Leveraging Data Governance To Improve Data Quality In Health Organizations
(Conference ID: CFP/438/2017)

Viola Vivian Busobozi

Department of Informatics

Tshwane University of Technology

BusoboziVV@tut.ac.za

Billy M. Kalema

Department of Informatics

Tshwane University of Technology

kalemabm@tut.ac.za

Abstract-

Literature shows that there is limited guidance for organizations to have effective data governance (DG) that could lead to high quality data needed for decision making. This paper seeks to report on the role of data governance in improving the quality of data within organizations. To achieve this objective, the study was underpinned by Contingency theory and Differed theory of action to test the constructs that were sought relevant for the designing of a data governance model. Data was collected from four health Non-governmental organizations providing HIV/AIDS services in South Africa and was analyzed quantitatively. Results of the study showed that, environment, tasks, data governance and structures, data quality management and technology are significant in improving data quality. On the contrary, data governance strategy, individual factors and deferred actions constructs were found to be insignificant. This study contributes to the ongoing debate of using data governance to enhance data quality. The study recommends that to recognize data quality improvement data governance teams need to consider business and technical data expectations in their organizations.

Key words- Data quality, Data governance, Data quality improvement, Data quality management

1. Introduction

Health organizations are non-profit making institutions that support and deliver healthcare services to the public. They also support the clinical programmes, intended for health research, training, mentoring, counselling to the public on behalf of the state. Because of the nature of their work, data quality is an important aspects as it helps in making informed decisions necessary for making correct diagnostic, treatment and support to patients.

The Department of health South Africa's e-health strategy (2012) indicates that health institutions in the country are implementing a national health information system but still face the challenge of data quality. Botha et al. (2015)add that, the South African health information system (HIS) use parallel data that lack integration across disease program areas, putting the Monitoring and Evaluation (M&E) system at a risk of confusion during reporting. More still, health workers responsible for data collection from patients, duplicate and mix up datasets hence causing redundancies. Friedman and Smith (2011)noted that health organizations also face the challenge of over budgets due to extra resources required for data cleansing of numerous duplications and profiling that causes wastage and ineffectiveness. They further allude that this results into inappropriate decision making that causes patients distress and loss of confidence. Mphatsweet al. (2012) also observed that even though many health institutions are using HIS, these systems' are still challenged with data accuracy and reliability that impacts health service delivery.

It isessential for healthcare organizations to have and maintain data quality with appropriate governance so as to improve the operational and strategic decision making processes. In the bid to adhere to this call, of the studyBotha etal. (2015)also identified data governance as one of the most data quality challenge faced by the health sector. Their study indicated that the many elements that make up data governance positions it to be a serious challenge for data quality. These elements include but not limited to, lack of assignment of data responsibilities, administration, roles ambiguity in relation to tasks, missing written quality policies, managers lack of emphasis on the importance of data quality, ineffective organizational procedures and assessment. The World Health organization report (WHO, 2003)also indicate that having such challenges as impedements of data quality in health institutions, lead to extra work, production problems, loss of revenue, higher costs, impacts on quality of care and leads to privacy and security issues.

Kiwanuka et al. (2015)confirm there many challenges faced by HIS desipite of the fact that they are widely used in health organizations. They put it that HIS are challenged by data quality issues, poor information flow and the integration of Monitoring and Evaluation (M&E) system.

2

From their analysis they indicated that many of challenges are as a result of lack of clarity in roles concerning data creation and use, poor management focus, lack of systematic and standard measures, poor written procedures and lack of guidance or framework for data quality. Consequently, Weber et al. (2008)also add that the missing standards for data requirements, lack of ownership and guidance has highly compromised the creation of high quality data in organizations. According to Mutale et al. (2013), health institutions also face the challenge of lack of effective communication that cause poor coordination and cooperation among their business units and such compromise the quality of data. They alluded that un- streamlined communication also impacts on the healthcare institutions' reporting system and leads to poor decision making. Additionally, Qazi and Ali (2011) allude that health workers use a one way reporting system and they often don't get feedback from the health management information system (HMIS). They noted that in many instances the captured data into the HMIS is either not used or misused, which leads to the risk of generating inappropriate reports for decision making and may contribute to the manifestation of errors in patient's service delivery.

It is estimated that 5% of organizations' data are of poor quality and causes 10% average cost impact on the organizations' annual revenue (Friedman, 2011:26). Poor data quality is categorized as causing direct and indirect impacts on financial losses; reduced customer, supplier, employee confidence and satisfaction; reduced productivity due to increased workload or efficiency, increased risk and compliance(Loshin, 2010).According to Murakwani and Sethi (2015),costs are as a result of nonconformance to compliance, business and data requirements. Further, these costs affect the internal and externalcosts of an organization.

Kahn et al. (2015)assertthat in healthcare, there is misclassification of key elements for administrative billing systems, which leads to a biased estimate of treatment effects more especially when data capture process and analysis are not properly done. This negatively impacts clinical decision support, research and patient's safety. Further, they suggest the development of a cost-benefit and business case with data collection and analytics activities as well as their inherent costs for the process of data quality assessment, monitoring and governance. As a result, administrative and clinical data quality for clinical care, outcome and research could be improved. Furthermore, emphasized that a formal cost-benefit analysis could be used to model the return on investment (ROI) for data quality assessment, organizational and scientific risks associated with the use of poor data quality from different stakeholders. To counter the above challenges, Ladley (2012)suggested that there should be a culture of interdepartmental collaboration which can be achieved through business user engagement with the aim of improving data quality.

The purpose of this paper is to report on how data governance could be leverage to improve data quality in health institutions. Data for this study was analyzed quantitatively and based on the results a data governance model was designed.

2. Data Governance and Data Quality

a) Data

Data is defined as a representation of facts. It can be in form of digital, textual, numerical or graphical(Gomm, 2008). Further, it added that data is a raw material (unprocessed data) that is used to produce information (processed data) when put in context that gives meaning. However, processed data can be considered as a raw material by other systems or users. Data has to go through a defined process that produces an information product for example reports, file, single numbers, images or verbal phrases. Data has the characteristics of being collected, measured, analyzed and visualized.

b) Data quality

Data quality explains the state of data that fits to be used by its consumers and conforms to the user's requirements (Mustafa et al., 2016). Data quality must be considered in relation to user's objectives, goals and in a specific context, assessed and measured using its dimension (Kahn et al. 2012; Chen et al., 2014). Data quality is critical in ensuring that appropriate conclusions are drawn from information captured and integrated into organization's reports needed for decision making (Ledikwe et al., 2014).

c) Data quality Management

Data quality management is a business function that develops and executes the acquisition, control, protection, delivery, storage, enhancement and presentation of high quality data (Geiger, 2004; Wende & Otto 2007). Data quality management involves the implementation of decisions made during data governance (Alhassan et al., 2016).

d) Data quality improvement

Data quality improvement is shifting from the undesired to the desired new state of data quality (Batin et al., 2009). It involves the selection of quality process, plan for implementation, examining the impact and standardization. Batin et al. (2009) further noted that data quality Improvement involves two steps whereby step one includes evaluation of costs, assignment of process and data responsibilities, selection of strategies and techniques as well as identifying the causes of errors; and step two involves process control, design of data improvement solutions, process redesign and improvement, monitoring and management.

e) Data governance

Russom (2008) defines data governance as an organization structure that can either be a committee or board that creates and enforces policies and procedures for the use of business and technical management of data across the organization. Data governance defines decisions on the roles of managing, policies and procedures that control data assets (Alhassan et al., 2016; Holmes, 2016). On the other hand, Wende and Otto (2007) noted that the policies guidelines and standards must be consistent with the organization's mission, strategy, values, norms and culture. Furthermore, Friedman (2011) argues that Data governance is a continuous improvement process which benefits every business unit and helps to ensure business information is reliable, complete, consistent, current, correctly interpreted and manifested at all levels of management (Geiger, 2004).

f) Role of Data governance

Data governance improves accountability, communication, coordination and allows data and application software to communicate for data integration (Dail et al., 2015). Data governance creates an environment for informal and social interaction which leads to management alignment and integration across all business units (Espinosa &Armour, 2016). Data governance gives the organisation the ability to realise value of data by getting to know the cost of that which is inappropriate or of poor quality (Khan et al., 2015). Data governance provides a well-defined approach for balancing value creation, risk exposure and cost (Tallon, 2013). Data governance creates a sense of data ownership, unites business objective, and designs information policies by making sure all stakeholders see one true version of data (Information builders inc., 2011).

3.Related Work

Geiger (2004)identified a data quality challenge of failure of organizations to recognize that they have a Data quality problem. Data quality issues are hidden and persistent, they can exist unnoticed for some time and even propagated to other systems or business units due to increased connectivity. For example, in the HMIS incomplete datasets from ART may affect those from Tuberculosis (TB) system or unit. Chen et al. (2014)add that Data quality issues are hidden in other areas which may lead to ignorance of data management thus creating unawareness of available data quality problems that continue to hinder public health practices. Geiger (2004)Comments that to solve any problem it has to be recognized first that it exists. Hence, organizations remain in denial about the quality of data ending up making inappropriate decisions with data that is not accurate, valid, and consistent or complete thus compromising the

efficient operations of the business processes. This as the result ends up with reduced ROI in enterprises or profit making organizations.

Khatri and Brown (2010)designed a framework that could be adapted by organisations with growing need of making their data an asset. The framework consists of five decision domains of data principle,metadata, access, quality and lifecycle. With key organisation assets of human, financial, Physical, Information product information, IT and relationship asset, can be used to ensure data quality. The framework is appreciated for its contribution to the body of knowledge and successfully used by authors like Alhasaan et al. (2016). However, it has a limitation of not being adaptable by all organisations by the virtual of the nature of environment and culture they operate in. For instance Begg and Caira (2012) used the framework on 10 Small Medium enterprises (SME) but it could not be adaptable to their business environment and culture. So this confirms the view of Wende (2009) who argues that one size does not fit all.

Kukemuller (2011) asserted that data quality is of value within its context of use and the value of an information product is influenced by its quality. That being the case, organizations use IT tools to fix data quality issues. IT tools used include data warehousing, customer relationship management (CRM), chain supply management, enterprise resource planning (ERP), master data and many other enterprise systems. All the above information systems generate huge amounts of data, nonetheless, it is not a matter of capturing and storing data that goes unmanaged. This is because unmanaged data results into poor quality data and increased costs of business operations (Niemi, 2013). For the case of healthcare institutions it is the HMIS used as a tool for quality improvement. However, IT software tools just improve data quality only relevant for analytics purposes.

Young and McConkery (2012) highlight the achievements and goals of Data Governance Advisory Group . The researchers say that is responsible for providing advice to senior management on data governance policies, standards and strategic approaches; data quality initiatives; privacy; architecture and integration requirements; compliance and security; data ware housing and business intelligence priorities. They further mentioned the members that make the committee of the group include; stakeholders from planning, quality and reporting, finance, student and academic services, human resource, finance, information technology (IT), research services, library, facilities management, external relations and corporate services. The researchers found out that monthly meetings provides a forum to discuss data quality and governance issues throughout the lifecycle from collection, processing, reporting and decision making. Their study recommended data governance groups to focus on policy, standards and strategy, data quality, privacy, compliance and security, architecture and integration, data warehouses, business intelligence and management alignment. They caution that the metrics

identified in their study may differ among institutions and sectors and need to be tested for relevancy in the respective environments. However, much as their study was exhaustive enough and provided several factors that inform data governance, it lacked an underpinning theory that could be used to guide future research.

Niemi (2013) asserts that, data is never used due to uncontrolled redundancy, lack of datause policies andprocedures. Thus making organizations to remain in confusion since managers do not know how much the available data costs and its importance. More so, data quality issues include frequently misinterpreted data that cannot be shared amongst business units.

Nahar et al. (2013)identified data quality issues such as duplication, missing information, formatting, and inaccurate profiling reflect in computational intelligence more especially when data are not pre-processed cleaned through cleansing, verification, formatting and updating. These result in clients' or patients' distress, wastage of money and increased organizational risks. Chen et al. (2014)indicate that data quality is influenced by technical, organizational, behavioral and environmental factors. However Cappiello et al. (2013)argued that if organizations continue to rely on technology, the more data and information quality will remain a concern. However, Cappiello et al. (2013)suggests that through data governance, the above challenges can be minimized when business processes are well described in relation to the ability to identify data requirements.

Chen et al. (2014)reviewed 39 publications about data quality assessment in public health information system. Out of the review it was noted that data collection and use was given least attention. And yet according to Niemi, (2013)data collection and use are themost critical stages in data life cycle that need governance for improved data quality.

Ledikwe et al. (2014)discovered that HMIS are of complex setting where data from different sources and datasets are stored waiting to be retrieved when demanded. The complex nature of HMIS makes data integration a challenge for data quality without governance. Furthermore, in healthcare organizations, there are short term projects whose delivery focuses on programs funded at functional business unit level. The teamsresponsible do not account for how the program data may be used by others. As a result the communication requires data to flow within the systems (functional business units) which have connection points that mustcross strict project boundaries due to bureaucracy.According Wittwer (2000),bureaucracy is a data quality obstacle, if there is an absence of direct authority for the accountability of timeliness, accuracy and an appropriate data integration environment.

Khan et al. (2015)conducted a research on transparent reporting of data quality in distributed data networks. The researchers found out that reporting the strength and weaknesses of data sets at each level of data life cycle, may improve transparency and trust, as well as unearthing unintended negative consequences of revealing internal data quality problems which negatively affect data contributors who may withdraw from a data sharing network. The researchers recommend that there is need to have a culture that embraces transparency as a means of improving data quality. This study can bridge this gap through data governance. For instance, if problems are identified earlier, then it creates a basis for quality improvement. Ladley (2012)asserts that it is the role of data governance to resolve a data issue before it gets out of hand. Oracle inc. (2011)indicates that reporting at each level can help track "fit for use" by using service level agreements. In addition, Loshin (2013) recommends the use of a data quality score card to assess the level of data quality.

Cai and Zhu (2015)identified the challenges of data quality and its assessment in big data era. These researchers found out that data quality not only depends on its dimensions but also on business environment, processes and users. The researchers further pointed out that data producers are not its users and that makes it difficult to measure quality. They proposed a hierarchical data quality standard from the users' perspective which involves data quality dimensions, elements and indicators. It also formulated a big data assessment process with a feedback mechanism which consists of elements from data collection goal setting, determining quality dimensions and elements, indicators, formulating evaluation baseline, actual data collecting, cleaning, assessment generating reports, analysis, mining and finally output results and this process is continuous. The researchers recommended that further studies should focus on data governance because it drives high quality data which is a precondition for Big Data analysis. However, much their study provided a thorough literature review, it lacked empirical testing and an underpinning theory that could inform further research.

Botha et al. (2015)affirmed that creating quality data that is fit for use in the healthcare institution is still a challenge. Qazi and Ali (2011)also, observed that data is collected for the purpose of only generating reports. Cai and Zhu, (2015)noted that data producers are not the users, and data users cannot improve their own data produced within their systems or functional business units. For example, data entrants into the HMIS do not have the incentive to maintain high quality data because they focus on entering data quickly without rejection by the system. As a result when data is summarized, standardized, integrated and subjected to another system or used in another context, data quality issues emerge.

Paoline et al. (2016)adapted the use of procedures and processes that allowed project and regulatory requirements that can protect patients, their data and health care systems in a Patient

8

Outcome Research to Advanced Learning (PORTAL). The researchers found out that data governance is one of the governance priorities identified and can address issues of overseeing procedures to request and use data, ensuring data quality and integrity, addressing conflict of interest, developing and maintaining transparency of activities and results in defining guidance related to data access and use. Theresearchers recommended that governance should not work in a vacuum, it must be aligned with organization policiesand create a culture of trust and collaboration. This study incorporated culture and trust under the construct of environment in the research model as these factors have been widely acknowledged to inform data governance.

Espinosa and Armour (2016)designed a framework for coordination and governance of Big Data analytics. The researchers used the coordination theory to support their study in which they pointed out that structural (data ownership rights, steering committee), operational (data retention, access rights, data protection, storage and migration policies) and relational (awareness and education on data practices, communication) practices are vital for Big Data analytics governance. Much as the framework of their study was based on related literature, it lacked empirical validation and such limits its use to inform other studies. More still, their study focused on effective coordination and governance for improved Big Data analytics practices and little was done in relation to data governance and its role to improve data quality which is the key aspect of the current study.

Holmes (2016)identified that health personal information is increasingly becoming important for a number of users in business operations, quality improvement and research. This requires high standards of security, privacy and proper use of personal health data in order to preserve trust. The researcher holds that despite the existence ofhealth information technology of Big Data and distributed clinical research, there is still limited access and short supply of health data governance. He further recommended that to maximize the utility and availability of data sets, data owners and policy makers should discuss data access policies as a means of improving its governance. Further still, the researcher recommended that data governance design, implementation and functions should be uplifted from being an afterthought or an add-on issue in the health sector. The researcher noted that data governance is a central challenge in the health sector that needs to be handled independently in its own field.

Holmes (2016) and Ladley (2012)observed that many organizations don't have data governance departments or units. They argue that data governance issues are virtually handled by business and IT department thus depriving the ability for data governance to enable carry out the resolution, monitoring and directing data quality issues. Both Holmes (2016) and Ladley (2012)studies lacked underpinning theories that could empirically be used to support organizations in data governance or in the improvement of data quality.

Alhasaan et al. (2016)analyzed 31 peer reviewed papers on data governance activities using Khatri and Brown (2010)data governance framework of five decision domain. Using a content analysis method, they found out 8 major data governance action areas that includes; data roles and responsibilities, policies, processes and procedures, standards, strategy, guidelines, technologies and requirements. They recommended that these areas need to be empirically tested in future data governance research. This study filled this gap.

A. *Theoretical foundations and Research model*

To embrace the factors of DG identified from the literature, this study utilized Contingency Theory (Vroom & Yetton, 1973) and Deferred Theory of Action (Patel, 2006).From Contingencytheory, a Management information system (MIS) contingency model's constructs were used (Weills&Oslon, 1989). Constructs used included Strategy, Structure, Environment, Technology, Tasks and individual. From contingency theory one construct was used which was Deferred action. Finally data quality management was the only external construct that was used. The research model is demonstrated in Figure I and is supposed to be inserted here.

B. *Hypotheses Development*

Basing on the research model in figure 1, the study suggested the Hypotheses (H) and the construct definition as explained below.

a) Data governance Strategy - this construct focuses on the availability of strategy, policies, principles, procedures, data sets, data collection goals that must be in line with organization strategy. From this definition hypothesis (H1) was developed.

H1. Data governance strategy when mediated by data governance influences data quality improvement.

b) Data governance structure - this is viewed in terms of decision making authority on data quality issues which can either be centralized or decentralized. That is, who decides on what, when and where about data principles, architecture, metadata, quality and lifecycle. From reviewed literature hypothesis (H2) was developed

H2. Data governance structure when mediated by data governance influences data quality improvement.

c) Environment - this construct focuses on the culture and norms, the ability to change to new innovations, different business unit's integration internally within the organization. Internal is how systems interact to ensure data quality and externally the policies on data quality. From reviewed literature hypothesis (H3) was developed

H3. Environment when mediated by data governance influences data quality improvement.

d) Individuals - this looks at roles, responsibilities, accountability and the way employees perceive data governance for data quality improvement. From reviewed literature hypothesis (H4) was developed

H4. Individuals when mediated by data governance influence data quality improvement.

e) Tasks (Business processes) - these are data quality activities that should be in line with Data governance processes, user requirements and organizational needs. From this understanding hypothesis (H5) was developed

H5. Tasks when mediated by data governance influence data quality improvement.

f) Technology – this looks at data technologies used to support the way people work but not the way it works thus making it easy to extract the needed information by the users. From reviewed literature hypothesis (H6) was developed

H6. Technology when mediated by data governance influences data quality improvement.

g) Deferred Action– this focuses on future actions that can be done to improve data quality through data governance. From reviewed literature hypothesis (H7) was developed.

H7. Deferred action when mediated by data governance influences data quality improvement.

i) Data Quality Management - this focuses at ensuring that high quality data is achieved at operational level and responds to the gap identified by data governance for improvement. From this definition and literature reviewed hypothesis (H8) and (H9) were developed.

H8. Data Quality Management when mediated by data governance influences data quality improvement.

H9. Data Quality Management directly influences data quality improvement.

h) Data governance – practices that allow the assessment, measurement, reporting on, reacting to and monitoring, controlling, coordinating data within the organization. These practices further assist in coordinating and controlling data quality processes by providing feedback to the functional contingencies, institutional processes at strategic level and to data quality management at operation level about current data quality requirements. Based on this understanding hypothesis (H10) was developed.

H10. Data governance influences data quality improvement.

j) Data quality improvement – this is a dependent variable that is emergent and unpredictable. However it can be made predictable with proper data governance practices. This construct's hypothesis was not developed because it is this study's point of focus.

3.**METHODOLOGY**

Following the research model constructs, the questionnaire actual development took precedence.The questionnaire development used a 5 Likert scale where 1= Strongly Disagree, 2= Disagree 3=Neutral 4= Agree and 5= Strongly Agree.During the process of questionnaire design, the construct of the research model formed the sections of the questionnaire whereas their attributes were used to formulate the measuring items. 200 questionnaires were distributed to employee within selected health nongovernmental organizations in South Africa.

Out of the distributed questionnaires, 163 were returned and 152 usable. This gave a response rate of 81.5% and a usable rate of 92.8%. Data from usable questionnaires was screened, coded and entered into SPSS version 22.0 for analysis. The overall reliability of the measuring instrument was 0.794. According to Pallant (2005), this study's questionnaire met the required value for it to be considered reliable. This is because the results fell in the required range of 0.7.

The reliability of constructs was tested independently before the presentation of their results.9 constructs were tested for their reliability and 8 of them meet Pallant (2010)recommended value of 0.7 and above values. But only 1 construct of environment that emerged 0.444 Cronbach Alpha coefficientsthat didnot meet the threshold of 0.7. However, SPSS software gave the option of deleting 1 item on themetrics in order to improve the reliability of the Environment construct other than losing it all. So Item Envt6 which rose to 0.465 was still below theaccepted threshold of 0.7 was deleted. The environment construct was not eliminated because it did not reach 0.7 the allowed level because McCrae et al. (2011)argued that, even though the threshold values of 0.7 is not reached values close to 0.5 can give meaningful interrelated reliability and internal consistency that may substitute Cronbach's Alpha reliability. Basing on the above explanation the environment construct was left to be considered for further analysis because its reliability of 0.465 could be rounded off to the nearest hundredth to make it 0.5.

4. RESULTS

Upon the completion of reliability tests, the collected data was further analyzed to determine the correlation between   constructs.

A. Correlation

Correlation analysis was carried out to determine the relationship and interdependencies between constructs.The correlation output showedthat data governance strategy was significant with deferred action with a Pearson correlation of 0.338 at 0.001 level (2-tailed). Data governance structure was significant to individual, data governance, tasks, deferred action, data quality management and data quality improvement with a Pearson correlation of 0.746, 0.483. 0.430, 0.379, 0.489 and 0.505 respectively at 0.001 level (2-tailed). Individual was significant to data governance, tasks, deferred action, data quality management and data quality improvement with a Pearson correlation of 0.396, 0.329, 0.306, 0.460 and 0.536 respectively at 0.001 level (2-tailed). Data governance is significant to task, deferred action, data quality management and data

quality improvement with a Pearson correlation of 0.523, 0.479, 0.626 and 0.536 respectively at 0.001 level (2-tailed). Tasks are significant to deferred action, data qualitymanagement, data governance structure, individual, data governance and data quality improvement with a Pearson correlation of 0.437, 0.639, 0.430, 0.329, 0.523 and 0.323 respectively at 0.001 level (2-tailed).

Further still, deferred action was significant to data quality management and data quality improvement respectively at 0.001 level (2-tailed). Data quality management is significant to data governance structure,individual, data governance, deferred action and data qualityimprovement with a Pearson correlation of0.489, 0.460, 0.626,0.501and 0.582 respectively at 0.001 level (2-tailed). Deferred action is significant to data governance structure, data governance strategy, individual, data governance and tasks with a Pearson correlation of 0.379, 0.338, 0.306, 0.479, and 0.437 respectively at 0.001 level (2-tailed). Data quality improvement is significant to data governance structure, individual, data governance, tasks, deferred action and data quality management with 0.505, 0.412, 0.536, 0.324, 0.355 and 0.582 respectively at 0.001 level (2-tailed). However the output results from table 4.2 show there is insignificant correlation of environment and technology to data quality improvement. According to table 4.2 presentations above, there was no a Pearson correlation results of 0.05 level (2-tailed). This implies that the relationship of the constructs was strongly correlated with a Pearson correlation of 0.01 level (2-tailed).

B. Regression

Basing on the correlations results, this study used a multiple regression to determine a set of variables of data governance strategy, structure, individual, tasks, deferred action, data quality management data governance, environment and technology with the aim of predicting the outcome on improving data quality as a dependent variable and having a more exploration of interrelationship amongst set variables which made an ideal investigation from the real life generated from 152 participants results feedback from the collected data. Hence it provided information about a proposed research model of improving data quality through data governance (Pallant, 2005).

As demonstrated in Table I, results shows thatdata quality management construct contribution to data quality improvement was the highest with a positive percent value of 47% (β=0.470)and a significant value of(p=0.000<0.05). Followed by data governance structure, data governance with a percent value of 29.6%, 28.6% with beta value of(β=0.296), (β=0.286) and a significant value of(p=0.003<0.05,p=0.001<0.05) respectively,and data governance strategy contributed positively with 10.6% (β=0.106) but was not statistically accepted with a p= 0.111>0.05. Constructs of technology, tasks, environment contributed negatively with -0.12% (β=-0.124) 0.23%, (β=-0.234), -0.18% (β=-0.182) with a significant contribution of (p=0.043>0.05, p=0.005<0.05, p=0.003<0.05) respectively.

As Pallant (2005) argues that the + and – sign before a figure doesn't determine the strength of a construct relationship but shows the direction in which it moves. So technology, tasks and environment constructs showed a significant relationship towards improving data quality through

13

data governance model.On the other hand some constructs were shown not to have had a significant contribution to the overall model of improving data quality through data governance. These included individual and deferred action with a percent value of -0.7%, -0.04% and a beta value of β=-0.073, β=-0.049 and a significant value of p=0.428>0.05, p=0.531>0.05 respectively.Table I is supposed to be inserted below

Table I shows high significant correction. With that note multicollinearity was tested. This was measured using variance inflation factor (VIF) and tolerance shown in table 4.4. Pallant (2005)set a cutoff point of multicollinearity where VIF is above 10 suggests that there is a high correlationamongst independent variables and a Tolerance of less than 10 shows the presence of multicollinearity. Basing on the regression results above, there was no multicollinearity because tolerance was above 10 and VIF below 10.

C. Hypotheses Testing

On the basis of the results, the ten hypotheses were tested. Table II shows the extracts of the results of the tested hypotheses showing their significance at p=0.05.Table II is supposed to be inserted below

5. DISCUSSION CONCLUSION AND RECOMMENDATIONS

A. *Limitations*

There was limited literature on data governance from published articles. So the study used some white paper like SAS. In SAS white paper practitioners like Russom (2008) from TDWI (2015) and Loshin (2013) from knowledge integrity Inc. had tangible information that was found useful in this study.

B. *Recommendations and Future Research*

Data quality is subjective in nature. To realize that there is a data quality improvement, data governance teams need to consider data quality and business expectations (Batini et al. 2009). Data quality expectations are measured basing on the validity of data values, for example conflicting data values, duplicated records, missing data as well as their linkages and unusable data. On the other hand business expectations are measured using the processes of organization performance, productivity, and efficiency for example data governance teams should be in position to realize the decreased throughput due to data quality errors, the time spent reworking failed business processes.

It is suggested that future research in data governance and data quality be carried out in different industries both public and private since this study focused only on health nongovernmental organizations. Further, this study recommends future research to use a qualitative method of data collection in order to get in depth understanding of data quality and data governance through interviews.It should be noted that data governance practices are not data rules for ensuringimproved data quality as rulesareinflexible and hard to follow as well as maintain (Ladley, 2012).It is recommended to use principles which are core beliefs that create a link between policies, processes and behaviors for information asset management. Since principles are used on a daily basis and they form data governance habits and norms hence forming a culture. Although data governance strategy, deferred action and individuals were statistically insignificant at 0.05 standard coefficient value, they were significant at 0.01 level (2 tailed). So the study recommends that health organizations should consider all factors identified to influence data quality improvement.

C. *Contribution of the study*

i. Theoretical contribution

Theoretically this study contributed to the field of data quality and data governance in managing data. It fills the gap of the data governance factors that specifically influence data quality improvement. In the review of literature it was identified that despite the number of studies done on improving data quality in health institution and various tools used, there is still a challenge in ensuring high levels of data quality. Further, the recommendation for ensuring data governance in the health organization by researchers such as Qazi and Ali (2011), Mphatswe et al. (2012);Nahar et al. (2013);Ledikwe et al. (2014), Chen et al. (2014);Kahn et al. (2015);Botha et al. (2015) has been addressed.

i. Practical contribution

The overall practical contribution of this study may assist M&E professionals in improving data quality.Data governance model canprovide guidance to M & E teams to ensure effective data governance practices that could led to high quality data. In such a way contributing to data quality assurance, data auditing and M &E activity tools or frameworks already in existence. Furthermore, the high quality of data achieved through data governance as a result could make M & E reporting effective, efficient and strengthen organizations' competitive advantage. Further still, organisations could reach their business goals and objectives due to improved data quality which aids big data analytics and business intelligence.

D. *Conclusions*

With the emerging computing trends of big data analytics, there is a need for data governance practices to be established in organizations. Big data result into business intelligence and analytics. Due to inaccurate data, and inability to consolidate data from business units, organizations cannot generate meaningful business intelligence. More so unstructured data make it hard to be mined. According toLabouser and Matheus(2017),data quality challenges accrue from the "fitness for use". That is to say once data is subjected to another presentation with the purpose for fitness for use, it is when challenges come up. This is because of the dynamic nature of data. For that reason data quality must be considered in relation to user's objectives, goals and in a specific context (Kahn et al., 2012; Chen et al., 2014;Cai& Zhu, 2015).

Organizations need to pay attention to data governance practices to improve their data quality. This is seconded byMphatswe et al. (2012);Nahar et al. (2013);Chen et al. (2014);Kahn et al. (2015); Botha et al. (2015). Organizations lack a coherent, centralized approach to handling data quality issues. This creates a challenging environment to having quality information or data when each business unit has different standards and methods of data management.

Data governance creates a data quality assurance by creating the ability to protect businesses from serious negative impacts. This is done through early identification of data quality errors before any physical impact is made within the business (Loshin, 2013). Further, it establishes trust in the created data and provides confidence to the organization that it can be used as a competitive advantage (Information builder's inc., 2011).

Russom (2008)argues that data governance is implemented with 4Ps. That is people, procedures, policies and processes. This implies that people collaborate withinbusiness units to create procedures and policies whichaltogether comeupinto a data governance process. Hence data governance tasks are organizational and interpersonal.This study has identified data governance factors that need to be given attention for the improvement of data quality in health nongovernmental organizations. If leveraged,it can serve as a practical guide to M & E professionals in organizations. When the spotted factors are taken care of, the number of data quality challenge will be reduced. "Data quality improvement is a prerequisite for high quality health services" (Metter et al., 2008).

Tables and Figures

Table I: Multiple regression Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	3.283	.546		6.012	.000		
DGStru	.221	.072	.296	3.065	.003	.379	2.635
DGStra	.104	.065	.106	1.605	.111	.806	1.241
IND	-.048	.060	-.073	-.795	.428	.418	2.394
DG	.205	.059	.286	3.472	.001	.520	1.924
TECH	-.173	.085	-.124	-2.040	.043	.956	1.046
Tasks	-.182	.064	-.234	-2.855	.005	.525	1.905
ENVT	-.218	.073	-.182	-2.996	.003	.962	1.040
DefA	-.043	.069	-.049	-.627	.531	.570	1.754
DQM	.376	.075	.470	5.040	.000	.406	2.464

a. Dependent Variable: DQI

Based on the regression and correlation results, the set hypotheses were then tested

TABLE II. HYPOTHESES TESTING

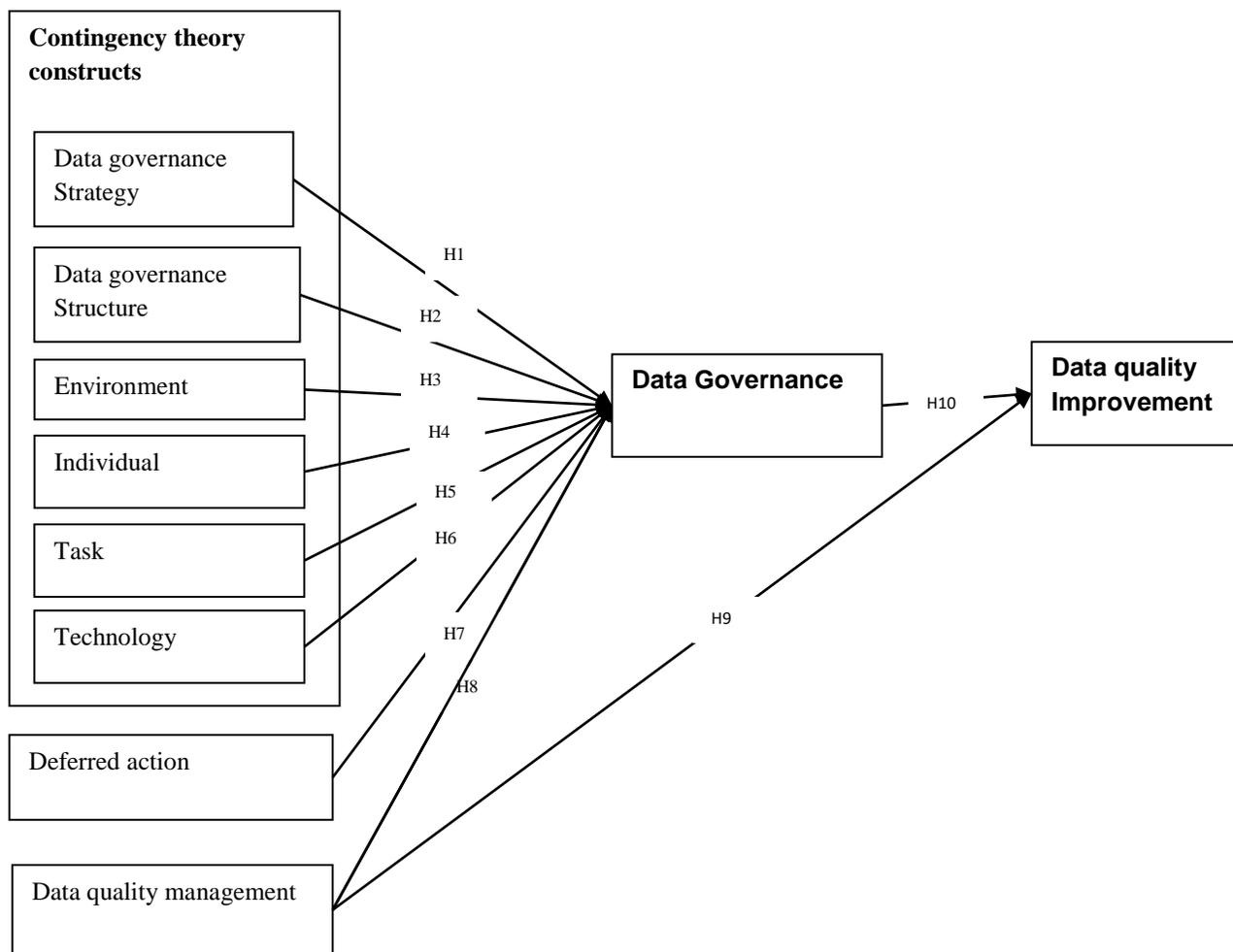| Constructs	Hypotheses	Significance (P Value)	Results
Data governance strategy	H1. Data governance strategy when mediated by data governance influences data quality improvement.	P =0.111>0.05	Rejected
Data governance structure	H2. Data governance structure when mediated by data governance influences data quality improvement.	P=0.003<0.05	Accepted
Environment	H3. Environment when mediated by data governance influences data quality improvement.	P=0.003<0.05	Accepted
Individual	H4. Individuals when mediated by data governance influences data quality improvement.	P=0.428>0.05	Rejected
Tasks	H5. Tasks when mediated by data governance influences data quality improvement.	P=0.005<0.05	Accepted
Technology	H6. Technology when mediated by data governance influences data quality improvement.	P=0.043>0.05	Accepted
Deferred action	H7: Deferred action when mediated by data governance influences data qualityImprovement.	P=0.531>0.05	Rejected
Data quality management	H8: Data quality management when mediated by data governance influences data quality Improvement	P=0.000<0.05	Accepted
	H9: Data quality management directly influences data quality improvement.	P=0.000<0.01	Accepted
Data governance	H10: Data governance influences data quality improvement.	P=0.001<0.05	Accepted

Figure I: Data governance research model

REFERENCES

[1] Alhassan, I., Sammon, D., and Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision Systems*, Vol. 25, No.1, pp. 64–75

[2] Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. (2009).Methodologies for data quality assessmentandimprovement.*ACMComput.Surv*, Vol.41, No.3, pp. 16.

[3] Begg, C and Caira, T. (2012).Exploring the SME Quandary: Data Governance in Practice in the Small to Medium-Sized Enterprise Sector.*The Electronic Journal Information Systems Evaluation,*Vol.15, No. 1.Pp. 3-13

[4] Botha, M., Botha, A.,and Herselman, M.(2015).*Prioritized e-health data qualitychallenges: An application on the Tiered ART Monitoring Strategy*. Steyn, J., Van Belle, J.P. (Eds.). 2015. Beyond development. Time for a new ICT4D paradigm?Proceedings of the 9th IDIA conference, IDIA2015, Nungwi, Zanzibar: pp 217-231

[5] Cappiello, C. Caro, A. Rodriguez, A. Caballero, I.(2013). An Approach To Design Business Processes: Addressing Data Quality Issues. *Association for Information System AIS Electronic Library (AISeL) ECIS 2013.Proceedings of the 21st European Conference on Information Systems, 6-8 June 2013 at Utrecht, Netherlands.*

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

[7] Chen, H., Hailey, D., Wang, N. and Yu, P.(2014).A review of data Quality Assessment Methods for Public Health Information System. *International Journal of Environmental Research and Public Health,* Vol. 11, No. 5, pp.

[8] Chen, H., Yu, P. and Wang, N. (2012). Do we have the reliable data? An exploration of data quality for AIDS information system in China. Stud. Health Technol. Inform. Vol.192, pp.1042-1042.

[9] Department of Health South Africa(2012).eHealthStrategy South Africa 2012-2016.[Online] Available from http://www.doh.gov.za/docs/stratdocs/2012/eHealth_Strategy_South_Africa_2012-2016.pdf[Accessed on: 18/6/2016]

[10] Espinosa, A.J. andArmour, F.(2016). The Big Data Analytics Gold Rush: A Research Framework for Coordination and Governance. *49th Hawaii International Conference on System Sciences*.5-8 Jan 2016 at Grand Hyatt Kauai-Kauai, USA.

[11] Friedman, T. (2011).*Data Quality Makes the Big Time: How We Got Here, and Why We'll Never Be Done*. The Fifth MIT Information Quality Industry Symposium, July 13-15, 2011.

[12] Friedman,T. andSmith, M. (2011).Measuring the Business Value of Data Quality. Gartner, Inc. | G00218962

[13] Geiger, J.G. (2004).The Most Critical Initiative You Can Implement.Data warehousing management and quality. Intelligent Solutions, Inc., Boulder, co. pp. 098-29.

[14] Gomm, R. (2008). Social Research Methodology A Critical Introduction. Palgrave Macmillan, 2rd Edition.

[15] Hahn, D., Wanjala, P.,and Marx, M.(2013). Where is information quality lost at clinical level? A mixed-method study on information systems and data quality in three urban Kenyan ANC clinics.Glob. *Health Action 2013*, No.6,pp.1–10.

[16] Holmes, J. H. (2016).Privacy, Security, and Patient Engagement: The Changing Health Data Governance Landscape.*EGEMS* (Wash DC).Vol.4, No. 2.

[17] Information builder, (2011).Seven Steps to Effective Data Governance.A white paper, iWay software. [Online] available from http://www.enterprisemanagement360.com/wp-content/files_mf/white_paper/wp_iway_7steps.pdf[Accessed: 15/04/2016]

[18] Kahn, M. G., Jeffrey, S. Brown, Alein, T. Chun, B., Davidson, N., Daniella M., Patrick B. Ryan, Lisa M. S., Nicole G. W., Andrew E. W., and Meredith, N., (2015). Transparent Reporting of Data Quality in Distributed Data Networks. eGEMs (Generating Evidence & Methods to improve patient outcomes), Vol. 3, No. 1.

[19] Khatri, V. AndBrown, C. V. (2010). Designing data governance.*Communications of the acm.*Vol.53, No. 1.

[20] Kiwanuka, A., Kimaro, H.C., and Senyoni, W. (2015).Analysis of the acceptance process of district health information systems (dhis) for vertical health programmes: a case study of TB, HIV/AIDS and malaria programmes in tanzania.*EJISDC*, Vol.70, No.8,pp.1-14

[21] Kokemueller, J. "An empirical investigation of factors influencing data quality improvement success" (2011).*AMCIS 2011 Proceedings*,5 august2011 at Detroit, Michagan.

[22] Labouseur A. G. and Matheus C. C. (2017).An Introduction to Dynamic Data Quality Challenges.*ACM Journal of Data and Information Quality*, Vol. 8, No. 2, pp. 3

[23] Ladley, J. (2012). Data Governance: *How to Design, Deploy and Sustain an Effective Data Governance Program.*United States of America, *ELSEVIER* 2012.

[24] Ledikwe, J.H., Grignon, J., Lebelonyane, R., Ludick, S., Matshediso, E., Baraedi,W.S., Sharma, A. and Semo, B. (2014).Improving the quality of health information: a qualitative assessment of data management and reporting systems in Botswana.*BioMed central the open accesspublisher*, Health research policy and systems, Vol. 12.

[25] Loshin, D. (2010). Evaluating the business impact of poor data quality. Knowledge intergrity incorporated. Business Inelligence solutions.

[26] Loshin, D. (2013).Building a Data Quality Scorecard for Operational Data Governance SAS White Paper.

[27] McCrae, R.R., Kurtz, J.E., Yamagata, S. and Terracciano, A. (2011).PersSocPsychol Rev. Internal Consistency, Retest Reliability, and their Implications.*Journal of personality and social psychology,* Vol. 15, No.1,pp. 28–50.

[28] Mettler, T., Rohner, P. and Baacke, L., (2008). Improving Data Quality in Health Information Systems - A Holistic Design-oriented Approach. In: European Conference on Information Systems, Galway, Ireland.

[29] Mphatswe, W. Mate, K.S., Bennett, B., Ngidi, H., Reddy, J., Barker, P.M. and Rollins, N. (2012).Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa.Bulletin of the World Health Organization.World Health Organization, Vol.90, No.3, pp 176-182.

[30] Mutale, W., Chintu,N., Amoroso, C., Awoonor-Williams,K.,Phillips,J.,Baynes, C.,Michel,C., Taylor, A., andSherr, K. (2013).Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative. *BMC Health services Research*, Vol.13,No.9, pp. 12

[31] Murakwani, M. W. and Sethi, M. (2015). Big Data Quality: Early Detection of Errors in Process Flow using Alignments and Compliance Rules.*International Journal of Science and Research (IJSR).*Vol. 4, No. 6.

[32] Nahar, J., Imam, T., Kevin, Tickle, S., and Garcia-alonso, D. (2013). Issues of Data Governance Associated with Data mining in Medical Research: Experiences from an Empirical Study. Health Information Governance in a Digital Environment, E.J.S. Hovenga and H. Grain (Eds.) IOS Press.

[33] Niemi, E. (2013).*Enterprise architecture quality attributes: A case study.* Conference paper presented at 46th Hawaii International Conference on System Sciences 2013 on 7-10 January 2013 at Wailea, Hawaii, USA.

[34] Oracle Inc, (2011). All the Ingredients for Success: Data governance, Data quality and Master Data Management. The white paper, Hub solution design.

[35] Paolino, A.R, M.A., Elizabeth A. McGlynn, Lieu, T., Andrew F.N., Prausnitz, S., Michael, A.,Horberg, D.,Arterburn,E., Michael, K. G., Reesa, L. Laws, J.F., and Steiner, M. (2016). Building a Governance Strategy for CER: The Patient Outcomes Research to Advance Learning (PORTAL) Network Experience.*EGEMS*(Wash DC), Vol. 4, No.2.

[36] Pallant, J. F. (2005). SPSS survival manual.A step by step guide to data analysis using SPSS for windows (Version 12). Allen & Unwin. 2rd edition.

[37] Pallant, J. (2010). SPSS Survival Manual: a step by step guide to data analysis using SPSS for Windows. New York: McGraw-Hill.

[38] Patel, N.V. (2006). Organisation and Systems Design: Theory of Deferred Action. Palgrave Macmillan, pp. 253.

[39] Qazi, M.S., and Ali, M. (2011). Health Management Information System utilization in Pakistan: Challenges, Pitfalls and the way forward. *Bioscience Trends*, Vol.5, No. 6. pp. 245-254.

[40] Russom.P. (2008). Data governance strategies: Helping your organization comply, transform and integrate. Second quarter report. The Data Warehousing Institute (TDWI) best practices report.

[41] The Data warehousing Institute (TDWI), (2015).*Transforming data with intelligence*.TDWI conference in San Diego.20-25 September, 2015 at the Manchester Grand Hyatt.

[42] Vroom V.H. and Yetton P.W. (1973). *Leadership and decision-making*. Pittsburgh: University of Pittsburg Press.

[43] Tallon, P. P (2003). Corporate Governance of Big data: Perspective on Value, Risk and Cost. IEEE Computer Society, Vol. 13.

[44] Wang, R. Y., and Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems,* Vol.12, No.4.

[45] Wang, R. Y., Lee, Y.W., Pipino, L.L., and Strong, D.M. 1(998). Manage Your Information as a Product, *Sloan Management Review*, Vol.39, No.4.

[46] Weber, K., Otto, B., and Österle, H. (2009). One size does not fit all–A contingency approach to datagovernance. *Journal of Data and Information Quality (JDIQ)*, Vol.1, No.1.

[47] Weber, K., Cheong, L. K., Otto, B., andChang, V. (2008) *Organizing Accountabilities for Data Quality Management:A Data Governance Case Study.* Conference paper presented at IWI - Institute of Information Management on 28 October 2008 at St. Gallen.

[48] Wende, K. (2007). *A model for data governance-Organizing accountabilities for data quality management*.Paper presented at the 18th Australasian Conference on Information Systems, Toowoomba, Australia.

[49] Wende, K., and Otto, B. (2007).A contingency approach to data governance.*Iciq,pp.163*–176.

[50] Weills, p. and Oslon M. H (1989). An Assessment of Contingency Theory of Management Information Systems. *Journal of management information system (JMIS).* Vol 6, No. 1, pp. 59-86

[51] Wittwer, E. (2000). Data Quality Challenges for a State DOT. [Online] available from https://ntl.bts.gov/lib/10000/10000/10041/009.2prs.pdf Accessed on 28/6/2016

[52] World Health Organization. (2003). Improving data quality: a guide for developing countries. [Online] Available from http://www.wpro.who.int/publications/docs/Improving_Data_Quality.pdf Accessed on 23/5/2016.

[53] Young, A. and McConkey, K.(2012). Data governance and data quality – is it on your agenda or lurking in the shadows? *Journal of Institutional Research*, Vol.17, No.1, pp. 69-77.