Craig S. Mullins
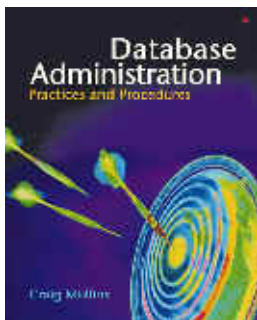
April 2007

The DBA Corner
by Craig S. Mullins

The Return of the Repository?

There was a day, oh so long ago, when data management professionals believed in that big repository of all corporate metadata... But somewhere along the way, the concept seems to have withered and died.

Let's take a step back and define what I mean. I'll call it the big "R" Repository. Some of you know what I'm talking about and are probably chuckling under your breath. Perhaps you even used IBM's Repository Manager/MVS or the RelTech, ummm, Platinum, no, CA Repository... or maybe even the Rochade, err Viasoft, no ASG-Rochade Repository. Others are wondering what the heck I'm talking about.

It used to be that data architects (or data administrators) were proponents of the big "R" Repository. This Repository was to be the central hub of all corporate metadata. If it is data the company cares about, then it should be defined in the Repository. Only when all appropriate definitions are documented will we understand our data and manage it accordingly.

Think about it: for any piece of data to be understood, metadata is required. At a minimum, we need a definition to know whether 12 (simple data) is a shoe size, an IQ, or a month, right? Without metadata, data has no identifiable meaning – it is merely a collection of digits, characters, or bits. Simply stated, metadata makes data usable.

So building a Repository – a database of metadata for our data elements – is not a bad idea at all. But the project to build that initial database of metadata can be troublesome to initiate and maintain. First of all, there is that big honking project to scan everything and document it and get it into the Repository. This will be time-consuming and probably expensive.

Then, as time progresses and "things" change, the Repository will get out of date unless proper procedures are instituted to update the Repository as changes and new projects roll out. This requires centralized management and significant up-front planning and coordination to succeed. So, if we started our Repository project in the 80's or early 90's, we were most likely derailed sometime shortly thereafter. We all remember the decentralized, client/server hey-days of the 1990s, don't we?

Today, the common wisdom is that big "R" Repository is dead. Instead, little "r" repositories are being populated in various tools. But without some central record of correctness, how can we be sure of our metadata? It is the same problem we have with data today. The OLTP database on mainframe DB2 holds customer data - as does that Oracle workgroup

database implemented in Unix by the business unit, and that data mart over there in Teradata has customer data, too. Is it all the same? When it is different how do we know what is correct and what is not?

Properly exploited, repository technology offers many benefits. The metadata in the Repository can be used to integrate views of multiple systems helping developers to understand how the data is used by those systems. Usage patterns can be analyzed to determine how data is related in ways that may not be formally understood within the organization. Discovery of such patterns can lead to business process innovation.

The primary benefit of a Repository is the consistency it provides in documenting data elements and business rules. The Repository helps to unify the "islands of independent data" inherent in many systems. It enables organizations to recognize the value in their systems by documenting program and operational metadata that can be used to integrate the legacy systems with new applications. Furthermore, a repository can support a rapidly changing environment such as those imposed by Internet development efforts on organizations.

Today's trends are driving is a resurgence in the importance being placed on metadata. Data governance initiatives, regulatory compliance demands, and even data warehousing contribute to the surge as they all require accurate metadata to succeed. A Repository can help aid in the maintenance and upkeep of your metadata.

Although the history of the Repository as a software product is interesting, it is littered with unsuccessful initiatives and products. But what is the future of the repository? I'm a fan of centralization, but I'm not convinced that the big "R" centralized Repository can make a comeback, even though it should. So we'll need to find creative ways to stitch together our little "r" repositories. Or create a stealth big "R" Repository using the metadata features in our data modeling tools.

The bottom line, though, is that regulatory compliance will continue to consume cycles and therefore increase our need for accurate, up-to-date metadata. So what will your company do to document and share metadata? It is a question worth asking.

From Database Trends and Applications, April 2007.

Home.