

A DOMAIN-SPECIFIC APPROACH TO EVALUATE INFORMATION QUALITY

*Mysore Ramaswamy, College of Business, Southern University, Baton Rouge, LA 70813,
mysore@acm.org*

ABSTRACT

Ease of access to information and its sheer abundance are some of the characteristics of this information age. As a direct consequence, assuring the quality of information used in managerial decision-making has become a critical issue. A clear understanding of the quality dimensions along with which information has to be measured is necessary to improve the usefulness of information systems that are used in different levels of management. In this article, we present a domain-specific view of information quality that could lead to a better comprehension of the dimensions of data-information-decision chain.

INTRODUCTION

In a recent study by the Automotive Aftermarket Industry Association (AAIA) Electronic Commerce Committee, it is estimated that data errors cost \$1.7 billion annually for suppliers and distributors. The Internet has provided information access to a degree unimagined earlier. This abundance of information also makes it imperative that we set some quality standards for information that is used in making decisions as well as data that are used in creating or processing information.

Current literature on data and information quality considers a lot of dimensions or attributes to measure the quality of input that goes into the decision-making process. In this article we take a domain-specific view of data-information-decision in order to conceptualize the quality dimensions of its components. The rest of the paper is organized as follows. The next section introduces data quality concepts. Information quality is discussed in section 3. This is followed by an analysis of the domain specific nature of information quality. The conclusions and future research directions are presented in the last section.

DATA QUALITY

The quality of data plays a dominant role in the final decision-making process. The decision task determines the quality dimensions of data [1]. A data quality problem exists if a given set of reported data cannot provide the information needed for decisions [2]. Dimensionality of data has been well researched [5] [6] [2] [4]. One stream of research has generated data quality attributes by conducting surveys. Even though this methodology is useful in getting a broad perspective on the issue, it is not very helpful in designing modifications to specific information systems to improve their yield quality. Wang [6] refers to data product similar to a tangible product that is produced from raw materials. Managerial decisions exhibit a large degree of diversity when compared to physical products, thereby yielding little help in designing/modifying information systems.

Galway and Hanks [2] categorize data into three groups: operational, conceptual, and organizational. Table 1 compares the problems that can occur among these types of data.

Operational Data	Conceptual Data	Organizational Data
Missing Data, Invalid Data, Inaccurate Data	Imprecise or Ambiguous Definition of Data	Disconnects between organizations that generate and/or use data

Table 1. Comparison of Data Problems [2]

From the point of view of dimensionality of data, the above comparison broadly points towards accuracy as the main attribute to monitor. Ambiguous definitions or mismatch between data definitions at different organizations finally contribute to data inaccuracy.

INFORMATION QUALITY

Information is the link between data and decision. Some researchers use the terms ‘data quality’ and ‘information quality’ interchangeably [4] [1]. Information is the result of processing data. The processing involves data and models. Therefore, in order to precisely identify the quality dimensions, these have to be considered separately. Zmud [7] suggests the following dimensions of information: quantity, reliability, timeliness, and format quality.

The dimensionality of information has a direct bearing on how it is used in decision-making process. Clearly, this depends on the level of management at which the decisions are made as well as the specific decision domain for which the information is sought.

A DOMAIN-SPECIFIC VIEW OF INFORMATION QUALITY

Current literature on information quality does not treat the components of data-information-decision chain separately for evaluating quality [3] [4] [1]. We illustrate the need to analyze the components separately by the following example.

Consider the information needed to make a decision regarding production schedule for the subsequent quarter in a manufacturing plant. Quantity to be produced (Q) depends on the following: estimates of demand for various products (ED), inventory of finished goods (IFG), and in-process inventory (IPI). ED in turn, depends on sales data (SLD) of previous time periods as processed by sales managers using appropriate forecasting models in their decision support systems.

Now, let us look at the factors that affect the quality of the parameters mentioned above. Good quality production schedule Q, really refers to a production schedule that uses the resources in an efficient and effective way. What are the dimensions of such a decision? Completeness would imply that all the products to be produced are covered, and relevancy refers to efficient and effective use of resources.

ED, the estimates of demands of various products to be manufactured are the outputs of information processing models such as forecasting the inputs for which are SLD, sales data for previous time periods. Quality dimensions of ED are accuracy and objectivity. Here the accuracy of ED also depends upon the appropriateness of the models used in the decision support systems. Quality dimension objectivity is defined to include both the appropriateness of model as well as lack of bias. SLD has to be necessarily accurate. IFG, and IPI have the same quality dimension of accuracy. Table 2 summarizes our analysis on the quality dimensions for the various components in this specific decision domain.

Level	Element	Quality Dimension(s)
Decision	Q	Completeness, Relevancy
Information	ED	Accuracy, Objectivity
Data	IFG	Accuracy
Data	IPI	Accuracy
Data	SLD	Accuracy

Table 2. Quality Dimensions for the Example Data-Information-Decision Chain

From the above discussion, we note the domain specific nature of quality dimensions that are appropriate for the different components of data-information-decision chain.

CONCLUSION

The quality of decision gets affected at different levels of management by the quality of data and/or information that is used. In order to improve the quality of any component of data-information-decision chain, it is necessary to clearly conceptualize the dimensionality of those components. In this article, we have suggested a framework where the decision domain is used to define the quality dimensions of the various components that are used in making that decision.

Future research in this area will consist of building a framework that integrates levels of management involved in the decision-making process as well as the different components of data-information-decision chain.

REFERENCES

- [1] Cai, Y. et al. "Evaluating Completeness of an Information Product," *Proceedings of the Ninth Americas Conference on Information Systems*, 2003, pp. 2274-2281.
- [2] Galway, L. A. and Hanks, C. H. *Data Quality Problems in Army Logistics*. Rand, Santa Monica, CA, 1996.
- [3] Klein, B. D. "Some Determinants of User Perceptions of Information Quality on the World Wide Web," *Proceedings of the Ninth Americas Conference on Information Systems*, 2003, pp. 2282-2286.
- [4] Wand, Y. and Wang, R. Y. "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM*, (39:11). 1996, 86-95.
- [5] Wang, R. Y., et al. "A Framework for Analysis of Data Quality Research," *IEEE Transactions on Knowledge & Data Engineering* (7:4), 1995, pp. 623-640.
- [6] Wang, R. Y., and Strong, D. M. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (12:4), 1996, pp. 5-34.
- [7] Zmud, R. W. "An Empirical Investigation of the Dimensionality of the Concept of Information," *Decision Sciences* (9), 1978, pp. 187-195.