

The Role of Semantic Metadata in Improving Information

ABSTRACT

An important measure of the quality of information is how well its users understand the information. Machine-readable metadata has traditionally been an important contributor to understanding structured data. It has also been a key to the application of model-driven techniques to data integration over the past fifteen years, enabling the current crop of graphical data mapping tools. These techniques and tools have been responsible for some progress in making integration of structured data less costly and error-prone.

However, integration costs are still too high, consuming a large percentage of IT budgets. Meanwhile, macroeconomic forces and technical realities are relentlessly increasing the demand for data integration. The bottleneck centers on the difficulty integration analysts face when they try to understand the meaning – that is, the semantics – of the data they are integrating.

Mainstream data and message format definitions contain only one kind of machine-readable metadata, namely metadata about the syntax of the information. Machine-readable syntactic metadata makes it possible for current state-of-the-art, model-driven data integration tools to generate data transformation code, once the integration analyst has entered a mapping from one format to another into the tool. However, this metadata does not provide integration tools with a basis to help integration analysts decide *what* the mapping between two formats should be. This is a key reason that integration work is still so labor intensive. Moreover, misunderstandings about the semantics of data and message definitions lead integration analysts to make subtle errors that can be costly and difficult to track down. Machine-readable semantic metadata helps to address these limitations to a modest yet useful degree.

This presentation will explain the nature and structure of machine-readable semantic metadata that is now coming on line in standards for the manufacturing, retail, consumer products, and financial services industries and is being increasingly used by ERP vendors, opening up possibilities for a new generation of tools that make integration analysts substantially more productive and accurate. It will describe how this new kind of metadata improves information quality by making data semantics more transparent to both humans and machines. It will also discuss how semantic metadata can be applied to unstructured information.

BIOGRAPHY

David S. Frankel

Lead Standards Architect - Model Driven Systems
Technology Strategy Group
SAP Labs, LLC



David is a member of SAP's Technology Strategy group. He focuses on standards strategy for model-driven systems and semantic metadata, with an orientation to ERP financials and the banking sector.

David has over 30 years of experience as a programmer, architect, and technical strategist. He is recognized as an authority on the subject of model-driven systems and semantic metadata. He has published two books and dozens of trade press articles, and has co-authored a number of industry standards, including UML and the ISO 20022 core methodology for financial networks.