Establishing an Enterprise Vocabulary Management Solution in a Fast Growing Publishing Firm using Topbraid Suite™

Solution Context and Requirements:
Enterprises use explicit enterprise vocabularies to:
- Enhance findability of data and information assets
- Communicate and integrate data across the supply chain
- Standardize publication of company product information to external partners and customers.

In today’s dynamic business environment, enterprises are feeling a tension between the desire to have a single, centralized, master vocabulary system and the current reality that vocabularies are managed in a disconnected, distributed manner. The most common storage choice for business vocabularies is spreadsheets, trapped and hidden on user desktops. In addition, some enterprise systems include their own vocabulary capabilities, but use of them simply results in a proliferation of disconnected vocabularies, subverting the value of a controlled vocabulary in the first place.

This tension has driven many organizations toward an Enterprise Vocabulary Management solution, where controlled vocabulary is an integral part of enterprise-level information management. Rather than focusing on particular capabilities of vocabulary tools, Enterprise Vocabulary Management focuses on the modularity and connectivity of vocabulary assets.

Challenges:
- Vocabularies from different systems are available in contrasting and often idiosyncratic forms.
- Different groups don’t communicate while developing taxonomies, resulting in overlapping coverage and differences in representation.
- Terms from one vocabulary often need to link to another. How does my notion of “Customer” relate to yours?
- External vocabulary resources are valuable, but typically are not designed in a way that is compatible with enterprise needs.
- Spreadsheets are a common tool but have a number of technical limitations including size, structure, and linkage.
- Taxonomy tools are often built into other apps (CRM, BPM, SOA, etc.), with limited and disconnected functionality.
- Vocabulary Management is treated as an isolated activity, though it has impact throughout the enterprise.

Corporate case study: NichePub, a growing niche publishing company sees addressing these challenges as essential to their survival. As a result of a series of mergers and acquisitions, they own several industry-established vocabularies, each one being managed in its own software system. They use these vocabularies to tag their publications, providing added value over their competitors in the ease with which their materials can be accessed and used. They also recognize the importance of public vocabulary sources which they would like to integrate with their own vocabularies. NichePub elects to investigate capabilities of semantic web solutions.

Semantic Web Technology Capabilities / Benefits for Vocabulary Management:
Enterprise Vocabulary Management has emerged as an essential facet in effective business information management. While creating centralized reference vocabulary and taxonomy repositories is one common approach, often the governance model is complex and unsustainable. To support the business needs of complex, distributed enterprises, Enterprise Vocabulary Management practices and enabling technology/tools are evolving from Reference-only (text, models) to Interoperable (standards based) to Executable (federated, model-based, semantic-enabled). Semantic web technology makes it easier to aggregate and analyze vocabularies through its capabilities for expressing, querying and federating information.

Commitment to the Semantic Web standards as the basis of the enterprise information infrastructure brings the following benefits:
The Semantic Web standards Resource Description Format (RDF) and RDF Schema provide the infrastructure for creating a web of information, whether on the public internet or within enterprise intranets. Like the familiar World Wide Web, the Semantic Web is easily extensible, breaking out of information silos once and for all, as vocabularies are available as resources on the web for anyone to reference and use. Each vocabulary element has a Uniform Resource Identifier (URI) which uniquely identifies it as a web resource.

The Semantic Web standard Web Ontology Language (OWL) provides a logical language for describing semantic relationships between data sources. Connections that previously had to be accomplished with custom software can now be described in a standard, declarative way. OWL also allows for creation of relationships that are more flexible than current thesaurus standards.

Enterprise Vocabulary Management: Corporate case study (continued)

As outlined, the NichePub company has established itself as the premier reference for industry-specific tagging of public documents. Millions of documents have been tagged with terminology that they maintain. Due to recent M&A, they own many of the frequently-used vocabularies in the industry.

But all is not well. Management of these vocabularies is the bread-and-butter of their business, but they are bogged down by the plethora of software systems used to manage them. Organizationally, they have to train personnel on several systems. Operationally, they have to maintain several vocabularies with different ways to control versions, update modes, permission models, and so on.

In terms of the content, the models have overlapping coverage; how can changes in one vocabulary be reflected in another? How can they even record that a term in one vocabulary has the same meaning as a term in another? Finally, they recognize the value of some external vocabularies that are maintained in the public domain. How will these external resources relate to their own vocabularies?

It isn't possible organizationally or industrially to come up with a single, master vocabulary that will serve all needs. Too much of the industry is entrenched in each vocabulary component. NichePub chooses a solution that uses Semantic Web standards and the capabilities of TopBraid Suite to rescue vocabularies from their idiosyncratic source forms, transforming each into a linked data set on the Semantic Web. This Vocabulary Web resolves their vocabulary management problems. While there is no single master vocabulary, relations between terminology elements in various vocabulary components are recorded just as hyperlinks in the World Wide Web. Changes in terms are available immediately to any component that references them. External vocabulary components are integrated in the same way as internal ones. The semantic web structure provides a uniform, connectible but distributable, representation of all vocabulary assets.

Benefits of Enterprise Vocabulary Management Solutions built with TopBraid Suite:

TopBraid Suite is an integrated, complete lifecycle semantic application development product suite that allows users to design, assemble, deploy, and collaborate within a new generation of dynamic business applications. Building on the capabilities and benefits that come from Semantic Web standards, it provides out-of-the-box support for the following Five Key Capabilities of an Enterprise Vocabulary Management solution:

- **Vocabulary Processing:** Standard hierarchical, associative and equivalency relationships; repositioning and numbering of terms; crosswalk mapping and graph capabilities.
- **Automatic processing:** Ability to create validation rules and automated script processing via SPIN and SPARQLMotion.
- **Import/Export:** Import/Export from RDBMS, RDF Store, SPARQL endpoints, spreadsheets (CSV), XML, RDF and OWL.
- **Merging:** RDF standard universal identifiers provide easy "hooks" for merging vocabularies.
- **Systems Integration:** Integrate with existing enterprise or vocabulary management systems via API.

With these capabilities, TopBraid Suite provides flexible, customizable solutions for managing taxonomies and business vocabularies in support of content delivery, search, navigation, data integration and disambiguation.