

# A Framework to Construct Data Quality Dimensions Relationships

Payam Hassany Shariat Panahy, Fatimah Sidi\*, Lilly Suriani Affendey, Marzanah A. Jabar, Hamidah Ibrahim and Aida Mustapha

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Selangor, 43400, Malaysia; payam\_shp49@yahoo.com

<sup>2</sup>fatimah@upm.edu.my, suriani@upm.edu.my, marzanah@upm.edu.my, hamidah@upm.edu.my, aida@upm.edu.my

## Abstract

Data and information obtained from data analysis is an essential asset to construct and support information systems. As data is a significant resource, the quality of data is critical to enhance data quality and increase the effectiveness of business processes. Relationships among all four major data quality dimensions for process improvement are often neglected. For this reason, this study proposes to construct a reliable framework to support process activities in information systems. This study focuses on four critical quality dimensions; accuracy, completeness, consistency, and timeliness. A qualitative approach was conducted using a questionnaire and the responses were assessed to measure reliability and validity of the survey. Factor analysis and Cronbach-alpha test were applied to interpret the results. The results show that the items of each data quality dimension and improvement process are reliable and valid. This framework can be used to evaluate data quality in an information system to improve the involved process.

**Keywords:** Data Quality Dimension, Framework, Relationship, Validation, Information Systems, Factor Analyzing.

## 1. Introduction

Data is the primary foundation in operational, tactical and strategic decision-making activities. As data is a critical resource in all applications within organizations, businesses and government agencies, quality of data is essential for managers and decision makers to resolve performance-related issues [12, 38, 5]. Data has multi-dimensional concept which can be measured by different dimensions such as currency, completeness, and accuracy [39, 16]. These dimensions are features to measure and manage data and information quality across different domain and their measurement metrics vary in different contexts [18]. Data quality dimensions are depending on each other and analyzing the dependency structure among them is an effective way to solve related problems in information systems. At present, beside of an existing

large variety of techniques for accessing and improving data quality such as using business rules, record linkage and similarity measures [5], there are some methodologies on data quality dimensions and improvement such as Total Information Quality Management (TIQM) and Total Data Quality Management (TDQM). These methodologies focus on finding the causes of none quality by measuring and improving activities on a set of quality dimensions (e.g., accuracy, timeliness, completeness) and analyze the dependencies among dimensions by using the properties of entropy. In fact, these methodologies measure and analyze the quality dimension indirectly by applying their techniques on the limited attributes on the database [21]. The data quality techniques have been previously used to find and evaluate the problem complexity but, proposing framework is the first step to develop methods for extracting subjective information from the database and reduce

\* Corresponding author:

Fatimah Sidi (fatimah@upm.edu.my)

data inconsistencies. Our research is more than measuring quality dimensions indeed it is dealing with proposing a data quality framework that can be used to find dependencies among dimensions. This framework is designed to create and manage a perfect database for real-world systems and domain as well as it supports database providers and users at every step from design, execution, analysis and improvement to assess and overcome database problems. It can yield multiple useful applications based on the real-world data toward system design and data production to identify data deficiency and inconsistency. However, there are different dependent variables for assessing information system in general and data quality in particular, but there is a little empirical or theoretical foundation for the choice [21]. This study is compared with previous study and demonstrated valid, reliable and final items for improvement process construct with more statistical analysis and results [35].

The goal of the present study is, to write a review on data quality dimensions, to propose a framework for data quality dimension in an effort to improve data processing in information systems and; to use qualitative approach by conducting a questionnaire and measures the reliability and validity of the survey scale. This study will survey on critical dimensions; accuracy, consistency, completeness, and timeliness, which are considered as a fundamental dimension and improvement process in information systems.

## 2. Literature Review

Data is described as a real world object that can be stored, retrieved, and elaborated during a software process and communicated through the network [9]. It is group of attributes and relations that refers to “raw material of information”. Data can be classified into different classes, for example implicit vs. explicit and elementary vs. aggregated. Classification of implicit and explicit data identifies data into three categories, which are structured, unstructured, and semi-structured [5]. Meanwhile, classification of elementary and aggregated data is based on the strictness to measure and to achieve data quality. Elementary data means data is managed by operational process and is represented by atomic phenomena in the real worlds such as gender and age. Aggregated data means elementary data is aggregated through application of some aggregation function, for example average income for taxpayers in a specific city [27]. Data quality refers to usage of the data while quality of producing data is judged

by the users in the system [39]. Common definitions of data quality that are adapted extensively include “quality of data to meet customer needs” or “appropriate for use” [36, 37]. Meanwhile, information quality is often defined as “fitness for all purposes in the enterprise that requires it” or “consistently meeting knowledge worker and end-customer expectation” [13]. Study of data quality provides an opportunity to better understand the nature of data. Researchers believe the fact that data quality is significant resources that have effects on the business success and overall efficiency of organizations [1].

To achieve high quality data, different strategies and techniques are employed, which can be divided into two types: data-driven and process-driven. Data-driven strategies improve the quality of data by directly modifying data values, whereas process-driven strategies improve the quality of data by redesigning the process of producing or modifying the data. Overall, process-driven has shown better performance in the long run as compared to data-driven strategies because it focuses on removing the cause of data quality problems. Data-driven strategies are also more expensive than process-driven strategies whether in the long run or within a short period [5].

In addition, the assessment also plays an essential role especially during data integration [41]. This is where data quality methodology is necessary to guide the architecture and implementation [23].

Based on the ISO standard, quality refers to “the totality of the characteristics of an entity that bear on its ability to satisfy standard implied needs” [17]. Data quality dimensions refer to “a set of data quality attributes that represent a single aspect or construct of data quality” [29]. These dimensions offer a way to measure and manage data and information quality [28].

According to Wand [42], there are three initial categories to identify appropriate data quality dimensions, which are data quality, information systems, and accounting and auditing. Meanwhile, Christy et al. [43] categories data quality differently; into four categories which are intrinsic DQ, accessibility DQ, contextual DQ, and representational DQ. Moreover, data quality has other dimensions such as objectivity, believability, reputation, and value added. Data quality dimensions are represented using taxonomy to create and manipulate the data for improving information and its process [40]. Process improvement refers to apply effective strategy for reducing costs, decreasing times cycle in order to improve quality that met user satisfaction [6]. Information systems work properly, if constant

improvement process be implemented. In fact, the concept of quality improvement provides significant assistance to address and solve related problem to information system.

The literature concludes that research on data quality has put significant efforts on data quality dimensions in order to improve data quality. Achieving high quality is not only depended on the accurateness of data but also it depends on several dimensions. At present, an information system is considered effective only when the data quality is high because making correct decisions and improvement in the process highly depends on the quality of the data [7]. Data quality dependency model among data dimensions can be divided into three major categories: perfect dependency, partial dependency, and independency [11]. Dependencies among the dimensions are essential for improving process quality [4] and understanding exist relations between dimensions is essential to promote effective and comprehensive knowledge discovery. Trade off analysis [15, 33, 26, 2, 4] and logical interdependence analysis [15] are examples of dependency analysis among different dimensions. Some quality dimensions have tradeoff relationships; which means one dimension can be affected by some deficiencies of other dimensions. The examples of this relationship are between completeness vs. consistency and timeliness vs. accuracy. The first trade off mentions; if the information becomes better over time, its accuracy will be decreased (negative effect). Another trade-off between completeness and consistency is that, if data is more complete, its consistency is lower or it has lack of consistency [44, 14, 11]. Overall, past researches have shown that very minimum work exists to study the structure of data quality dependency, hence there is needed to consider more dimensions to improve the quality of data.

According to the literature, product quality refers to data quality and service quality refers to information quality and is related to service delivery process [41]. Nowadays, several quality assessments and models exist, that most of them are utilized in the context of data warehousing and business domain with the aim of solving quality problems as a service quality [23]. These methodologies do not fit into the information system thus to improve process quality in the information systems we should look at data quality as a product quality not service quality. Therefore, an appropriate framework is needed to identify and reduce inconsistencies of the data by computing data quality dimensions dependencies for improving the involved process in the information systems.

This study develops and validates a questionnaire for improvement process, which is basically a process redesign

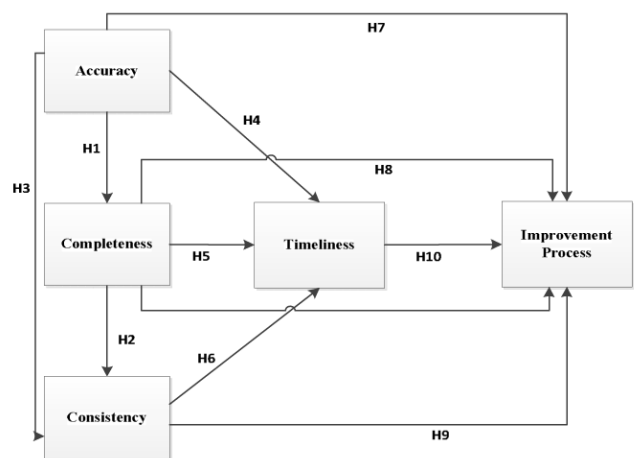
to identify causes of errors. According to Batini et al. [5], improvement process is part of improvement activities. In addition, we propose a framework with focusing on critical data quality dimensions and improvement process to find the relationship among them. The effectiveness of the improvement process is supported by reconstructing process as the primary step for improving process and making correct decision.

### 3. The Proposed Framework and Hypothesis

Barone et al. [4] points that while data quality dimensions can be strongly or weakly related to each other, effective dependency is able to improve data and process quality. Effective dependency is achieved by selecting appropriate dimensions and identifying the correlation among the dimensions. The process involves measuring the attributes for data quality dimensions and recovers the dependency structure among them to extract knowledge [20, 32]. While there is no clear consensus on the rightful number of data quality dimensions and their relationships, there are some fundamental dimensions that have to be considered [5]. These fundamental dimensions include the accuracy, consistency, completeness, and timeliness, which in turn effect on improving processes.

A framework is proposed as it is shown in Figure 1 to attempts to find dependency among data quality dimensions and their relationships with the improvement process in an information system.

Based on Figure 1, there are five variables: four fundamental data quality dimensions and improvement process.



**Figure 1.** A framework for data quality dimensions.

The relationships are represented by related hypothesis developed to determine dependencies among the independent variables and dependent variable including their dependencies with control variable, which is improvement process. The developed hypotheses are as follows:

- H1: Accuracy has a significant positive relationship with completeness.
- H2: Completeness has a significant positive relationship with consistency.
- H3: Accuracy has a significant positive relationship with consistency.
- H4: Accuracy has a significant positive relationship with timeliness.
- H5: Completeness has a significant positive relationship with timeliness.
- H6: Consistency has a significant positive relationship with timeliness.
- H7: Accuracy has a significant positive relationship with improvement process.
- H8: Completeness has a significant positive relationship with improvement process.

- H9: Consistency has a significant positive relationship with improvement process.
- H10: Timeliness has a significant positive relationship with improvement process.

Based on the above hypothesizes; we propose that the presented dimensions are related to improvement process. Table 1 adopted from Sidi et al. [34] compiles different definitions for four major quality dimensions that are critical to the improvement process.

Figure 2 shows the schematic diagram of the relationships between data quality dimensions and improvement process. The three main categories of this diagram are derived from the literature review. We consider three types of data quality dimensions: accuracy, consistency and completeness as independent variable and another dimensions; timeliness as dependent variable. Then, we consider the relationship of the timeliness as a moderator variable with improvement process as a control variable. Our schematic diagram concentrates on the relationship between data quality dimensions and improvement process in information systems. However, the main focuses of this

**Table 1.** Definition for critical data quality dimensions

#### Definition

##### Accuracy

Data are accurate when data values stored in the database correspond to real-world values [3, 5].

The extent to which data is correct, reliable and certified [39].

Accuracy is a measure of the proximity of a data value,  $v$ , to some other value,  $v'$ , that is considered correct [31, 5].

A measure of the correction of the data (which requires an authoritative source of reference to be identified and accessible [28]).

##### Completeness

The ability of an information system to represent every meaningful state of the represented real world system [3, 5].

The extent to which data are of sufficient breadth, depth and scope for the task at hand [39].

The degree to which values are present in a data collection [31, 5].

Percentage of the real-world information entered in the sources and/or the data warehouse [19, 5].

Ratio between the number of non-null values in a source and the size of the universal relation [30, 5].

All values that are supposed to be collected as per a collection theory [25, 5].

Information having all required parts of an entity's information present [8, 5].

##### Consistency

The extent to which data is presented in the same format and compatible with previous data [39].

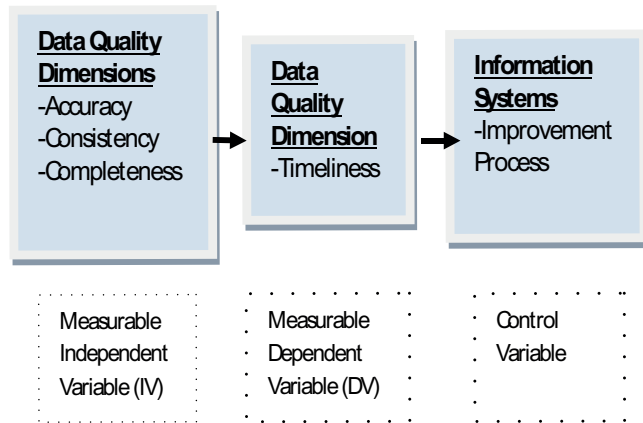
Refers to the violation of semantic rules as defined over the set of data [5].

##### Timeliness

The extent to which age of the data is appropriated for the task at hand [39].

The delay between a change of a real world state and the resulting modification of the information system state [42, 5].

Timeliness has two components: age and volatility. Age or currency is a measure of how old the information is, based on how long age it was recorded. Volatility is a measure of information instability the frequency of change of the value for an entity attribute [8, 5].



**Figure 2.** The research model relationship.

study is, applying statistical methods to assess the reliability and validity of the scales for data quality dimensions and constructing factor for the improvement process.

## 4. Research Design and Method

To illustrate the relationship of data quality dimensions and improvement process in the information system, we considered four major quality dimensions which are more efficient and effective than other dimensions. The proposed methodology to conduct this study is as follows:

- To construct a literature review on the relationship between data quality dimensions
- To construct a framework among four critical data quality dimensions
- To construct a draft questionnaire to evaluate the proposed framework
- To design and implement the survey
- To modify the questionnaire and to develop a questionnaire for the improvement process

This study provides a better understanding of data quality dimension relationships based on set of theoretical and statistical finding. Qualitative methods and evaluation of the degree of association or relationship between the dimension variables is carried out with the aid of a questionnaire. According to Creswell [10] and Lotfi [24] using a questionnaire to describe the opinions and characteristics of entire sample or population of people is a common and well known method of data collection in different research. In our study items for four data quality dimensions were adopted [21] and the items for measuring improvement process with regard to the information sys-

tem was self-developed from the literature. The assumption is that data quality dimension will improve process quality in information systems.

In this survey, simple random sampling method is applied to select respondents who were familiar with information systems in the period of 3 weeks. To ensure that respondents are aware to accurately respond to the questions, we focused on the academic groups. The nine-Likert scale is used to assess each item which one (1) represented as “Completely Disagree” and nine (9) as “Completely Agree”. From, 50 surveys just 37 surveys were completed and returned. The questionnaire, had 5 sections; A) Demographic information, B) Accuracy, C) Completeness, D) Consistency, E) Timeliness, F) Improvement Process (Appendix A). To perform statistics analysis SPSS ver. 20 was used to assess Cronbach- alpha and factor analysis for validity and reliability of the construct’s scales.

To prove that the survey scale is consistent and reliable, the Cronbach-alpha test was carried out to specify amounts of random errors that exist in the instruments. Cronbach-alpha determines reliability of the variables and its value is between 0 and 1 which 0 indicates a lack of reliability of the questions and value near to 1 indicates a high reliability. Based on the rule of thumb if Cronbach-alpha be more than 0.70 the instrument is reliable and acceptable. According to the Table 2, the reliability of coefficient for each individual variable was: 0.899 for accuracy, 0.897 for completeness, 0.915 for consistency, 0.794 for time-lines, and 0.810 for the improvement process. The total Cronbach-alpha is 0.944 that is beyond the minimum acceptable level of 0.70 in general and 0.60 for exploratory research [45]. Therefore, the overall Cronbach-alpha reliability coefficient scores indicate that the instrument is reliable and all items for each construct have relatively high consistency.

Content validity and construct validity are two major types of validity. Content validity is assessed by experts to

**Table 2.** The Cronbach-alpha for research variables

Variable	No of Items	Alpha
Accuracy	4	0.899
Completeness	6	0.897
Consistency	4	0.915
Timeliness	5	0.794
Improvement process	7	0.810
Overall Cronbach-alpha		0.944

extent to which questions or items within an instrument are relevant and representative of the construct. In our study, a pilot survey was sent to three experts to evaluate content validity for the items of improvement process for any clarifying and rewording. After some modifications and obtaining the feedback from experts and pre-testing the questionnaire was finalized. Therefore, we concluded the improvement process questions have content validity and can be used as an instrument during this study.

Construct validity was applied with conducting factor analysis extraction for measuring the validity of the instruments. Among different available methods for extracting factor, we applied “*Principal Components Analysis*” (PCA) and “*Principal Axis Factoring*” for extracting underlying factors from the variables. “*Principal Component Analysis*” is a method of data reduction or factor extracting based on correlation matrix of the variables involved. We applied this method to construct variable’s measurement with selecting “*Varimax Rotation*” technique that is a technique to simplify the interpretation of the factors. “*Varimax*” method is an orthogonal rotation method that minimizes the number of variables that have high loading on each other. Moreover, we selected scores coefficient greater than 0.5 to make factor matrix be more reliable with considering *Eigenvalue* greater than 1 and Kiaser-Meyer- Olkin (KMO) greater than 0.50 for measuring sample of adequacy. The reason is if *Eigenvalue* be less than 1, the solution cannot be calculated for the redundancy problems [22]. The value of the factor loading shows correlation between variables and a factor, thus the items with high loading on the same factor are closely associated with that factor.

The result of Principal Component Analysis is shown in Table 3, Table 4 and Appendix A. The details of this table are as follows:

- The KMO for items of accuracy dimension was 0.743 that indicates the sample size was adequate and there were sufficient items for each factor. All items of accuracy had factor loading greater than 0.7 which indicated all items can be loaded in the same factor. Also, the percentage of variance extracted for first factor was 77.834% with Eigenvalues 3.113 that showed all items can be matched with related variable and can be grouped on one factor.
- The KMO for items of completeness dimension was 0.838 that indicates the sample size was adequate and there were sufficient items for each factor. All items of accuracy had factor loading greater than 0.6 which

**Table 3.** Result of Principal Component Analysis for four critical data quality dimension

Component	Item	Factor loading	Eigenvalues	% of variance
Accuracy	AQ1	0.939	3.113	77.834
	AQ2	0.872		
	AQ3	0.914		
	AQ4	0.797		
Completeness	ComQ1	0.888	4.303	71.724
	ComQ2	0.835		
	ComQ3	0.844		
	ComQ4	0.790		
	ComQ5	0.822		
	ComQ6	0.875		
Consistency	Con Q1	0.935	3.186	79.642
	ConQ2	0.731		
	ConQ3	0.945		
	Con Q4	0.940		
Timeliness	TQ1	0.862	3.000	60.01
	TQ2	0.693		
	TQ3	0.646		
	TQ4	0.838		
	TQ5	0.812		

**Table 4.** Result of Principal Component Analysis for improvement process

Component	Item	Factor loading	Eigenvalues	% of variance
Improvement Process	IFQ3	0.511	2.998	59.593
	IFQ4	0.696		
	IFQ5	0.913		
	IFQ6	0.861		
	IFQ7	0.823		

indicated all items can be loaded in the same factor. Also, the percentage of variance extracted for first factor was 71.724% with Eigenvalues 4.303 that showed all items can be matched with related variable and can be grouped on one factor.

- The KMO for items of consistency dimension was 0.814 that indicates the sample size was adequate and there were sufficient items for each factor. All items of accuracy had factor loading greater than 0.7 which indicated all items can be loaded in the same factor.

Also, the percentage of variance extracted for first factor was 79.642% with Eigenvalues 3.186 that showed all items can be matched with related variable and can be grouped on one factor.

- The KMO for items of timeliness dimension was 0.766 that indicates the sample size was adequate and there were sufficient items for each factor. All items of accuracy had factor loading greater than 0.6 which indicated all items can be loaded in the same factor. Also, the percentage of variance extracted for first factor was 60.01% with *Eigenvalues* 3.000 that showed all items can be matched with related variable and can be grouped on one factor.
- The KMO for items of improvement process after two times repeating PCA and iteration was 0.609 which indicates the sample size was adequate. We repeated PCA for this variable two times since the output from iteration 1 and 2 compute 2 *Eigenvalues* greater than 1 which indicated 2 components can be extracted for this variable and initial factor solution was based on the extraction of 2 components. At the first iteration, item 2 (IFQ2) was removed since its communalities was 0.293 which is less than 0.5. At the second iteration, output indicated that item 1 (IFQ1) should be removed. Thus, after two iterations the percent of variances extracted were %59.953 with Eigenvalues 2.998.

Next, Principle Axis Factoring (PAF) was performed to examine and confirm the consistency of improvement process constructs to look at the underlying factor pattern. We repeated PAF for this variable two times since again the output from iteration 1 and 2 compute 2 Eigenvalues greater than 1. At first iteration, item 2 (IFQ2) did not load in any factors because of low loading thus it was removed and at the second iteration item 1 (IFQ1) was removed since 2 components extracted for this variable. Moreover, item 3 (IFQ1) had loading less than 0.5 which showed this item should be clarified and reworded for the next study. The result of Principal Axis Factoring is shown in Table 5.

Hence, this study provides the confirmatory evidence of scales validity of four data quality dimensions and improvement processes variable. Although, the analyzing of reliability and validity of these items was examined with more sample size in [35], we applied more statistical methods; Principle Axis Factoring and Principal Component Analysis techniques in this study to compute the number of fix items for improvement process construct, because this construct was self-developed and needed more analyzing.

**Table 5.** Result of Principal Axis Factoring for improvement process variable after second iteration

Component	Factor1
Improvement Process 3	-
Improvement Process 4	0.572
Improvement Process 5	0.945
Improvement Process 6	0.938
Improvement Process 7	0.740

The result proved validity and reliability of these items for this construct with mentioned methods. As a conclusion, the result of factor analyzing showed the items have high validity in construct and demonstrated good statistical properties to test the hypothesized research model in the future with identifying associated problems .

## 5. Conclusion

This study identifies relationship amongst data quality dimensions while providing primary empirical support to develop a framework for data quality dimensions and improvement process. Focusing on four significant quality dimensions: accuracy, consistency, timeliness, completeness can enhance process quality in information systems. A qualitative approach was applied to assess the reliability and validity of the scales by applying factor loading methods for the evaluation of the relationship between dimensions and improvement process. The objective of this study was to develop a reliable and valid instrument for the relationship between data quality dimensions and improvement process. The items for improvement process were identified, developed and validated through different statistical techniques on the data gathered from the survey. Content validity for improvement process, reliability and construct validity were applied to the all 5 constructs. The results of the Cronbach-alpha test showed all construct are reliable, as the value of KMO provided good evidence that the sample size was adequate and there were sufficient items for each factors. Moreover, the result of factor loading extraction demonstrated that all constructs are unidimensional. Thus, the overall results show that this framework can be adopted and used effectively in information systems within organizations and industrials to evaluate relationships among data quality dimensions, with the goal of enhancing data quality.

As more dimensions exist, considering another dimension can be effective on the improvement process. Future work would concern on validating the proposed framework

via using empirical methods and advanced statistical techniques. This is specifically to extend the analysis by analyzing the population via confirmatory factor analysis or Structural Estimation Modeling (SEM). Therefore, the ultimate goal will be to improve the proposed framework and to enhance the process involved in data quality in the information systems.

## 6. Acknowledgement

This research is supported by the Fundamental Research Grant Scheme (FRGS 03-12-10-999FR) from the Ministry of Higher Education, Malaysia.

## 7. References

- Al-Hakim L (2007). Information quality management: theory and applications, Igi Global. Idea Group Publishing, Hershey, USA, London, UK, 119–144.
- Alizamini F G, Pedram M M et al. (2010). Data quality improvement using fuzzy association rules, *Electronics and Information Engineering (ICEIE)*, International Conference On Electronics and Information Engineering (ICEIE), 2010, V1-468–V1-472. [doi:10.1109/ICEIE.2010.5559676]
- Ballou D P, and Pazer H L (1985). Modeling data and process quality in multi-input, multi-output information systems, *Management Science*, vol 31(2), 150–162.
- Barone D, Stella F et al. (2010). Dependency discovery in data quality, *Proceedings of the 22nd international conference on Advanced information systems engineering (CAiSE'10)*, Pernici B (Ed.), Springer-Verlag, Berlin, Heidelberg, 53–67.
- Batini C, Cappiello C et al. (2009). Methodologies for data quality assessment and improvement, *ACM Computing Surveys*, vol 41(3), 1–52, doi:[10.1145/1541880.1541883].
- Calvo-Manzano J A, Cuevas G et al. (2012). Methodology for process improvement through basic components and focusing on the resistance to change, *Journal of Software: Evolution and Process*, vol 24(5), 511–523, doi:10.1002/smr.505.
- Berner E S, Kasiraman R K et al. (2005). Data quality in the outpatient setting: impact on clinical decision support systems, *AMIA Annual Symposium Proceedings*, vol 2005, 41–45, American Medical Informatics Association, [PMCID: PMC1560426].
- Bovee M, Srivastava R P et al. (2003). A conceptual framework and belief-function approach to assessing overall information quality, *International Journal of Intelligent Systems*, vol 18(1), 51–74.
- Carey M J, Ceri S et al. (2006). *Data-Centric Systems and Applications*, Springer, Verlag Berlin Heidelberg, [doi: 10.1007/978-3-540-76452-6].
- Creswell J W (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*, Chapter 1, 2<sup>nd</sup> Edn., London: Sage Publications, Inc, 14.
- De Amicis F, Barone D et al. (2006). An analytical framework to analyze dependencies among data quality dimensions, *Proceedings of the 11th International Conference on Information Quality (ICIQ)*, Cambridge, MA, USA, 369–383.
- Eckerson W (2002). *Data Warehousing Special Report: Data quality and the bottom line*, Applications Development Trends.
- English L P (1999). Seven deadly misconceptions about information quality, *Information Impact International, Inc.*, Tennessee, Brentwood, 1–8.
- Fisher C, Eitel L E et al. (2012). *Introduction to information quality*, AuthorHouse, USA, 126.
- Gackowski Z J (2005). Informing systems in business environments: A purpose-focused view, *Informing Science Journal*, vol 8, 101–122.
- Heinrich B, Kaiser M et al. (2007). How to measure data quality? A metric-based approach, *Twenty Eighth International Conference on Information Systems*, Montreal, 101–122.
- Heravizadeh M, Mendling J et al. (2009). Dimensions of business processes quality (QoBP), *Business Process Management Workshops*, Springer, Berlin Heidelberg, vol 17, 80–91.
- Huang H, Stvilia B et al. (2012). Prioritization of data quality dimensions and skills requirements in Genome annotation work, *Journal of the American Society for Information Science and Technology*, vol 63(1), 195–207, doi:10.1002/asi.21652.
- Jarke M, Lenzerini M et al. (2003). *Fundamentals of Data Warehouses*, SIGMOD record, Springer-Verlag, vol 32(2), 55–56. [ISBN: 3-540-42089-4].
- Katerattanakul P, and Siau K (1999). Measuring information quality of web sites: development of an instrument, *Proceedings ICIS'99 of the 20th international conference on Information Systems*, Charlotte, North Carolina, United States, Association for Information Systems, Atlanta, GA, USA, 279–285.
- Lee Y W, Strong D M et al. (2002). AIMQ: a methodology for information quality assessment, *Information & Management*, vol 40(2), 133–146, doi:10.1016/S0378-7206(02)00043-5.
- Leech N L, Barrett K C et al. (2008). *Statics, spss For Intermediate*, 3rd Edn., New York, London: Lawrence Erlbaum Associates Inc. Publishers, New Jersey, 58.
- Li Y, and Osei-Bryson K-M (2010). Quality factory and quality notification service in data warehouse. *Proceedings of the 3rd workshop on Ph.D. students in information and knowledge management - PIKM '10*, New York, New York, USA: ACM Press, 25–32. doi:[10.1145/1871902.1871907].



24. Lotfi Z, Shahnorbanun S et al. (2013). A Product Quality-Supply Chain Integration Framework, *Journal of Applied Sciences*, vol 13(1), 36–48, doi:[10.3923/jas.2013.36.48].
25. Liu L, and Chi L N (2002). Evolutional data quality: A theory-specific view, *Proceedings of 7th International Conference on Information Quality (ICIQ 2002)*, Cambridge, Boston, MA, 292–304.
26. Madnick S E, Wang R Y et al. (2009). Overview and Framework for data and information quality research, *Journal of Data and Information Quality (JDIQ)*, vol 1(1), 1–22. [doi:10.1145/1515693.1516680].
27. Maguire H (2007). Book Review: *Data Quality: Concepts, Methodologies and Techniques*, Batini C, Scannapieco M, *International Journal of Information Quality*, Springer, vol 1(4), 444–450. [ISBN: 13 978-3-540-33172-8].
28. McGilvray D (2008). *Executing data quality projects: Ten steps to quality data and trusted information*. Morgan Kaufmann, Elsevier, Burlington, MA, USA. [ISBN: 978-0-12-374369-5].
29. Milano D, Scannapieco M et al. (2006). Design and Implementation of a Peer-to-Peer Data Quality Broker, *Interoperability of Enterprise Software and Applications*, Springer, London, 289–300, [DOI: 10.1007/1-84628-152-0\_26].
30. Naumann F (2002). Quality-driven query answering for integrated information systems, *Lecture Notes in Computer Science*, Springer, Verlag Berlin Heidelberg, vol 2261, [ISBN 978-3-540-43349-1].
31. Redman T C, and Blanton A (1997). *Data quality for the information age*, 1st Edn., ACM Digital Library, Artech House, Inc., Norwood, MA, USA, [ISBN:0890068836].
32. Sadeghi A, and Clayton R (2000). The quality vs . timeliness tradeoffs in the BLS ES-202 administrative statistics, *Federal Committee on Statistical Methodology*, 1–7.
33. Scannapieco M, Missier P et al. (2005). Data Quality at a Glance, *Datenbank-Spektrum*, Citeseer, vol 14, 6–14.
34. Sidi F, Shariat Panahy P H et al. (2012). Data quality: A survey of data quality dimensions, *2012 International Conference on Information Retrieval and Knowledge Management*, 300–304, doi:[10.1109/InfRKM.2012.6204995].
35. Shariat Panahy P H, Sidi F et al. (2012). Discovering Dependencies among Data Quality Dimensions: A Validation of Instrument, *Journal of Applied Sciences*, In press.
36. Strong D M, Lee Y W et al. (1997). 10 Potholes in the road to information quality, *Computer*, vol 30(8), 38–46 [doi:10.1109/2.607057].
37. Strong D M, Lee Y W et al. (1997). Data quality in context, *Communications of the ACM*, vol 40(5), 103–110, [doi:10.1145/253769.253804].
38. Tee S W, Bowen P L et al. (2007). Factors influencing organizations to improve data quality in their information systems, *Accounting and Finance*, vol 47(2), 335–355. doi:[10.1111/j.1467-629X.2006.00205.x].
39. Wang R Y, and Strong D M (1996). Beyond accuracy: What data quality means to data consumers, *Journal of management information systems*, vol 12(4), 5–33.
40. Wang K Q, Tong S R et al. (2008). Analysis of data quality and information quality problems in digital manufacturing. *2008 4th IEEE International Conference on Computing & Processing (Hardware/Software); Engineering Profession, Management of Innovation and Technology*, 2008, ICMIT, IEEE, Bangkok, Thailand, 439–443, [doi:10.1109/ICMIT.2008.4654405].
41. Wei-Liang C, Shi-Dong Z et al. (2009). Anchoring the Consistency Dimension of Data Quality Using Ontology in Data Integration. *2009 Sixth Web Information Systems and Applications Conference*, 201–205, doi:[10.1109/WISA.2009.32]
42. Wand Y W (1996). Anchoring data quality dimensions in ontological foundation, *Communication of the ACM*, vol 39(11), 86–95.
43. Christy S, Rajakumari S B et al. (2010). Quality data representation in web portal—A case study, *Trendz in Information Sciences & Computing (TISC)*, IEEE, 230–232.
44. Heinrich B, Kaiser M et al. (2007). How to measure data quality? – a metric based approach by. *Twenty Eighth International Conference on Information, Montreal, Canada. Systems*, vol 4801, 101–122.
45. Nunnally J C, and Bernstein I (1994). *Psychometric theory*, 3<sup>rd</sup> Edn., McGraw\_Hill Inc: New York.

## Appendix A

### Appendix A: Questionnaire Items with factor loadings.

Response ranges were measured from 1 “Completely Disagree” to 9 “Completely Agree”. Items labels with “(R)” were reverse coded. All factors, with exception of the “Improvement Process”, are adapted from Lee et al. [21].

### Accuracy

- AQ1: This information is correct. (0.939)
- AQ2: This information is incorrect. (R) (0.872)
- AQ3: This information is accurate. (0.914)
- AQ4: This information is reliable. (0.797)

### Completeness

- ComQ1: This information includes all necessary values. (0.888)
- ComQ2: This information is incomplete. (R) (0.858)
- ComQ3: This information is complete. (0.844)

ComQ4: This information is sufficiently complete for our needs. (0.790)

ComQ5: This information covers the need for our task. (0.822)

ComQ6: This information has sufficient breadth and depth for our task. (0.875)

### Consistency

ConQ1: This information is consistently presented in the same format. (0.935)

ConQ2: This information is not presented consistently. (0.731)

ConQ3: This information is presented consistently. (R) (0.945)

ConQ4: This information is represented in a consistent format. (0.940)

### Timeliness

TQ1: This information is sufficiently current for our work. (0.862)

TQ2: This information is not sufficiently timely. (R) (0.693)

TQ3: This information is not sufficiently current for our work. (R) (0.646)

TQ4: This information is sufficiently timely. (0.838)

TQ5: This information is sufficiently up-to-date for our work. (0.812)

### Improvement Process

IF1: The overall aims and targets of the information system is improvement. (item was removed after second iteration)

IF2: The process and people involved in the process should have opportunities for improvement. (item was removed after first iteration)

IF3: Quality improvement in the process is obvious. (0.511)

IF4: This information or data should be modified or redesigned for quality improvement. (0.695)

IF5: This information improves our organization's process. (0.913)

IF6: This information will provide value to our organization. (0.861)

IF7: This information or data has resulted in overall quality improvement for consultations. (0.823)