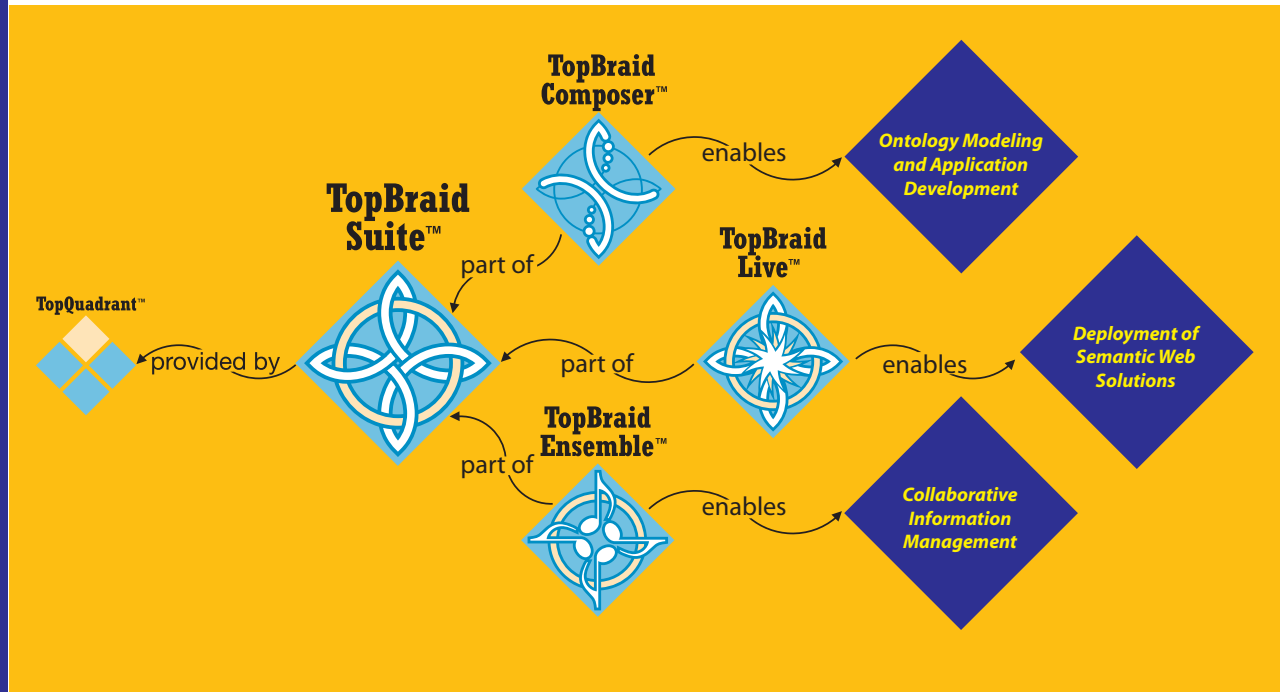




## Semantic Web Solutions at Work in the Enterprise



### ABOUT TOPQUADRANT

Established in 2001 with the mission to bridge the gap between business collaboration needs and enabling technology through semantic products and services, TopQuadrant offers the industry's leading platform for semantic-enabled applications — TopBraid Suite. Over 300 customers worldwide use TopQuadrant solutions to enable systems and people to fuse relevant information from diverse sources, put knowledge into context, collaborate effectively, and make better decisions. *Visit us at [www.topquadrant.com](http://www.topquadrant.com) or inquire at [info@topquadrant.com](mailto:info@topquadrant.com).*



# Semantic Web Solutions at Work in the Enterprise

- **INTRODUCTION AND OVERVIEW**

**Information Overload — An Overriding Business Challenge**

**Semantic Web Technology — A Foundation for  
Restoring Information Relevance**

**The Role of TopBraid Suite™ —  
To Support Information-Integration-Intelligence Solutions**

- **KEY BUSINESS REQUIREMENTS FOR AN  
INFORMATION-INTEGRATION-INTELLIGENCE SOLUTION**

- **SOLUTION COMPONENTS**

**Standards based data representation**

**Business Ontologies**

**TopBraid Suite Capabilities**

- **TOPBRAID SUITE IN USE —  
SOME SAMPLE CUSTOMER APPLICATIONS**


- **CONCLUSION**

- **MORE DETAILS — COMPARING TRADITIONAL DATA  
REPRESENTATIONS WITH SEMANTIC STANDARDS**

- **ENDNOTES**

# Introduction and Overview

## Information Overload — An Overriding Business Challenge



The amount of digitized information is growing at unprecedented rate. By 2007 the size of individual databases at many organizations reached up to hundreds and in some cases thousands of terabytes. For example, in 2004 AT&T had 11 exabytes (10<sup>7</sup> TB) of wireline, wireless and Internet data. This is an equivalent amount of data to that held by 1 million Libraries of Congress. Wal-Mart had 500 terabytes of transactional data and was adding 10<sup>7</sup> transactions per day. On average, the size of transactional databases doubles every five years with core databases doubling every two years. Data reporting and analysis warehouses (OLAP stores) triple in size every three years. On the web, by 2007 there were 29.7 billion pages, roughly five pages for every man, woman, and child on the planet. In 2006 alone, the size of the information created or replicated worldwide was 161 exabytes (10<sup>8</sup> TB)<sup>1</sup>.

**We are inundated with information.** Billions of pages and exabytes of data sound like an overwhelming proposition, but, unlike people, computers are designed to process large amounts of data. And they grow more powerful and less expensive every day. Shouldn't then this ongoing information explosion be welcome news? Arguably, the more information that exists in digital form, the better our understanding can become about our environment, our customers, and our business. To some extent this is happening. **More people can get more information more quickly than ever before.** But, increasingly, it feels like we are drowning rather than swimming in information. Why? There are many reasons, but here are a few of the most critical:

- 1. Information is fragmented across many sources** — as a result, we, the end users, constantly find ourselves collecting and manually integrating information from an increasing number of sources in order to have a complete view
- 2. Aggregating related and relevant information is challenging** — with the expanding amount of information, there's also an increasing number of contexts in which it has been collected and a corresponding growing number of different vocabularies used to describe it.
- 3. The quality, relevance and freshness of information is often unclear** — with the increasing number of potentially pertinent data sources it is becoming more time consuming to screen out outdated, irrelevant or conflicting information from different sources

A common solution in the enterprise is to create data reporting and analysis warehouses carefully designed to accommodate information in the identified data sources and the type of queries the information user may need. The processes for design and loading of warehouses and data marts are not intended to accommodate rapid change. As a result, they tend to be fairly static and do not meet the requirements of today's dynamic environments where the number of data sources and information they contain constantly grows and the types of queries desired by the users change. So, just as organizations find themselves with an expanding number of silo transactional sources, increasingly they also find themselves with silo data warehouses.

## Semantic Web Technology — A Foundation for Restoring Information Relevance

Computers should be able to carry the burden of aggregating information in a coherent, consistent and meaningful way to enable the users of information to query and explore cohesive bodies of data in order to gain insights within the context of decisions we must make. To do this, computers need to know the structure of the different data sources, how the data they contain is represented and what it is about. ***This information, often called metadata or schema, is available in current IT systems and data stores. It is expressed in a variety of different ways, including:***

- Design, or schema, of relational databases
- Structure of XML documents
- Structure of spreadsheets and various files
- Content of special tables and files expressing local or global standards for referring to commonly used entities such as geographical places, product names and lines of business<sup>2</sup>
- Tags describing unstructured information including user entered tags and the key words generated by search engines

In order for computers to integrate data from different sources in a consistent way, they must be able to process, integrate and correlate the respective metadata.

Until a few years ago, there were no standards for describing and inter-relating metadata. The World Wide Web Consortium (W3C)<sup>3</sup>, the body that develops and governs standards that made it possible to interlink documents on the Web, has developed a simple yet robust set of standards for describing and connecting metadata and aggregating data. The standard development was driven by a vision for connecting a web of data, much as documents (web pages) have been connected on the Internet and intranets. The vision is referred to as a Semantic Web or a Data Web (sometimes even Web 3.0); the technologies that implement the standards are called Semantic Web technologies.

Semantic Web technologies and products based on them have matured in recent years and are successfully working today in government, financial services, healthcare, life sciences and other industries. ***By making it possible to automate the correlation of disparate information specific to the enterprise, solutions based on Semantic Web standards are creating for their users the next meaningful competitive advantage from technology.*** They leverage IT investments in existing systems and data to provide context specific views on aggregated information. They integrate the data while being able to preserve as needed the distinct, context-specific vocabularies, meaning and intent of the underlying data sources. They do all of this in a dynamic and agile way enabling actionable insights and discoveries that are possible only when the right information is brought together in the right way to the right users as and when needed in a cost effective way.

## The Role of TopBraid Suite™ — To Support Information-Integration-Intelligence Solutions

A wide range of valuable applications and solutions can be created using Semantic Web standards, technology and products to resolve critical business problems caused by information explosion and fragmentation. These solutions will obviously vary in details and complexity, but nearly all share a common solution pattern and core value proposition captured by the simple word sequence: *Information-Integration-Intelligence*. The meaning is straightforward and connotes the activity of addressing information overload and fragmentation through effective integration capabilities to produce results that deliver intelligence to the information consumers. For simplicity, we refer to the fundamental Information-Integration-Intelligence pattern as the **I-I-I** solution.

TopBraid Suite (TBS) exploits the W3C Semantic Web standards to provide an open, flexible, configurable and standards-based platform with multiple integrated capabilities to rapidly construct **I-I-I** solutions and to produce other semantic-enabled business capabilities. The TopBraid Suite of products natively implement and take advantage of the W3C standards to help enterprises put control and decision power for information integration and intelligence into the hands of the people who need it most — business users.

# Key Business Requirements for an Information-Integration-Intelligence Solution

Whether it is an aerospace engineer working to determine the cause or impact of a system failure, a scientist researching the outbreak of an epidemic, financial services professionals making investments or policy decisions, or a sales and marketing executive wanting to understand all the interactions his company has had with customers or prospects in a given sector — these seemingly unrelated professionals performing different tasks, have one key thing in common. To make the right decisions these days, they nearly always need to pull together relevant information from different sources.

These sources may include corporate and departmental systems. Often they also include external sources ranging from pay-for data feeds and information and/or data feeds and databases from partnering organizations to freely available information from numerous sources on the World Wide Web. Each year business users become more proficient and comfortable with the use of technology. More and more people build their own personal systems and tools. These commonly include spreadsheets, personal contact management systems, reference files and materials. Bringing together and correlating this variety of information in the context of a decision becomes time consuming, error prone and is sometimes, not a practically doable chore. Yet the quality of the business decision and actions directly depends on how effectively this task is performed.

***An I-I-I solution as implemented using TopBraid Suite puts business users in control of both information integration and use. The solution addresses the following key requirements:***

1. Support for standards based data representation
2. Access to data from existing internal and external sources as needed and when needed
3. Aggregation of information based on the common and related aspects found in the sources
4. Ability to be driven by the context and frame of reference as defined by the information user
5. Adherence to company policies
6. Continuous enrichment of the data usefulness for the next user

Note that the last five requirements focus on the needs of the information user and the enterprise he works for. The first requirement, on the other hand, is technology centric. It is an enabling foundation for making possible the five business requirements. In the next section of the paper, we focus initially on examining this first requirement in more detail. We outline advantages of the *standards based data representation* as well as provide a simple example of a *business ontology* and how it may be used in an application. We then describe the three integrated products within *TopBraid Suite* in terms of their capabilities and use for building and deploying **I-I-I** solutions.

# Solution Components

## Standards based data representation

All information integration products use information models describing how sources being integrated relate to one other. Some approaches implement such models in an ad hoc way, for example, by embedding business concepts and connections between them in the data transformation scripts. Others support a more scalable integration approach by building canonical models, often in the form of *business ontologies*, enterprise object models or unified information models. These types of models describe concepts important to the enterprise and relationships between them. They can represent all information that is being captured about customers, products and enterprise activities. The data in the individual data sources is then described by connecting them to the canonical model to indicate and map the information they contain.

**Prior to the development of the Semantic Web standards, there was no standard language to represent business ontologies.** Each vendor had to build their own proprietary approaches for representing this information. Therefore, an organization would spend considerable time and resources expressing business knowledge pertaining to the area being integrated, only to find it locked within a proprietary solution. Often different products and vendors are being used to integrate different aspects of the enterprise information. Each integration solution created using the proprietary technology becomes yet another silo of data that is expensive and time consuming to integrate across the isolated silos. Integration using Semantic Web standards has key advantages for the enterprise:

- Business ontologies represented in a standard compliant way can be used across products and applications similar to how HTML is usable across browsers, HTML editors and other tools
- Semantic Web standards have been developed from the ground up to support re-use and extensibility. This is a necessity not only in the World Wide Web, but also in the enterprise environment. Modular models can import and extend each other so that two departments can share common company-wide concept definitions, while each preserving their unique business differences. This enables a practical, incremental *distributed but connectable strategy* for developing solutions.
- With growing industry adoption, using standards to represent enterprise models ensures that the work will be future-proof.

TopBraid Suite fully supports W3C standards for representing and connecting information.<sup>4</sup> TopQuadrant, as an active member of W3C, directly contributes to the development and evolution of standards.

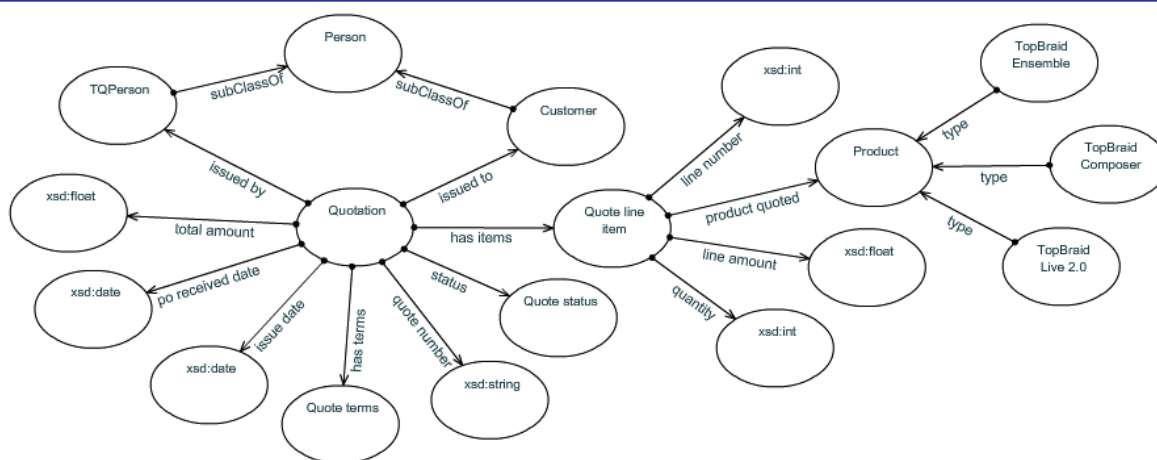
## Business Ontologies

Throughout this paper we are using the words 'ontology' and 'information or data model' interchangeably. The term data model is well known and understood by information technology professionals. The word 'ontology' is a newer term in computer science. It originated in philosophy as a way to talk about the concepts and relationships in some area or domain. For example, when we define the concepts and relationships needed to describe customers of a company and their interactions with a company, we are developing an ontology (or data model) needed to support the customer relationship management (CRM) domain. As part of the work, we may find it necessary to describe a company's organization structure. This brings us into the organization domain. Thus, an ontology built for CRM purposes may be re-usable in whole or in part in other areas — human resources, supplier management, etc.

For ontologies to be accessible to computers they need to be defined in a computer processable language. The W3C standard languages for ontologies are called RDF (Resource Description Framework), RDF Schema<sup>5</sup> and OWL<sup>6</sup> (Web Ontology Language). RDF/OWL ontologies can include:

- **Individuals** — individual data objects such as “TopQuadrant” or “California”
- **Classes** — these are sets of individuals that have some common characteristics such as “Company”, “State”
- **Properties** — these are characteristics of the data objects including data attributes such as “age” or “price” and relationships<sup>7</sup> that connect objects such as “works for” and “created by”

Any individual, class or property included in the ontology is called a ‘resource’. Each resource has a globally unique identifier, so that it can be uniquely referenced. Following standard practice on the Web, resource identifiers are URIs<sup>8</sup>. To illustrate the above components of a business ontology, the diagram below, generated using a TopBraid Suite visualization component, shows a fragment of an ontology TopQuadrant uses to generate product quotes for our customers.



*TQPerson*, *Person*, *Customer*, *Quotation*, *QuoteStatus*, *QuoteTerms*, *QuoteLineItem* and *Product*<sup>9</sup> are classes. Members of these classes are the individual data items such as specific products, people and quotes. *TopBraidEnsemble*, *TopBraidComposer* and *TopBraidLive\_2.0* are members of the *Product* class. The ontology describes, for each class, properties of its members that are of interest in the context of ontology use. In the context of quote generation we are interested in capturing information about product prices and version numbers, but we are not interested in the platform requirements for each product.

The Semantic Web languages, RDF, RDFS and OWL, have been created with distribution and aggregation in mind. This means that not only data, but the schema or model can be modular and distributed in its representation. A different ontology may not be concerned with the product prices, but, instead, be focused on the platform requirements or tools and languages used to create software products. Its definition of what a product is or what information concerning the product can exist would be different from the ontology shown above. Both are valid views and represent valid sets of information. Semantic Web standards allow us to easily combine and aggregate these different views — as and when needed. This is supported through the fundamental merge capability of RDF. Ontologies can import one other. A rich modeling vocabulary exists for connecting and relating resources within and between models. It includes statements like *subClassOf*, *subPropertyOf*, *equivalentClass*, *sameAs*, *differentFrom* and *disjointWith*. In addition to this basic vocabulary with predefined semantics, users can create custom connections.

Semantic Web standards, in their support for creating models of information, are often compared to the UML (Unified Modeling Language) models, database models and XML Schema, but they provide important advantages. For readers interested in the technical details on how these data representation approaches compare, see the table in the More Details section at the end of the paper — it outlines similarities and differences between relational databases,

UML models, XML Schema and RDF/OWL ontologies. A few of the key distinguishing features of the semantic web standards — RDF/OWL — are:

- Global identifiers enabling cross-linking and referencing across models
- Flexible graph-based data model for ease of aggregation and distribution
- Expressive semantics based on the precise and sound computational model

Description of RDF/OWL ontologies is not complete without talking about inferencing. Inference is the process of deriving a conclusion based solely on what one already knows<sup>10</sup>. Conclusions are inferred using a software program called a reasoner. The semantics of RDFS and OWL are defined in terms of what types of conclusions can be drawn based on what types of facts. Because these languages offer standard semantics, an ontology can be processed in the same way (the same conclusions will be drawn) by different reasoners. A typical use of inferencing is to classify a data resource (find what class it is a member of) or to derive additional properties of a data resource based on the available information.

For example, in the TopQuadrant product quotes application, we may want to be able to find all the quotes that are for the entire suite of products. We could create a new class of quotes — *CompleteSuiteQuote*, a subclass of *Quote* and define it as quotes that contain all three products. The inference engine will examine all members of the class *Quote* and for those that include all three products, insert an additional fact stating that they are also members of *CompleteSuiteQuote* class. Thanks to this inference we can now query for such quotes by simply asking for all resources that are members of *CompleteSuiteQuote* class. If TopQuadrant were to introduce another product, we could adjust the definition of the *CompleteSuiteQuote* class and continue to use the existing query which would now produce different results.

TopBraid Suite supports a number of inference engines ranging from the reasoning services built into RDF databases (such as Oracle Spatial 11G) to specialized inferencers such as a descriptions logics reasoners and rule engines.

## TopBraid Suite Capabilities

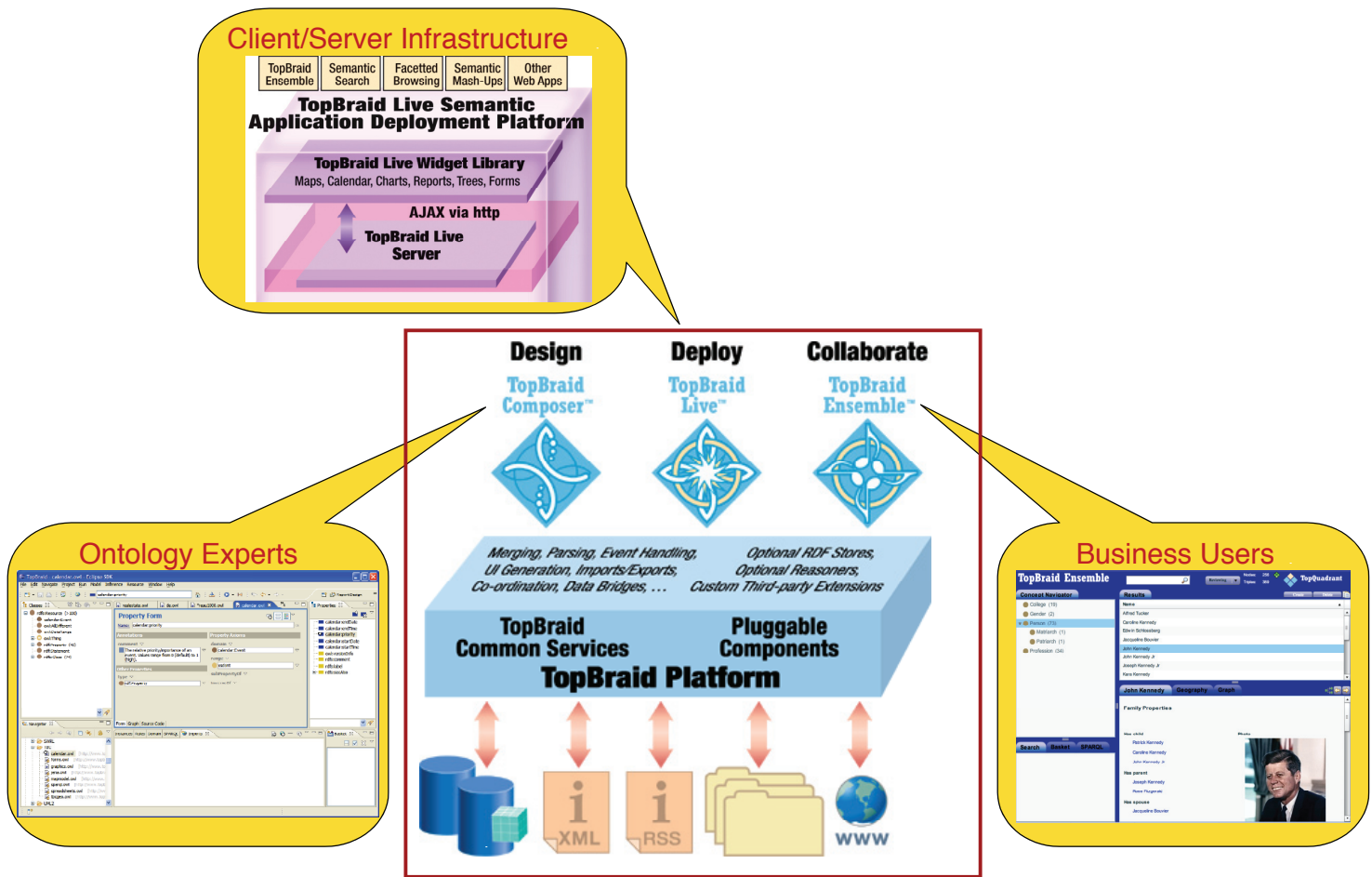
As illustrated in the figures on the front cover and next page, TopBraid Suite encompasses an integrated set of products with the tools, technologies and capabilities necessary to build and deploy enterprise I-I solutions.

**TopBraid Composer™**<sup>11</sup> is the leading professional development tool for building and testing ontologies. Users of Composer are typically information technology experts — modelers and developers.

**TopBraid Live™** is a platform for deploying ontology based applications. These are the applications that use ontologies at run time to, for example, dynamically integrate data, provide model driven user interfaces or to drive search and discovery. TopBraid Composer and TopBraid Live are key components of TopBraid Suite. Data merging, inferencing and integration are some of the services that are supported by both, Composer (for development) and Live (for deployment).

Additionally, TopBraid Live includes a comprehensive set of Flex user interface/application widgets that can be used to create ontology based Rich Internet Applications with little to no programming. The third component of the suite, **TopBraid Ensemble™**, is an example application built using these widgets. When developing ontologies, it often becomes necessary to involve the end users of the solution. TopBraid Ensemble makes it possible for the end users to contribute RDF content without having to use a technical tool such as Composer. Ensemble is a multi-user web application that is fully ontology driven and runs on TopBraid Live platform. Business users can utilize their web browsers to develop and extend controlled vocabularies and/or create any data objects described in the ontologies.

**TopBraid Suite** has a very extensive and continually growing set of capabilities for *importing information* from many common data source (RSS, spreadsheets, databases, etc.), *integrating and processing information* (through queries, rules, visual scripting of workflow using SPARQLMotion, etc.), and for exporting and/or displaying results in user applications (XML, html, forms, maps, calendars, trees, interactive graphs, BIRT charts, etc.). More detailed information for each of the TBS products is available at: <http://www.topquadrant.com/topbraid>.



In this whitepaper, we frame the use of several key capabilities of TopBraid Suite for addressing the requirements two through six for successful information integration and construction of I-I-I solutions.

### ***Access to data from existing internal and external sources, as and when needed***

TopBraid Suite provides 'out of the box' adaptors for access to a variety of data sources including relational data-bases, XML files, UML class models, RSS feeds, spreadsheets and e-mail content. Most of the existing data sources already have a predefined structure. Irrespective of how this information is expressed, TBS adaptors will automatically extract the structure into a model expressing how the data source represents enterprise data. TBS user can elect to either import the data or establish a dynamic connection to the underlying data source.

Using drag and drop import capabilities, data models and data from the different sources can be brought together. Once the data access is in place, users can immediately search and query information across the multiple data sources.

With the growing number of services and data feeds available over the internet, business decisions can benefit from supplementing enterprise data with the relevant outside information. For example:

- When planning sales visits, you may want use the geocoding service to establish the geographic location of each customer based on their addresses.
- When analyzing air traffic patterns, you may want to include weather condition information from a weather service.
- In determining whether a company is a good investment, its latest tax filing from Edgar online or news articles that mention it, may need to be considered.

TopBraid Suite makes it as easy to connect to the external data sources (including web services calls, RSS and XML data feeds), as it is to connect to internal information.

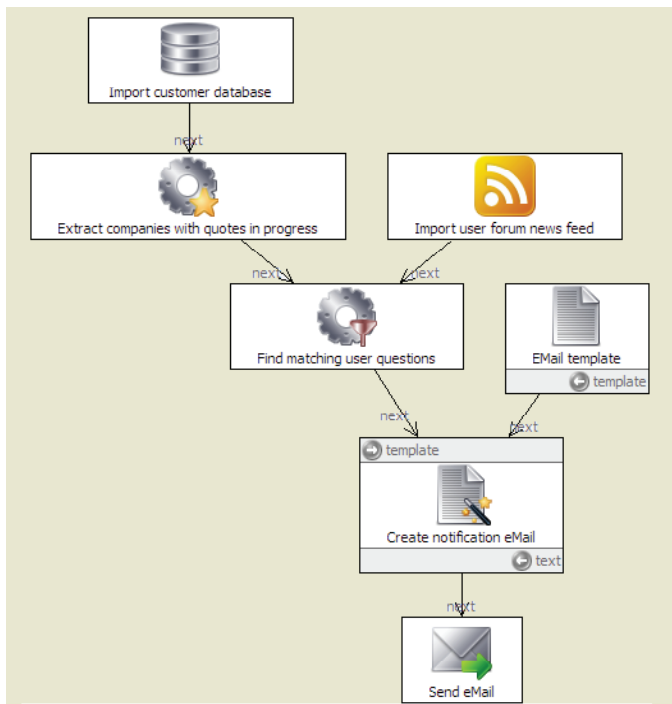
### ***Aggregation of information based on common and related aspects found in the sources***

Accessing the data from different sources in one place and with a single query is only the first step.

It is typical for data sources to reflect the differences related to their intended use and organizations they support. Some may be the result of legacy solutions that reflect differences in processes and requirements that may no longer apply yet are still embedded in the structure and content of the data. Others reflect valid and important business differences between the viewpoints, requirements and vocabularies of different departments and lines of business. There are also technology-based differences. For example, the design of a relational database structure will reflect not only the nature of the information stored in the database, but also individual database designer decisions on how to optimize the data access for the needs of a particular application.

**TopBraid Suite solves the problem of reconciling the differences in the data through capabilities for:**

- Extending the existing information models by 'building on top' — leaving the underlying data and structure intact
- Enabling a user to describe connections between data objects, their relationships and attributes — this can be done using simple connections such as 'subclass' or 'same as' and, when needed, using sophisticated rules describing relationships in the data
- Creation of the data transformation pipes, through an easy to use visual editor, that apply sequences of transformations defined by the user — such pipes are dynamic and can be executed on a scheduled basis, on request or triggered by an event
- Aggregating disparate information using TopBraid Suite's starter pack of transformation modules and rules supporting many of the common operations — these can be directly re-used or easily extended and customized



The diagram on the left shows a data transformation pipe expressed using SPARQLMotion™, a visual scripting language for semantic data processing supported by TopBraid Suite. Visual scripts can be displayed and edited graphically by people with minimal or no programming skills. End users can chain together simple processing steps to form complex processing pipelines. Data processing pipelines can be used to merge, search, query and mash-up data as well as to create a report or information dashboard. This example shows integration of the information tracked in TopQuadrant's quote generation system with questions posed on TopBraid User Forum. As shown in the workflow, TopQuadrant employees working with customers on product purchases want to know about questions and issues customers raise in the forum. Appropriate personnel get notified via e-mail about postings they should be aware of.

**Information processing driven by the context and frame of reference as defined by the information user**

Most data sources have a specific context embedded in them. For example, a database designed to support a call center will be organized around the process of tracking customer problems with the

company's products and their resolutions. A system used internally to track product requirements has a different set of users and a different context. Even though the information is clearly related conceptually, there will be differences in the structure of the data as the data sources collect and manage different data and they support different requirements. Even when the information is essentially the same, it is likely to be distributed in a different way or referenced in a different way. A product name may change as it moves from conception to implementation. Product parts and components tracked internally may be called differently from the names exposed to the customers and specified at a different level of granularity.

Using TopBraid Suite, an information user can define his own context and point of view. For example, they may want to see the impact of the requirements as they get implemented downstream in the types and number of problems reported by the customers.

TopBraid Suite uses the requested context definition to access the necessary data, populate the ontology and apply any necessary semantic processing to answer a user query. Semantic processing is performed using one or more of the inference engines incorporated into the product. This includes rules, queries and classification engines.

#### ***Adherence to company policies***

As the access to enterprise data becomes easier for business users to manage, the ability to control the rights to view and use data grows more critical. This includes the ability to define access based on business roles and policies.

TopBraid Suite provides flexible and extensible access control mechanisms for defining read and write rights. Custom access rights can be specified as an ontology model and can be at any level of granularity. For example, rights can be defined for a class of resources or for a single resource. The definitions can be based on the provenance of data, its scope and any other characteristic important to its management.

#### ***Enriching the usefulness of the data for the next user***

Any interaction a business user has with data has a potential to enrich it. More and more people are becoming familiar with the concept of tagging. We tag photos on Flickr, videos on YouTube and bookmarks on del.icio.us. We do it so that we and others can access the information more readily including information we post and information posted by others that shares the same or related tags.

Similarly, tagging can improve access and organization of enterprise data. TopBraid Suite supports creation of custom tags and associating them with the data objects. Tags can be shared across groups or be kept private. They can also be organized in hierarchies or clouds of related tags, so that the object tagged with 'Palo Alto' could be found when looking for 'Northern California' or 'South Bay'.

## **TopBraid Suite in Use — Some Sample Customer Applications**

This paper has described the advantages of standards based Semantic Web technologies to help enterprises integrate information in order to derive new insights and make better decisions — through creating and deploying I-I solutions. We conclude with a few examples describing how our customers are benefiting from this approach:

### ***1. A major retailer with an established name in appliances, lawn and garden, automotive, furniture and other products***

- **Business Requirement:** Need to give consumers an integrated way to deal with product features — use and care, warranties, service records, proofs of purchase, etc. for all the product lines
  - Hundreds of product lines with different features, new product appear regularly.
- **TopBraid Suite Solution provides:**
  - Flexible, model-driven application platform — product displays and user entry forms are generated from the model describing characteristics of different products and their cross dependencies
  - Seamless integration of many and varied product lines

## **2. Computer Task Associates ([www.ctg.com](http://www.ctg.com)) develops applications for the Medical/Healthcare informatics for its customers**

- **Business Requirement:** For improved outcomes and cost effectiveness, health care providers as well as patients require a seamless, integrated view of all health care information and services
  - Tests, available drugs, insurance information, clinic availability and other information is available on the individual basis, but not in the integrated way
- **TopBraid Suite Solution provides:**
  - A seamless health care solution for outcome-based medicine and for reduction of health insurance fraud.

## **3. Medical center at the University of Texas in Houston:**

- **Business Requirement:** Need to aggregate hospital admission data with other sources (such as weather and environmental news) for early detection of epidemics.
- **TopBraid Suite Solution provides:**
  - A way to bring data in different formats together taking into account different data management practices of each hospital.
  - Ability to browse and analyze aggregated information to determine trends and connections in the data
  - A dynamic, highly flexible warehouse. The end users using the business ontologies can specify the OLAP dimensions depending on the nature of the data. OLAP data structures are then generated on the fly from the merged data sources.

# Conclusion

The ongoing information explosion has created a critical, overriding business challenge. Information is fragmented across many sources; aggregating relevant information that is needed is increasingly difficult; and, the quality, relevance and currency of information is too often unclear. Semantic Web technology as realized and supported through W3C standards (e.g. RDF/S, OWL) provides a stable, capable foundation that is specifically designed and implemented for restoring and enhancing information connectedness and relevance.

TopQuadrant's TopBraid Suite of integrated products exploit the W3C Semantic Web standards to provide an open, flexible, configurable and standards-based platform to build and deploy solutions for a large class of information integration problems. These solutions will obviously vary greatly in details and complexity, but nearly all share a common pattern and core value proposition that is captured by the simple word sequence: Information-Integration-Intelligence (I-I-I).

The key requirements and components for building these types of essential solutions are mapped to the capabilities of the three TopBraid Suite products — TopBraid Composer, TopBraid Live and TopBraid Ensemble. These products are specifically designed to collectively provide the tools, technologies and capabilities necessary to build and deploy enterprise I-I-I solutions.

Example customer applications provided in this paper show how the TopBraid Suite of products is helping enterprises today to put control and decision power for information integration and intelligence into the hands of the people who need it most — business users.

Interested in learning more? We invite you to visit TopQuadrant web site [www.topquadrant.com](http://www.topquadrant.com) for more information and to obtain a free download of TopBraid Composer, our enterprise modeling tool.

# More Details — Comparing Traditional Data Representations with Semantic Standards

## Unique Identifiers

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>Records in the database can have locally unique identifiers</li> <li>There is no notion of a global identity. Relational database models are designed to support local queries and not the queries that must go across different data sources</li> </ul>	<ul style="list-style-type: none"> <li>UML tools maintain internal identifiers which are unique within the scope of a single model</li> <li>There is no notion of global identity. Linking models or cross-referencing across models is problematic.</li> </ul>	<ul style="list-style-type: none"> <li>Poor support for global references that could be used externally</li> <li>XPath could be used to build connections across elements, but the references break if the target document changes</li> </ul>	<ul style="list-style-type: none"> <li>Semantic web ontologies are based on the global uniform naming scheme — Uniform Resource Identifiers (URIs), the most significant subset of which is HTTP URLs of the Web</li> <li>Global identifiers make it possible to connect disparate schema and data information expressed using RDF/OWL. Queries can go across different sources.</li> </ul>

## Query Access

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>SQL is used as a query language for RDBMS</li> <li>Queries' WHERE clauses can not combine data and schema selection criteria</li> <li>Queries tend to be complex and contain business logic</li> </ul>	<ul style="list-style-type: none"> <li>There is no standard query language for UML models as they are intended for design-time description and not for run-time representation and manipulation of the data</li> </ul>	<ul style="list-style-type: none"> <li>XQuery is the standard language for XML</li> <li>Queries do not combine information from different XML documents with different schemas</li> <li>Queries tend to be complex</li> <li>Maintenance and reuse is difficult because queries contain business logic</li> </ul>	<ul style="list-style-type: none"> <li>SPARQL is used as a query language for Semantic Web data</li> <li>Queries' WHERE clauses often combine data and schema selection criteria</li> <li>Business logic can be represented in the ontologies making it possible for queries to stay generic and be very simple</li> </ul>

## Attributes and Relationships

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>Data attributes and relationships between the data objects are implicit. Typically, their meaning is embedded in the name of the column or in the name of the join table</li> <li>Data attributes and relationships are local to the tables and can not be re-used</li> </ul>	<ul style="list-style-type: none"> <li>Attributes and relationships are explicitly defined local to objects (classes) and can not be re-used</li> </ul>	<ul style="list-style-type: none"> <li>Relationships are difficult to define and reuse. XML is fundamentally a hierarchical model with parent-child relationships.</li> <li>Re-use of attributes requires the use of Attribute Groups which add complexity to the schema</li> </ul>	<ul style="list-style-type: none"> <li>Attributes and relationships are explicitly defined.</li> <li>Attributes and relationships are global and can be re-used. Each has a unique identifier and can be defined in terms of other properties (e.g., equivalent property or subproperty)</li> <li>It is possible to attach metadata to property definitions to make their meaning easier to understand for machines or human users</li> </ul>

## Flexibility

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>Inflexible. Business recruitments changes require modification of the database structures. These, in turn, require re-write of the queries and porting of the data.</li> </ul>	<ul style="list-style-type: none"> <li>UML models can be modified as needed. However, since they are not operational, impact on the run-time needs to be managed elsewhere.</li> </ul>	<ul style="list-style-type: none"> <li>Inflexible. Any modifications to XML Schemas could cause a serious ripple effect on the existing XML documents and queries</li> </ul>	<ul style="list-style-type: none"> <li>Flexible. Ontologies can often be extended and modified without impacting queries. No data porting is required.</li> </ul>

## Re-use and Extensibility

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>Reuse is limited to design patterns</li> </ul>	<ul style="list-style-type: none"> <li>UML models can not import or extend other UML models. Reuse is limited to architectures and design patterns</li> </ul>	<ul style="list-style-type: none"> <li>XML Schemas can import each other, but 'building on top' of the existing schema by extending it is possible in only a limited way</li> <li>XML documents are difficult to merge, even with XInclude.</li> </ul>	<ul style="list-style-type: none"> <li>Ontologies are highly re-usable and extensible. For example, a standard ontology of units of measure or industry classifications can be imported into another ontology and extended as needed</li> </ul>

## Expressivity

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>Database models have limited expressivity. Data objects (tables) can be defined only in terms of its columns and primary and foreign keys</li> <li>Cardinality is limited to the ability to express 1:1, 1:many and many:many relationships</li> <li>Any business rules would need to be expressed in either queries or program code (for example, stored procedures)</li> </ul>	<ul style="list-style-type: none"> <li>Data objects are defined in terms of their attributes and relationships. Inheritance is supported through subclasses</li> <li>Because roots of the UML models are in expressing logic of software programs, it is common for the UML model to specify methods of the objects (classes)</li> <li>Cardinality is limited to the ability to express 1:1, 1:many and many:many relationships</li> <li>The QCL specifications for expressing constraints and business rules are in the early stage of adoption and tool support is poor.</li> </ul>	<ul style="list-style-type: none"> <li>Data objects are defined in terms of their attributes and relationships.</li> <li>Support for inheritance is limited to single inheritance</li> <li>Extensibility, using inheritance, can only be restriction of properties or extension of properties and not both.</li> <li>XML is based on the hierarchical document model of parent-child relationships. Describing relationships that fall outside of the tree hierarchy requires complex workarounds</li> <li>The semantics of 'choice' are complex and can lead to imprecise specifications.</li> <li>The use of <i>SubstitutionGroup</i> may work with some Schema processors but relies on a liberal interpretation of the XSD Recommendation, which may lead to interoperability issues.</li> <li>Business rules or any logic are not supported and have to be expressed in the program code</li> </ul>	<ul style="list-style-type: none"> <li>OWL has a number of built-in logical constructs. These include ability to express equivalency between classes and properties and ability to constrain class definitions based on the value of properties. Inheritance is supported through subclasses. Multiple inheritance is commonly used</li> <li>By default, all relationships are many to many. Any integer value can be used as a constraining cardinality</li> <li>In addition to the built-in capabilities of OWL, other business rules can be expressed using RDF vocabularies</li> </ul>

## Computational Model

RDB	UML Model	XML Schema	RDF/OWL Ontology
<ul style="list-style-type: none"> <li>Relational calculus</li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>	<ul style="list-style-type: none"> <li>Description logics</li> </ul>

## ENDNOTES

<sup>1</sup>Statistics compiled by Dr. Michael L. Brodie, the Chief Scientist of Verizon Services Operations in Verizon Communications, in a presentation entitled "Computer Science 2.0: A New World of Data Management" given as an industrial keynote at the 33rd International Conference on Very Large Data Bases, September 2007.

<sup>2</sup>These are typically called taxonomies, look up tables or controlled vocabularies

<sup>3</sup>For more information, visit [www.w3c.org](http://www.w3c.org)

<sup>4</sup>Among standards supported by TopBraid Suite are XML, XML Schema, RDE, RDFS, OWL, SPARQL, SWRL, GRDDL and RDFa. We also support any and all standard vocabularies expressed using Semantic Web standards including (but not limited to) Dublin Core, SKOS and FOAF.

<sup>5</sup><http://www.w3.org/TR/rdf-schema/>

<sup>6</sup><http://www.w3.org/2004/OWL/>

<sup>7</sup>In OWL, attributes are called datatype properties and relationships are called object properties

<sup>8</sup>URI (uniform resource identifier) is a compact string of characters used to identify or name a resource. The main purpose of the URI is to enable interaction with representations of the resource over a network, typically the World Wide Web, using specific protocols. URLs, such as [www.topquadrant.com](http://www.topquadrant.com) are URIs and are used by the HTTP protocol to locate web pages. E-mail addresses, such as [info@topquadrant.com](mailto:info@topquadrant.com) are also URIs used by the mail protocol to route the e-mail.

<sup>9</sup>Although resource identifiers can not contain spaces, the ontology diagram shows names with spaces (e.g., Quote status instead of QuoteStatus). This is because the diagram displays labels – text strings associated with a resource using `rdfs:label` property.

<sup>10</sup><http://wn.wikipedia.org/wiki/Inference>

<sup>11</sup>TopBraid Composer Maestro Edition is an extended version of TopBraid Composer with additional capabilities such as SPARQLMotion.