Meta Data Repository Redux, Part 1 – Meta Data Use Case Detail and Diversity

Article published in DM Direct Newsletter
April 2, 2004 Issue

By *John Singer*

They say what goes around comes around and this certainly applies to the meta data information repository. Over the last 25 years, each great evolution of information systems technology and best practices has bred the same set of issues and the invariable vendor response - the meta data repository (or some derivation). What is this problem that keeps recurring? Simply put, as newer technologies are used to develop systems, the increase in complexity breaks the system management status quo. The answer always involves storing information about the system components in a central place for analysis - i.e., the meta data repository. It hasn't always been called this, but nonetheless, the basic approach to solving IT management complexity is the same. We are now witnessing this event playing out once more, but I'm getting ahead of the story.

This article seeks to examine the phenomena of meta data repositories, identifying common patterns of process and technology. Despite what appears to be great similarities, differing IT constituencies and problem domains suggest that a one-size-fits-all solution approach will not work. Finally, in part 2, I will define a strategy for successfully implementing an IT knowledge management program based on meta data repository concepts.

Decade	Enabling New Technology	Complexity and Integration Issue	Repository Technology	User
70s	Centralized DBMS	COBOL copy books and DBMS structures	Data dictionary	Programmer
80s	Graphical PC, relational DBMS	CASE tool design artifacts	CASE tool repository	Analyst
90s	Cheap disk, cheap processing	Data definition, movement and translation in the data warehouse	General purpose meta data repository	End user
00s	Internet	Multitier application architectures	Technical infrastructure repository	Operations
00s	EAI/ERM	Application-to-application data synchronization	Technical infrastructure repository	Systems
00s	Enterprise architecture	IT-to-business alignment	Zachman Framework Repository	Management

Figure 1: Meta Data Management Through the Decades

The Nature of the Problem

Ultimately, the organization will implement a "just good enough" approach to managing complexity. The problem is that each new successive wave of application innovation expands the linkages between system components, thus breaking the current "sufficient solution" put into place to meet the prior generation's system management problems. Figure 1 illustrates the enabling technology that led to the IT complexity

and, ultimately, initiated the meta data management response. In all cases, meta data from related toolset and infrastructure suites was integrated for the purpose of better understanding and managing the IT resource. Furthermore, the solution must provide the same generic types of meta data retrieval and analysis capabilities:

- **Definition analysis** - I have this thing, but I don't really know what it is or what it means. Every project starts with the software archeology phase where the developers and users try to understand the current system.
- **Impact analysis** - I make a change to one thing, and something else stops working. How can I figure out the impact of my changes?
- **Where used analysis** - I have a certain thing that is used in many places. Now I have to change this thing. How many do I have and where do I find them?
- **Difference analysis** - I have many instances of the same thing, how can I prove they are exactly the same? Another example: Something just stopped working - but I don't know what's changed in the environment.
- **Location/navigation analysis** - I have this thing, now I want to go look at it directly. Because meta data is often removed from the original source, the user may want to inspect the source directly.

The examples in Figure 1, while similar in nature, illustrate a growing diversification in the type of meta data and the intended audience. The "user" column now includes virtually everyone in the organization! From technical providers to senior management, everyone in the organization wants easier access to information that defines the IT environment. Given such a broad range of users it should not be surprising that their information needs vary.

This expanding universe of "everything IT" information requirements forces us to examine meta data complexity. This complexity can be measured across two important axes:

- **Detail** - Detail includes several different measures including:
  - Number of different classes of data
  - Number of instances of each class of data
  - Number of properties of each class of data

  Meta data can exist at different "roll ups," just like data in the data warehouse. For example, the identical database may be installed on multiple servers, each in a different location. Depending on the analysis being performed, you might want to know the count of all instances of the same database or just the fact that it exists. Further grouping of databases into subject areas would represent a higher level of abstraction (and less detail).

- **Diversity** - there are literally thousands of "classes" of meta data. Simply identifying every occurrence of every IT asset can produce millions of individual records in the repository. Diversity also includes relationships between IT components. Some analysis requires only closely related meta data (i.e., what columns are in a table), while other analysis requires chasing down long chains of relationships (i.e., what database servers support this business unit).

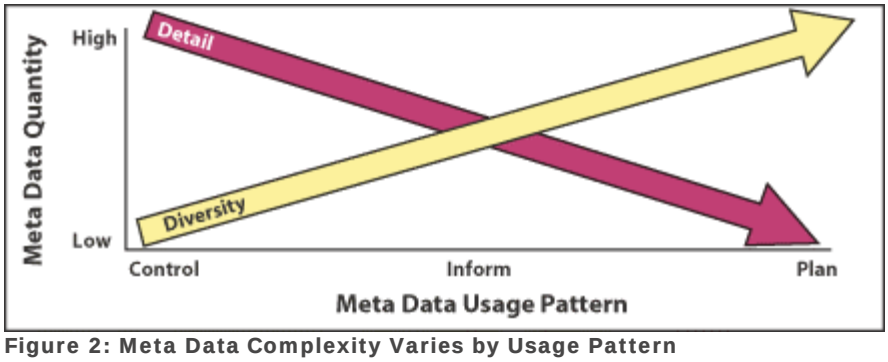

Figure 2: Meta Data Complexity Varies by Usage Pattern

Figure two illustrates the relationship between meta data detail and diversity. Closer examination of the broad range of meta data-related efforts in the context of the user community being served and the detail/diversity of meta data required leads to the observation of three distinct use cases for the application of meta data. These use cases have distinctly different patterns of meta data detail and diversity.

- **Control**. Focus is on families of related toolsets and infrastructure with the goal of improving some aspect of IT system management. Emphasis is on "impact analysis" and "difference analysis." Rapid problem resolution requires access to real-time meta data from multiple infrastructure component sources. These efforts occur within the technology domain team (i.e., the data architecture group creates a data dictionary, the systems management group installs inventory management).
- **Inform**. Here the focus is on "definition analysis" and transferring knowledge from the IT domain experts to the generalist. This includes application developers and systems analysts who start every project with the "software archeology" phase (i.e., researching existing code and database structures to determine how existing systems work). This isn't an issue if your systems and databases are all well documented. End users want to know the definition of data that appears on data warehouse reports, where the data came from and how it was transformed along the way. The problem is that much of this knowledge is locked up in tools or infrastructure (e.g., data models, DBMS catalogs, source code). Making this information available to a wide audience is the goal for these types of meta data management efforts.
- **Plan**. The goal is understanding how the business relates to IT across the entire range of IT assets (e.g., data, application, infrastructure). "Where used" analysis is the primary requirement. For example, if the company expands this product line, what systems will be impacted? Which systems support our customers? Identify processes that redundantly store data. IT strategic planning is increasingly looking to the application of meta data to better support investment decision making. These efforts are an outgrowth of the emphasis on enterprise architecture planning and create the added challenge of describing and documenting more abstract concepts such as business process and application system, which are really agglomerations of literally thousands of the individual IT components stored in the "control" and "inform" repositories.

Twenty-first century meta data management for IT has morphed into a knowledge and content management project with the goal of capturing, classifying and categorizing all things IT. Pursuing this goal will result in a better informed IT staff and will reduce time to market for new systems. IT to business alignment will improve as the IT process and deliverable becomes more transparent to the business and as the business can more readily measure the impact of spending decisions on the IT architecture. Critical to success is recognizing the three distinct meta data management patterns: control, inform and plan. No single vendor provides an overall solution, rather a mix of technologies and in-house developed solutions will be required. Part 2 of this article will examine the three meta data patterns and suggest an architectural approach for a comprehensive meta data solution.

Meta Data Repository Redux, Part 2: Crafting the Enterprise IT Knowledge Management Strategy

by John Singer

Summary: *Part 1 examined the phenomena of meta data repositories, identifying common patterns of process and technology. Part 2 defines a strategy for successfully implementing an IT knowledge management program based on meta data repository concepts.*

In Part 1 of this article, we explored the concepts of detail and diversity in meta data. We noted that meta data projects tend to cluster into three distinct groups based on their mix of meta data detail and diversity. These three use cases can be described as "control," "inform" and "plan."

The "control" use case typically involves an individual IT technical service provider group (e.g., DBA, security, network, systems management) working within its domain. Existing system management tools are leveraged as collectors of meta data for better control, reporting and analysis of the environment being managed. Data is captured at the most detail level possible; however, including only that meta data directly controlled by the technical domain (i.e., low diversity). Virtually every technical service department in your organization has a homegrown database to track and manage IT assets, yet none of them are integrated or even known outside their department. These efforts are typically run as system management improvement projects and not necessarily recognized as meta data projects at all. Unfortunately, these projects also represent the source system or the system of record for all other meta data related efforts.

The "inform" use case has at its core the desire to inform and educate people outside of the immediate technical domain. This form of altruism is typically driven by user demand for greater understanding of IT systems in their own terminology. In this use case you see the introduction of higher level of abstractions to assist the user in navigating through the forest to find the particular tree of interest (e.g., naming application or data subject areas of interest). Data is also less detailed with emphasis on names of objects and their definitions but without the myriad other properties of interest to the technician. Diversity goes up in terms of the relationships that exist between classes of data. For example, data warehouse-oriented repositories place a huge emphasis on tracing relationships from OLAP report output back to original source system of data.

The "plan" use case takes the granularity and diversity dimension to the extreme. Diversity now includes linking meta data across the entire IT ecosystem, from technical infrastructure to business strategy. However, the meta data must be presented at a very abstract level (i.e., with little detail). The "plan" use case also introduces new classes of meta data that describe the business itself in terms of business processes and strategy. Notoriously difficult to develop and maintain, the business must define itself in terms that can be stored and analyzed without too much detail making it impossible to maintain. Finally, you see "non-traditional" meta data such as IT financial data and project management data added to the mix.

Crafting a successful enterprise IT knowledge management strategy clearly requires more than a good meta-model and repository. One must consider the various access patterns, use cases and meta data complexity. Figure 1 illustrates the architecture for bringing all this together.

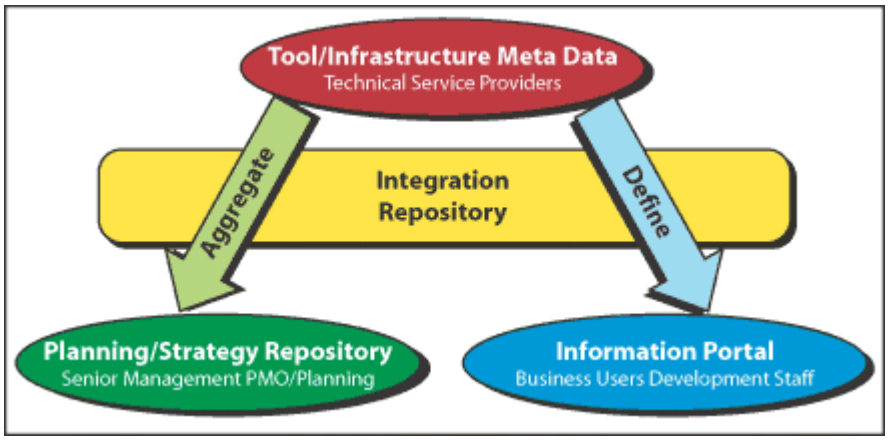

Figure 1: IT Knowledge Management Architecture

Tool / Infrastructure Meta Data

We are drowning in a sea of meta data. Every tool and infrastructure component creates and stores meta data. However, it is typically not accessible or even understandable outside of the immediate team of domain experts (i.e., the tool/infrastructure owners). There is little if any interchange of this information between domain experts, yet the need for this information exists across the organization. The highest cost information exchange is the hallway conversation needed to communicate across domain expert boundaries.

The obvious place to start is with existing tool and infrastructure meta data sources. The starting point for a more transparent IT organization is with the technical service providers who, on a day-to-day basis, build and implement the IT components driving the business. A strategy for strengthening meta data capture must start here and should include the following:

- Survey IT domain experts to identify and model available meta data. Look at consolidating tool suites and leveraging vendor supplied tool suite repositories.
- Establish accountability for providing quality meta data as a part of designing, building and implementing systems. This responsibility should rest with the domain experts responsible for implementing a particular component. You will likely find that valuable meta data for a given component is "optional" (e.g., the "comment" field in a table definition). Define minimum standard local meta data requirements and establish accountability for supplying it.
- Identify change control procedures for implementing or changing IT components. The process that changes an IT component should become the triggering event for capturing meta data, otherwise the meta data quickly becomes stale and is not trusted.

Focus "control" type use cases on these local tool or infrastructure suite repositories. This provides value to the domain experts as a reward for supplying higher quality

meta data. This solution, however, is rarely sufficient for reaching a broader audience.

Integration Repository

Moving beyond the "control" use case will require linkage and abstraction of meta data across expert domains. This will require the creation of a special purpose repository which might be homegrown or purchased. This repository serves two primary purposes:

- Lightly summarized and aggregated meta data from the tool/infrastructure repositories is stored here to facilitate further analysis.
- The IT taxonomy is established that categorizes, classifies and groups meta data into higher level abstractions. Every technical service provider group in IT (e.g., security, DBA, testing) has a list of the organization's applications. You can bet none of these lists match. Creating the common IT vocabulary is critical for linking meta data across the expert domains. These taxonomies provide the valid values for meta data that must be supplied at the tool/infrastructure level. Furthermore, the low level taxonomies provide the basis for easily relating detail technical meta data to the higher level of abstractions needed to support the "plan" use case.

Create a small central function for designing taxonomies that are applied across IT and to design meta-models at the tool level. However, this central group should not be responsible for filling in missing meta data. Accountability for providing quality meta data must reside at the source, while design of the meta-models should be consolidated to achieve the desired level of integration.

Planning/Strategy Repository

This most difficult aspect of this effort is capturing the business model in a manner that is useful. Loosely modeled concepts on PowerPoint charts are insufficient for this effort, yet most business process modeling tools require too high a level of detail to work. A carefully balanced approach that creates a simple meta-model for defining the business architecture coupled with an interview approach to gathering the data will result in enough usable information to perform planning processes. These business concepts must then be linked to the detailed data and application architecture meta data captured in the "control" repositories. Without the integration meta data described above, this will not be possible.

Information Portal Repository

The information portal repository is not necessarily a separate data store but represents a different look and feel. The data may come directly from various "control" oriented repositories, but requires the integration repository to create organization and structure. Users want customizable views; for example, the user interface would only list data elements associated with certain applications or data subject areas. A higher emphasis is placed on search capabilities and formatting information for viewing online.

21st century meta data management for IT has morphed into a knowledge and content management project with the goal of capturing, classifying and categorizing all things IT. Pursuing this goal will result in a better informed IT staff and will reduce time to market for new systems. IT-to-business alignment will improve as the IT process and deliverable becomes more transparent to the business and as the business can more readily measure the impact of spending decisions on the IT architecture. Critical to success is recognizing the three distinct meta data management patterns: control, inform and plan. No single vendor provides an overall solution, rather a mix of technologies and in-house developed solutions will be required. Finally, a standardized taxonomy of IT components is critical for enabling an end-to-end solution, where IT technologists and senior business strategists can finally speak the same language.

John Singer is a 24-year veteran information systems professional who has focused on data management activities including database administration, data administration and enterprise architecture in both staff and management roles. Currently working as a data architect at MasterCard International, Singer has experience in the pharmaceutical, healthcare, manufacturing, retail and criminal justice industries. He can be reached at john_singer@mastercard.com.