A FRAMEWORK FOR DATA QUALITY MANAGEMENT IN NIGERIAN HIGHER INSTITUTIONS

Egbokhare, F.A., Akpon-Ebiyomare, D.E and Chiemeke, S.C.
Department of Computer Science
University of Benin
Benin City, Nigeria

Correspondence: fegbokhare@yahoo.com

Abstract
This empirical research developed a questionnaire instrument as a tool for gathering data targeted at identifying critical success factors of data quality in tertiary institutions Information Systems databases (based on reviewed literature and findings reported in Akpon-Ebiyomare et al (2012)). Factors motivating the study include stakeholder perceived data quality success factors in Nogerian tertiary Institution databases. Based on the findings from data gathered from the research, a generalized framework for Data quality management in Nigerian Higher Institutions is proposed. The framework provides evaluation and improvement components that can be used to interact with the other components to ensure data integrity and hence quality data success at all times.

Keywords: Framework, Data Management, Critical Factors, Integrity, Evaluation and Nigerian Higher Institutions.

1. INTRODUCTION/BACKGROUND OF STUDY

Data is one of the most critical assets of any organization because the quality of data has a strong influence on the decision making processes. While wrong organizational decisions may not all be 100% attributed to data quality issues, Strong et al (1997) noted that the percentage contributed by poor data quality is quite high. As data quality awareness and requirements increased, researchers began to focus on data quality frameworks (Tayi andBallou,2008; Wang et al. 2006); data quality assessment (Wand and Wang 2006; English 2009); data quality management (Wang 2009; Fletcher, 2004) and data quality dimensions (Wang et al 2006; Pipino 2012). Several factors influence the quality of data in organizations (Akpon-Ebiyomare et al, 2012; Redman, 2006; Tayi and Ballou, 2008). Ballou et al 2002 identified four dimensions that are most pertinent to data values: Accuracy, Timeliness, Consistency and Completeness. Olson (2003) and Wang et al(2006)view accuracy and correctness as the most important dimensions because if data is not accurate, then the other dimensions are of little importance. Other approaches for assessing quality of data attempted to manage data in terms of definition, content and presentation, (English 2009).

Poor data quality can have adverse effects on organizations, for example Olson (2003) reported that Poor data quality management costs more than $1.4 billion annually in 599 surveyed companies and up to 88% of data-related projects fail, largely due to issues with Data Quality. Wang et al (2001) discovered that 70% of manufacturing orders are assessed as being of poor data quality while Data Quality issues accounted for nearly $600M losses for US companies in 2001. Redman (2006)estimated that poor data quality results in 8% to 12% loss of revenue in a typical enterprise, and informally estimated poor data quality to be responsible for 40% to 60% of expenses in service organizations.

Strong et al, (2004) observed that between 50% and 80% of computerized U.S. criminal records are estimated to be inaccurate, incomplete and ambiguous. Because of the imperfect nature of data therefore, the need for organizations to design frameworks for continuous improvement of data quality cannot be over emphasized.

Wang (2009) proposed frameworks for the assessment of data quality. Wang's framework takes into account the fact that there are different types of data and different consumers and users. The framework also recognizes that data is used for different applications. As such, the needs and quality requirements are different for the different data customers and applications. Akpon-Ebiyomare et al (2012) studied critical success factors influencing data quality in Nigerian higher Institutions. The research used the University of Benin as a single case study and obtained twenty-one (21) data quality factors out of which thirteen (13) were rated critical. The twenty-one factors obtained were exposed to other stakeholders/custodians of data at the International Conference of the Nigerian Computer Society at Abuja, Nigeria in 2011. The result obtained is used to propose a framework for data quality management in Higher Institutions in Nigeria.

2. MATERIALS AND METHOD

In order to identify critical success factors of data quality in tertiary institutions IS database, a survey instrument was developed. The instrument (questionnaire) was developed based on reviewed literature and the findings reported in Akpon-Ebiyomare et al (2012), with specific reference to Tertiary Institutions Information Systems. The questionnaire was fine-tuned to effectively address all the aspects of stakeholder perceived data quality success factors.

A pretest of the questionnaire was administered on five data quality stakeholders: one data supplier, two data consumers/users (Ph.D students), one data manager and one data custodian. The survey questionnaire comprised the 21 factors identified as capable of influencing the data quality.

Respondents were asked to rate each of those factors according to their perception on the importance of each factor in ensuring data quality in IS. The importance was based on a five- point Likert scale with '5' being the highest rating and '1' the lowest.

A total of 300 questionnaires were distributed to randomly selected stakeholders at the Nigerian Computer Society (NCS) 10th International conference that held at the International Conference Centre, Abuja-Nigeria in July, 2011. It was specifically indicated in the questionnaire that respondents must be stakeholders (employees or students) of tertiary institutions in Nigeria. It was not possible to determine immediately who amongst the conference participants were eligible (stakeholders of tertiary institutions) to respond to the questionnaires, since the participants were IT professionals from diverse organizations in Nigeria. We had to therefore inquire from the respondents if they fell within the stipulated category before administering copies of the questionnaire. Only 123 (41%) copies of the questionnaires returned were found to be complete and usable for the final analysis.

3. DATA ANALYSIS

Demographic information can aid the understanding and build the possible useful correlations with other survey findings.

Table 1: Demographic characteristics of Respondents

Measure	Item	Freq	%
Sex	Male	84	68.3
	Female	39	31.7
	Total	123	100.0
Age	18-25	23	18.7
	26-35	34	27.6
	36-50	35	28.5
	Above 50	31	25.2
	Total	123	100.0
Highest Qualification	Student	21	17.1
	Bachelor	30	24.4
	Post Graduate	72	58.5
	Total	123	100.0
Years Of I.T. Experience	Less than 5 yrs	30	24.4
	6-10yrs	21	17.1
	10-20yrs	30	24.4
	Greater than 20yrs	42	34.1
	Total	123	100.0
Primary Job Function/ Stakeholder Group	Data producers	73	59.4
	Data Consumers	24	19.5
	Data Managers	10	8.1
	Data Custodians	16	13.0
	Total	123	100.0

This section describes the demographic information of the respondents in order to highlight the important characteristics of the respondents. Table 1 shows the demographic characteristics of the respondents.

Table 1 shows that most of the respondents fell within the age bracket of 26 – 50 years (56.1%) with work experience of between 5 (five) to 20 (twenty) years. Also, their job definitions cut across the necessary data users/producers in organizations: data producers (59.4%), consumers(19.5%), Managers(8.1%) and custodians (13%). Factor analysis, a multivariate technique for reducing matrices of data to their lowest dimensionality by the use of orthogonal factor space and transformations that yield predictions and /or recognizable factors (AKinyokun and Chiemeke, 2006) was used to group and analyze the data.

Principal Component factor Analysis (PCA) is a commonly used method for grouping the variables under few unrelated factors. Variables with a factor loading ≥ 0.4 are grouped under a factor. The main application of factor analysis in this research was to classify the variables. Six factors were produced from the factor analysis procedure with total variance > 74.1%. The balance 25.9 could be as a result of various factors which include process or environmental factors.

To make interpretation easier and the results more meaningful, an **equamax rotation** was done. This gives fewer components (factors) that are uncorrelated with one another. Principal component factor analysis with equamax rotation was performed on the survey data and the result is presented in Table 2. Only factor loadings ≥ 0.4 were extracted.

Table 2: Rotated component matrix (only factors >=4.0 are shown)

Item	Component					
	1	**2**	**3**	**4**	**5**	**6**
q1	.800					
q2	.733					
q3						
q4				.842		
q5	.798					
q6		.803			.744	
q7			.787			
q8						
q9	.801					
q10						
q11			.728			
q12			.602			
q13		.757				
q14		.750				
q15						
p16			710			
q17						.638
q18		.810				
q19		.512				
q20	.756					
q21				.803		

The six factors namely Organizational Management, Technical characteristics, Implementation and Project Management, People-Related characteristics, Performance Evaluation and Work Environment are presented in Table 3 together with their loadings.

Table 3: Factors and their loadings

Factor 1	**ORGANISATIONAL MANAGEMENT** p1 - Management Commitment p2 - Data Quality Policies p5 - Organizational Structure p19 - Sufficient fund to execute Project
Factor 2	**TECHNICAL CHARACTERISTICS** p6 - Condition of IS information system hardware and software p13 - Input Controls p14 - Understanding of IS and Importance of DQ p18 - Continuous Power Supply
Factor 3	**IMPLEMENTATION AND PROJECT MANAGEMENT** p7 - Data quality controls & improvement p11 - Change management p12 - Internal controls (systems and processes) p15 - Team work.
Factor 4	**PEOPLE CHARACTERISTICS** p4 - Training and communication p21 - User/customer focus & involvement p8 - Information supplier quality management
Factor 5	**PERFORMANCE EVALUATION** p9 - Performance evaluation
Factor 6	**WORK ENVIRONMENT** p17 - Conducive physical environment

We combined these six factors with other organizational factors to design the framework in Figure 1 for data quality management in Nigerian higher institutions.

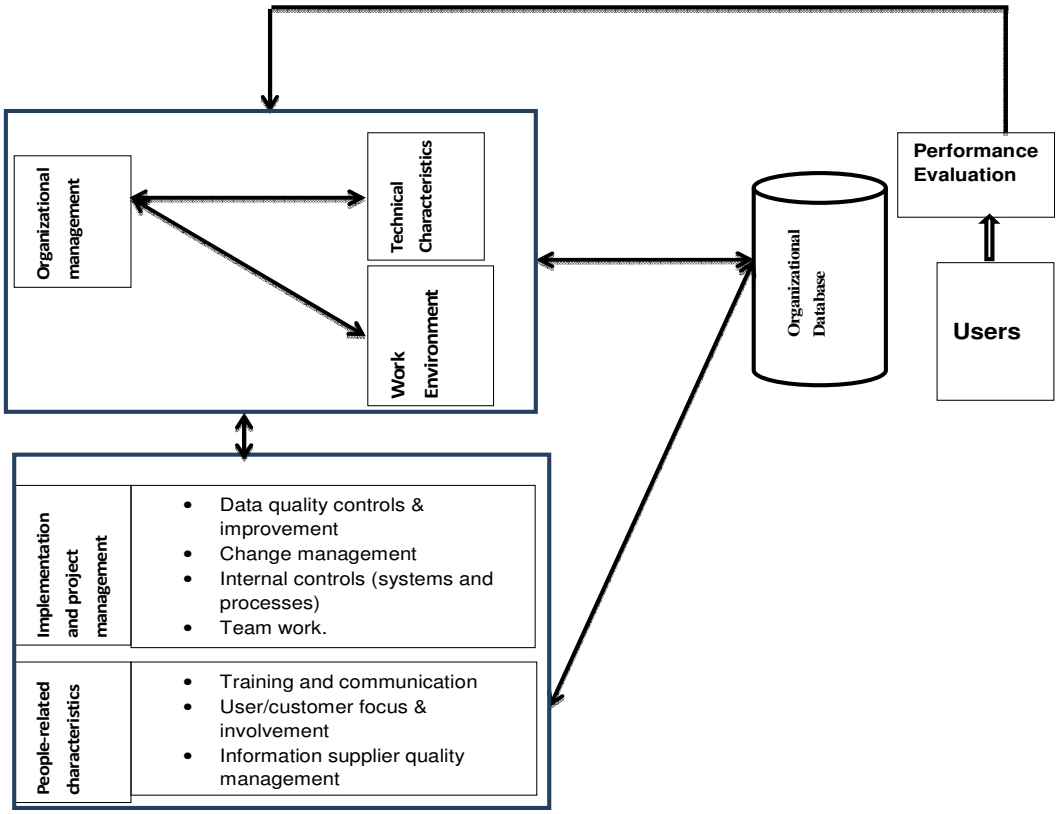

Figure 1: A Framework for Data Quality management in Nigerian Higher Institutions

4. DISCUSSION

The success of a data quality management system depends on the management of an organization. To obtain high quality work performance, an organisation must provide a good work environment for employees. A positive work environment makes employees feel good about coming to work, and this provides the motivation to sustain them throughout the day. Apart from the tools to do the work, a good work environment needs Lighting, adequate windows, air circulation and other elements that add to the quality of a workplace environment. Also, management has to keep abreast with the latest technology by providing the necessary hardware and software for data production. Also, the necessary controls are defined by the department established by management to handle all issues related to information production and management.

Most Nigerian Higher Institution have an Information and Communications Technology (ICT) section saddled with this task. During the implementation stage, management is required participate in the entire process. The areas that are considered very critical and important include integration/plan, scope, timeline, cost, quality, human resource management, communication, risk, and procurement. An organiz ation needs to focus on the stakeholders involved in the process of generating data upon which decisions are based. These include data suppliers, users and custodians. Empowering these category of stakeholders and taking into consideration their needs (training to improve skills and efficiency; incorporating their observations and complaints into the processes) has been known to improve the quality of output. The data produced is sent to the department responsible for preparing and organizing the data as an organizational resource (Database). The details of database management and definition of access levels is not discussed in this paper. An organization must develop formal systems to evaluate, track and monitor its data quality management schemes. This helps to send feedbacks from users to data producers on data performance and areas that need possible improvement.

5.0 CONCLUSION

The need for a generalized framework for Data quality management in Nigerian Higher Institutions has become necessary especially with the level of collaboration currently existing between most institutions. Using factors obtained from randomly selected stakeholders, a framework was proposed for data quality management. This framework was designed with some of the basic components considered by the authors for data quality management. It also provides evaluation and improvement components that can be used to interact with the other components to ensure data integrity and hence quality data success at all times. This research did not however study the details of database definition and management.

REFERENCES

Akinyokun, O.C. and Chiemeke S.C. (2006). IT Projects Performance Indices and Evaluation. Journal of Testing and Evaluation.34(6). ISSN: 0090-3973.

Akpon-Ebiyomare, D.E., Chiemeke, S.C. and Egbokhare, F.A. (2012). A Study of the Critical Success Factors Influencing Data Quality in Nigerian Higher Institutions.

Ballou, D.P., Chengalur-Smith, I.N., and Wang, R.Y. (2002).Sample-Based Quality Estimation of Query Results in Relational Database Environments. *IEEE Transactions on Knowledge & Data Engineering* 18(5) 639-650.

English, L (2009). Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. Wiley Computer Publishing, New York.

Fletcher et al. (2004). Achieving data quality: new data from information system earns trust of its users, Journal of AHIMA, 75(10), 22-26.

Olson, J. E. (2003). Data Quality.The Accuracy Dimension. San Francisco: Morgan Kaufmann Publishers.

Pipino, L., Lee, Y.W and Wang, R.Y. (2012). Data Quality Assessment. Communications of the ACM, 45(4), 211-218.

Redman, T. C. (2006). Data Quality in the Information Age: Artech House.

Tayi, G. K., and Ballou, D. P. (2008): Examining data quality'. Communications of the ACM, 41(2), 54-57.

Storey, V. C., and Wang, R. Y. (2004).Modeling quality requirements in conceptual database design. Paper presented at the International Conference on Information Quality, Massachusetts.

Strong, D.M, Lee Y.W, and Wang. R. Y. (1997). Data Quality in Context. Communications of the ACM 40(5), 103-110.

Strong, D., Lee Y., and Wang. R. (2004), The life of data quality projects. Working paper, MIT TDQM Research Program, E53-320, 50 Memorial Drive, Cambridge, 02139.

Strong, D., Lee Y., and Wang. R. (2004), The life of data quality projects. Working paper, MIT TDQM Research Program, E53-320, 50 Memorial Drive, Cambridge, 02139.

Wand, Y., and Wang, R. Y. (2006). Anchoring data quality dimensions in ontological foundations. Communications of the ACM, 39(11), 86-95.

Wang, R.Y., Ziad, M. and Lee Y.W. (2001). Extending the Relational Model to Capture Data Quality Attributes. IJIO 2(1), 19 – 35.

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

Wang, R. Y. (2009): Data Quality. Massachusetts. Kluwer Academic Publishers.

Computing, Information Systems & Development Informatics Journal

Call For Papers

The Computing, Information Systems and Development Informatics Journal (CISDI) provides a distinctive international perspective on theories, issues, frameworks and practice at the nexus of computing, information systems Developments Informatics and policy. A new wave of multidisciplinary research efforts is required to provide pragmatic solution to most of the problems the world faces today. With Computing and Information Technology (IT) providing the needed momentum to drive growth and development in different spheres of human endeavors, there is a need to create platforms through which breakthrough research and research findings that cuts across different discipline can be reported. Such dissemination to a global audience will in turn support future discoveries, sharpen the understanding of theoretical underpinnings and improve practices. The CISDI Journal publishes cutting edge research in computing, short communications/reviews and development informatics activities that appropriate design, localization, development, implementation and usage of information and communication technologies (ICTs) to achieve development goals. We also promote policy research that seeks to employ established (and proposed) legal and social frameworks to support the achievement of development goals through ICTs - particularly the millennium development goals.

The CISDI Journal is published four times in a year. Special issues are also published periodically from papers presented in conferences or other academic meetings. Technical reports are welcomed and published when available. Authors should submit manuscripts for consideration as e-mail attachment to the Managing Editor at info@cisdijournal.net or longeolumide@fulbrightmail.org. Submissions should not be longer than 5,000 words including abstract, keywords and references. **The CISDI Journal will publish research articles, short communications, empirical research, case studies, conference proceedings and reviews (including book reviews) in the following focus areas (and other allied themes):**

General Computing
* Hardware Technology
* Software Engineering
* Web Technologies
* Information Technology
* Information Systems
* Information Science
* Data & Information Management
* Information Security
* Business Computing
* Business Information Systems
* Computer Networks
* Artificial Intelligence
* Theory of Computation & Automata
* Software Metrics and Measurements
* Knowledge-based Systems
* Database Management
* Data Mining & Data Warehousing
* Knowledge Management e-Government portals
* Computer Forensic & Data Privacy
* E-Systems (Webocracy, e-Democracy, e-Learning, e-Commerce, e-Government & e-Health, e-Agriculture)
* citizen centric information systems
* Web-enabled knowledge management
* ICT enabled systems in the public and private service
* Internet Governance
* Information Systems Policy
* TeleHealth & Telemedicine & Telemarketing
* Design Structures & Annotations
* Computer Graphics & Games
* Multimedia & Mixed Media Systems
* E-Library and Virtual Library Systems
* Wireless Networks & Applications
* Economic Intelligence Systems
* Development Informatics
* Mobile Applications & Technologies
* Information Technology Policies
* Web Usage Ethics & Policies
* Enterprise Informatics and Policies
* Social Informatics, Social Media and Policies

Only electronic submissions are accepted. We welcome submissions on a rolling basis.

CISDI JOURNAL AUTHOR'S PUBLICATION TEMPLATE
Computing, Information Systens & Development Informatics
Paper Title (font Size 24)
Space (14 points font size)
Author's names (12 Points Font Size Bold)
Affiliations (10 Points Font Size)
E-mal Address and Phone Numbers (10 Points Font Size Italicize)

(Single column format for the above section please)

1. INTRODUCTION

All manuscripts must be A4 size paper, one inch top and buttom margin and 0.75 left and right margins. It should be written in English. These guidelines include complete descriptions of the fonts, spacing, and related information for producing your manuscripts for publication. Follow the style in this template for preparing your articles/papers/manuscripts for submission.

2. TYPE STYLE AND FONTS

Times New Roman 10 points is the acceptable font for typesetting the entire body of the manuscript. If this is available on your word processor, please use the font closest in appearance to Times. Pictures and graphics, preferably in Jpeg format or Bitmap format can be embedded as well as mathematical and scientific symbols. Equations should be numbered and acronyms/abbreviations should be explained /defined or written fully on first use.

3. CITATIONS IN THE BODY OF THE WORK

Referencing style in the body of the work should use square [7] braces. Works cited should be listed in the bibliography/referencing section at the end of the paper chronologically. Footnotes are not encouraged.

3.2 Main and Sub-Headings

Main Headings should be in UPPERCASE and proceeded by a point number. For Sub-Headings, use toggle case – or capitalize the first letter of each sentence/sub heading. Sub headings should be bold and italicized. Provide a space between each section and sub-sections.

3.3 Columns

Submission MUST follow the two column format depicted in this document. Equal width of 3.38cm and column spacing of 2.5cm. Lines are not allowed between columns. Where necessary or unavoidable, the two column format can be merged to a column to accommodate pictures, tables or graphics.

3.4 Paragraphing

The block structure is adopted for this journal.

4. TABLES AND FIGURES

Table headings/titles should be written above the tables. Figure headings should be written below the tables. These titles should proceed sequentially within the body of the text. Fig. 1 for the first figure, the next Fig 2 etc. Figures should not be numbered based on subsections of the manuscripts. Tables should also be labeled in the same sequence. Centralize Figures. Align Tables to the LEFT.

Table 1: How To Set Tables

Table Head	Table Column Head		
	Table column subhead	*Subhead*	*Subhead*
copy	More table copy[a]		

Source – (Adefolake, 2012)

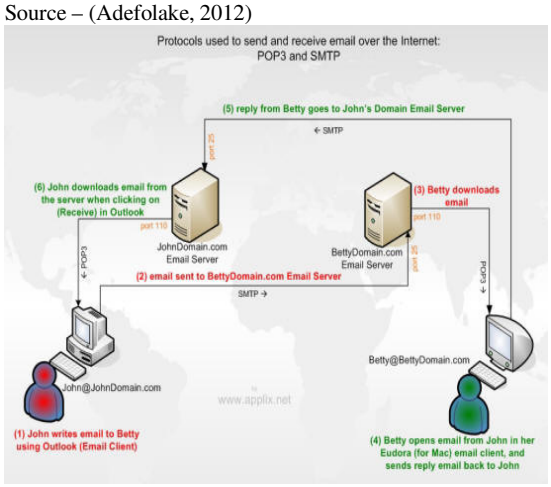

Fig.1. Example of a figure Caption

4.1 Units

The SI Unit is the acceptable Unit for specifying scientific measurements in this journal.

4.2 Equations

Authors are encouraged to use the equation editor facilities provided in MS Word to prepare equations for the manuscript. In very extreme situations, you may use either the Times New Roman or the Symbol font (please no other font) to prepare equations for publications. To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled. Equations should be numbered equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence. For example:

Align equations to the left.

Acknowledgement

A section on acknowledgement for sponsored research, collaborations, funds, grants/ other research material sources as well as individuals or organization that contributed to the success of the research.

REFERENCES

List and number all bibliographical references in 10 point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation as (Charles, 2009); for two authors (Laud & Laud, 2011); for more than three authors (Segun et al, 2010). Where appropriate, include the name(s) of editors of referenced books. Square braced referencing [3] is also allowed for authors who are more comfortable with that style of referencing.

Unless there are more than 3 authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should not be cited [4]. Papers that have been accepted for publication should be cited as "in press". Capitalize only the first word in a paper title, except for proper nouns and element symbols.

or papers published in translation journals, please give the English citation first, followed by the original foreign-language citation..

APA style referencing available online at http://www.apastyle.org/manual/index.aspx is also acceptable

Kindly visit the journal website at www.cisdijournal.net or contact the Managing Editor for additional information

[1] Eason, B. G. Noble, and Sneddon, K (2011) "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] Clerk , K and Maxwell, U (2010) A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Olumared, J., Kay, L., nd Bean, O. (2004) "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] Jauna, E. (2003). "Title of paper if known," unpublished.

[5] Electronic Publication: Digital Object Identifiers (DOIs):
Article in a journal:

Template Adapted from -
http://www.computer.org/portal/web/cscps/frmatting