

AnalyticsWeek article – Enterprise Data Modeling Made Easy

Enterprise data modeling has remained an arduous, time-consuming task for myriad reasons, not the least of which is the different levels of modeling required across an organization's various business domains.

Data modelers have to consider conceptual, logical and physical models, in addition to those for individual databases, applications, and a variety of environments such as production and post-production. Oftentimes, the need to integrate new sources or to adapt to changing business or technology requirements exacerbates this process, causing numerous aspects of it to essentially begin all over again.

Enterprise data modeling is rendered much more simply with the incorporation of semantic technologies—particularly when compared to traditional relational ones. Nearly all of the foregoing modeling layers are simplified into an evolving semantic model that utilizes a standards-based approach to harmonize modeling concerns across an organization, its domains, and data environments.

Moreover, the semantic approach incorporates visual aspects that allows modelers to discern relationships between objects and readily identify them with a degree of precision that would require long periods of time with relational technologies.

“Semantics are designed for sharing data,” **Franz** CEO Jans Aasman reflected. “Semantic data flows into how people think.”

Between Objects

The crux of the semantic approach to data modeling is in the

technology's ability to define relationships between data and their different elements. In a standards-based environment, data objects are given specific descriptions courtesy of triples that are immensely useful in the modeling process. "The most important thing in semantic modeling is that everything is done completely declaratively," Aasman revealed. "So instead of thinking about *how* you do things, you think about *what* you have." The self-describing nature of triples is integral to semantic models because it allows those models to determine the relationships between different data elements. "You are very explicit about the relationships between objects in your data, so semantic modeling is far more like object-oriented modeling than relational database modeling," Aasman said. Those relationships, which are easily visualized in an RDF graph, function as the building blocks of semantic models. Additionally, there are no schema limitations in a standards-based environment, which saves time and effort when modeling across applications, domains, or settings—which is required for enterprise data modeling.

Adjusting Models

The issue of schema is critical to conventional data modeling, particularly when incorporating additional requirements or new data types and sources. Semantic models are based on standards that any type of data can adhere to, so that there is "a standardized semantic model across many different data sources," **Paxata** Chief Product Officer Nenshad Bardoliwalla said. All data can conform to the conventions of semantic models. Thus, when updating those models with additional requirements, there are fewer steps that data modelers have to go through. In relational environments, if one wants to incorporate a new data source into a data model of three other data sources, one would have to make adjustments to all of the databases to account for the new data types. Frequently, that re-calibration pertains to schema. In a standards-based

environment, one would simply have to alter the new data source to get it to conform to the semantic model—which saves time and energy while expediting time to action. “Because we don’t have to redefine the schema all the time, it’s easier to use semantic technology,” Aasman remarked. “But it’s not impossible in a relational world.”

The Importance of Vocabularies

According to **TopQuadrant** CTO Ralph Hodgson, the precision of expression in semantic models—which allows organizations to model aspects of regulatory requirements and other governance necessities—is possible because the semantic model is “the model that is the most expressive thing that we have today. You don’t do that with an object model, you don’t do it in UML, you don’t do it with an entity relationship model, you don’t do it in a spreadsheet. You do it with a formalism that allows you to express a rich set of relationships between things.”

Nonetheless, enterprise data modeling is abetted in a semantic environment with vocabularies and systems for unifying terms and definitions throughout the enterprise. These technologies assist with the modeling process by ensuring clarity among all of the terms that actually mean the same things, yet are expressed differently (such as spellings, subsidiaries of companies, names, etc.). The result is that “you’re using the same word for the same thing,” Aasman maintained.

One can attempt to model most facets of terminology and their meanings. However, there are specific semantic technologies that address these points of distinction and commonality much faster to actually aid existent semantic models and ensure points of clarity between different data types, sources, and other characteristics of enterprise data modeling. “When we talk about how do you actually link and contextualize data and develop a data lake and its relationships, those taxonomies and vocabularies are actually central to being able to do that effectively,” **Cambridge Semantics** VP of Solutions and Pre-

Sales Ben Szekely observed.

Modeling Enterprise Data

Perhaps the easiest way to facilitate enterprise data modeling is with the incorporation of an organization's entire data into an RDF graph. Smart data lakes provide this capability in which all data assets are linked together in a graph with a comprehensive semantic model that quickly adds new sources and requirements. "The ability to do semantics is essentially the ability to create an enterprise graph of the entire enterprise and its information," Cambridge Semantics VP of Marketing John Rueter said. "Up until now it's been done at a departmental level." Facets of regulatory compliance, data governance, and other organizational particulars can all coalesce into such an inclusive model, which provides a monolithic framework for the fragmented concerns of the different layers of modeling that have traditionally monopolized the time of data modelers. In these instances, the majority of the preparation work for modeling is done upfront and simply requires that additions conform to ontological model requirements.

Enhanced Analytics

Enterprise data modeling is considerably simplified with smart data approaches. Modelers can largely account for all of the disparate layers of modeling in a single semantic model that is supported by requisite vocabularies and terminology definitions. Furthermore, that model is based on standards that allow additional sources or requirements to mesh with it by adhering to those standards. Improvements in analytics and data discovery are just some of the many benefits of this approach, which saves substantial time, effort, and cost. "You can ask the data what's the relationship between things, rather than making guesses and asking the data is your guess correct," Cambridge Semantics VP of Engineering Barry Zane commented.

When one considers that such data can encompass all enterprise information assets, the potential impact of such insight—both

for data modelers that facilitated it and for business users that perform better with it—is nothing short of transformative. “People in the world of semantics make sure that their models are entirely self descriptive and self explanatory,” Aasman stated. “There’s far more emphasis on being very clear about the relationships between types of objects.”