

CDI and MDM Are Broken, Part 2

Steve Putman, Lemon Curry Solutions Ltd.

In [Part 1](#) of this paper, I described the current state of technology-based solutions for the Master Data Management (MDM) market as “dead on arrival.” I described the MDM market in that manner because the systems offered appear to be overly complex and not flexible enough to adapt to myriad manifestations of the models related to the main subject areas of customers, products and other MDM subject areas. I also described a possible way forward out of this situation: the application of “semantic” technologies in the storage, indexing and retrieval of enterprise data.

In Part 2 of this paper, I will describe two extremes of MDM data storage, a possible middle ground between the two extremes, and the design considerations for constructing such a system. Armed with this knowledge, you should be able to make effective, practical decisions on the logical and physical design of your future MDM system.

Vendor Offerings

Major vendors in the MDM space including [Oracle](#), [SAP](#) and [IBM](#) have sophisticated products that leverage the most advanced technologies available, such as service-oriented architectures, concept hubs for storing master data, and highly capable messaging architectures. These products work well in a homogeneous (single-vendor) computing architecture – quite rare in the corporate world – and can be made to work in the more common heterogeneous environments with a great deal of customization. Please refer to [Part 1](#) of this paper for a detailed description of the benefits and limitations of current vendor offerings.

Semantic Technology

On the end of the spectrum furthest from past accepted practices, a new paradigm in both web- and data- based computing is being developed. Called Semantic technology, the basic advantage is the storage of the relationship of two objects in addition to the object itself, which allows the processing of these relationships by both humans and other computers. The goal is to make computer systems both more interoperable and flexible to new situations with common rule sets and definitions, enabling business operations to be more efficient and cost-effective. This technology has far-reaching implications to a wide variety of computing subjects. But those are beyond the scope of this paper, so we will focus here on its application to the MDM application space.

Semantic data storage disconnects the concepts of the relationship framework from the data that are inherent in relational database designs and allows for the separate storage of these concepts. The data are stored in a simple structure denoted by the [RDF](#) “triple,” based on the “subject-object-predicate” structure that you likely learned in high school English or logic classes before you moved on to computer programming courses. The “data model” in a semantic system is stored in a framework called an “ontology,” which represents the valid objects, rules and relationships for the system implementation. This storage technique is advantageous to an MDM system because it frees the system designers from the requirement of anticipating all possible

data and relationship combinations before the system is implemented. Both the data storage and structures are easily extendable compared to the relational model.

Semantic systems currently are being used for applications with huge amounts of data, such as fraud detection, pharmaceutical research and geospatial mapping. These are fundamentally different applications than typical ERP applications such as financial analysis, human resources and customer relationship management. It is estimated that commercial applications of semantic technology in corporate enterprises currently are “at least five to ten years away from market adoption.”(1) Therefore, companies wanting to take advantage of this technology are truly on the “bleeding edge” of the state of the art, with few pre-packaged tools available and even less external support. This alone makes a pure semantic solution unlikely in most enterprises.

The Middle Ground

The introduction of semantic concepts is well suited to an MDM implementation due to the enterprise-wide nature of successful MDM. However, the widespread adoption of semantic technologies is not a given primarily because it involves a fundamental shift in thinking by IT professionals in how data are stored and represented in databases. This might be too large of a change for many organizations to accept. In addition, the lack of significantly mature tools or packaged applications likely will mean that adoption in the enterprise will be slow.

If the future adoption of semantic ideas – and technology, eventually – is accepted as beneficial to your organization, several steps can be taken in your company to prepare for this change. Most of them have nothing to do with technology. The integration of these ideas into your software development and data management practices will give the enterprise the proper perspective to be successful in your MDM initiative, and will be helpful in other areas of operations as well, regardless of the adoption of semantic technologies.

For the purposes of this paper, consider a huge assumption – you have secured active and committed senior management involvement and commitment for your MDM initiative, along with an effective governance and stewardship program. This is essential for success for a large program, but there are many opportunities for delay or failure. Senior management support is also important for the move to semantic design principles because many of the activities required are changes in philosophy as well as changes in toolsets, which requires top-to-bottom buy-in to be successful.

The following activities will lead you in the direction of semantics:

- Document all processes thoroughly, paying particular attention to the relationships between objects at all levels of granularity.
- Make the tools “invisible.”
- Teach all people involved to think in terms of interconnected metadata, not specific tasks that are not relative to other activities.
- Learn to use or “mine” data rather than to simply collect it.

Document All Processes

This recommendation should be self-evident. Businesses train new employees on how to perform their jobs every week, and department veterans are treasured for their knowledge and skill at doing their jobs. However, it is rare to find a business that both trains its employees and thoroughly documents its processes to the point that an external group can understand them effectively. Knowledge is retained by employees, and changes in process are difficult to maintain

in existing documentation stores. The evidence of this phenomenon can be seen in the thriving practice of business analysis in IT shops, which would be unnecessary if thorough documentation were present.

MDM can be described as an implementation of “organizational semantics” - the common definition of business terms and concepts with the enforcement of relationships between them. To this end, commercial tools currently are being developed to record these relationships in computer-readable form. While these tools are being developed, enterprises can prepare for their arrival by performing this documentation manually. One opportunity for accomplishing this documentation lies in the analysis done in implementing a new system or upgrading an existing one. Part of this analysis includes mapping data elements from current into new, which is the same type of relationship data that semantic systems will need in order to operate. At a higher level, groups of tables in a database are related to each other in the same way. Once your relationships are documented in this fashion, you will have the beginnings of a hierarchical ontology.

I have found in my travels in the corporate world that virtually all business and data analysts already know how to perform this sort of analysis. In fact, most are doing it right now. We want to move from an environment where this analysis is done repeatedly over many years to one where it is done only for a new situation. The phrase to use in implementing this new way of thinking is “do it one more time.” The IT group's responsibility is to provide a mechanism so data related to existing processes can be reused.

Make the Tools “Invisible”

Two of the fastest ways to ensure failure of a documentation effort are either to impose tools that are too difficult to understand and use, or to add tools that duplicate the effort already being expended by staff. For example, as a consultant I usually create at least three timesheets per week when I am at a client location – one for the project, one for the time accounting system at the client, one for my internal accounting system. This duplication of effort is inherently inefficient and can breed resentment in the staff, which can doom the effort to failure.

Any system implemented to determine semantic relationships should be able to read activity within the enterprise and store data appropriately. The move to service-oriented architecture in the computer industry is an example of this type of system. A service could be programmed to watch for particular events within the system, such as a new customer placing an order or the storage of a new version of a project plan on a shared drive. The system then would record the relationships of the objects contained in the events along with a time stamp. In my example above, the storage of one timesheet could trigger events to automatically update systems both inside and outside of the organization. The technology exists to do this. What does not yet exist are fully-developed commercial applications that integrate these transactions in a meaningful way.

Interconnected Tasks and Reuse

One of the most difficult concepts for many employees to grasp is how their tasks fit into the fabric of the enterprise. The integration of individual tasks into general business processes falls to senior management, while the front-line worker simply does his job without knowledge of the effect of their work on others along the internal and external continuum the organization, or how the data generated by their tasks can be reused in other processes.

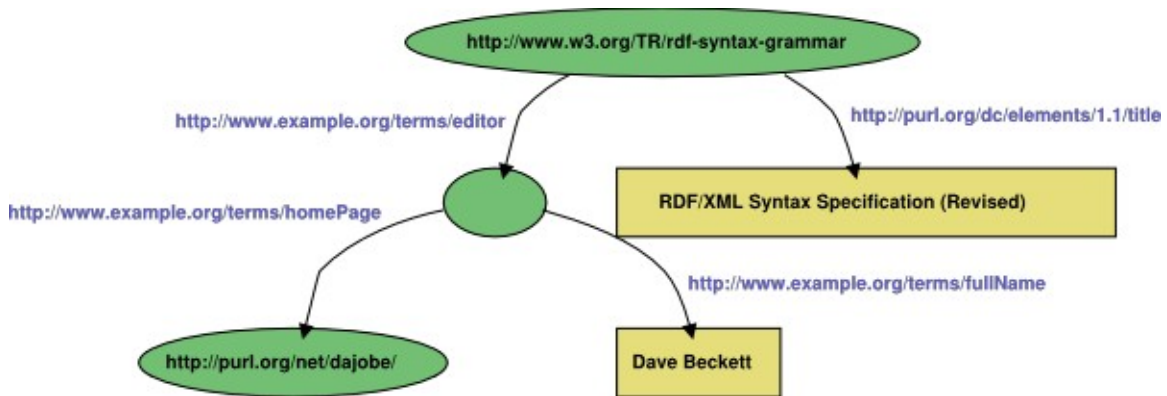


Figure 1. An RDF/XML Graph example (from <http://www.w3.org/TR/rdf-syntax-grammar/>)

Semantics technologies can help the entire organization understand how tasks interrelate and affect each other. The storage mechanism lends itself to producing relationship graphs such as in Figure 1, enabling senior managers and front-line employees to visualize relevant relationships of data elements and concepts. While it is possible to produce these drawings with current technologies, semantic graphs have the advantage of being produced from the underlying relationship data. They are updated instantly with any changes or additions without human effort required.

Conclusion

Semantic technologies are emerging as a powerful computing paradigm for recording relationships between data items and concepts in computer systems. MDM is an appropriate implementation of this technology since these systems record data common to the organization and maintain relationships between data items, systems and users. However, semantic technology is too immature at this time to implement in MDM systems due to lack of commercial tools and integration protocols. But I believe that it provides a more robust and easier-to-maintain solution than is being provided by the major MDM vendors. Your organization can benefit from preparing for the semantic wave by changing existing processes – even if the implementation of a semantic solution is not in the foreseeable future. So armed, you will be well positioned for a transition to a semantic system in the future.

(1) *Jorge Cardoso*, "The Semantic Web Vision: Where are We?" IEEE Intelligent Systems, September/October 2007, pp.22-26,
Seth Grimes, "Semantic Web Visions: A Tale of Two Studies," [Intelligent Enterprise](#) WebLog, October 17, 2007.