individual departments and others by the IT department. Any organisation which adopts the Emergent Data Governance model can however employ centralised, decentralised or hybrid. The same observation can be made on the coordination of decision-making authority which could be a hierarchical or cooperative approach. It is in this context that the Emergent Data Governance model is proposed to be used as a framework by organisations which intend to implement a data governance strategy.

6 Implications for Theory and Practice

While there is limited scholarly research on data governance, the concept is gaining traction in both public and private entities as data increasingly get recognition as an organisational asset. We contribute to data governance research by proposing an emergent data governance model for merged and collaborating institutions. The model has been constructed after a thematic analysis of data collected from one of the leading universities of technology through coalesced models. The use of thematic analysis as a theoretical foundation and a methodological approach for analysing data contributed to a better understanding of the institution's data. The main finding reported in this study is that access to quality data is not possible without an elaborate approach to data management.

7 Summary and conclusion

This study identified data-related challenges in the institution that are the results of poor data management. Challenges were categorised in themes: data principles, data access, data quality, data integration, metadata, data lifecycle, and design parameters. The emergent data government framework was developed and its constructs are the themes that emerged from the two models which guided data collection and analysis. Since the institution under study was established through a merger of independent technikons, the findings were deemed to be applicable to many other institutions within a merger or a collaborating environment.

References

1. Albrecht, B. and Pirani, J.A. (2009) Revitalizing data stewardship through risk reduction: Managing sensitive data at the University of Virginia. *EDUCAUSE Center for Applied Research, Case study*, <http://www.educause.edu/ecar> [15 November 2015]
2. Alhojailan, M. I. (2012) Thematic Analysis: A critical review of its process and evaluation. *West East Journal of Social Sciences*, 1(1):39-47
3. Badat, S. (2010) *The challenges of transformation in higher education and training institutions in South Africa*. In Development Bank of Southern Africa (DBSA) Report, 2010.
4. Boyatzis, R.E. (1998) Transforming qualitative information: thematic analysis and code development. USA: Sage Publications.
5. Burnard, P., Gill, P., Stewart, K., Treasure, E. and Chadwick, B. (2008) Analysing and presenting qualitative data. *British Dental Journal*, 204(2008):429–432, April 26.
6. Bryant, W. (2014) Data governance simplified. *Point B management consultants*, http://www.pointb.com/documents/latest-insights/BTS_Data_Governance_Simplified.pdf [15 February 2015].
7. Drucker, P. (2005) Academic analytics in the future of higher education. *EDUCAUSE Center for Applied Research*. <https://net.educause.edu/ir/library/pdf/ers0508/rs/ers05089.pdf> [02 August 2015]
8. Fisher, T. (2009) *The data asset: How smart companies govern their data for business success*. Hoboken, NJ: John Wiley & Sons.
9. Gillard, E., Saunders, S., Terblanche, J. and Sukel, M. (2012) *A review of four case studies in restructuring the South African Higher Education System*. African Higher Education.
10. Godfrey, J., Hodgson, A., Holmes, S. and Kam, V. (1997) *Financial Accounting Theory*. 3rd Edition. John Wiley and Sons, New York.
11. Institute of Directors (IoD). (2009) *King report on governance for South Africa*. Institute of Directors Southern Africa: Johannesburg.
12. Jansen, J., and Taylor, N. (2003) Educational Change in South Africa 1994-2003: Case studies in Large-Scale Education Reform. *Country Studies – Education Reform and Management Publication Series*, 11(1), October 2003.

13. Khatri, V. and Brown, C. V. (2010) Designing data governance. *Communications of the ACM*, 53(1):148-152.
14. Korhonen, J. J., Melleri, I., Hiekkänen, K. and Helenius, M. (2013) Designing Data Governance Structure: An Organisational Perspective. *GSTF Journal on Computing*. 2(4), January 2013.
15. Lee, Y. W., Pipino, L. L., Funk, J. D., and Wang, R.Y. (2006) *Journey to data quality*. Massachusetts Institute of Technology (MIT) Press, Cambridge, MA.
16. Moses, A. (2014) *The strategic role of the Chief Information Officer during post-merger at institutions of higher learning: A CASE STUDY*. [MTech Thesis, Cape Peninsula University of Technology]
17. Olson, J. (2003) *Data quality: The accuracy dimension*. San Francisco, CA: Morgan Kaufmann Publishers.
18. Poor, M. (2011) *Applying Aspects of Data Governance from the Private Sector to Public Higher Education*. Mount Wachusett Community College, 1277.
19. Redman, T. C. (2001) *Data quality: The field guide*. Boston: Digital Press.
20. Redman, T. C. (2004) Data Quality: Should Universities Worry? *EDUCAUSE Review*, 39(5):12–13. Available from: <http://net.educause.edu/ir/library/pdf/ERM0457.pdf> [Accessed 11 August 2014]
21. Redman, T. C. (2008) *Data Driven: Profiting from Your Most Important Business Asset*. Boston: Harvard Business Press.
22. Ruhode, E. (2016) E-Government for Development: A Thematic Analysis of Zimbabwe's Information and Communication Technology Policy Documents. *The Electronic Journal for Information Systems in Developing Countries (EJISDC)*, 73(7): 1-15
23. Russom, P. (2008) *Data Governance Strategies: Helping your Organisation Comply, Transform, and Integrate*. The Data Warehouse Institute Best Practices Report, pp. 1-30.
24. Russom, P. (2012) *Seven reasons why master data management needs data governance*. TDWI Checklist Report. TDWI Research.
25. Universities of Technology (UoT). (2008) *Position, role and function of Universities of Technology in South Africa*. South African Technology Network. March, <http://www.satn.co.za>
26. Van Grembergen, W. (2004) *Strategies for information technology governance*. Hershey: IDEA Group Publishing.
27. Wende, K. (2007) A model for data governance – organizing accountabilities for data quality management. In *Proceedings of 18th Australasian Conference on Information Systems*, Toowoomba, Australia 5-7 December 2007.
28. Wende, K. and Otto, B. (2007) A Contingency Approach to Data Governance. In *Proceedings of the 12th International Conference on Information Quality*, Cambridge, MA (USA).