

2011

A Framework to Enhance Decision Outcomes: Data Quality Perspective

Muhammad Al-Abdullah

Virginia Commonwealth University, mm.alabdullah@gmail.com

Heinz Roland Weistroffer

Virginia Commonwealth University, hrweistr@vcu.edu

Follow this and additional works at: <http://aisel.aisnet.org/sais2011>

Recommended Citation

Al-Abdullah, Muhammad and Weistroffer, Heinz Roland, "A Framework to Enhance Decision Outcomes: Data Quality Perspective" (2011). *SAIS 2011 Proceedings*. 15.

<http://aisel.aisnet.org/sais2011/15>

This material is brought to you by the Southern (SAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in SAIS 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A FRAMEWORK TO ENHANCE DECISION OUTCOMES: DATA QUALITY PERSPECTIVE

Muhammad Al-Abdullah
Virginia Commonwealth University
mm.alabdullah@gmail.com

Heinz Roland Weistroffer
Virginia Commonwealth University
Roland@Weistroffer.net

ABSTRACT

The relation between data quality and decision outcomes has been studied extensively in information systems research. In this paper, based on the published literature, a framework is developed that shows the key factors that affect decision outcomes and data quality management in organizations. These factors are differentiated according to their specific impacts: data quality information and process metadata enhance the awareness in the decision maker about the data used in decision-making; TDQM (total data quality management), management experience, and commitment are important factors for overseeing the data; and visualization mitigates information overload for the decision maker.

Keywords

Data quality, decision-making, decision outcomes

INTRODUCTION

The ever-increasing amounts of data and the importance of good data and information for business competitiveness has made data quality a critical issue for information systems research and practice. Poor data quality can result in huge losses for businesses. Information systems research has responded by developing techniques to control the quality of data.

Creating information is similar to manufacturing products: raw material (data) is processed into a finished product (information). It is generally understood that the quality of any processing system (including decision-making) is strongly impacted by the quality of its inputs (i.e. "garbage in, garbage out"). Thus, it is well accepted that poor data lead to poor decisions.

There has been considerable published research on how to enhance decision outcomes using data quality information. However, most studies have looked only at specific aspects of this problem, and the greater, holistic picture has not clearly emerged. In this paper we develop a framework describing the important factors that enhance decision outcomes and help manage organizational data effectively. These factors were collected from the published literature and categorized according to their area of impact.

Though usually information is defined as data in context, or as "useful" data, in this paper we do not strictly differentiate between data and information, as what constitutes information to one person may be merely data to someone else.

The paper will start with a literature review in the next section that explicates the effect of poor data on an organization, describes some of the causes of the problem, and also looks at some proposed solutions to the problem. In the following section then we introduce our framework. We conclude with a summary of the contributions of this paper as well as future research directions.

CONSEQUENCES OF POOR DATA QUALITY

Being aware of a problem is the first step in solving it, and for that reason, information quality problems and their effect on an enterprise have been studied extensively by information systems researchers. This section will start with a look at the consequences of poor data quality on an enterprise in order to illustrate the importance of the problem to be researched.

A major study on this topic was published by Redman (1998). The author divides the impacts of poor data quality into impacts on the operations, impacts at the tactical level, and strategic impacts. At the operational level, poor data leads to customer dissatisfaction, increased cost, and lower employee job satisfaction. Customers expect no errors from the organization, and simple mistakes with customer orders, information, and statements will lead to dissatisfaction and harms the credibility of the business. Errors in data also increase the operational cost of the organization, as data cleansing, and error checking and fixing consumes employees' time and efforts, resulting in increased costs. Finally, poor data affect employees'

job satisfaction: for example, a front desk customer service employee will be frustrated and show reduced productivity when dealing with erroneous customer data.

At the tactical level, poor data will wreck the integrity of the decision-making process, especially when decisions are based on large quantities of data. Poor data also creates inter-departmental mistrust in an organization as each department may blame the other for the errors in the data that is used by both. At the strategic level, poor data leads to flawed strategy development and implementation, with long lasting consequences. Inaccuracies and errors in any one project mean that any modifications to the project and assessments of the project will also be flawed.

When discussing data quality and the consequences of poor quality data, it is important to define what data quality means. In information systems research there are generally two aspects of data quality: the (objective) characteristics of the data itself, and the (subjective) user perceptions of the data. Objectively, data quality is defined as meeting specific requirements and specifications (e.g. accuracy) (Kahn and Strong, 1998). Subjectively, data quality is determined by the fit of the data to the users' needs (Wang and Strong, 1996). A more detailed list of data quality definitions is provided by Ge (2009).

US businesses reportedly are losing almost \$611 billion a year due to poor data quality (Shankaranarayanan et al., 2006). As knowing the reason for a problem is generally a first step towards solving it, we will now look at possible causes of poor quality data, as reported in the literature. We will also discuss some proposed solutions, focusing on decision outcomes as impacted by data quality.

From a technical perspective, data warehousing is a basic element for DS. According to Inmon (1996), a data warehouse is “a collection of Integrated, Subject-Oriented, Non Volatile and Time Variant databases where each unit of data is specific to some period of time. Data Warehouses can contain detailed data, lightly summarized data and highly summarized data, all formatted for analysis and decision support.” Singh et al. (2010) produced a state-of-the-art classification in order to organize the reasons for data deficiencies and non-availability or reach-ability problems in data warehousing and to formulate a descriptive classification of these root problems. The authors summarize the possible causes of poor data quality in data warehouses, and they decompose the problems according the various stages in building a data warehouse. The dimensions used in their analyses are: completeness, consistency, validity, conformity, accuracy, and integrity. The paper is helpful for data warehouse developers to avoid many of the causes of poor data quality.

Another approach in information system research for dealing with data quality issues and their impact on decision-making is to enhance the original data with data quality information. Such metadata would make decision makers aware of the possible flaws in the data they use in decision-making, and perhaps make them use the data more appropriately and cautiously, thereby mitigating the impact on the decision outcomes. Chengalur-Smith et al. (1999) studied the effect of adding information about the quality of the data on decision outcomes, looking at the types of information to be added, the decision strategies that may benefit from such information, and the kind of environment in which such information would be helpful. They found that data quality metadata is helpful in the case of simple decisions but may be harmful in the case of complex decisions. They also found that the metadata is helpful when the decision maker needs to select a best alternative, but not when comparing large numbers of alternatives. Shanks (2001) looked at the impact of adding data quality tags (Shanks, 2001). Two decision strategies were compared: additive and elimination by attributes. The results showed that data quality tagging was beneficial for the elimination by attributes strategy but not for the additive strategy.

The information to be added in the aforementioned papers was intrinsic information (tags), which is information about the data to be used. Another trend of research on this matter is looking at the effect of adding process metadata on the decision outcomes. Shankaranarayanan et al. (2006) suggest that adding process metadata, i.e. information about the processes, stores, and deliverables that the data will go through, should also be helpful. Furthermore, Ziad et al. (2003) proposed a framework of managing the data for the decision making process. With the technologies available for accessing and developing data, and considering the dynamic environments that require the decision maker to act/react in a timely manner, it may be helpful to provide managers with tags about the data (data-metadata) as well as the mediated processes (process metadata). Shankaranarayanan and Cai (2006) illustrate the use of the IPview software to create an IPMAP. An IPMAP is similar to a data flow diagram; however, it captures the sequence of manufacturing steps that creates a product. In this case, the product is an information product, but the steps are similar: there is input, processing, and output. The authors use IPMAP diagrams and propose a DS framework that provides quality information to the decision maker. They refer to quality as completeness, claiming that completeness is a very important dimension of data quality that outweighs other aspects.

Another factor in information systems research concerning the effect of data quality on decision outcomes is the importance of the context in data quality information. Strong et al. (1997) studied the contextual data quality patterns and concerns. They defined contextual patterns with regard to how well the data matches the task context. Contextual patterns include relevancy, value-added, timeliness, completeness, amount of data, and the representation of the data.

Watson (2007) added to this by looking at the believability of the data. He suggested that when people do not believe in the data they have, they do not use it. And if the data is poor, the decision maker will not believe in it and thus will not use it. But even high quality data may not be sufficiently believable. Hence, the author tried to determine the factors that enhance the believability in the data and found those to include: source of data, timelines, the processes that manufactured this data, and how well the data matches user notions.

The final dimension related to data quality effect on decision outcomes, as found in the literature, deals with the decision maker's expertise, involvement, and knowledge about the decision task.

In order to integrate objective and contextual dimensions of data quality, Wang (1998) proposed a theoretical model based on the *heuristic systematic model* (HSM). The same study also found that the characteristics of the decision maker play an important role with respect to the decision outcomes. Similarly, Raghunathan (1999) investigated the impact of information quality and decision maker quality on decision outcomes. He found that information quality only enhances the decision outcomes if the manager has good knowledge about the decision task; otherwise, the decision outcome will likely be weak. Shankaranarayanan and Even (2004) also found that the information quality and the data and process quality metadata will not be sufficient if the manager's quality is poor. They concluded that data quality metadata might harm the decision outcomes and add complexity to the decision task if the decision maker is not sufficiently knowledgeable about the decision task. Such metadata is overhead information and the decision maker may trade off cognitive effort and sacrifice decision accuracy.

Also, there seems to be consensus in the literature that to reduce the effect of data overload, data should be visualized. One way to do that is to use the analytical decision making model. This model is a well-structured, computerized and mathematical model. It represents the different alternatives and the possible decision outcomes (Zhu et al., 2007).

Furthermore, management commitment and the existence of a quality champion strongly influence the data quality of an organization (Watts et al., 2009). Tee et al. (2007) propose applying *total data quality management* (TDQM). TDQM includes four main iterative phases, viz. define, measure, analyze, and improve.

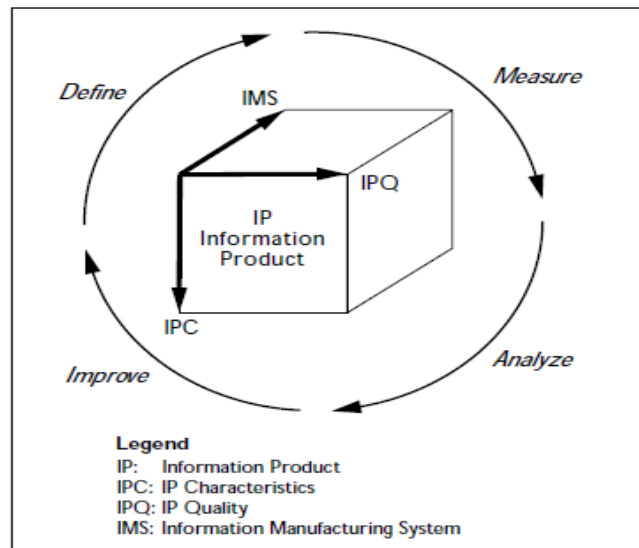


Figure1. TDQM methodology (Tee et al., 2007)

Looking at Figure 1, the *define* phase consists of:

1. Defining the functionalities of the information product (IP) desired by the consumer.
2. Defining the IP units, components, and relations between the components.
3. Defining the information manufacturing system.

During the *measure* phase, the organization needs to develop the metrics that will be used to measure the information quality. Then comes the *analyze* phase in which the IP team will investigate the major causes of the information quality problems. Finally, in the *improve* phase the IP team defines the key areas of improvement. Multiple iterations of these phases may occur during which data quality is measured and improved, leading to a sustainable solution.

THE DATA QUALITY FRAMEWORK

According to the previous discussion, this paper proposed a framework, as shown in figure 2, which organizes the factors that have been proved to enhance the decision outcome. In this framework, the factors are grouped according to their purpose.

To enhance the quality of the data of an organization, some managerial responsibilities have been proposed. As mentioned before TDQM is an important approach that can be deployed to manage the data.

There is a similarity between product manufacturing and information manufacturing. Both are processing systems that develop products or information from raw materials. Thus, information research proposed using corresponding models to manage the quality of information products. Logically it makes sense to apply the TQM procedures on data to solve the quality problems of the data repositories.

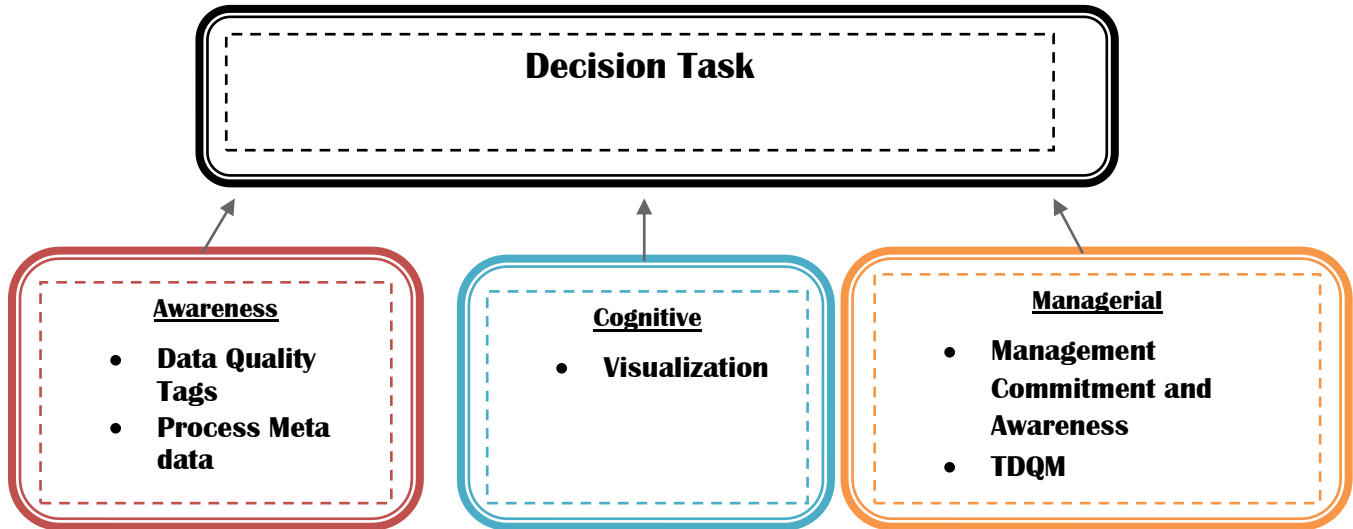


Figure 2. Data quality framework

TDQM methodology has been shown to be successful when management shows commitment. Policies can be developed when top management presents awareness and interest in the data as an important asset. Furthermore, management should look at data as a strategic asset that may provide a competitive advantage. Thus, management commitment and awareness is listed as an important factor to enhance data quality.

Data tags and process metadata, if attached with the actual data, are proved to be important factors for better decisions. When they are provided to the decision maker, they will create awareness about the data used in the decision making process. This awareness illustrates the amount of uncertainty of the decision task. It may also be presented in a way to perform some sensitivity analyses and to develop alternative strategies for the organization.

Process metadata can also be a data management step, as decision makers see the data stores, processes, and their quality. Any defects in the data may be tracked and dealt with. Defects of data are not necessarily derived from entry or stored data. When data is processed, it may be also become defective. Thus, processes may also be reengineered for the sake of better data quality.

Data quality is not the only factor for better decisions. At times a decision maker may be provided with the metadata without this resulting in a better decision. Information systems research suggested two reasons for this: first, the way this information is presented to the decision maker is important. It has been proved that if metadata information is visualized, a decision maker will utilize them more. However, if they are provided in a text form, for example, they will add overload to the decision maker and hence, they are likely to be ignored. Second, if the information is provided even in a visualized form, they may still result in information overload and be ignored, if the decision maker's experience and knowledge about the decision task is low. In other words, visualization and decision maker experience and knowledge are completing each other.

Even if the factors are categorized for different reasons they still complement each other. For example, even if visualizing the metadata information is listed as a cognitive solution and the manager's experience is listed as a managerial process to manage the data, they still complete each other and require each other.

As mentioned earlier in this paper, proper context of the data in reference to the decision task is important for quality decision outcomes. In our proposed framework, we assume that proper context of data is given. When having a data quality team dealing with the data and managing it using the TDQM methodology, the team will define where and why the data will be used. So, when the data or information is presented to the decision maker, the data quality team will consider what is needed and how it should be presented.

CONCLUSION

The problems of data quality management and decision support have been widely studied, and based on the published literature, it is generally accepted that better and more accurate decisions require better quality data. Poor quality data has been shown to lead to financial losses and loss in organizational trust by consumers.

This paper introduced a framework that relates data quality aspects to decision support, based on the published literature. This framework should be useful to decision makers in assessing the potential outcomes of their decisions and ultimately lead to better decision support and better decision-making. The review of the literature also shows that providing decision makers with information (metadata) about the data may improve decision outcomes.

REFERENCES

1. Chengalur-Smith, I., Ballou, D., and Pazer, H. (1999) The impact of data quality information on decision making: an exploratory analysis, *IEEE Transactions on Knowledge and Data Engineering*, 11, 6, 853-864
2. Ge, M. (2009) *Information quality assessment and effects on inventory decision-making*. PhD thesis, Dublin City University
3. Inmon, W.H. (1996), *Building the Data Warehouse*, 2nd edition, Wiley, New York.
4. Kahn, B.K. and Strong, D.M. (1998) Product and service performance model for information quality: an update, *Proceedings of the 4th International Conference on Information Quality*, Boston, Massachusetts, USA
5. Raghunathan, S. (1998) Impact of information quality and decision maker quality on decision quality: a theoretical model and simulation analysis, *Decision Support Systems* 26,4, 275-286
6. Redman, T.C. (1998) The impact of poor data quality in the typical enterprise, *Communications of the ACM*, 41, 2, 79-82
7. Shankaranarayanan, G., and Cai, Y. (2006) Supporting data quality management in decision-making, *Decision Support Systems*, 42, 1, 302-317
8. Shankaranarayanan, G., and Even, A. (2004) Managing metadata in data warehouses: pitfalls and possibilities, *Communications of the Association of Information Systems (CAIS)*, 14, 13, 1-49
9. Shankaranarayanan, G., Even, A., and Watts, S. (2006) The role of process metadata and data quality perceptions in decision making: an empirical framework and investigation, *Journal of Information Technology Management*, 17, 1, 50-67
10. Shanks, G. (2001) The impact of data quality tagging on decision outcomes, *Proceedings of 12th Australasian Conference on Information Systems*, December 4-7, Coffs Harbour, Australia
11. Singh, R., Punjab, P., and Patiala, P. (2010) A descriptive classification of causes of data quality problems in data warehousing, *International Journal of Computer Science Issues*, 7, 3(2), 41-50
12. Strong, D.M., Lee, Y.W., and Wang, R.Y. (1997), Data quality in context, *Communications of the ACM*, 40, 5, 103-110
13. Tee S.W., Bowen, P.L, Doyle, P., and Rohde, F.H (2007) Factors influencing organizations to improve data quality in their information systems, *Accounting and Finance*, 47, 2, 335-355
14. Wang, R.Y. and Strong, D.M. (1996) Beyond accuracy: what data quality means to data consumers, *Journal of Management Information Systems*, 12, 4, 5-34
15. Wang, R.Y. (1998) A product perspective on total data quality management, *Communications of the ACM*, 41, 2, 58-65
16. Watson, H. J. (2007) The believability of data, *Business Intelligence Journal*, 12, 4, 4-8
17. Watts, S., Shankaranarayanan, G., Even, A. (2009), "Data Quality Assessment in Context: A Cognitive Perspective", *Decision Support Systems*, 48, 1, 202-211
18. Zhu, B., Shankar, G., and Cai, Y. (2007) Integrating data quality data into decision-making process: an information visualization approach, *Proceedings to the Human Computer Interaction International Conference*, Beijing, July 8, 366-369
19. Ziad, M., Shankaranarayanan, G., and Wang, R. (2003) Managing data quality in dynamic decision-making environments: an information product approach, *Journal of Database Management*, 14, 4, 14-32