Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 7 No. 4, December, 2016 - www.cisdijournal.net

A Framework for Data Quality Assessment and Improvement

Akpon-Ebiyomare D.E, Chiemeke S.C, & Egbokhare F.A
Department of Computer Science
University of Benin
Benin City, Nigeria
Email: ebiyomare77@yahoo.com;
Tel: 08023400716

ABSTRACT

A good number of organisations in Nigeria, in both private and public sector, rely on information systems for their daily operations, strategic planning and decision making. These organisations have information systems to either keep track of stock in sales and distribution organisation, keep track of transactions in financial institutions, register and maintain records of clients and employees of an organisation, register and maintain records of citizens of a country, or maintain academic and personal information of students in institutions of learning. The information generated by these organisations information systems have the potential of being used to provide greater understanding of their clients, intended users, processes, and the organisation itself. If the database used to generated the information contain deficient data models which are used as input, any generated output could be misleading. There has been reports of poor data quality generated by information systems. This has resulted in the development of a number of data quality frameworks but these frameworks were environment specific. The objective of the study is to adapt one of the existing data quality frameworks to make it suitable for use by organisations in Nigeria. The study determined the critical dimensions of data quality and the processes and activities responsible for the deficiency of the dimensions by means of interview and questionnaire at different stages. Respondents were all four categories of data quality stakeholders. The outcome were used to adjust the framework meant for adaptation.

Keywords: Data quality, framework, domain independent, database, data standardisation

CISDI Journal Reference Format
Akpon-Ebiyomare D.E1, Chiemeke S.C.2, and Egbokhare F.A3 (2016): A Framework for Data Quality Assessment and Improvement. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 7 No 4. Pp 103-114
Available online at www.cisdijournal.net

1. INTRODUCTION

Many organisations in Nigeria, in both private and public sector, rely on information systems for daily operations as well as strategic planning. According to Chiemeke and Akpon-Ebiyomare (2011), these organisations have information systems to keep stock (inventory), keep track of financial transactions (e.g. banks), register and maintain records of clients and employees (e.g. insurance companies) register and maintain records of citizens (government agencies), or maintain academic and personal information of students (educational institutions). For better results from the use of information technology, the quality of data used to generate information need to be of high quality to enable the user to effectively and efficiently make a decision or execute a task. Quality data is data that are "fit for use" from the perspective of the information consumer (English, 2009; Vaziri, 2012). Data that is without quality are unreliable and could even be dangerous in domains like healthcare and aviation (Batini, 2009). The rapid growth of the use of information systems has caused a growth in the volume of data and the number of data sources in both size and scope and consequently has significantly increased the complexity of data management. As the size of data accumulated grows, it gets more and more challenging to maintain, and this has resulted in poor data quality issues (Loshin, 2008).

The data quality issues include data standardisation, incorrect or contradictory data values, database contamination, untimeliness of reports, irregular or poor updates, duplicates, incomplete or blank fields, file backup issues, missing records / files and metadata issues (Scannapieco et al., 2005). Redman (2012) has identified the high cost of poor data quality. He identified that organisations may lose upwards of 10% of revenue due to poor operational data, together with other serious consequential effects. The cost of poor data quality to affected organisations is more than merely financial. Trust is lost from valuable customers (both internal and external), potential customers and sales are missed, operational cost increased through rework, workers lose motivation, long-term business strategy is hindered and business re-engineering is impeded. Vaziri et al. (2012); Loshin, 2011; and Chiemeke and Akpon-Ebiyomare (2011) emphsized that to avoid the pitfalls of poor quality data, organizations need to incorporate data quality assessment and improvement activities into their operations in order to periodically measure the state of their data to determine its quality.

The outcome of the assessment would determine the direction for improvement to take. There are a number of existing data quality frameworks developed for other environment. Considering that a data quality framework that is suitable for organisations in one environment might not be suitable for another (Vaziri, 2012; Klein, 2011). This study developed a framework for use by organisations in Nigeria to determine the quality of their information system data and carry out improvement irrespective of domain or sector.

1.1 Statement of Problem

With more organisations adopting information technology to improve or enhance their organisational processes in Nigeria, the issues of data quality management issues is on the increase. These data quality issues are summarised as shown in Table 1.

Table 1: Data quality issues in Nigeria's organisations

S/N	DATA / INFORMATION GENERATED	DATA QUALITY ISSUES
1	Driver License	Delay in issuance of drivers license (Timeliness issues); poor quality service (Responsiveness issues).
2	Population database	Un-updated database used to generate population statistics. Incorrect spelling of names and addresses. Database contamination. Discrepances over population figures. Outragous figures from regions with clearly low population (Guardian, June 17, 2014).
3	Voters register database	Omission of voters' names, inaccurate spelling of names and addresses; names in wrong voting units, duplication of voter names. (Nigeria Vanguard, March 23, 2014; Daily Independent, March 19, 2014; Punch, March 14, 2014).
4	Bank loan defaulters database	Un-updated database used to generate published bank debtor list (Nigerian Vanguard, August, 2015; Nigerian Guardian, August 2008; Sept 2015).
5	Voters card	Incorrect entries in voter card, missing names on voter register, un-updated voter database. Delayed release of voter card and election registers (timeliness issues).
6	National Examination result	Delay in releasing exam results; wrong spelling of names; allocating grades for subjects not registered for nor taken in exam (Uwadiae and Adelakun, 2008; Daily Times, 7th Sept 2011).
7	National Identity Card	Wrong spelling of names; Undue delay in release of ID cards. ID cards not ready for collection in over one year after registration.

These issues of data quality in Nigerian organisations necessitated studies by Egbokhare et al. (2013); and Chiemeke and Akpon-Ebiyomare (2011) on factors influencing the quality of data in information system databases and strategies for improvement. The studies indicated that there is a gap between data consumers expectations and the quality of data that they consume. The consequences of these data quality problems include (Su and Jin, 2007): Lost Opportunities (for data consumers), increased user service costs, user dissatisfaction, lost revenue, operational inefficiencies, poor Regulatory Compliance, poor Decision Making, lost business opportunities and system credibility.

1.2 Objective of Study

The objective of this study was to develop a framework suitable for organistions in Nigeria to assess and enhance the quality of information system database product irrespective of domain.

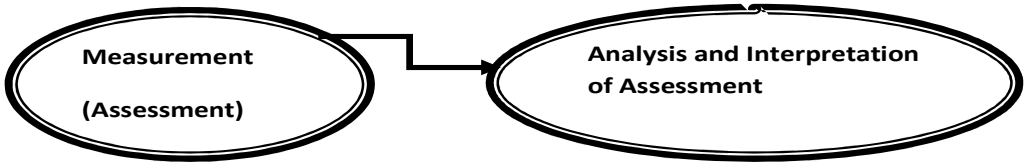
2. MATERIALS AND METHODS

In the course of reviewing the literature, fourteen data quality (DQ) frameworks were studied among which three domain independent frameworks were identified and singled out for in depth study. The three frameworks are the AIMQ (Lee et al., 2002); DQA (Pipino, 2009) and the TDQM (Wang and Strong, 1997). The AIMQ framework was chosen for adaptation by the study for the many reasons, chief among which are indicated in Table 2. (with asterisked serial numbers).

Table 2: Features of the AIMQ, DQA and TDQM frameworks

S/N	Characteristics / Features	TDQM	AIMQ	DQA
1	Domain Independence	Yes	Yes	Yes
2	Contains the 3 recommended stages	3 stages	2 stages	3 stages
3*	Allow extensibility to dimensions	No	Yes	No
4	Allow measurement of data quality	Yes	Yes	Yes
5	Based on the 'fitness for use' approach	Yes	Yes	No
6	Measure information system product	No	Yes	Yes
7*	Incorporate all four DQ stakeholders requirements	No	No	No
8*	Prioritize DQ dimensions	Fixed and predefined	Dimensions are weighted. weak / strong / relevant ones determined	Fixed and predefined
9	Identify causes of DQ problems	Yes	Yes	Yes
10	Has Improvement component	Yes	No	Yes
11*	Based on Data, Product or process oriented improvement	Data oriented only	Product oriented only	Data oriented only
12	Time to implement	Long	Short	Medium

The framework developed is an enhanced AIMQ framework making it more effective in assessing and improving data quality. This was achieved by incorporating the process oriented improvement methodology into the proposed framework. This involved identifying and redesigning the processes which create, modify and store data. It inserted checks and control procedures, as well as correcting defective processes. The AIMQ framework is a product-oriented framework that improves data quality from the perspective of the data consumers. A combination of the two methodologies provided a more robust framework. The AIMQ framework is shown in Figure 1.



PHASE 1: Measurement

1. **Input** – Data quality classified according to PSP/IQ Model
   Measurement of quality – Creation of a questionnaire for measuring IQ along specified important dimensions
   Output – Subjective quality assessment

PHASE 2: Analysis and Interpretation of assessment

Input – Assessment results
Measurement of quality – Use of IQ Gap Analysis Technique, aggregates of the dimensions ;
Performs a bench mark Gap analysis by considering best practices;
.
Output – Improvement directives.

Fig. 2: The AIMQ Framework Phases (Lee et al, 2002)

3. DATA GATHERING AND ANALYSIS

3.1 Data Gathering

The proposed framework was primarily based on case studies. Three organisation participated in the study. The study used both questionnaire and interviews to solicit information from the four categories of data quality stakeholders. The study determined the critical success factors of data quality using the result of questionnaire data analysis. It determined the absence or presence of processes and controls in place at the case organisations as well as the process deficiencies in the organisations using the interview data. The case organisations used as case study are as shown in Table 2.

Table 2: Case study organisations

S/N	Organisation	Org.Code	Product for assessment
1	Independent National Electoral Commission (INEC),	A	Voters register ; Voters card,
2	Federal Road Safety Commission (FRSC)	B	Driver's licensing
3	National Identity Card Management Commission (NIMC)	C	National ID card

3.2 Data Analysis

Both quantitative and qualitative approaches were employed within the current study. The nature of the data and the relationship between the method and the research objectives are considered as the basis for selecting the right method for analyzing data. Statistical tests were conducted using SPSS in order to provide results used to develop the framework. The quality of data in the domain under study was first assessed to determine its current state and the extent of any DQ deficiencies. The questionnaire data collected was analysed by employing statistical analysis methods. The descriptive analysis of the data was used to provide a summary of the responses using means, frequencies and standard deviations of the responses (Cooper and Schindler, 2003).

4. RESULTS

The analysis of thw data revealed the dimensions that are important to data consumers, as well as the current state of data quality were identified. Table 3 shows the critical data quality dimension. These are the result of the analysis indicating dimensions of data quality that data consumers used to judge the quality of data.

Table 3: Critical dimensions from data consumers perspective

S/N	Dimension	MEAN VALUE	LEVEL OF IMPORTANCE
1	Accuracy	4.21	Extremely Important
2	Availability	4.05	Extremely Important
3	Responsiveness	4.06	Extremely Important
3	Timeliness	4.12	Extremely Important

Table 4 shows the interpretation of the descriptive statistics which is the state of the critical dimensions based on data consumers perspective.

Table 4: Interpretation of the descriptive statistics

S/N	Dimension	MEAN VALUE	LEVEL OF PERFORMANCE
1	Verifiability	1.43	Poor
2	Responsiveness	1.92	Poor
3	Sufficient data	2.00	Poor
16	Timeliness	2.05	Poor
4	Relevancy	2.08	Poor
5	Reliability	2.34	Poor
6	Usefulness	2.35	Poor
7	Reputation	2.52	Moderate
8	Accessibility	2.55	Moderate
9	Believability	2.66	Moderate
10	Understandability	2.69	Moderate
11	Efficiency	2.71	Moderate
12	Consistency	2.75	Moderate
13	Availability	2.79	Moderate
14	Value-added	2.92	Moderate
15	Security	3.02	Good
17	Accuracy	3.53	Good
18	Completeness	3.55	Good
19	Duplication	3.78	Good

4.3 Interview Data Analysis Result

The objective of the interview was to gather data on the data environment and to understand the perception or perspective of the data managers, producers and custodians of the organisations. The interview helped to identify the root causes of the data quality problems in the organisations. Figure 2 shows the integrated activity diagram of the case organisations with the gray symbols indicating problem points.

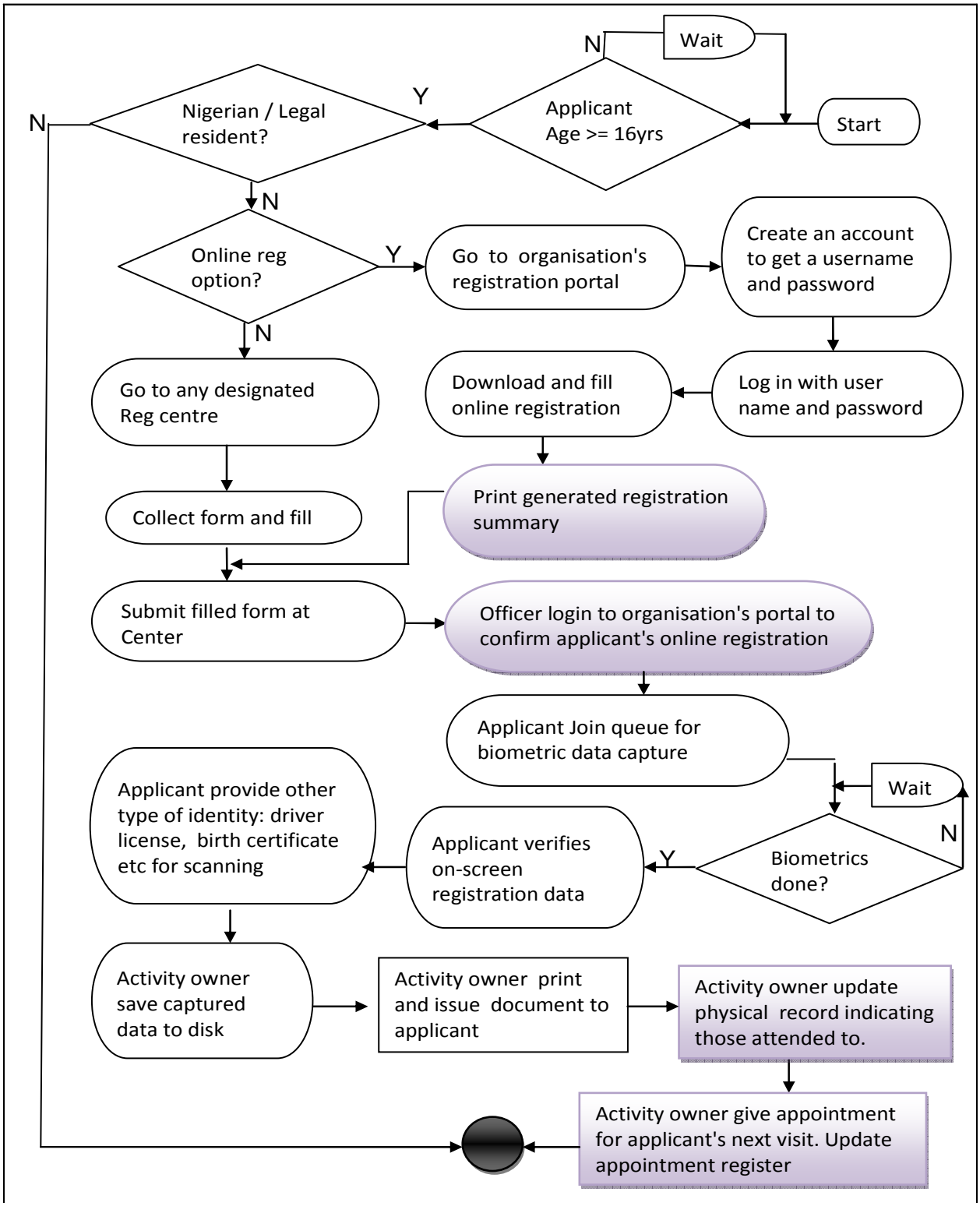


Fig. 2: Integrated Activity Dagram for Data Capture Activities

Figure 3 shows the integrated data flow map of the case organisations with the grey symbols indicating data destination points that are either missing or problematic and negatively influencing the quality of data.

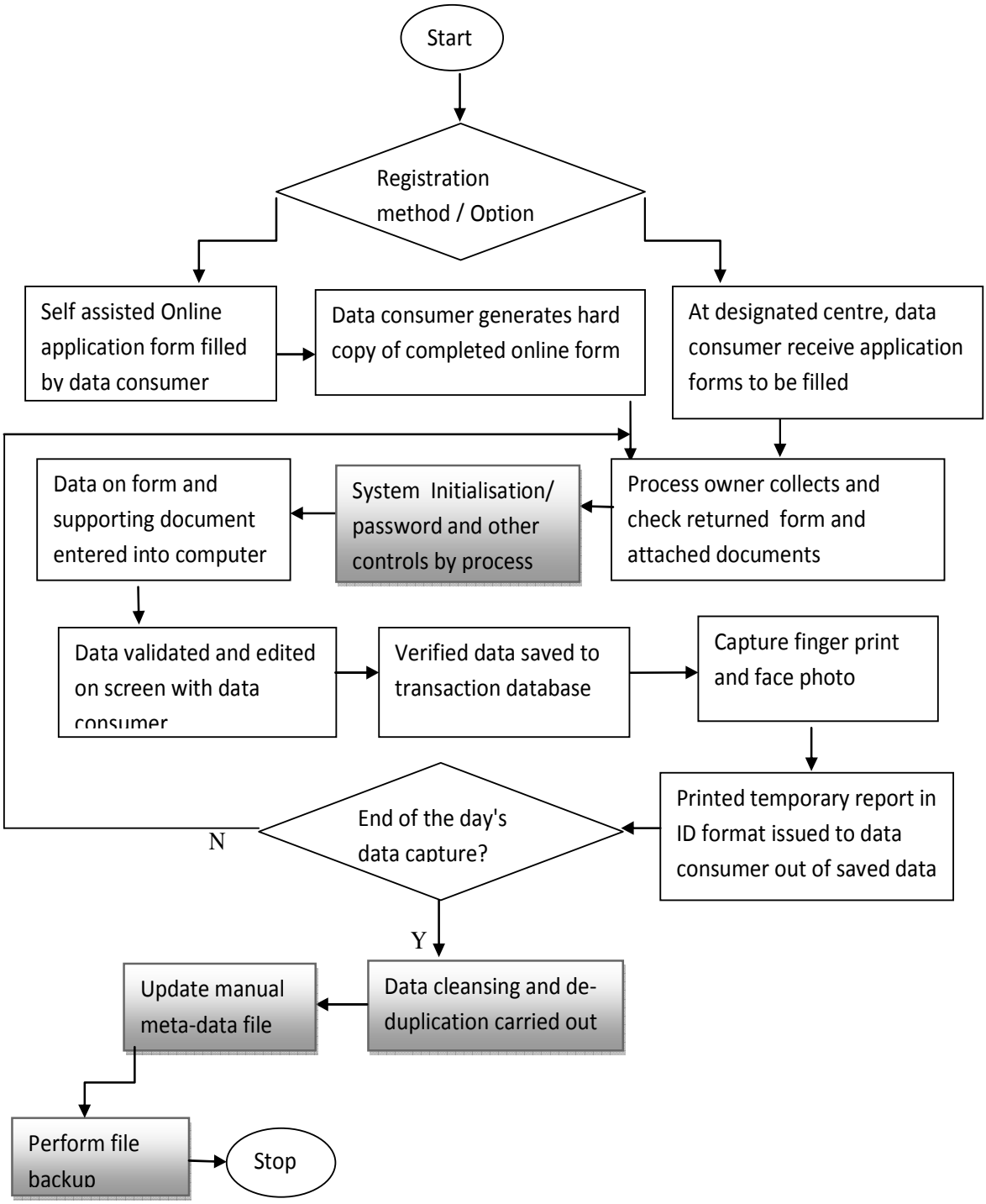


Fig. 3: Integrated Data Life Cycle and Controls in Place at Organization A,B and C.

The critical look at the data life cycle of the organisations under study showed the critical processes and activities that negatively affect the critical dimensions which ultimately affect the quality of data. Based on the identified critical data destination points missing in the cause of generating processed data, the study developed data improvement check list and then the framework of data quality. The checklists are shown in Tables 5, 6,7,and 8. The framework developed by this study is as shown in Figure 4.

Table 5: DQ Access /Process Control Checklist

DQ ACCESS /PROCESS CONTROL CHECKLIST					
S/N	**Control Elements**	**Fully In practice**	**Partially**	**Not at all**	**Comment**
1	Login password				
2	Logout from system				
3	Detection of double registration at point of data capture				
4	Data dictionary for document data.				
	Daily update of meta-data file				
5	Batch files at end of each work day				
6	Daily or weekly sign up for registration materials collected				
7	Daily or weekly sign up for registration materials returned (indicating used and unused consumables)				
8	Report on the day's performance to trace errors to process owners				
9	Up to date documentation for use by process owners				
10	Standard manual and electronic data submission forms and procedures exist				

Table 6: Timeliness dimension improvement checklist

TIMELINESS CHECKLIST					
Critical Dimension	**Process Element**	**Yes**	**Not all**	**Partially**	**Comment**
Timeliness	i. Check equipment (printer, scanner, data entry machine etc) at end of day as part of preparation for the next day. ii. Have enough staff to handle work on ground iii. Check to ensure control during the entry and handling process. iv. Control are added to important fields to avoid typo errors. v. Data captured are in backed up in batches. vi. Process is investigated from time to time to know how and whether captured data is batched, backed up and transferred. vii. Increase computer systems and personnel to reduce work overload. viii. Periodic change in the technology used for creating and storing data. ix. Change in the number and capability of the employees assigned to the different processes in the data production line.				

Table 7: Availability dimension improvement checklist

AVAILABILITY CHECKLIST					
itical nension	Process Element	Yes	Not at all	Partially	Comment
ilability	i. Generated reports are available for users; ii. Data consumers are aware of how to get available reports. iii. There are personnel whose main duty is to ensure proper execution of this task; iv. Progress of availability of reports is monitored. v. Request for regular updates of all collections documented. vi. Employ extra staff or outsource any overload of work				

Table 8: Responsiveness dimension improvement checklist

RESPONSIVENESS CHECKLIST					
Critical Dimension	Process Element	Yes	Not at all	Partially	Comment
Responsiveness	i. Personnel in place whose main duty is to attend to complaints from data consumers. ii. A mechanism in place for users of data to feedback data quality issues iii. Mechanism in place to tract complaints and responses to ensure they were resolved. iv. Request for regular updates on data consumer complaints . v. Rate data consumer satisfaction. vi. Create awareness amongst data consumers on how and whom to contact if they have complaints relating to the data they consume. vii. Data consumers are aware that if they are not satisfied, they could contact a higher authority. viii. The contact details of whom to contact when data consumers are not satisfied are readily available. ix. Employ extra staff or outsource any overload of work				

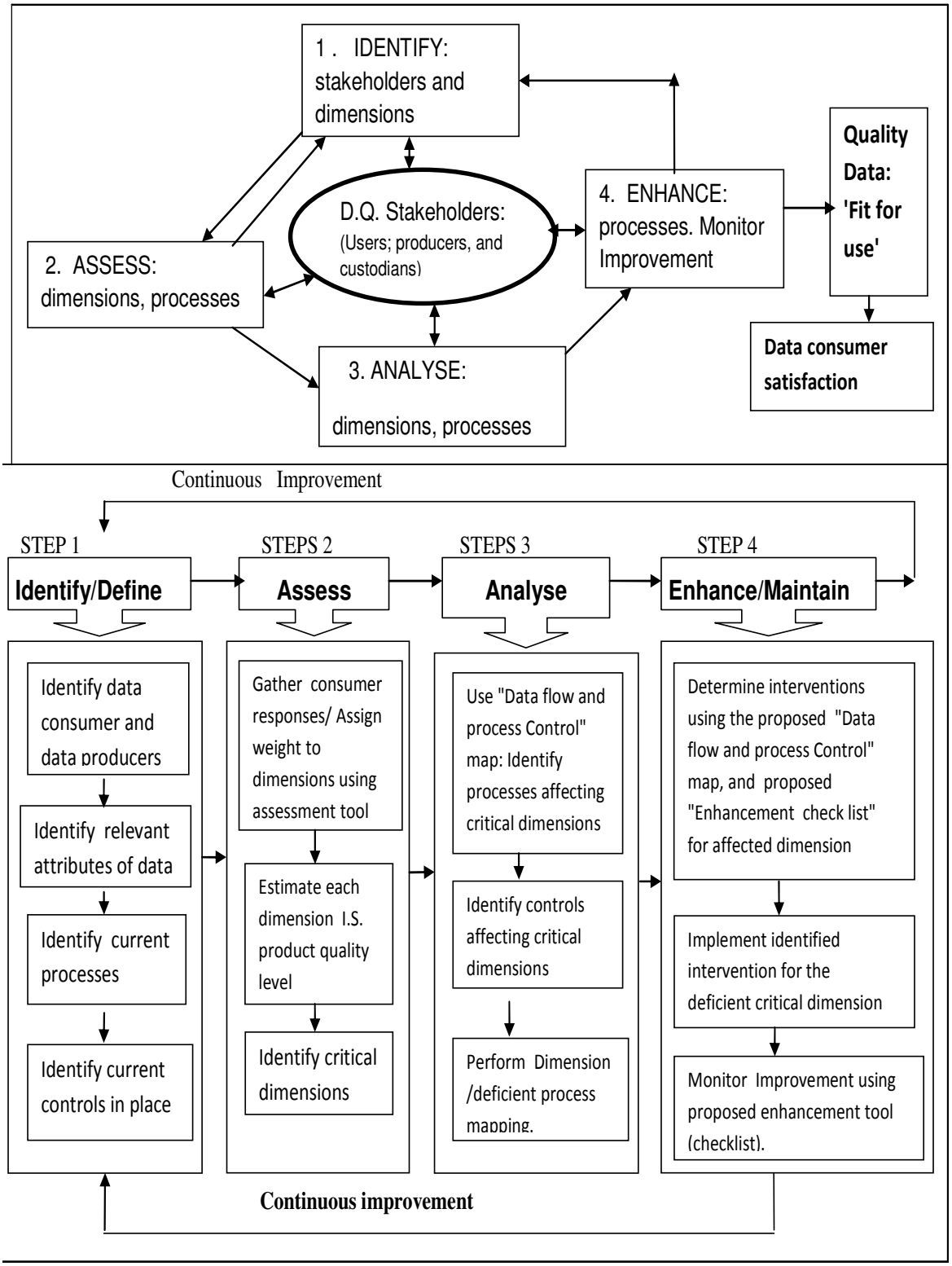


Fig. 4: Proposed Data Quality and Improvement Framework Developed by this Study

5. CONCLUSION

The study developed a generic framework for data quality assessment and enhancement for use by organisations in Nigeria. The framework is based on adaptation of the AIMQ framework by Lee et al. (2002). The framework identified two new dimensions (availability and responsiveness) that are peculiar to organisations in the Nigerian environment and were not considered in the development of the AIMQ framework but were found to be critical dimensions based on Nigerian data consumers perception. The framework incorporated input from all four categories of data quality stakeholders making an improvement of the AIMQ that is based only on data consumer perception of data quality. the framework is both process and product based as against the AIMQ that is only product based.

6. SUGGESTIONS FOR FURTHER STUDIES

As one direction for furthering this research work, the framework could be implemented in more categories of Nigerian organisation to determine its applicability in all sectors.

REFERENCES

1. Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. (2009). "Methodologies for data quality assessment and improvement. ACM Comput. Surv. 41, 3, Article 16.
2. Bulletpoint, (1996). "Automatic evaluation of metadata quality in digital repositories". International Digital Library 10(2–3):67–91. doi:10.1007/s00799-009-0054-4 CrossRef
3. Chiemeke S. C. and Akpon-Ebiyomare D. (2011). "High Quality Database Content: Success Factors and Improvement Strategy. In Proceedings of the 11th International Conference of the Nigerian Computer Society (NCS). Vol 22.
4. Daily Independent Newspaper (2014). "Election Fraud: Why are Election Registers Missing". Daily March 19, 2014. pp2. www.*independentnig.com*
5. Daily Times Newspaper. (2011). "WAEC and Confusing Results" www.dailytimes.org. Published 7-09- 2011. Accessed: 09-05-2014. Pp 4. *www.dailytimes.com.ng*
6. Egbokhare, F.A. , Akpon-Ebiyomare D., and Chiemeke S.C (2013). "A Framework For Data Quality .Management In Nigerian Higher Institutions". The Journal of Computing, Info Systems Development, Informatics and Allied Research. **4(1).**
7. Klein, B. D. (2011). "User Perceptions Of Data Quality: Internet And Traditional Text Sources". The Journal Of Computer Information Systems; Vol. 41 No. 4, 9–18.
8. Lee, Y.W., Strong, D. M., Kahn, B. K., and Wang, R. Y. (2002). "AIMQ: A Methodology for Information Quality Assessment".Journal of Information Management. 40(2): 133–460.
9. Loshin, D. (2011). "The Practitioner's Guide to Data Quality Improvement". San Francisco: Morgan Kaufmann Publishers.
10. Nigerian Punch newpaper (2015). " Drivers' license: FRSC warns Nigerians to stop patronising touts ". 21st August 2015. Pg 6. *www.punchng.com.*
11. Nigerian Punch Newspaper (2014). "A.P.C. Knocks INEC For Poor Voters Registrantion". March 14th 2014, Pp 12. *www.punchng.com*
12. Nigerian Guardian Newspaper (2014). "Lack of Data: Bane of Nigeria's Development." June 17th, 2014, Page 21. *www.guardian.ng.*
13. *Nigerian* Guardian *Newspaper (2008). "Bank Debtors and CBN Directive to Publish Names of Bank Loan Defauters" Pp1. www.guardian.ng*
14. *Nigerian* Vanguard *Newspaper (2015). " Anger, Denials, Threaths Trail Published Lists of Bank Loan Defaulters" August 10, 2015. Pp 4. www.vanguardngr.com*
15. *Nigerian* Vanguard *Newspaper (2014). "Biometric Voters Registeration: Problems Encountered". March 23, 2014. Pp 2. www.vanguardngr.*
16. Pipino, L. L., (2009). "Information Quality Assessment. In Encyclopedia of Database Systems". Springer Science and Business Media, LLC.
17. Redman, T. C. (2012). "Get Responsibility for Data Out of IT". Harvard Business Review
18. Su, Y. and Jin, Z. (2007). "A methodology for information quality assessment in the designing and manufacturing processes of mechanical products. In Proceedings of the 9th International Conference on Information Quality (ICIQ). pp:447–465.

19. Uwadiae, I. and Adelakun A. A. (2007). "The Challenges Of Ensuring Quality Assurance In Assessment- The Case of West African Examination Council" http://www.iaea.info/documents/paper_30e44070.pdf. retrieved on 27/3/2015.
20. Vaziri, Y., Reza. M. and Mehran M. (2012). "Towards a Practical State Reconstruction for Data Quality Methodologies: A Customized List of Dimensions". 2nd International Conference on Computer Science, Engineering and Applications (ICCSEA-2012). Springer Berlin; Heidelberg, 825–835.
21. Wand, Y., and Wang R.Y. (2012). "Anchoring data quality dimensions in ontological foundations". Communications of the ACM 39 (11).
22. Wang, R. and Strong, D. (1997). "Beyond accuracy: What data quality means to data consumers". Journal of Management Information Systems (MIS). Vol. 12, No. 4.

An Evaluation Of A Framework Of Data Quality Assessment And Improvement

Akpon-ebiyomare D.E[1], Chiemeke S.C.[2], and Egbokhare F.A[3]
[1,2,3]Department of Computer Science, University of Benin, Nigeria
Email: ebiyomare77@yahoo.com;
Tel: 08023400716

ABSTRACT