

Development of Data Quality Dimensions from User's Perspective Framework

Mohd Zafrol Abdullah
Faculty of Computer Systems & Software Engineering
University Malaysia Pahang
Pahang, Malaysia
mohdzafrol@gmail.com

Ruzaini Abdullah Arshah
Faculty of Computer Systems & Software Engineering
University Malaysia Pahang
Pahang, Malaysia
ruzaini@ump.edu.my

Abstract— High quality data is an important asset in numerous business and organizations. The quality of data, i.e., the ability of data to meet user requirement can have a tremendous impact in an organization to develop an astounding data quality satisfaction subsequently provide a better platform to achieve top service in organizations. The assessment of data quality dimensions must consider the degree to which data satisfy users' needs. Therefore, it is important to develop a specific framework to assess data quality dimensions specifically to measure the degree of user's satisfaction and judgement of the data to obtain a correct interpretation of data quality assessment result. This paper proposes a conceptual framework of data quality assessment from user's perspective that draws the assessment specifically to measure user requirements and satisfactions. This framework can be evaluated and will be used to improve and extend knowledge of relationship between data quality dimensions and its assessment from user's perspective.

Keywords— Data Quality framework, User's requirement, Data Quality Dimensions, User's Perspective,

1. INTRODUCTION

Data is the most critical resources in various information systems as its quality gains competitive advantages. The effects of poor data quality negatively decrease the quality of information which is needed for decision makers in various level of user to make good decisions and for management purpose. Unfortunately, "more than 25% of critical data in the world's top company is flawed and almost 13.6% - 81% pieces of information needed for clinical decisions were missing all the time [25]. According to an earlier study, "the total cost of poor data quality" is between 8% to 12% of their revenues. Furthermore, 67% of managers think that the satisfaction of their customers suffers from poor data quality [26]. These figures impressively illustrate the relevance of data quality today. In short, poor data quality can negatively effect on economic, financial and management process and furthermore decrease the trust between users and the system.

To understand the challenges of data quality it is first required to define the term. Data define as a language, mathematical or other symbolic surrogates which are generally agreed upon to represent people, objects, events and concepts [1][2]. Therefore, we can describe data in its simplest form as a representation of objects or phenomena in the real world. Thus, when it comes to the discussion of quality data, we can say that poor quality data is a result of poor representation of the real world. According [3] information is an outcome of processed data. Therefore, in the context of information systems, these representations of real world are moderated by the needs of the system users., and hence the reference framework to evaluate the representation is the set of data user needs. In the early years, researchers began to study quality issues,

especially for the quality product, and a series of its definitions, for example, “Conformance to requirements” [4]; “fitness for use” [5] [6]. The most prevalent definition of data quality can be succinctly summed up as “fitness to use”, i.e., how well do the data serve the data consumer’s purposes [15]. The Total Data Quality Management group of MIT University led by Professor Richard Y. Wang has done in-depth research in the data quality area and proposed that “data quality judgement depends on data consumer” [7]. In general, two widely data quality definitions accepted are “fitness for use” and “conformance to requirements”. The fitness for use definitions preferred among product designers and marketing but it is difficult to measure since consumer expectations may change over time. Meanwhile, conformance to requirements definitions favored among producers and custodians since specifications can be defined and measured.

2. DATA QUALITY FROM USER’S PERSPECTIVE

Data quality assessment and its dimensions have been broadly discussed by many researchers. Current literature review appears to validate that data quality assessment although multi-dimensional but identical in most of the assessment structure. Many of the researchers provided and developed data quality framework specifically in their particular area. Researcher mostly focused on data quality dimensions associated to data values to assess the quality of data. Assessment and evaluating of data quality framework generally emphasis on data values, however researchers mostly do not consider the purposed use of each data. To assess which degree data that surpass quality requirements, it is essential to evaluate data along the process in which they are involved and under the user’s perspective by considering the expectations of the user who requests the data. For example, the data may be incomplete but it meets the requirement of the users or in some case the data was complete but did not meet user’s requirement. Therefore, the believability and trust of the user depends on the degree the data meet user’s requirement. Therefore, the believability and trust of the user depends on the degree the data meet user’s requirement. In general, user’s trust has been documented as one of the most critical aspects affecting the success of any online applications or systems and its environment [18].

Data Quality has a contextual and multi-dimensional concept. Therefore, data quality can be measured by different dimensions and structure. Dimensions selected and defined depends on the types of data and the focus of the research. Quality can be measure from the product itself or from the users of the product. In term of data quality assessment, quality of the data can be assessed from the data or from the users of the data. Researchers use different term and name to highlight on both perspective but the definitions and its classifications still refer to the product perspective and user’s perspective. Reference [1] proposed data quality dimensions from Declarative Perspective and Perceptual Perspective.

Declarative Perspective: Focuses on user independent characteristics of data which explains data itself like measures comparing the data with real world object and its representation as data or characteristics imposed by the operational aspects of organizations.

Perceptual Perspective: Focuses on user dependent characteristics of data such as effective usability of data for intended purpose and users’ judgement about the fitness for use.

Meanwhile, Kahn¹⁶ used difference term “conform to specifications and meets or exceeds consumer expectations” but the focus of the dimension still the same. The focus of the dimensions was based from the definition itself.

Conform to Specification:

Product Quality- The characteristics of the information supplied meet the standards.

Service Quality- The process of converting data into information meets the standards.

Meet or Exceeds Consumer Expectations:

Product Quality- The information supplied meets consumer task needs.

Service Quality- The process of converting data into information exceeds information consumer needs.

3. DEVELOPING CONCEPTUAL FRAMEWORK OF DATA QUALITY DIMENSION FROM USER’S PERSPECTIVE

This section discuss methodology used to develop data quality dimension from user’s perspective conceptual framework. The flow shows how the process used to develop the framework. data quality dimensions been classified and grouped from various field and domain into specific data quality dimensions for user’s perspective. Figure 1 illustrated the process of developing the proposed data quality dimensions’ framework.

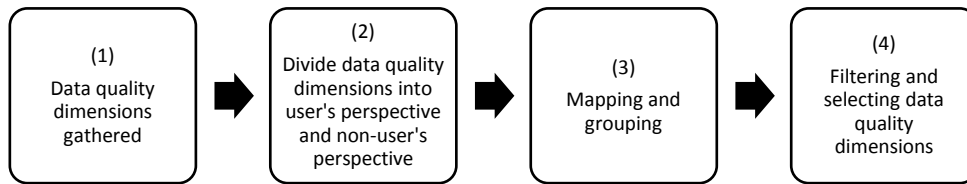


Fig.1. Process of Filtering and Mapping Data Quality Dimensions

The literature review begins with repeated searches on articles in the nature of data quality to gain insights about data quality dimensions from user's perception. The literature search was performed in English language articles focus on data quality, data quality assessment, data quality dimensions. The phrases 'information system', 'data quality attributes', 'information quality' and 'education management system' were also used in order to captures articles that may not have been search correctly. The literature review therefore created a basis for addressing research questions and research objectives of this paper. The result from the articles searched was extracted and presented in Table 1.

Table 1 Data Quality Dimension

Data Quality Dimensions gathered	
[19]	Accessibility, interpretability, usefulness, believability
[20]	Performance, features, reliability, conformance, durability, serviceability, aesthetics
[21]	responsiveness, courtesy, consistency, convenience, accessibility, accuracy, completeness, time and timeliness,
[22]	Believability, value added, timeliness, security, reputation, relevancy, objectivity, interpretability, accessibility, amount of information, completeness, concise representation, consistent representation, ease of manipulation, free-of-error, ease of understanding
[6]	Accuracy, completeness, consistency, timeliness
[23]	Accuracy, completeness, consistency, correctness, timeliness.
[13]	Correctness, completeness, flexibility, simplicity, implement ability, understandability, integration.
[14]	Accuracy, completeness, consistency, relevance, timeliness, usability, Interpretability, provenance, Priority, confidentiality, secure access
[15]	timeliness, definition/documentation, metadata, Accuracy, integrity, credibility, consistency, completeness, auditability, fitness, readability, authorization, structure, accessibility
[9]	Prerequisites of quality, Integrity, Methodological soundness, Accuracy and reliability, Serviceability, Accessibility
[7]	Objectivity, accuracy, believability, reputation, appropriate amount of data, timeliness, relevancy, value-added, completeness, representational consistency, ease of understanding, concise representation, interpretability, access security, accessibility
[12]	Relevancy, understandability, believability, interpretability, reputation, concise representation, value-added, price, documentation, reliability, security, customer support, timeliness, completeness, verifiability, objectivity, availability, amount of data, consistent representation, latency, accuracy, response time
[16]	Consistent representation, concise representation, completeness, free-of-error, security, timeliness, understandability, relevancy, interpretability, appropriate amount, objectivity, believability, accessibility, ease of manipulation, reputation, value-added
[1]	Performance, feature, reliability, conformance, durability, Completeness, accuracy, time and timeliness, Service ability, aesthetics, perceived quality, Courtesy, consistency, accessibility, convenience, responsiveness

Then these dimensions were divided into two categories which are data quality dimensions from user's perspective and non-user's perspective. This study focuses on data quality dimensions from user's perspective only therefore it is essential to eliminate data quality dimensions from others perspective. Outcomes and flow of the process showed in the Figure 2.

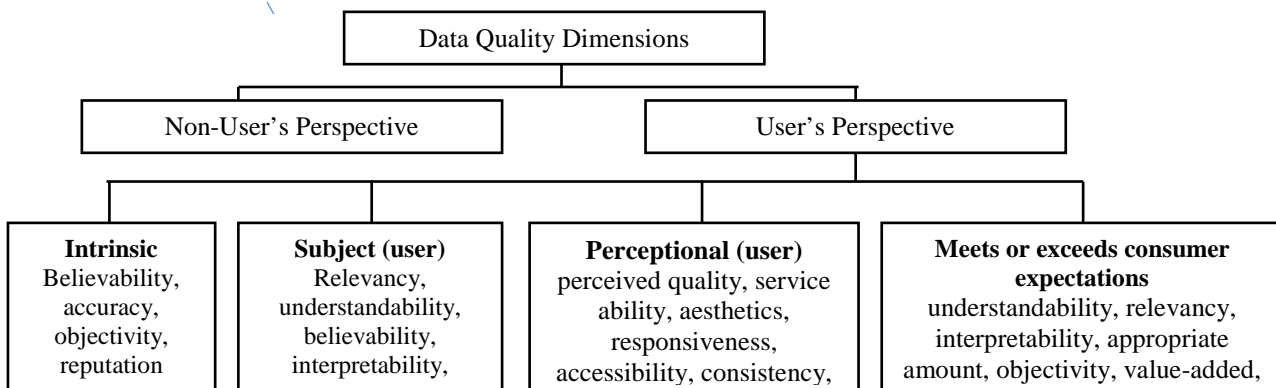


Figure 2 Divide Data Quality Dimensions

Data quality widely known as a multi-dimensional and it is important to select appropriate and specific dimensions to assess the quality. Although all these dimensions listed for user's perspective, it is important to narrow and choose the relevance dimensions and eliminate redundancies. Most of the researchers have settled and viewed in favour of reduction of data redundancies to ensure high quality and integrity of data before it can be processed in supporting decision-making [24]. Only dimensions selected by two or more researchers were considered. The filtering and selecting process shown in the Table 2.

Table 2. Filtering and Selecting

Filtering and Selecting					
Data Quality Dimensions	[1]	[16]	[7], [10]	[12]	Total selected
Believability		/	/	/	3
Accessibility	/	/			2
Appropriate amount		/			1
Reputation		/	/	/	3
Ease of manipulation		/			1
Value-added		/		/	2
Relevancy		/		/	2
Understandability		/		/	2
Interpretability		/		/	2
Objectivity		/	/		2
Courtesy	/				1
Consistency	/				1
Convenience	/				1
Responsiveness	/				1
Serviceability	/				1
aesthetics	/				1
Perceived quality	/				1
Concise representation				/	1
Accuracy			/		1

From the filtering and selecting process there are eight aspect of data quality dimensions clearly need to be used to assess from user's perspective. These dimensions **believability, accessibility, reputation, value-added, relevancy, understandability, interpretability** and **objectivity** considered the main dimensions of data quality assessment from user's perspective. This classification and selecting process fits into our framework and cover all aspect of user's perspective balancing comprehensiveness of dimensions with the nature of user in the systems. Figure 3 show data quality conceptual framework from user's perspective.

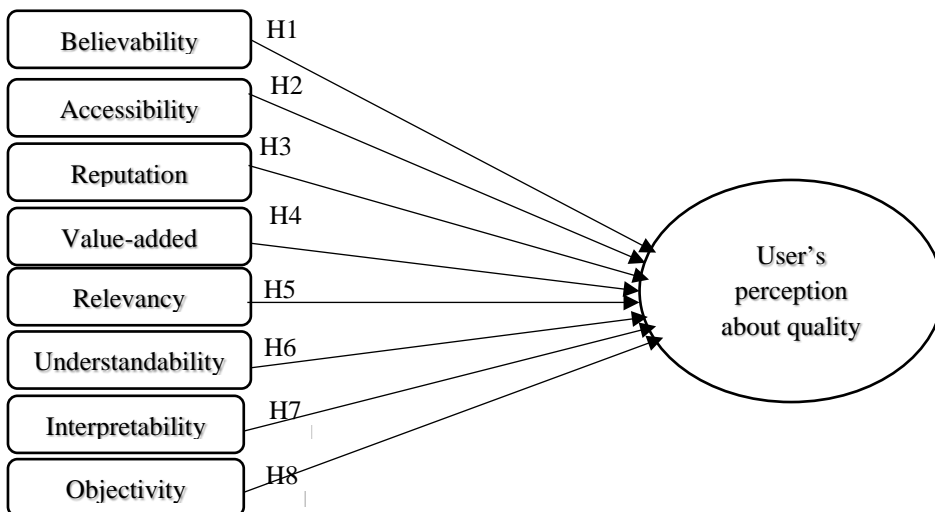


Figure 3 Data Quality Dimension Conceptual Framework

Figure 3 shows proposed a conceptual framework of data quality assessment and its dimensions from user's perspective. The dimensions along with its definitions was shown in Table 3.

Table 3. Data Quality Dimensions from User's Perspective.

Data Quality Dimensions from User's Perspective	
Dimensions	Definitions [5], [12]
Believability	Degree to which data is credible and true.
Accessibility	The degree to which the data is retrievable.
Reputation	Data contents or source are kept in high consideration
Value-added	Data provides a competitive advantage.
Relevancy	How usable, applicable or interesting the data is.
Understandability	Extent to which data are clear without ambiguity and easily comprehend.
Interpretability	The extent to which the data meaning is explained.
Objectivity	Extent to which information is unbiased, unprejudiced and impartial.

4. CONCLUSIONS

In this paper, we surveyed data quality assessment structure to develop broader understanding on data quality assessment from various domain. We summarized data quality dimensions with emphasized on its term and structure. We identified, grouped, filtered and select frequently cited data quality dimensions specifically focus by previous researcher on user's perspective.

A proposed conceptual framework of data quality assessment and its dimensions from user's perspective indicates that assessment from user's perspective, although intangible but manageable. Thus, it is vital to develop a specific data quality assessment framework and select the dimensions that focus on the user instead of the data or product. We believed that experimental research is needed to evaluate this model and find which dimensions of data quality are more related to user's perspective in order to improve data quality.

This study will provide a specific framework to evaluate or assess data quality from user's perspective. To improve data quality from user's perspective, it is essential to assess the quality of data from user's perspective instead of evaluating the quality of data itself. Since this conceptual framework consist of eight data quality dimensional that covers all the dimensions and specifically focus on user's perspective available from the literature, the result will be more accurate, details and comprehensive in showing the outcome of data quality assessment from user's perspective. We intend to validate and used this conceptual framework to assess data quality from user's perspective in Malaysian Educational Management Information System (EMIS). This study will contribute to content validity on how eight data quality dimensions (Believability, Accessibility, Reputation, Value-added, Relevancy, Understandability, Interpretability and Objectivity) will impact on data quality assessment specifically from user's perspective. Nevertheless, there are some limitations of this framework particularly on type of users and pervasive use of data in the systems. Although these factors are not directly considered in the framework, they will be measured in the research to evaluate and measure the impacts of these factor in the assessment of data quality from user's perspective.

ACKNOWLEDGMENT

REFERENCES

- [1] Jayawardene, V., Sadiq, S., & Indulska, M. (2013). An Analysis of Data Quality Dimensions. ITEE Technical Report No. 2013-01, 1, 1–32.
- [2] Morris, C. W., & Ders. (1971). Foundations of the Theory of Signs. In Writings on the General Theory of Signs (pp. 13–71). University of Chicago Press.
- [3] Wang, R. Y. (1998). Total Data Quality Ma. Communications of the ACM, 41(2), 58–65.
- [4] Wang, R. Y., Kon, H. B., & Madnick, S. E. (1993). Data Quality Requirements Analysis and Modeling. Proceedings of IEEE 9th International Conference on Data Engineering, (April), 670–677.
- [5] Laranjeiro, N., Soydemir, S. N., & Bernardino, J. (2016). A Survey on Data Quality: Classifying Poor Data. Proceedings - 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing, PRDC 2015, 179–188.
- [6] Tayi, G. K., & Ballou, D. P. (1998). Examining data quality. Communications of the ACM, 41(2), 54–57.
- [7] Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Source Journal of Management Information Systems, 12(4), 5–33.
- [8] Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A Review of Data Quality Assessment Methods for Public Health Information Systems. International Journal of Environmental Research and Public Health, 11(5), 5170–5207.
- [9] UNESCO-UIS. (2014). Assessing education data quality in the Southern African Development Community (SADC).

- [10] Sidi, F., Panah, P. H. S., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2012). Data quality: A survey of data quality dimensions. *Proceedings of the 2012 International Conference on Information Retrieval & Knowledge Management*, 300–304.
- [11] Panahy, P., & Sidi F. (2013). A Methodology to Explore Rules and Methods for Data Quality Dimensions toward Improvement the Quality of Databases.pdf. *Journal of Applied Sciences*.
- [12] Knight, S., & Burn, J. (2005). Developing a framework for assessing information quality on the World Wide Web. *Informing Science Journal*, 8(3), 159–172.
- [13] Moody, D. L., & Shanks, G. G. (2003). Improving the quality of data models: empirical validation of a quality management framework. *Information Systems*, 28(6), 619–650.
- [14] Almutiry, O., Wills, G., Alwabel, A., Crowder, R., & Walters, R. (2013). Toward a framework for data quality in cloud-based health information system. *Information Society (I-Society)*, 2013 International Conference on, 153–157.
- [15] Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, 14, 2, 1-10.
- [16] Kahn, B., Strong, D., & Wang, R. (2002). Information quality benchmarks: product and service performance. *Communications of the ACM*, 45(4), 184–192.
- [17] Slone, J. P. (2006). Information quality strategy: An empirical investigation of the relationship between information quality improvements and organizational outcomes, (October), 1–283.
- [18] Al-sharafi, M. A., & Arshah, R. A. (2016). The Impact of Customer Trust and Perception of Security and Privacy on the Acceptance of Online Banking Services: Structural Equation Modeling Approach, 171–177.
- [19] English, L. P. (1999). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits* (Computer Science). John Wiley & Sons, Inc.
- [20] Garvin, D. a. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(87603), 101–109.
- [21] Russell, R. S., & Russell, R. S. (2003). *Operations management / Roberta S. Russell, Bernard W. Taylor III*. 1. MANAJEMEN OPERASI, 2. MANAJEMEN PRODUKSI, *Operations Management / Roberta S. Russell, Bernard W. Taylor III*, 2003(2003), 1–99.
- [22] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211.
- [23] Liaw, S. T., Rahimi, A., Ray, P., Taggart, J., Dennis, S., de Lusignan, S., ... Talaei-Khoei, A. (2013). Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature. *International Journal of Medical Informatics*, 82(1), 10–24.
- [24] Mohammed Adam Taheir Mohammed, Wan Maseri Wan Mohd , Ruzaini Abdullah Arshah, M. Mungad, Edi Sutoyo and Haruna Chirum. (2016). Analysis of Parameterization Value Reduction of Soft Sets and its Algorithm. *International Journal of Software Engineering & Computer Systems (IJSECS)*, ISSN: 2289-8522, Volume 2, pp. 51-57.
- [25] Fan, W., & Wenfei. (2015). Data Quality. *ACM SIGMOD Record*, 44(3), 7–18.
- [26] Kaiser, M., Klier, M., & Heinrich, B. (2007). Association for Information Systems AIS Electronic Library (AISeL) How to Measure Data Quality? -A Metric-Based Approach How To Measure Data Quality? A Metric Based Approach, 12–31.