Chapter II

Using a Metadata Framework To Improve Data Resources Quality

Tor Guimaraes
Tennessee Technological University, USA

Youngohc Yoon
University of Maryland, Baltimore, USA

Peter Aiken
Defense Information Systems Agency, USA

The importance of properly managing the quality of organizational data resources is widely recognized. A metadata framework is presented as the critical tool in addressing the necessary requirements to ensure data quality. This is particularly useful in increasingly encountered complex situations where data usage crosses system boundaries. The basic concept of metadata quality as a foundation for data quality engineering is discussed, as well as an extended data life cycle model consisting of eight phases: metadata creation, metadata structuring, metadata refinement, data creation, data utilization, data assessment, data refinement, and data manipulation. This extended model will enable further development of life cycle phase-specific data quality engineering methods. The paper also expands the concept of applicable data quality dimensions, presenting data quality as a function of four distinct components: data value quality, data representation quality, data model quality, and data architecture quality. Each of these, in turn, is described in terms of specific data quality attributes.

The importance of a company-wide framework for managing data resources has been recognized (Gunter, 2001; Sawhney, 2001; Stewart, 2001). It is considered a major component of information resources management (Guimaraes, 1988). The

complexity of data resources management is increasing as computer applications become more accessible to mobile users (Nesdore, 2001) and organizations attempt to extract more value from their data (Webb, 1999). As the volume, importance, and complexity of data management increases, many organizations are discovering that imperfect data in information systems negatively affects their business operations and can be extremely costly (Brown, 2001). Results from a survey indicate fifty percent of IS managers reported losing valuable data in the last two years and at least twenty percent with losses costing $1 million or more (Panettieri, 1995). Another survey reports 70% of the IS managers having their business processes interrupted at least once due to imperfect data (Wilson, 1992). Still another study showed that the nature of the problems associated with defective data ranges widely, from damaged files and lost data accounting for 23 percent of the responses, cost overruns (17%), conflicting reports (16%), improper regulatory reporting (13%), improper billing (9%), poor decisions (7%), delivery delays or errors (6%), and others (9%) (Knight, 1992).

We believe imperfect data can result from practice-oriented and structure-oriented causes. Practice-oriented causes result in systems capturing or manipulating imperfect data (i.e., not designing proper edit checking into data capturing methods or allowing imprecise/incorrect data to be collected when requirements call for more precise or more accurate data). Operational in nature, practice-oriented causes are diagnosed bottom-up and typically can be addressed by the imposition of more rigorous data handling methods. Structure-oriented causes of imperfect data occur when there exists a mismatch between user requirements and the physical data implementation designed to meet the requirements. The imperfections are inadvertently designed into the implementation. Correcting structural causes more often requires fundamental changes to the data structures and is typically implemented top-down. Structural problems result when a user cannot obtain desired results due to lack of access and/or lack of understanding of data structure, as opposed to getting an incorrect value or representation.

Adopting an organization-wide perspective to data quality engineering integrates development activities using data architecture. Failure to develop systems as coordinated architecture components results in fragmented data resources whose definitions apply at best within system boundaries. One additional consequence is that data interchange among company systems and those of partner organizations is more difficult. Structurally defective data results in unfavorable outcomes such as: 1) providing the correct response but the wrong data to a user query because the user did not comprehend the system data structure; 2) organizational maintenance of inconsistent data used by redundant systems; or 3) data not supplied at all due to deletion anomalies (i.e., storing multiple facts in the same physical entity).

Previous studies of data quality have addressed practice-oriented causes of imperfect data with data quality engineering methods such as those reported by English (1996) and Broussard (1994). Less guidance has been available to organizations interested in addressing the problems creating structurally defective data and how it relates to the comprehensive dimensions of data quality engineering.

Related Content

Fuzzy Based Project Time-Cost Optimization Using Simulated Annealing Search Technique
Khan Md. Ariful Haqueand M. Ahsan Akhtar Hasin (2014). *International Journal of Information Technology Project Management (pp. 90-103).*
www.irma-international.org/article/fuzzy-based-project-time-cost-optimization-using-simulated-annealing-search-technique/111178

ERP Implementation in State Government
Ed Watson, Sylvia Vaught, Dan Gutierrezand Dan Rinks (2003). *Annals of Cases on Information Technology: Volume 5 (pp. 302-318).*
www.irma-international.org/article/erp-implementation-state-government/44549

A Computer-Aided Conceptual Cost Estimating System for Pre-Stressed Concrete Road Bridges
Nikolaos Fragkakis, Sergios Lambropoulosand John-Paris Pantouvakis (2014). *International Journal of Information Technology Project Management (pp. 1-13).*
www.irma-international.org/article/a-computer-aided-conceptual-cost-estimating-system-for-pre-stressed-concrete-road-bridges/111171

Deep Stacked Autoencoder-Based Automatic Emotion Recognition Using an Efficient Hybrid Local Texture Descriptor
Shanthi Pitchaiyanand Nickolas Savarimuthu (2022). *Journal of Information Technology Research (pp. 1-26).*
www.irma-international.org/article/deep-stacked-autoencoder-based-automatic-emotion-recognition-using-an-efficient-hybrid-local-texture-descriptor/282708

Learning Technology Management While Teaching Technology Management: A Trial of Distance Learning in Higher Education
Linda L. Brennanand Victoria E. Johnson (2000). *Annals of Cases on Information Technology: Applications and Management in Organizations (pp. 39-60).*
www.irma-international.org/article/learning-technology-management-while-teaching/44627