Assessment of the Database Product Quality of Nigerian Organisations

Cheimeke S.C. & Akpon-Ebiyomare D.E
Department of Computer Science
University of Benin
Benin City, Nigeria
Email: schiemeke@yahoo.com, ebiyomare77@yahoo.com
G.S.M: 2348023400716

ABSTRACT

Data begins its journey in information systems as codes of real life events and thereafter, they are processed for the purpose of improving the efficiency and effectiveness of the output from that organization' information system. It is especially so for organisations that rely heavily on data, as high quality information is an extremely valuable asset as well as one of the keys to their success. Many of these organizations have rightly invested hugely on information technology with the objective of enhancing the quality of their information product to increase or maintain customer satisfaction, revenues and/or profits. Despite the huge investments, reports have it that very significant percentage of poor quality data is being reported in organisations of all types both in the public and private sectors. When poor data quality results in customer dissatisfaction, there can be a direct negative impact on the organisation resulting in loss of customers, lives, trust and profit. Orgnisations therefore need to improve on and maintain high quality data to remain competitive. To achieve and maintain consistent high data quality, organisations need to first assess the quality of their data. The outcome of the assessment of quality points the direction for the required improvement and also provide means of comparison in order to determine the progress of any improvement initiative. This paper is to determine the quality of information generated by Nigerian organisations information systems. It focuses on the assessment across multiple data quality dimensions based on the expectation and experiences of data consumers of the information generated by public sector organisations' information system. Data was collected by means of a survey using questionnaire. The survey was performed on a random sample of 276 participants from a population of persons who consume any one of the reports of the case organisations studied. The data collected was analysed by employing statistical analysis methods. The outcome of the analysis gave the critical dimensions of data quality and the 'fitness for use' of the data produced by the organisations.

Keywords: Data, Dimensions, Assessment, Data Quality, Improvement, Data consumers.

1. BACKGROUND OF STUDY

The information generated by organizations could be used to generate greater understanding of their clients, intended users, processes, and the organisation itself. But, if the quality of the data used to generate the information is poor, any relationship found in the generated report could be misleading (Vaziri, 2012). This would also create dissatisfaction from the consumers (internal and external) whose information is contained in the information system. Poor data quality (DQ) related issues include (Klein et al, 2011): untimely reports; contaminated reports due to poor security of data; inconsistency of the content of reports due to inconsistency of the content of databases files used to generate reports.

Others include outdated data in reports as a result of poor or lack of regular updates; incomplete reports due to missing data fields or records. The causes of these data quality issues are numerous but they could be categorised into: ineffective management systems, out-of-control processes or other human behavior factors during data capture and production. Information quality problems can only be solved by implementation of an effective and efficient information quality management, so that the information production process can be controlled. Managing data quality involve assessments of the outcome of the data production process and this can be used to determine the level of quality of output from the system. Data quality assessment involves determining values for the dimensions of data quality and to understand the state of data in the database in order to improve it (Ahmed, 2011). To commence the improvement of data quality, organisations need to first determine their position on the data quality curve as this would help to determine any progress made in the course of their data quality improvement drive (English, 2009). Data Quality need to be measured in order for an organisation to identify precisely where they are and where they want to go with their data management infrastructure. Within organizations in Nigeria, people usually provide ad hoc solutions for DQ problems and these solutions are found after the problem has surfaced rather than using preventive measures (Akponewhe, 2013).

A typical example is the discovery of over four million duplicate records in the voters register (database) (Nigerian Punch,2014; Nigerian Vanguard, 2014). So also is the issue of bank loan defaulters publish in 2008 and 2015 (Newspapers: Nigerian Guardian, Punch and Vanguard). Contained in the databases used to generate the reports of these and similar organisations that rely heavily on data are ghost entries, un-updated entries, incomplete entries and omitted records (Akponewhe, 2013). Though not all of these issues are attributable to data quality, reports have it that a worrisome percentage is as a result of data quality issues. Assessment of the quality of organisation's operational data makes it easy to carry out and monitor improvements. Data quality assessment is the starting point from where the appropriate action to be taken to enhance quality is determined (English, 2009). This would enable comparison to be made in future to determine the level of improvement achieved.

2. STATEMENT OF PROBLEM

The problem of poor data quality processed by information systems is widespread in commercial and government environments. There is therefore a growing interest on the quality of data captured and subsequently produced after processing by organisations that have adopted information technology for the purpose of providing timely, accurate and reliable information to data consumers. Poor information quality has impact on various organisation aspects. Literature has identified the high costs of poor data quality: Redman (2008) identified that firms may lose upwards of 10% of revenues due to poor operational data, together with other consequences that include poor decisions and strategies. A Gartner report by Fisher (2009) stated that 75% of the revenue generated by organisations will be reduced and the cost of producing quality data will increase if organisations fail to include quality assurance in their operations.

In Nigeria, studies conducted by Chiemeke and Akpon-Ebiyomare (2011); Akpon-Ebiyomare et al (2012); and Egbokhare et al, (2013) reported that the data in our databases have issues. They identified the critical success factors of data quality and framework for data quality. According to Davenport et al. (2010), organisation that focus on information management are successful because they take the quality of their data that generate information used for decision making as very critical asset. This view was supported by Harris (2007). To avoid the pitfalls of poor quality data, Otto & Huner (2009); and Foley and Helfert (2010) stress that organizations need to incorporate data quality assessment and enhancement into their data operations and to periodically measure the state of their data from time to time to determine its quality. They further stated that assessing data quality to determine its level is something that organisations whether big or small should consider as very important to their success as it will give insight into the state of the quality of data they produce.

3. OBJECTIVE

The purpose of this study was to assess the quality of data generated by organisations information system. The assessment is based on the importance data consumers place on the dimensions of data quality and the "fitness for use" of the information generated for the consumption of the consumers. The assessment is based on 19 frequently occurring dimensions harvested from thirteen reviewed data quality frameworks (Table 1).

Table 1: 19 frequently occurring dimensions harvested from thirteen reviewed data quality frameworks

NS	DIMENSION	DH	DQQ	CDEP	AIMQ	DQArv	CED	ESAT	AIMQ	DBA	IQM	TDM	TDDM	DQ2
1	Accessibility	√	√	√	√				√	√			√	
2	Accuracy	√	√	√	√	√	√	√	√	√	√	√	√	√
3	Amount of data				√			√	√		√		√	
4	Availability	√	√	√		√	√	√		√	√			
5	Believability	√	√		√	√		√	√	√	√			
6	Completeness	√		√	√	√	√	√	√		√	√	√	√
7	Consistency	√		√	√	√	√	√	√	√	√	√		√
8	Duplication				√	√	√	√	√	√	√			
9	Efficiency	√			√	√		√	√				√	
10	Relevancy				√		√	√	√	√				
11	Reliability	√			√	√		√		√		√		√
12	Reputation			√	√	√	√	√	√	√	√		√	
13	Security				√		√	√	√		√		√	
14	Timeliness	√	√	√	√			√	√		√	√	√	√
15	Understandability				√	√	√	√	√	√	√		√	
16	Usefulness	√			√			√	√	√				
17	Responsiveness		√		√			√	√			√	√	
18	Value added	√	√		√	√	√				√		√	√
19	Verifiability	√	√			√	√			√		√		

4. METHODOLOGY

There are techniques and methodologies in place to assess and improve the quality of data (Cronemyr and Danielsson, 2007). The quality of data in a database can be assessed using either process, structural, or outcomes variables. Since none is more superior than the other (Donabedien, 2003), this study applied the outcome variable approach. Using this approach, data was gathered using questionnaires from consumers of the information system reports generated by three intensive data usage organisations with presence in all the states of Nigeria. The questionnaires were distributed by means of electronic mail and manual distribution using random selection method. The number of usable responses received were 276. The questionnaire on one section required the participants to rate the dimensions of data quality based on the importance they placed on them (the dimensions) on a Likert scale of 1 to 5. Another section of the questionnaire required respondents to rate the quality of information produced by the organisation which they consume, on a scale of 1 to 5. Demographic data on the participants were also collected. The items in the questionnaire had been used in a number of studies for which reliability were calculated and acceptable (Lee et al, 2002; Wang and Strong, 1996). This study used the Cronbach alpha coefficient method to assess internal consistency (Hair et al, 2007). The sample gave a Cronbach alpha value of 0.969 indicating good reliability of the instrument. The data collected was analysed by employing statistical analysis methods. Summary of the findings from the study was prepared and dimensions that are important to DQ as well as the current state of DQ dimensions were identified.

5. RESULTS

5.1 Demographic Statistics
In terms of data consumer roles with respect to the organisations they responded for, all three of the targeted groups (organisations A, B, and C) are represented. Respondents for organisation A make up nearly half the sample size of 158 with 57.25%. Organisation B constitute 26.08% of the sample, and organisation B respondents make up the remaining 16.67%. Independent *t*-tests were conducted on all the target items in the survey to determine whether there were significant differences in the responses:

a) between those who completed the survey through email and those through manually administered survey.
b) between the genders.
c) different parts of the country.
d) different Occupation,
e) different stakeholder role

No item posed any significant difference in response at $p = .01$. except for V (different stakeholder role). 6.0% of the target items were found to have significant differences at $p = .01$. Closer inspection revealed a systematic pattern in which data producers who work in the case organisations happen to be responding as data consumers (they also consume the information system product under study) rated the quality of information in their systems higher than information consumers and the DQ dimensionsused for the rating was different from. Because of this pattern, it was determined that separate analysis would be required to assess the implications of these differences. Table 2 shows the summary of the demographic characteristics of the survey participants.

Table 2 :Demographic characteristics of respondents

S/N	Measure	Item	Frequency	%
1	Organsation consumer report	A	158	57.25
		B	72	26.08
		C	46	16.67
		Total	276	100
2	Gender:	Male	197	71.4
		Female	79	28.6
		Total	276	100
S/N	Measure	Item	Frequency	%
3	Years of consuming information system report of the organisation	Under 1 year	71	25.7
		1 to 5	114	41.3
		6 to 10	42	15.2
		11 to 20	28	10.2
		Over 20	21	7.6
		Total	276	100
4	Age of respondent	18 - 29	51	18.5
		30 - 39	113	40.5
		40 - 49	58	21.0
		50 - 59	47	17.0
		60 and above	7	3.0
		Total	276	100
5	Occupation of respondent	IT Professionals	23	8.3
		Private sector employees	51	18.5
		Civil servants	97	35.2
		Self employed	32	11.6
		Students	64	23.1
		Others	9	3.3
		Total	276	100
6	Highest Qualification	Secondary sch or equivalent	41	14.9
		Diploma/First degree	148	53.6
		Post graduate degree	87	31.5
		Total	276	100

5.2 Importance Rating and Critical Dimensions of Data Quality
Determining the critical data quality dimensions as perceived by the consumers provided us with the data quality attributes most important to the data consumers.

Table 3: Dimension importance ratings

S/N	Dimension	N	Minimum	Maximum	Mean	Std Deviation	Variance
1	Accessibility	276	1	5	2.42	.879	.773
2	Accuracy	276	1	5	4.21	.649	.920
3	Appropriate amount of data	276	1	5	2.40	.885	.783
4	Availability	276	1	5	4.05	.067	1.331
5	Believability	276	1	5	3.05	.923	.853
6	Completeness	276	1	5	2.74	.852	.727
7	Consistency	276	1	5	2.60	1.205	1.331
8	Duplication	276	1	5	2.25	.944	.890
9	Efficiency	276	1	5	3.20	1.154	1.331
10	Reliability	276	1	5	2.93	1.374	1.889
11	Relevancy	276	1	5	2.82	1.351	1.825
12	Responsiveness	276	1	5	4.06	.991	.983
13	Reputation	276	1	5	2.68	1.293	1.671
14	Security	276	1	5	3.43	1.004	1.008
15	Timeliness	276	1	5	4.12	1.284	1.649
16	Understandability	276	1	5	2.48	1.052	1.106
17	Usefulness	276	1	5	2.64	1.125	1.269
18	Value-added	276	1	5	2.72	1.619	2.622
19	Verifiability	276	1	5	2.59	1.690	1.365

A dimension mean was computed as the average of the responses to all of the items in the survey instrument. From Table 3, we identified which items were considered as highly important, moderately important or not important to data consumers as shown in Table 4.

Table 4: Level of importance interpreted

S/N	Dimension	MEAN VALUE	LEVEL OF IMPORTANCE
1	Duplication	2.25	Low
2	Appropriate amnt of data	2.40	Low
3	Accessibility	2.42	Low
4	Understandability	2.48	Average
5	Verifiability	2.59	Average
6	Consistency	2.60	Average
7	Usefulness	2.64	Average
8	Reputation	2.68	Average
9	Value-added	2.72	Average
10	Completeness	2.74	Average
11	Relevancy	2.82	Average
12	Reliability	2.93	Average
13	Believability	3.05	Average
14	Efficiency	3.20	Average
15	Security	3.43	Average
16	Availability	4.05	Extremely Important
17	Responsiveness	4.06	Extremely Important
18	Timeliness	4.12	Extremely Important
19	Accuracy	4.21	Extremely Important

5.3 Performance Ratings (Fitness for Use) of I.S. Product
Table 5 shows the performance ratings. Performance was rated on a scale of 1 to 5 :

Table 5: Performance ratings by data consumers

S/N	Dimension	N	MINIMUM	MAXIMUM	MEAN	STD DEVIATION
1	Accessibility	276	1	5	2.55	1.160
2	Accuracy	276	1	5	3.53	.988
3	Appropriate amnt	276	1	5	2.00	.867
4	Availability	276	1	5	2.79	1.397
5	Believability	276	1	5	2.66	1.155
6	Completeness	276	1	5	3.55	1.177
7	Consistency	276	1	5	2.75	1.180
8	Duplication	276	1	5	3.78	1.048
9	Efficiency	276	1	5	2.71	1.174
10	Reliability	276	1	5	2.34	1.074
11	Relevancy	276	1	5	2.08	1.301
12	Responsiveness	276	1	5	2.22	1.338
13	Reputation	276	1	5	2.52	1.471
14	Security	276	1	5	3.02	1.280
15	Timeliness	276	1	5	3.05	1.065
16	Understandability	276	1	5	2.69	1.244
17	Usefulness	276	1	5	2.35	1.301
18	Value-added	276	1	5	2.92	1.310
19	Verifiability	276	1	5	1.43	1.366

Table 6: Performance Ratings interpreted

S/N	DIMENSION	MEAN VALUE	LEVEL OF PERFORMANCE
1	Verifiability	1.43	Poor
2	Responsiveness	1.92	Poor
3	Appropriate amount	2.00	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
16	Consistency	2.75	Moderate
13	Availability	2.79	Moderate
14	Value-added	2.92	Moderate
15	Security	3.02	Moderate
12	Timeliness	3.05	Moderate
17	Accuracy	3.53	Good
18	Completeness	3.55	Good
19	Duplication	3.78	Good

5.5 Discussion and Findings

From Table 4 that showed the importance ratings of the dimensions, data consumers rated 4 dimensions as the most important dimensions that they used to judge the quality of data that they consume. In Table 5 consumers rated the data quality dimensions (fitness for use of the output of the information system) based on the quality of data they consume. The findings from the analysis of their responses on importance and fitness for use showed that:

I.   Of the 4 most important dimensions that consumers used to judge the quality of data, only 1 of them is among the dimension rated as "Good" (average) in terms of performance or fitness for use (Table 6).

II. For the ratings for performance (fitness for use of the information system product), the highest ratings of the four critical dimensions Of the four dimensions they rated as critical dimensions, Accuracy dimension has the highest rating.

Table 6: Performance rating of the data quality dimensions

	Most important Dimension (used to judge data quality)	Perfomance Rating (Fitness for use)
1	Accuracy	Good
2	Availability	Moderate
3	Timeliness	Moderate
4	Responsiveness	Poor

iii. Only three dimensions out of the 19 dimensions got the mean rating just above poor performance. (table 6)

iv. No dimension was rated "excellent" or "very good" in terms of 'fitness for use'.

In comparing the expectations of data consumers with the rating of the performance of the organisations, the findings indicate that data consumers are not satisfied with the quality of data they consume: As a result of the dissatisfaction:

i. Many eligible Nigerians shy away from going through the application process to acquire the I.S. reports produced by the organisations .

ii. Some resort to multiple registration to acquire the information system report when they have challenges of getting their data in the database of the organisation updated to reflect any new data like marital status, or number of children. This results in multiple records of the same individual in the database. As result, any report generated from the database will not reflect the true real world situation.

iii. Timeliness issue: Touts take advantage of the deficiency of the timeliness dimension to offer applicants of some of the organisation shorter time to get their required information system report. This touting results in loss of revenue that should accrue to government.

iv. Duplication is a critical dimension from the perspective of data producers but discussions with data consumers showed that they did not seem to care how many times their names appeared in the organisations database so long as it does not affect the accuracy or timeliness of the report they require.,

v. On the part of the organisations, poor quality data lead to loss of productivity and time used to rework data that is too poor to be used or removing errors like duplications from data.

vi. Could cost the organisation lot of money. A recent case is that of a former legislator whose name was recently published (Guardian newpaper - August 2015) as a bank loan defaulter. The person instituted a law suit of N500 million.

6. CONCLUSION

Good quality data is fundamental to the success of any organisation whether big or small. Data quality is especially critical for organisations that rely heavily on data for its operations. Organisations in Nigeria need to pay attention to the quality of data that enter their information system. Data is to a product and not a by-product. The quality of a product is as good as its raw material. A number of factors contribute to poor data quality but the biggest of them has to do with the people who handle the data, the controls in place and the technology used. It has been reported variously that up to 85% of errors in information system databases are introduced at the point of data capture. As such, organisations need to ensure that their data capture personnel are qualified and the number of workers are sufficient for the job at hand. Organisations in Nigeria are known to give target that must be met and as a result, the employees responsible for data capture concentrate more on quantity (number of records entered) over quality (number of records that are free from error). Some organisations are known to use ad-hoc data capture staff that are replaced on a yearly bases. It may be a cheap method of getting work done but at the expense of data quality. The use of suitable and adequate equipment is critical for the achievement of quality data. This in addition to controls of the processes that produce data. In working on the processes to improve them the first place to start is the data source. This is to identify the root cause of the data quality issues which could include:

• Not enough data validation and edits
• Not understanding the meaning of a data element
• No adequate metadata update
• No domain definitions
• No reconciliation process
• No data verification process
• Poor data entry training or use of inexperienced staff
• Inadequate time for data entry (work overload)
• No incentive for quality data entry

7. SUGGESTIONS FOR FUTURE RESEARCH

Considering that data is a critical asset in many organisations, and deficiencies with the quality of this data have been reported to created situations that include resulted missed opportunities and in extreme cases loss of life (Woodall et al, 2013). The study have established that the quality of data in Nigerians organisations information systems master data which is used to generate report fall below the expectations of those who are expected to consume the reports. As a preventive measure, further research in this area that would culminate in the development of a framework for data quality assessment and enhancement is recommended. The recommended framework is expected to be used inhouse by data quality managers, data custodians or any IT professional saddled with the responsibility ensuring quality data. There are a number of data quality assessment and enhancement frameworks in the literature but whether these frameworks are suitable for use by Nigerian organisations is to be determined by further research. Based on the literature the framework for data quality improvement is influenced by a number of factors which include: the organisation's location, structure, culture and orientation of the local people; the types of information systems that use, modify, and manage the data; the software used to process the data; and the processes that create and update the data (Cervo and Allen, 2011). future research to develop a data quality framework that is practicable and usable by Nigerian organisations to assess and enhance data quality is here recommended.

REFERENCES

1. Akpon-Ebiyomare, D., Chiemeke, S.C. and Egbokhare, F.A. (2012). "A Study Of The Critical Success Factors Influencing Data Quality In Nigerian Higher Institutions, African Journal of Computing and ICT. Volume 5, No 2. March Edition.
2. Carey M.J. Ceri S., Benstein P., Dayal U., Faloutsos C., (2006) "Data-centric systems and applications." Springer, Berlin Heidelberg, Germany.
3. Cervo, D., and Allen, M. (2011) Master Data Management in Practice: Achieving True Customer MDM, John Wiley & Sons Inc., New Jersey, ISBN 978-0-470-91055-9
4. 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). Held in Abuja, Nigeria.
5. Cronemyr P. and Danielsson M (2007). 'Process Management 1-2-3: A maturity Model and Diagnostics Tool. Journal article. Total Quality Management and Business Excellence, 24(7-8) 933-944.
6. Donabedian, Avedis, (2003). 'An Introduction to Quality Assurance in Health Care, Oxford University Press: Oxford.
7. 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. *Vol 4, No 1.*
8. Harrington, H. J. (1991). Business Process Improvement. New York: McGraw-Hill.
9. IBM (2007). The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance
10. JIG (2013). "Improving business performance through improved data quality". Joint Information Group / MOD / LSC Group 2013. Version 3 of the Data Quality Assessment Tool (DQAT). The JIG is a subgroup of the National Defense Industries Council.
11. Lee, Y. W., Strong, D. M., Kahn, B. K. and Wang, R. Y. (2002). "AIQM: A Methodology for Information Quality Management". Information and Management, 4: 133-146.
12. Madnick S.E., Wang R.Y., Lee Y.W., Zhu H. (2009). " Overview and framework for information quality research. Journal of data and infor,ation quality, Vol1.
13. Paulk, M. C., Curtis, B., Chrissis, M. B., and Weber, C. V. (1993) Capability Maturity Model, Version 1.1. *IEEE Software,* Vol. 10, No. 4, July, pp. 18-27.
14. Ponzio F. J. (2010) Industry Insight DATAMRI: "A user Tool for Assessing Data Quality". http://www.symbolic.com/Industry-Insight/DATAMRI
15. Suhardi A. R., Gunawan, I.G.N.; Dewi, A.Y. (2014). "Designing process maturity measure of Total Information" Quality Management (TIQM). IEEE Conference Publications
16. Woodall P., Parlikad A. and Ajith K. P.. (2013). "Data Quality Assessment: The Hybrid Approach": University of Cambridge - Service Alliance www.cambridgeservicealliance.org