ISO 8000-61 DATA QUALITY MANAGEMENT STANDARD, TDQM COMPLIANCE, IQ PRINCIPLES

(Completed Research Paper)

Abdul Ghani Mohammed, Aziz Eram, and John R. Talburt,
University of Arkansas at Little Rock
ggmohammed@ualr.edu, axeram@ualr.edu, jrtalburt@ualr.edu,

ABSTRACT

The adverse effect of poor quality data is a widely-recognized problem in large and small companies, ranging from finance and health care to manufacturing and processing. Poor data quality can increase operational costs and risk and decrease customer satisfaction. Many approaches to data quality management have been proposed to help organizations address this problem. Recently the International Organization for Standardization (ISO) developed and published ISO 8000 Part 61, Data Quality Management: Process Reference Model. The purpose of this paper is to give an overview the standard and how it incorporates generally accepted information quality principles. The paper also provides an analysis of how the ISO 8000 standard incorporates the Total Data Quality Management (TDQM) methodology developed by the MIT Chief Data Officer and Information Quality Program.

Key Words: Data Quality, Information Quality, Data Quality Management, ISO 8000, Total Data Quality Management, Data Governance, Total Quality Management

INTRODUCTION

Organizations worldwide are beginning to understand and treat their data as an important resource and asset. While it is true that most organizations now recognize the importance of data, they have yet to fully realize the true potential of data [9]. Making real-time decision of the basis of data requires having the right amount of data at the right time and the right place for it to support the decision [14]. However, many organizations lack this capability.

Data can be classified based on structure as well as function [10]. Almost every activity carried out on a day-to-day basis will have some data associated to it. Very often the executives are concerned that with data being spread across the organization in large volume how can they possibly attempt to monitor, control, interpret, analysis and use it for decision making [12] [15]. The ISO 8000 – 61 [11] standard is an international standard developed by ISO Technical Committee 184 on Automation Systems and Integration, Sub-Committee 4 on Industrial Data, Working Group 13 on Data Quality. The standard defines the characteristics of information and data that determine its quality, and provides methods to manage, measure and improve the quality of information and data. In essence, it is put forward by ISO as an international "best practices" model for data quality management.

ISO 8000-61 DATA QUALITY MANAGEMENT

A data quality structure in an organization's setting comprises the processes, departments, resources that perform their activities with the sole aim of improving and maintaining data quality. It is also commonly referred as Data Quality Management.

The scope of the ISO 8000-61 standard includes

- The fundamental principles of Data Quality Management
- The structure of a data quality management process within an organization setting.
- Definitions of the lower level processes required for data quality management
- The relationship between data quality management and data governance
- The requirements to implement Part 61

The ISO 8000-61 standard recognizes data quality management is founded on three principles

- Data quality is achieved through processes – these are processes that create, update, and use the data, processes that must be well-defined, repeatable, and reliable.
- Continuous improvement – data quality management requires goals and measurements and analysis to keep improving the processes and the quality of the data they produce
- People – Everyone in the organization has some data quality management responsibility, some more directly than others. Two important stakeholders are the user of the data who determine data quality goals and top management who must not only oversee the process and set the vision, but must also provide the resources that enable other to accomplish their tasks.

Part 61 prescribes a basic structure for data quality management in an organization has three major components. These are the Implementation Component, Data-Related Support Component, and the Resource Provision Component as show in Figure 1 [11].

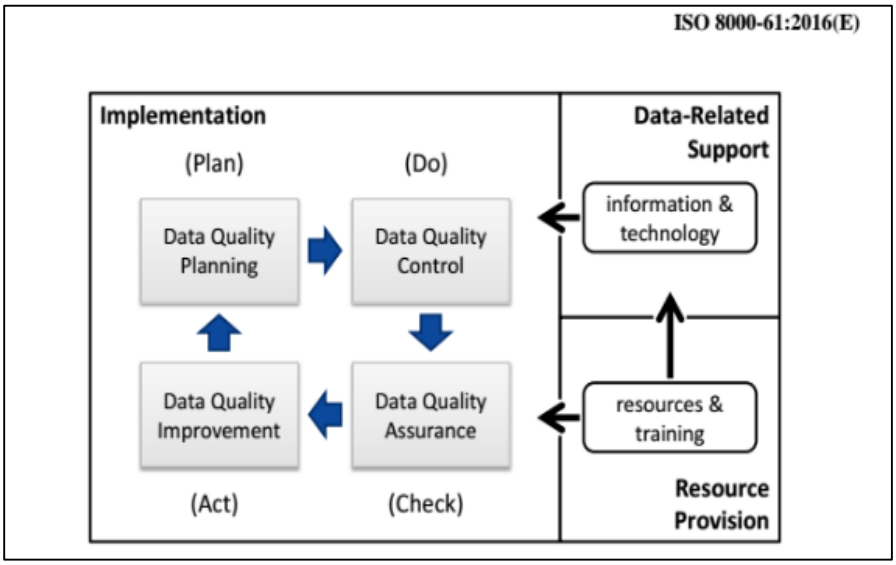

Figure 1: Data Quality Structure- ISO 61 [11]

Implementation Component

At any given point of time if the executives decide to implement something in their setting they need to have a plan on how the change is going to be carried out effectively up to that point where it becomes an integral part and becomes a way of work. This is based on the Plan-Do-Check-Act cycle often referred to as the PDCA cycle [5] [7] originally proposed by Walter A. Shewart, recognized as the father of statistical quality control. The cycle later adopted and made popular as the Plan-Do-Study-Act or PDSA cycle by William Edwards Deming. The entire implementation of data quality management is based on repeating this cycle which in turn leads to continuous improvement of data quality.

The 4 phases in the implementation of Data Quality

- **Data Quality Planning**: Corresponds to the Plan phase of PDCA. This step is very crucial in determining the scope, requirements, strategy, policies, standards, procedures and the steps

involved in implementing these within the organization [13]. It is essential that we take these into consideration and improve them as we proceed further in the cycle.

- **Data Quality Control**: Corresponds to the Do phase of PDCA. Quality control often refers to improvement of the products quality [11] [18]. In this scenario, the product is an information product. The different ways to ensure that an information product is of good quality will be addressed in the current phase. It includes provision of well-documented data specifications and work instructions. The actual data processing, and the monitoring and control of these processes. These activities contribute to ensuring the information product being produced is of high quality.
- **Data Quality Assurance**: Corresponds to the Check (or Study) phase of PDCA. Quality assurance often refers to the improvement of the products quality by improving the processes that revolve around the product [11] [18]. It deals with process improvement. This phase includes a review of data quality issues, provisioning the measurement criteria, the actual measurement of data quality and process performance, and evaluating the meaning of measurement results.
- **Data Quality Improvement**: Corresponds to the Act phase of PDCA. This step focuses on the improvement of processes and data quality. Two important activity of this step are root cause analysis to prevent future errors, and data cleansing to repair existing errors. Root cause analysis should lead to solution development the implementation of data processes improvement to prevent reduce the amount non-conforming data [11].

Taken together and repeated, these steps comprise a continuous data quality improvement cycle.

Data Related Support Component

This is present to make sure that the implementation phase is being carried out effective and smoothly. It is there to provide support for the data architecture, data transfer, data operations and data security. It involves the technology acquisition and continuous support that is required for the implementation of data quality in an organizational setting [11].

Resource Provision Component

The purpose of this is to provide and control the organizational resources required for the performance of Implementation and Data-Related Support. It involves data quality organization management including implementing and managing organizational unit, committees, and managing the data, documents, and other artifacts they produce. It also involves provisioning these units with suitably qualified and trained personnel [11].

COMPLIANCE WITH ISO 8000-61

The normative (compliance) is Section 10 of the ISO 8000-61:2016(E) and states “The organization shall prepare documentary evidence of the implementation of data quality management in accordance with this part of ISO 8000. This evidence shall identify the activities performed and the outcomes achieved, as per clauses 6 to 8.” Note that Clauses 6, 7, and 8 are the detailed specifications of the three components of Implementation, Data-Related Support, and Resource Provision, respectively.

DATA QUALITY MANAGEMENT AND DATA GOVERNANCE:

DQM contributes to the processes, roles, standards and metrics of data governance [19].

- DQ Strategy Management: By establishing goals and objectives
- DQ Policy, Standards and Procedures Management: By doing the same.
- DQ Implementation Planning: Implementation plans and their evaluation

- Data Quality Organization Management: Roles and responsibilities
- Human Resource Management: Developing relevant knowledge and skills of personnel.

All these criteria's serve as an input for the scope of data governance. When you have such a quality focused setup in place it is essential on how you are going to assess the implementation being carried out in your organization [8]. The questions that arise are: How are you going to assess the effect? There are still organizations that are freshly new to the concept of data quality. This field of study is fairly new to a lot of professionals. We all know when there is a change in the corporate sector there are people who believe in the change and respond to it in a positive manner and then there are some who respond to it in a negative manner. As Thomas Redman puts it "The hard issues are often soft" but it can be managed [15]. This standard also provides ways on how we can monitor and keep track on how we are moving forward in the positive direction. [10]. What if 5 months from now or 5 years from now an organization has a question on how they have grown or improved in their data's quality? How are their employees handling the data quality initiative? Do they believe in the change?

To implement the data quality management process model in an organizational setting it is essential that the organizations prepare a documentary evidence for each step of the implementation of each of the process model.

This evidence would contain the following facts:

- Data specifications, results of DQ measurements, log of non-conformities and log of RCA and actions taken are evidence how DQM is carried out [3] [6].
- Extent to which data quality is improved.
- Assigned roles and responsibilities for DQM across organization [17].
- Resources used in performing DQM.
- How data quality management being applied anywhere in the organization is creating or using the data.

Have a log and documentation of all these facts would help you keep track of your growth.

INFORMATION QUALITY

Information quality is all about maximizing the value and minimizing the risk of an organization's information assets and ensuring the information products produced by them will create value for, and meet the expectations of the customer who use them [24]. There are seven (7) fundamental principles of Information Quality that ensure that it is being implemented and practiced in an acceptable manner [25].

Principle 1: Realization that Information Quality and Governance is a business function [9] [21]. It is an on-going enterprise program, not a one-time project, a team in the IT department, or even a department in IT.

- Needs a dedicated Chief Data Officer the same as other business functions. For example, a Chief Information Officer for IT systems, a Chief Financial Officer for accounting, and Chief Operating Officer to oversee business operations.

Principle 2: Information only produces value when it is used for some purpose, i.e. applied.

- Data when is processed in a form that is needed by an application is known as Information Product (IP). The product that is developed should be in alignment with the business need and need to be of good quality [1] [21].
- People often mistake RESULTS with VALUE but that is not the case. They are often short sighted and fail to see the long-term benefits that will arise through their investment [20] [21].

Principle 3: Information Quality is contextual in nature and not absolute [21] [16]. The same information can have different levels of quality when it is applied for different purposes.

Principle 4: Information needs to be managed at the product in the Information System and not as a byproduct [13] [21].

Principle 5: For Information Quality to be successfully implemented and practiced in an organization IQ management must be supported at the highest levels of the organization [1]. People should be aware how their data can be used as an organizational asset. Ownership needs to be assigned to data stewards who would take responsibility for the data and make sure that the data is of good quality when being applied [21].

Principle 6: (The B-I-T Principle) Business needs should drive the information needs which should drive the technology need in the organization [16] [21].

Principle 7: Information has many dimensions or characteristics that would help us determine if it is of good quality or not. A dimension assessment is important for us to realize the rate of quality of our data. There are many frameworks that can be used for this assessment one of which is the Wang- Strong framework of Dimensions [3] [4] [21].

ISO 8000 - 61	Information Quality Principles						
DQM Structure	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Data Quality Planning	**X**	**X**	**X**	**X**	**X**	**X**	
Data Quality Control		**X**	**X**	**X**	**X**		**X**
Data Quality Assurance			**X**		**X**		**X**
Data Quality Improvement		**X**	**X**	**X**	**X**	**X**	**X**
Data Related Support (IT)						**X**	
Resources and Training	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Table 1: Integration of the Seven Information Quality Principles into ISO 8000-61

The intersection matrix shown in Table 1 lists the major components of the Part 61 DQM Structure as the rows of the matrix, and each column represents one of the Seven Principles of IQ. The letter "X" in a cell indicates the principle represented by the column is implemented by the Part 61 structural component in the row.

For example, the first principle talks about the realization that Information Quality and Governance is a business function. This realization takes place in the initial stages of the establishment of a Data Quality Management Structure. It is realized and accepted by the executives in the organization and in the Data Quality Planning phase they make amends for a continuous and sustainable data quality management structure that is in alignment with the business needs of the organization. Similarly, all the other principles that are integration play a unique role in a certain phase or several phases through the course of implementation.

TOTAL DATA QUALITY MANAGEMENT

The methodology of Total Data Quality Management (TDQM) was proposed by Richard Wang [1] as a way to adapt the principles, guidelines, and practices of Total Quality Management (TQM) in manufacturing to the production of information products. Table 2 [1] shows the analogy between the production of a manufactured product by a factory and the creation of an information product by an information system.

	Product Manufacturing	Information Manufacturing
Input	Raw Materials	Raw Data
Process	Assembly Line	Information System
Output	Physical Products	Information Products

Table 2: Analogy of Product Manufacturing with Information Manufacturing [1]

In TQM quality control is done by a physical examination of the product [6] [2]. But that sort of a quality check is not possible for an information product. Even though data and information are not kinesthetic, they nevertheless have intrinsic characteristics or dimensions that can be measured.

Many data quality dimensional frameworks have been proposed by both practitioners and academicians. One of the most well-known frameworks is the Wang-Strong Framework of 16 dimensions of data quality developed from research done by Richard Wang and Diane Strong [1] [4] [21]. Measurable dimensions of data quality are key to the implementation the TDQM data quality management methodology. The first step in TDQM is to define measurable quality goals. The objective of TDQM is to continuously improve information product quality through repeated measurements and analysis. TDQM also recognizes that information products have a life cycle similar to the life cycle of a manufactured products. Information products are introduced, improved, and evolve. However, over time an information product will be become less relevant, and at some point, it will be retired.

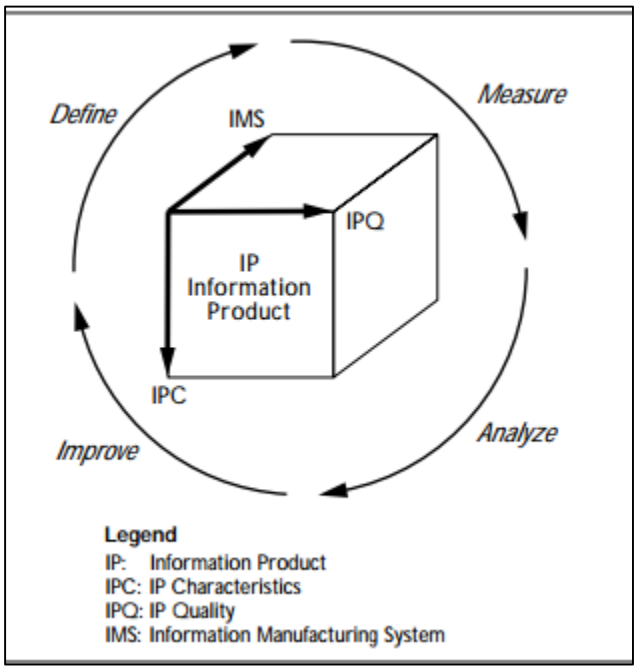

Figure 2: TDQM Methodology [1]

The schematic of the phases in TDQM is shown in Figure 2 [1] [21]. The phases are:

Define: It is identified in three steps, and they are:

> ***Step 1:*** It determines the data product characteristics. The data product characteristics can be described using two levels. The highest level describes the characteristics of the total data product and the lowest level describes each product attribute individually. [1]. Because not all data products

can be assessed in one-time effort it is useful to focus on the most important data product or attribute at a time.

Step 2: It determines the requirements for the data products. The quality dimensions for data products can be determined by answering these questions: How important each team member thinks the dimension is important, the perceived level of quality in dimension and the expected level of quality in a dimension.

Step 3: It determined the data manufacturing process. The data manufacturing process consists of data flows from the supplier to the user of the data, including certain processing activities and quality checks.

Measure: The measurement phase determines the quality of the dimensions identified in the define phase. This involves two steps.

Step 1: In this step, we select proper metrics. In determining a metric for a data quality dimension, the team should keep in mind underlying business rules, norms and laws that might have contributed to a dimension's importance.

Step 2: It measures and presents data. Once the metrics are determined, measurement can be conducted. Several types of charts can be used to display the results.

Analysis: The goal of this phase is to find the root causes for the problems in different dimensions. There are three methods to find root causes: the cause and effect diagram, the interrelationship diagram and the current reality tree [23].

Improvement: This can be identified in four different steps, and they are:

Step 1: To generate solutions. This will help us to identify possible redundant processes and redundant data sources.

Step 2: To select solutions. It is necessary to know what impacts the solutions might have, the cost to design and implement the solution, the resources required and the reduction of risk.

Step 3: To develop an action plan. The team should assign team members or other employees to these actions to make sure they are executed.

Step 4: To check progress

When data are passed through these phases, they are transformed into an information product. The TDQM phases are based on the Plan-Do-Study-Act (PDSA) phases of manufacturing TQM developed advocated by quality pioneers Walter Shewart and Edwards Deming [2] [22] [23]. The PDSA phases are:

Plan: In this phase, we list our action steps along with person(s) responsible and time line. For example: Recruit team, draft an aim statement, describe current context and process, identify the problem and, identify causes and alternatives

Do: In this phase, we start to implement our action plan. To try out the test on a small scale and based on that the team should be able to document problems, unexpected efforts and general observations.

Study: In this phase, we set aside time to analyze the data and study the results. We determine whether the plan result in an improvement or not, was the action worth the investment, do you see trends and were there unintended side effects.

Act: In this phase we refine the change, based on what was learned from the test. Describe what modifications to the plan will be made for the next cycle from what you learned.

In applying the TDQM methodology, an organization must clearly articulate the Information Product (IP) in business terms, establish an IP team consisting of senior executives, teach data quality assessment and management skills, and institutionalize continuous IP improvement.

Define IP: To define IP characteristics and IQ requirements. This phase produces two key results: a quality entity-relationship model that defines the IP and its IQ requirements, and an information manufacturing system that describes how the IP is produced.

Measure IP: To measure the development of IQ metrics. These can be basic IQ measures such as data accuracy, timeliness, completeness and consistency.

Analyze IP: The measurements results are investigated by IP team to find root cause problems for current IQ problems. The methods and tools for performing this task can be simple or complex.

Improve IP: In this phase, the IP team identifies key areas for improvement such as aligning information flow and work flow with the corresponding information system and realigning the key characteristics of the IP with business needs.

CONCLUSION

In this paper, we have shed light on how the MIT TDQM methodology's iterative cycle which is Define, Measure, Analyze, Improve (DMAI cycle) uses the same fundamentals that were proposed in the Shewhart-Deming PDCA cycle that underpins manufacturing TQM. The Data Quality Implementation component of ISO 8000 – 61 is an implementation of the PDCA cycle. It is explained how the ISO 8000 standard ingests TDQM within its implementation phase explaining the criticality of TDQM for the standard. It is also made clear in this paper how the standard has an integration of the principles of information quality within them which is quite essential for the functions sustainability.

REFERENCES

1. Wang, R. Y. (1998, February). A Product Perspective on Total Data Quality Management. Retrieved from Communications of the ACM
2. Dahlgaard, J. J., & Mi Dahlgaard-Park, S. (2006). Lean production, six sigma quality, TQM and company culture. *The TQM magazine*, *18*(3), 263-281.
3. Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: product and service performance. *Communications of the ACM*, *45*(4), 184-192.
4. Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & management*, *40*(2), 133-146.
5. Johnson, C. N. (2002). The benefits fo PDCA. *Quality Progress*, *35*(5), 120.
6. Yusof, S. R. M., & Aspinwall, E. (2000). Total quality management implementation frameworks: comparison and review. *Total quality management*, *11*(3), 281-294.
7. Gupta, P. (2006). Beyond PDCA-a new process management model. *Quality progress*, *39*(7), 45.
8. Mosley, M., Brackett, M. H., Earley, S., & Henderson, D. (2010). *DAMA guide to the data management body of knowledge*. Technics Publications.
9. Birol, C. A. (2014). Data quality management: what are the requirements and are companies ready to embrace them? (Doctoral dissertation).
10. Kim, S., & Lee, C. (2013). The process reference model for the data quality management process assessment. *Journal of Society for e-Business Studies*, *18*(4).
11. ISO/TS 8000-61, Data quality — Part 61: Process Reference Model.
12. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, *12*(4), 5-33.
13. Redman, T. C., & Blanton, A. (1997). *Data quality for the information age*. Artech House, Inc..
14. Berman, F. (2008). Got data? A guide to data preservation in the information age. *Communications of the ACM*, *51*(12), 50-56.
15. Redman, T., & Grieves, J. (1999). Managing strategic change through TQM: learning from failure. *New Technology, Work and Employment*, *14*(1), 45-61.

16. Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ)*, *1*(1), 4.
17. Rosenbaum, S. (2010). Data governance and stewardship: designing data stewardship entities and advancing data access. *Health services research*, *45*(5p2), 1442-1455.
18. Wagner, G. (1995). Basic approaches and methods for quality assurance and quality control in sample collection and storage for environmental monitoring. *Science of the total environment*, *176*(1), 63-71.
19. Wende, K. (2007). A model for data governance-Organising accountabilities for data quality management.
20. 18. Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. British journal of psychology, 55(4), 429-432
21. Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2009). *Journey to data quality*. The MIT Press.
22. Samson, D., & Terziovski, M. (1999). The relationship between total quality management practices and operational performance. *Journal of operations management*, *17*(4), 393-409.
23. Wijnhoven,, F., Boelens,, R., Middel, R., & Louissen, K. (2007). TOTAL DATA QUALITY MANAGEMENT: A STUDY OF BRIDGING RIGOR AND RELEVANCE. Retrieved May 31, 2017, from http://doc.utwente.nl/76558/1/Wijnhoven07total.pdf
24. Talburt, J. & Zhou, Y. (2015) *Entity information life cycle for Big Data: Master data management and information integration.* Morgan Kaufmann
25. Talburt, J.R. (2011) *Entity resolution and information quality*. Morgan Kaufmann