

Dimensions of Database Quality

John A. Hoxmeier

Colorado State University, USA

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 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.

BACKGROUND

While data quality has been the focus of a substantial amount of research, a standard definition does not exist in the literature (Wang & 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, that is, "conformance to requirements" and "fitness for use." 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 (Crosby, 1995; Juran, 1989). 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. Un-

fortunately, neither of these definitions is adequate for the purposes of assessing database quality. A database must also be judged by how closely it represents the world of the data consumer (the model), its ability to respond to both routine and unanticipated requests within the domain it is expected to manage (the behavior), and maintain this representation over time. The framework presented herein expands on work previously proposed (Hoxmeier, 1997) and incorporates data quality dimensions put forth by several prominent data quality researchers (Ballou & Pazar, 1995; Krogstie, Lindland, & Sindre, 1995; Lindland, Sindre, & Solvberg, 1994; Orr, 1998; Storey & Wang, 1994; Strong, Lee, & Wang, 1997; Wand & Wang, 1996; Wang, 1998; Wang, Kon, & Madnick, 1993; Wang et al., 1994; Wang, Storey, & Firth, 1995; Wang, Strong, & Guarascio, 1996; Wang, Strong, Kahn, & Lee, 1999). The framework is important because it expands the definition of strict data quality to that of a broader context of database quality and incorporates the importance of process management.

THE CHALLENGE

Many database applications are ultimately unsuitable to the consumer. The process must incorporate three conceptually distinguishable domains: the modeling, the performance, and the enactment domains. Designers attempt to conceptualize the problem domain into a suitable physical model. The proposed physical model is subject to many performance constraints including the physical representation, the network topology, system configuration, and system administration. Finally, and what may be the most difficult to administer, the information is presented to the consumer for interpretation and enactment. The representation of the database after each of these domain layers all contribute to the quality of the solution by the information consumer. The critical elements below are the bases for the discussion on database quality dimensions.

- The cycle process must be managed toward a successful outcome.
- The model itself must represent a usually diverse and fuzzy problem domain.

Dimensions of Database Quality

- The quality of the data in the database must be of sufficient grade.
- The application must behave or have the ability to behave in a way the consumer understands.

To ensure a quality database application, should the emphasis during model development be on the application of quality-assurance metrics (designing it right)? It is hard to argue against this point, but there are a significant number of studies and anecdotal evidence that suggests that a large number of database applications fail, are unusable, or contribute to negative organizational consequences (Abate, Diegert, & Allen, 1998; Redman, 1998; Stackpole, 2001; Standish Group, 1997; Wand & Wang, 1996). The Data Warehousing Institute estimates that businesses lose billions each year attributable to bad data (Eckerson, 2002; Trembley, 2002). A quality process does not necessarily lead to a usable database product (Arthur, 1997; Hoxmeier, 1995; Redman, 1995). There are also many examples of database applications that are in most ways well formed with high data quality but lack semantic or cognitive fidelity (the right design; Motro & Rakov, 1999). Additionally, determining and implementing the proper set of database behaviors can be an elusive task.

While researchers have developed a fairly consistent view of data quality, there is little available in the literature on the evaluation of overall database quality including other considerations such as semantic fidelity (model), behavioral, and value factors.

A Database Quality Framework

It is proposed that through the hierarchical framework presented in Figure 1, one can consider overall database quality by assessing four primary dimensions: process, data, model, and behavior. Portions of the hierarchy draw heavily from previous studies on data and information quality, and documented process quality standards (Arthur, 1997; Department of Commerce, 2004; Wang, 1998). A dimension is a set of database quality attributes or components that most data consumers react to in a fairly consistent way (Wang et al., 1996). Wang et al. define data quality dimension as a set of data quality attributes that represent a single data quality abstract or construct. The use of a set of dimensions to represent a quality typology is consistent with previous quality research (Dvir & Evans, 1996; Strong et al., 1997; Wang et al., 1996). The framework presents the four dimensions in a dimension-attribute-property hierarchy.

Process Quality

Much attention has been given over the years to process quality improvement. *ISO-9000-3, total quality manage-*

ment (TQM), quality function deployment (QFD), and the capability maturity model (CMM) are approaches that are concerned primarily with the incorporation of quality management within the process of systems development (Dvir & Evans, 1996; Herbsleb, 1997; Hill, 2003; Schmauch, 1994). *Quality control* is a process of ensuring that the database conforms to predefined standards and guidelines using statistical quality measures. *Quality assurance* attempts to maintain the quality standards in a proactive way. In addition to using quality control measures, quality assurance goals go further by surveying the customers to determine their level of satisfaction with the product. Conceivably, potential problems can be detected early in the process.

Database Data Quality

Data integrity is one of the keys to developing a quality database. Without accurate data, users will lose confidence in the database or make uninformed decisions. While data integrity can become a problem over time, there are relatively straightforward ways to enforce constraints and domains and to ascertain when problems exist (Moriarty, 1996). The identification, interpretation, and application of business rules, however, present a more difficult challenge for the developer. Rules and policies must be communicated and translated and much of the meaning and intent can be lost in this process.

Data Model Quality

As has been presented, data quality is usually associated with the quality of the data values. However, even data that meet all other quality criteria is of little use if they are based on a deficient data model (Levitin & Redman, 1995). Data model quality is the third of the four high-level dimensions presented above. Information and an application that represent a high proportionate match between the problem and solution domains should be the goal of a database with high semantic quality. Representation, semantics, syntax, and aesthetics are all attributes of model quality (Levitin & Redman; Lindland et al., 1994).

The database design process is largely driven by the requirements and needs of the data consumer, who establishes the boundaries and properties of the problem domain and the requirements of the task. The first step in the process, information discovery, is one of the most difficult, important, and labor-intensive stages of database development (Sankar & Marshall, 1993). It is in this stage where the semantic requirements are identified, prioritized, and visualized. Requirements can rarely be defined in a serial fashion. Generally, there is significant uncertainty over what these requirements are, and they only

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/dimensions-database-quality/14354

Related Content

Information Management Challenges in Autonomous Vehicles: A Systematic Literature Review
Adrija Ghansiyal, Mamta Mittal and Arpan Kumar Kar (2021). *Journal of Cases on Information Technology* (pp. 58-77).

www.irma-international.org/article/information-management-challenges-in-autonomous-vehicles/277657

Factors Influencing the Use of Decision Support Tools of Enterprise Resource Planning Systems
Emad M. Kamhawi (2010). *Information Resources Management: Concepts, Methodologies, Tools and Applications* (pp. 218-237).

www.irma-international.org/chapter/factors-influencing-use-decision-support/54481

Artificial Intelligence Applications in Tourism

Carey Goh, Henry M.K. Mok and Rob Law (2009). *Encyclopedia of Information Science and Technology, Second Edition* (pp. 241-247).

www.irma-international.org/chapter/artificial-intelligence-applications-tourism/13580

PROLOG

Bernie Garrett (2009). *Encyclopedia of Information Science and Technology, Second Edition* (pp. 3147-3151).

www.irma-international.org/chapter/prolog/14040

Consistent Queries over Databases with Integrity Constraints

Luciano Caroprese, Cristian Molinaro, Irina Trubitsyna and Ester Zumpano (2009). *Encyclopedia of Information Science and Technology, Second Edition* (pp. 691-695).

www.irma-international.org/chapter/consistent-queries-over-databases-integrity/13650