

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, where they typically remain 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, and so on), with limited and disconnected functionality.
- Vocabulary Management is treated as an isolated activity, though it has impact throughout the enterprise.

Corporate case study

AtlasPub, a fast-growing publishing company sees addressing these challenges as essential to their success. 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. Vocabularies are used to tag publications, providing added value over AtlasPub's competitors in the ease with which materials can be accessed. Recognizing the importance of public vocabulary sources AtlasPub wants to integrate them with their own vocabularies. Their current home grown tools and information representation don't offer the flexibility AtlasPub needs. They don't support connecting and reusing terms across vocabularies or allow customizability for the different lines of business. AtlasPub elects to investigate capabilities of Semantic Web solutions.

Semantic Web Technology Capabilities / Benefits for Vocabulary Management

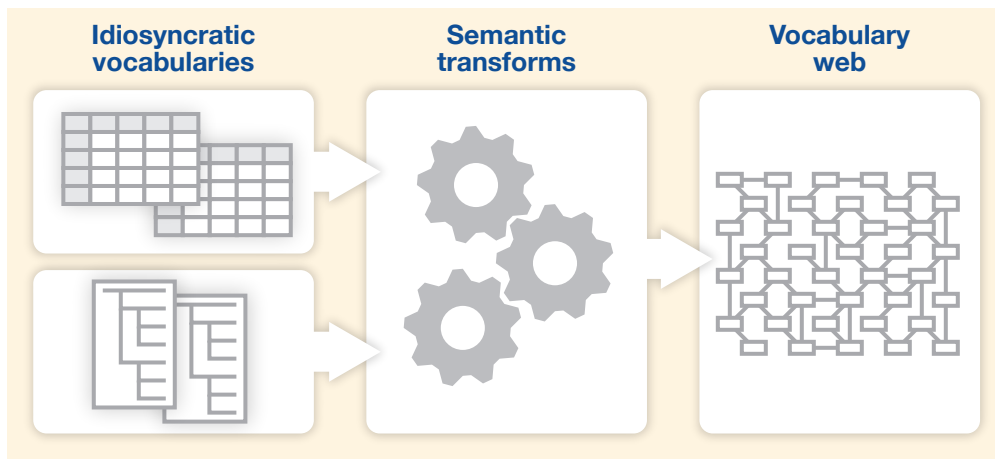
Enterprise Vocabulary Management has emerged as an essential facet in effective business information management. While creating centralized reference vocabulary and taxonomy repositories is one common approach, often the governance model is complex and unsustainable. To support the business needs of 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.



TopQuadrant is a premier semantic web solutions company with offices in Alexandria, VA; at NASA Research Park, CA, and an affiliate in Seoul, South Korea. Please contact us at:

TopQuadrant, Inc.
330 John Carlyle Street
Suite 180
Alexandria, VA 22314-5760

www.topquadrant.com
Phone: 703 299 9330
Fax: 703 299 8330



In terms of the content, the models have overlapping coverage; AtlasPub wants to ensure that changes in one vocabulary are reflected in another. This requires support for reuse and cross-reference across vocabularies. As they recognize the value of some external vocabularies that are maintained in the public domain, they need to be able to relate the external resources to their own vocabularies.

Use of the Semantic Web standards for 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 element has a Uniform Resource Identifier (URI) which uniquely identifies it.
- The Semantic Web standard Simple Knowledge Organization System (SKOS) provides an extensible language for describing structured terminology systems like taxonomies and thesauri. Based on RDF, it inherits the power of flexibility and distribution. SKOS extends RDF with specific notions of linkage between structured vocabularies, making it particularly useful for enterprise vocabulary management.
- The Semantic Web standard query language SPARQL provides much more than a query facility. In the context of SPARQLMotion and SPIN, it provides a powerful mechanism for modeling complex relationships between terminology. Connections that previously had to be accomplished with custom software can now be described in a standard, declarative way.

As outlined, the AtlasPub company has established itself as the premier reference for tagging of public documents specific to the industry sector it serves. 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.

Management of these vocabularies is the bread-and-butter of their business, but they were bogged down by the plethora of software systems used to manage them. Organizationally, they had to train personnel on several systems. Operationally, they had to maintain several vocabularies with different ways to control versions, update modes, permission models, and so on.

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 terminology. AtlasPub chose TopBraid Suite from TopQuadrant because it allows them to rescue vocabularies from their idiosyncratic source forms, transforming each into a linked data set on the Semantic 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. Customizability of TopBraid Suite means that AtlasPub can easily implement support for the different workflows and business rules required by each of their departments.

Benefits of Enterprise Vocabulary Management Solutions built with TopBraid Suite™

TopBraid Suite is an integrated, 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 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:** Direct integration with existing enterprise or vocabulary management systems via API.

Based on the TopBraid Suite platform, TopQuadrant has developed an Enterprise Vocabulary Management solution that offers specialized, yet extensible models for taxonomy development, governance and version control. Also provided are customizable rule and web application templates for common vocabulary development activities

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.