

I D C E X E C U T I V E B R I E F

Managing Metadata in the Coherent Information Environment: Essential for Success in the 21st-Century Enterprise

May 2006

Adapted from [Capturing Meaning in Metadata Management Software: Semantic Prerequisite to the Coherent Information Environment](#) by Carl W. Olofson, IDC #33057

Summary

This Executive Brief discusses how large organizations are struggling under the weight of diverse, overlapping, and inconsistent databases that hold their key information assets, and it examines how these enterprises need to rationalize those assets into a computing environment that delivers a consistent means of accessing information in an easy and unambiguous manner. This document explains how metadata repository management and the dynamic IT environment play critical roles in creating the emerging coherent information environment (CIE). Implementing a CIE will facilitate coordinated operations across diverse organizations and thus enable better decision making on the ground.

What Makes Information "Coherent"?

Data is the raw material of IT systems, and information is the output. Data itself is meaningless and therefore worthless without the availability of definitions and the ability to locate and retrieve data in a meaningful format and context. Doing so renders the data as information, which has real value. When information is put to work by information workers, the result is knowledge applied to solve problems, to develop the enterprise's expertise, and to achieve goals effectively and productively.

To reach this result, large enterprises of all types — businesses, governments, educational institutions, and research organizations — are encouraging the development of CIEs. These emerging CIEs are built from IT and allow users to query, search, and otherwise work with information that is consistent, reliably true, and reflective of all the information assets of an enterprise.

Yet these CIEs must be founded on valid and comprehensive data; therefore, interest in data integration and the artifacts of that integration, such as data hubs, operational data stores, and data warehouses, is growing. What's lacking is a semantic system that applies meaning to the data. Without meaning, the data is held hostage to application programs and reports within stovepipes and cannot be read, blended, and interpreted more broadly. Semantic metadata brings to the table the attribute of meaningfulness necessary to relate data as information to content. Thus, such metadata can enable managed data and content integration, query, search, and active use by both people and systems.

The Semantics of Data

Data is a term we apply to streams of values that are delivered to computer programs to drive their operations. Data is normally kept in a structured format, either in flat files or in a managed database. For coherent information environments to produce high-quality information, the data must be made understandable to humans. To do so requires using semantics to interpret the data, which in turn permits the data to yield information that can be combined with information in the form of content. Such semantics may be defined in metadata, which is then stored in metadata repositories for subsequent use in coherent information environments.

Unlike content, which is rendered in human language with its own embedded semantics, data has no semantics associated with it and therefore no automated way to be processed or associated with other data on the basis of meaning. Such semantics must be provided by an additional resource, which is semantic metadata associated with the data and stored in a metadata repository.

When the semantics of data are available, that data can be more reliably transformed and inserted into different management contexts without compromising data quality. This means streamlining such activities as data consolidation, data warehouse generation, peer-to-peer heterogeneous database, synchronization, intelligent data archiving, and multiple database access through a generalized facility (also known as federated or virtual database management).

Once the semantics of data can be applied to handling data, the data can be searched, converted, formatted, and associated with other data more efficiently to produce the information that workers need to do their jobs. To do this, metadata, which defines the data, must be created and then stored in repositories.

Structured data provides few clues to its semantics, which must be captured in the metadata. This metadata not only enables better data integration and query management but can facilitate complex operations that involve a mix of data and content as well. Metadata repositories, intended to provide complete data and data system documentation as well as business system context and data management support, were once consigned to the dustbin of history. Now they may be emerging once again as key enablers of the coherent information environment.

Benefits of a CIE

High-quality metadata enables semantically transparent data. Metadata management repositories store that high-quality metadata, thus ensuring that IT can deliver a coherent information environment. Data semantics are managed by the system in a set of structures designed to ensure that data is accepted and presented to the repository in a manner consistent with its meaning and that rules are followed to enforce such consistency in data processing, storage, and retrieval.

Metadata management repositories capture and apply the semantics of IT system components in general and structured data in particular.

When all these elements are put together, they create the dynamic IT environment required to support the CIE.

Dynamic IT requires on-demand or real-time data access with the ability to deliver just-in-time information that is useful, actionable, and capable of fueling event-driven IT actions. This means putting all the information currently bottled up in relational tables, formatted files, discrete online content documents and systems, and email in service to the organization's mission. Key indicators of interest in dynamic IT development include:

- The growing software trends around data integration (including data migration, data warehousing, data consolidation, and enterprise information integration)
- Information governance and control (including data or information hubs, master data management, and intelligent archiving)
- Ever more sophisticated requirements for online content management (including database searches, intranets, extranets with documents in business file formats, streaming and file-based media such as audio and video, podcasts, and blogs)

What information workers really want is the *Star Trek* computer — that is, the ability to ask a question and get an answer without having to log into several different systems, use multiple applications, or know how to use and manipulate various query tools and syntaxes. They want to use common business terminology to issue commands to the computer and, in turn, for it to perform tasks that are explainable in user (or business) terms.

Many piecemeal attempts have been made to deliver a *Star Trek*-like information service. They have often used some limited metadata to capture data formatting and transformation rules. Most of these solutions involved proprietary technology and, to the extent that metadata is used, an internal or proprietary format. At the same time, online content and structured data have been treated as if they exist in completely separate, parallel universes.

Data is requisite. When its definition is understood or readily available and is presented in a meaningful context, it can yield information. Yet that information may be inappropriate because it was created to drive applications, not necessarily to inform or empower information workers.

Metadata — that is, the data that holds application data definitions as well as their operational and business context — plays the most critical role in data or application design and development. It provides an intelligent operational environment that is driven by business meaning. Only by presenting data together with its definitions in a meaningful context, supporting any random combination of queries and actions and indivisibly coupling it with online content — in short, building the dynamic IT environment — can the *Star Trek*-like computer functionality of a CIE be achieved.

Success Strategies

For the past 25 years, metadata management repository vendors have provided the benefits of coherent, centrally managed data (and later, processes) based on business definitions, values, and priorities, delivering IT systems that are more responsive to the needs of the enterprise, easier and cheaper to maintain and develop, more reliable, and more integrated. Yet, for the past 25 years, they have been largely ignored, even as IT systems have become massively larger, much more fragmented and decentralized, much more expensive to maintain, and frighteningly ill-documented and chaotic. What has happened?

In part, this situation has been the result of shifting IT system architectures and technologies, attempts to decentralize (for expediency and under the false expectation of cost reduction), and the repository product vendors' inability to keep up with market needs. Moreover, the solutions developed thus far have not accelerated the use of development or data analysis tools directly, which has meant that human effort, arcane disciplines, and the possibility of error have remained as constants.

One key strategy required for success is to make the software do the system administration work, relieving information workers of the onerous task of learning more rules regarding data, process, definition, and usage. Another strategy is to debunk the notion that metadata management repositories offer no measurable return on investment (ROI). This can be accomplished by assessing:

- The cost savings realized from more efficient application and data development and management
- The reduced risk accrued from better-quality data for applications and warehouses, and concomitantly higher-quality information for information workers

Considerations

Metadata management technology is nearly complete with respect to the needs of organizations for understanding the data managed in the enterprise — what it means, how it relates to business processes and to the various components of even very complex IT environments, how to transform it, how to ensure its quality, and how to evolve it. Three challenge areas remain for metadata repository solution providers:

- Embrace online, unstructured content and blend it with structured data, including information about such things as its semantics.
- Support integration based on semantic affinities rather than simple matches. Most IT environments are very heterogeneous, and acknowledging that one central metadata repository is impractical means the various repositories need to function cooperatively.
- Provide more in the way of runtime services. Enable semantic mediation for Web services and dynamic query or search resolution based on business terms, and provide a business context for system monitoring and notification services.

Conclusion

The newest generation of metadata management software has made it much less painful and far more practical to integrate repositories that already conform to standards with the most popular modeling, development, and reporting and analysis tools. To ensure success, solution providers should consider the following:

- Make the tool integrations seamless so that tool users automatically find the contextual definitions, and ensure that their work automatically becomes available for others.
- Make the repository invisible so that information workers really don't know it's there.
- Guarantee that best practices for ensuring data consistency, based on its semantics, are built into the environment; as users create definitions, they should quickly discover if a definition for the data, process, service, business function, or whatever, already exists.
- Create an "aha!" moment for information workers so they can discover with unexpected joy that their jobs have been made much easier.
- Partner! The metadata management repository is a component in the system, and an invisible one at that. It must be implemented as part of a solution with the software platform, data integration software and systems management software vendors, and systems integrators.

Achieving these levels of abstraction and information transparency in metadata management software can truly enable IT departments to finally deliver the *Star Trek* computer — the one with a coherent information environment.

COPYRIGHT NOTICE

The analyst opinion, analysis, and research results presented in this IDC Executive Brief are drawn directly from the more detailed studies published in IDC Continuous Intelligence Services. Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. Contact IDC Go-to-Market Services at gms@idc.com or the GMS information line at 508-988-7610 to request permission to quote or source IDC or for more information on IDC Executive Briefs. Visit www.idc.com to learn more about IDC subscription and consulting services or www.idc.com/gms to learn more about IDC Go-to-Market Services.

Copyright 2006 IDC. Reproduction is forbidden unless authorized.