IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB7994

Chapter II

Dimensions of Database Quality

John A. Hoxmeier
Colorado State University, USA

ABSTRACT

Databases are a critical element of virtually all conventional and e-business applications. How does an organization know if the information derived from the database is any good? To ensure a quality database application, should the emphasis during model development be on the application of quality assurance metrics (designing it right)? A large number of database applications fail or are unusable. It is evident that a quality process does not necessarily lead to a usable database product. A database application can also be well-formed with high data quality but lack semantic or cognitive fidelity (the right design). This chapter expands on the growing body of literature in the area of data quality by proposing additions to a hierarchy of database quality dimensions that include model and behavioral factors in addition to process and data factors.

INTRODUCTION

The ultimate objective of database analysis, design and implementation is to establish an electronic repository that faithfully represents the conceptual and logical model of the manageable aspects of a user's information domain. Enterprise and Web-enabled databases must satisfy a wide set of demands and constituents. Software engineering in general and database development in particular can be a complex, complicated process. There is probably no other product development process that faces the same amount of uncertainty, which may account for the high failure rate of software projects.

This chapter appears in the book, *Developing Quality Complex Database Systems: Practices, Techniques and Technologies* by Shirley Becker. Copyright © 2001, Idea Group Publishing.

For the purposes of this discussion, a database is defined as a self-describing collection of data that represents a model of an information domain. The database server manages the data, and the application presents the information to the consumer in some form of programmed behavior. Many internal and external factors must be considered during the development lifecycle, including, but not limited to, historical and future information requirements, the diversity of the data consumer community, organizational integration requirements, security, cost, value, ownership, performance, interface issues and data integrity. These factors contribute to the success of a database application in both quantitative and qualitative ways, and determine the real or perceived quality of the database application. *Process* and *data* quality are quantitative database management factors that are fairly well documented and understood, albeit underutilized. However, data *model* and *behavioral* considerations include important qualitative factors that contribute to overall database quality. A database is more than the collection of instances of the data managed by the server. Data quality, while important, is just one element of assessing overall database quality.

This chapter expands on the growing body of literature in the area of data quality by proposing additions to a hierarchy of database quality dimensions that include model and behavioral factors in addition to the process and data factors. While data quality has been the focus of a substantial amount of research, a standard definition does not exist in the literature (Wang and Madnick, 2000). The International Organization for Standardization (ISO) supplies an acceptable definition of data quality using accepted terminology from the quality field. These standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose. Applying the term "database quality" in this context would build on the ISO definition of quality, i.e., *conformance to requirements* and *fitness for use* (1993). ISO 8402 as a quality management and quality assurance metric provides a formal definition of quality: the characteristics of an entity that represent its ability to satisfy stated and implied needs. This definition is consistent with the notion of "customer satisfaction" prevalent in the quality literature (Juran, 1989; Crosby, 1995). Thus, a database can be defined to be of the required quality if it satisfies the requirements stated in a specification, and the specification reflects the implied needs of the user. Therefore, an acceptable level of quality has been achieved if the database conforms to a defined specification, and the specification correctly reflects the intended use. Unfortunately, neither of these definitions is adequate for the purposes of assessing database quality. Because there are so many functions that are transparent to the user, user satisfaction itself is not a sufficient condition for assessing database quality. Applying quality techniques to the development process is also a frequently prescribed method of ensuring software quality. While the requirement definition phase of the system development lifecycle is critical to the success of an application, doing a good job of defining requirements is not sufficient in the implementation of a successful database application. A database must also be judged by how closely it represents the world