TITLE: "STAGNATION IN DATA ADMINISTRATION"

by Tim Bryce
Managing Director
M. Bryce & Associates (MBA)
P.O. Box 1637
Palm Harbor, FL 34682-1637
United States
Tel: 727/786-4567
E-Mail: timb001@attglobal.net
WWW: http://www.phmainstreet.com/mba/
Since 1971: *"Software for the finest computer - the Mind"*

"Most organizations use a DBMS as nothing more than an elegant file access method. Consequently, the opportunity to share data and integrate systems is lost."
- Bryce's Law

INTRODUCTION

I recently visited a "Meta-Data Conference" in Orlando conducted by the DAta Management Association (DAMA). I was primarily attending to see the latest vendor offerings and to see what, if anything, has changed. Frankly, not much.

I heard about "information management," "meta-data stores," "business rules," data dictionaries and repositories, and lots of other gobbledegook. Previously, the emphasis in Data Administration over the last few years was on how to migrate data from your legacy systems to new platforms. However, the literature I collected seems to be less concerned with data migration and more concerned with INTEGRATING both old and new applications. To which I have to give a big, *"DUH!"*

REPOSITORIES

My attention primarily focused on the latest offerings for cataloging and cross-referencing information resources. This now goes under a variety of names, be it "Meta-Data" (meaning "data about data"), Data Dictionaries, Encyclopedias, and Repositories. For the sake of this article, I will use the term "Repository" to represent all of these products. Here is what I found:

SCOPE - these products are primarily concerned with data resources only (data elements, records, files). Programs and proc libraries are also allowed but little else. Few, if any, catalog enterprise-wide systems, business processes (sub-systems), manual processes, enterprises, business functions, human and machine resources, skills, projects, business objectives, information requirements, etc. Their scope is on nothing but data resources, which is fine if you are a Data Engineer or a DBA, but please stop kidding us that this is an enterprise-wide solution for everything from the womb to the tomb.

DEFINITION - Individual components can be easily recorded, either through web based input screens or through some "harvesting" tools to scan data libraries and populate the Repository accordingly. Data components are typically identified by program label as opposed to a business name; e.g., "CustNo" versus "Customer Number." Further, I couldn't find anything supporting multiple physical views of data. For example, a Data Element such as "Ship Date" has one logical definition ("The date when the product was shipped to the customer") but may be expressed physically many different ways; for example, in COBOL we refer to it as "SHIP-DATE" and in C we refer to it as "ShipDate"; further, we may elect to store the data as YYYYMMDD, yet distribute it to users as DDMMYY. Because of the inability to support multiple physical definitions, multiple versions of the the same component are inevitably created and the chance to share and re-use the same data element is complicated.

I also didn't see a definition checking facility whereby I might create an initial entry (skeletal) and the computer reminds me to finish the definition later on (we used to refer to this as "status check").

REDUNDANCY - I saw little to prevent data redundancy. Sure, there were some slick search routines but I saw nothing to automatically stop or warn the user they are creating duplicate entries. Years ago we used the concept of "Logical Attributes" and "Data Taxonomy" to examine the uniqueness of a data element (I'll describe this another time). However, I didn't see anything like this in the products I looked at. Nor was there anything to check a data element by redundant physical characteristics. I thought this was very odd as I assumed we had progressed past this point.

RELATIONSHIPS & IMPACT ANALYSIS - all of the products I looked at did a good job establishing relationships between components, be it data-to-data or data-to-records, files, etc. Consequently, adequate "impact analysis" features are provided (cross referencing the use of resources).

INTERFACES - Although it appeared some of the Repositories had some form of export facility, they were all rather vague in terms of importing data from third party products. It was also unclear as to whether these tools could be used to drive other products such as program

generators, 4GLs, programmer workbenches, etc. This is rather disappointing as I had hoped to see an "open architecture" for such tools.

OTHER CONSIDERATIONS - There appears to be an effort to establish cross platform solutions of the various Repositories, but they are not quite there yet. Further, these products are fine in English but I saw little consideration for cross cultural applications (e.g., Chinese, Japanese, Arabic character sets).

HISTORY

In 1971 we created the first such repository in "PRIDE" using a manual technique. In 1974, we automated it and called it "PRIDE"-LOGIK which eventually evolved into the "IRM Repository." Our competitors at the time were Data Manager from MSP (U.K.) and Data Catalog from Massachusetts. Both products are long since gone, but I see little in today's offerings to differentiate them from yesteryear. Sure, today's offerings now run on PC's and have fancier graphical front ends to them, but they are still primitive offerings that do little to help you manage data resources. It would take a Herculean effort by a company to bring about true data integration using these tools. From this perspective, I was horribly disappointed in the progress the industry has made in this regard.

For more information on "Establishing an IRM Repository," see:

http://www.phmainstreet.com/mba/pride/spir.htm

VERNACULAR

The catch phrases at the conference were "Business Rules" and "Information Management." Whereas the former is a reincarnation of Entity Relationship (ER) modeling from the 1980's, the latter is plain and simply a fallacious concept. Here again is another example of the sloppiness of the terminology and thinking in our industry. "Information Management" is what Joseph Goebbels practiced in Nazi Germany where he controlled what people heard and saw. "Data Management" is a more apt title but again, we must be able to differentiate between "data" and "information."

This sloppiness is again due to the lack of standards and is an indication the industry is still in its infancy and not yet a legitimate profession.

PARTS DEPARTMENT ANALOGY

Look, it is all rather simple; in order for Data Administration to succeed (or "Data Resource Management" as we prefer to call it) we must take the position we are the "Parts Department" to our product assembly lines (our "systems"). As such, the intent is to inventory our parts so we know where they are and how to access them in the most expedient and cost effective means possible. Further, we have to be able to classify our parts so we know them to be unique and non-redundant, thereby making them available for multiple assemblies. A good example of this is years ago when GM was able to slip a Pontiac engine into a Cadillac. This is only possible through the standardization of parts.

In order to promote both standardization and inventory, you need a "bill of materials processor" (BOMP) to catalog and cross-references the various parts to the assemblies they belong in. BOMP, therefore, is the great-granddaddy of all "meta-data stores", repositories, or whatever you want to call it. It is simple and it works. It just demands a different perspective in our work effort.

Consider this; the first true "Data Management" organization was formed at the Quaker Oats Company in Chicago, Illinois in 1965. At the time, the company was trying to conquer a mammoth Management Information System (MIS) spanning several divisions within Quaker. As the development staff began to study the requirements for the project, it became obvious there was a considerable number of data elements to be shared and re-used. Consequently, Quaker's Management created a "Data Management" group dedicated to sharing and re-using data, thereby assuring systems integration. This group did not have such luxuries as a Data Base Management System (DBMS) or any of today's data base design tools. Nevertheless, Quaker's "Data Management" group met the challenge and created a world-class MIS that was fully integrated. This happened not because of any particular tool or technique, but because they took the posture of a "Parts Department" as mentioned previously. Quaker's MIS may not be the flashiest by today's standards, but it would be the envy of any true Data Administrator 40 years later.

CONCLUSION

I am always amazed how this industry makes mountains out of mole-hills. Was I impressed by the products I saw

(continued from page 2)

at DAMA? Hardly. But that is not the point. I was more disheartened than anything else. Basically, I believe the Data Administration people have stagnated for the last twenty years. Quite simply, it has gone to sleep. It is still

mired in programming and cannot see the big picture of managing all of an enterprise's information resources. They are more concerned with technical facade than substance. In this way, I believe they are still rearranging the deck chairs on the Titanic. What a shame.

END

"PRIDE" Special Subject Bulletins can be found at the "PRIDE Methodologies for IRM Discussion Group" at:

http://groups.yahoo.com/group/mbapride/

You are welcome to join this group if you are so inclined.

"PRIDE" is the registered trademark of M. Bryce & Associates (MBA) and can be found on the Internet at:

http://www.phmainstreet.com/mba/pride/pride.htm