Craig S. Mullins
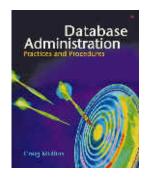
June 2008

DATABASE
TRENDS AND APPLICATIONS

The DBA Corner
by Craig S. Mullins

The Growing Importance of Metadata

What is metadata? Well, let me begin to answer by asking another question: Have you ever watched the *Antiques Roadshow* program on television? People bring items to professional antique dealers to have them examined and evaluated. The participants hope to learn that their items are long-lost treasures of immense value. The antique dealers always spend a lot of time talking to the owners about their items. They ask questions like "Where did you get this item?"

and "What can you tell me about its history?" Now, the item is sitting right there in front of them, yet they ask these questions. Why? Because these details provide knowledge about the authenticity and nature of the item. The dealer also carefully examines the item looking for markings and dates that provide clues to the item's origin.

Users of data must be able to put it in context before the data becomes useful as information. That's the role of metatdata. The simplest definition of metadata is "data about data." But to be a bit more precise, metadata describes data, providing information like data type, length, textual description, and other characteristics of the data. So, for example, metadata allows the user to know that the customer number is a five digit numeric field, whereas the data itself might be 56789.

So, using our *Antiques Roadshow* example, the item being evaluated is the "data." The answers to the antique dealer's questions and the marking on the item are the "metadata." Value is assigned to an item only after the metadata about that item is discovered and evaluated.

Metadata characterizes data. It is used to provide documentation such that data can be understood and more readily consumed by your organization. Metadata answers the who, what, when, where, why, and how questions for users of the data.

There are two basic types of metadata: technology metadata and business metadata. Technology metadata describes the technical aspects of the data as it relates to storing and managing the data in computerized systems. Business metadata describes aspects of how the data is used by the business, and is needed for the data to have value to the organization. So, knowing that the LICNO column is a positive integer between 1 and 9,999,999 is technology

metadata. Knowing that the LICNO column is the practitioner license number for certified course instructors, must be unique and every instructor can have one and only one license number is business metadata. Both are required, but the biggest need for improvement in most organizations involves business metadata capabilities.

The need for metadata has always existed. But its importance is growing rapidly in today's IT world. Regulatory compliance imposes requirements on data management and the data must be organized and classified in order to determine which regulations apply to it. Without metadata, classification is impossible – at least proper, correct classification.

Consider, for example, the many regulations dealing with long-term data preservation and retention. Different types of data must be retained for different durations. Without accurate metadata, the ability to tie your data to the appropriate regulation, and thus the correct retention period, is impossible.

Other regulations, such as PCI and HIPAA, contain stipulations that companies must protect personal information by implementing procedures to ensure that confidential data is accessed by legitimate sources only. This can require the implementation of detailed database auditing capabilities to track who did what to which piece of data when. But such stringent auditing is not required on all pieces of data. Once again, metadata is required to classify the data into the proper category in order to implement the appropriate controls for the appropriate data.

A wise organization will develop a metadata strategy to collect, manage, and provide a vehicle for accessing metadata. A sound metadata strategy should address the following:

- A policy for how metadata is used in the organization.

- Procedures for identifying and defining data ownership and stewardship.

- Identification of the types of metadata that needs to be collected.

- A fundamental description of the purpose for each type of metadata that is identified. In other words, the metadata strategy should provide a clear and concise reason why each piece of metadata is required by the organization.

- Methods for the collection and storage of metadata (typically using a repository).

- Methods for accessing the metadata.

- A security policy to enforce the data stewardship procedures, as well as to enforce security policies on metadata access.

- Identification of metadata sources, both internal and external.

- Measurements to gauge the quality and usability of metadata.

From Database Trends and Applications, June 2008.

Home.