

October 2005

<http://www.intelligententerprise.com/showArticle.jhtml?articleID=171000640>

## **Start Making Sense: Get From Data To Semantic Integration**

**Current integration methods including ETL, EAI and EII can't deliver on the goal of creating a single view of the truth. In fact, composite apps and service-based architectures will only underscore the limitations of current practices. Semantic integration promises a richer understanding of metadata and clearer context, but to embrace this emerging approach, we'll have to revise some cherished data warehousing concepts.**

By Neil Raden

Business intelligence (BI) and data warehousing professionals: Your comfort zone is shrinking. Dramatic changes lie ahead. Conventional applications are giving way to standards-based Web services. The lines between operational and analytic systems are blurring. And as organizations try to make sense out of what's contained in their myriad data stores, standard data integration and presentation techniques are proving costly and difficult.

The bottom line is that current practices aren't adequate. Business agility depends on the ability to assemble, disassemble and rearrange application components. These actions require a comprehensive understanding of not only data representation ("syntax"), but also data's meaning and its relationships to other data and information — that is, the "semantics." In this article, we'll look at how we're moving from data integration toward higher-level semantic integration.

## From ETL to Chaos

Over the past 15 years, we've seen a sequence of integration technologies and methodologies emerge, flourish and hang on. First, extract, transform and load (ETL) delivered data integration and movement for data warehouses. Then enterprise application integration (EAI), along with message-oriented middleware, opened the door to business-to-business Web commerce. Now there's a surge of interest in enterprise information integration (EII) among those looking to do real-time operational reporting and other time-sensitive activities. By supporting the delivery of queries to the data sources rather than waiting on ETL and data movement steps to get the data into the warehouse, EII addresses a weakness of conventional data warehousing when it comes to real-time objectives.

While each approach has its attributes, this collection of technologies can't be amalgamated to provide the level of information integration most organizations need. Analysts and vendors frequently suggest that ETL, data warehouse, EAI, EII and other integration tools are complementary. In other words, even if you apply them separately, you'll ultimately arrive at a complete solution. They're mistaken.

Each technology crevice between the different tools requires a separate modeling and mapping effort, leaving organizations with multiple models. ETL demands a target database schema. EAI requires agreement on a canonical form among the applications. And with the less mature EII, technical demands vary from vendor to vendor — from a simple set of views to a full model developed in Unified Modeling Language (UML). From a management perspective, each integration solution tends to operate from within a different technology stack and, therefore, carries unique design constraints, tuning characteristics and vendor upgrade cycles.

Finally, integration tools generally don't expose their metamodels or use other means of communicating. Certainly, none integrates data beyond the syntactic level. The tools manage semantics — information that focuses on conveying the *meaning* of data and information

## Executive Summary

When you can't effectively integrate information, valuable opportunities are lost — and fraud, regulatory or security threats go unrecognized. The right information in the right hands at the right time can lead to the right action by the right people at the right moment.

Unfortunately, current techniques limit the breadth and depth of integration. Valuable data and content lie buried in technology silos that are difficult to bridge.

For enterprise content integration (ECI), potential solutions include choices that go beyond mere connectivity to support sophisticated indexing, search and virtual repositories. XML, JSR 170 and other standards figure prominently. ECI implementation is still at the early adopter stage; vendors are jockeying for future dominance as organizations place increasing priority on cross-functional business processes and real-time information delivery.

Integration at a higher level — at something approaching natural language — holds great promise, especially as "Semantic Web" standards evolve. Semantic integration is a key strategic direction for both structured data and unstructured content integration because it focuses on meaning: that is, how pieces of information relate to each other. This is significant not only to human use of information but also to computer-based processes, applications and Web services that must respond in real time to customer activity and other business events.

## Debriefing

