TopQuadrant Technology Briefing

Semantic Integration
Strategies and Tools

Table of Contents

1. Executive Summary *3*

2. A Need to Integrate and a Need to Manage *4*

2.1 The Most Common Solution Strategy **4**

2.2 Semantic Solutions **5**

Figure 1: Illustration of a Unified View of Billing and Contractual Databases 6

3. Semantic Integration Vendors *7*

Table 1: Overview of Semantic Integration Vendors 10

4. Capabilities of Semantic Integration Platforms: *11*

Table 2: Comparison of Capabilities Offered by Vendors 12

Table 3: Support for Open Standards of Semantic Integration Products 14

Figure 2: Positioning of Vendor's Solutions within the Semantic Integration Space 14

5. Standard Languages for Knowledge Modeling *15*

Figure 3: Tree of Knowledge Technologies 15

Table 4: View of Knowledge Modeling Standards and Marketplace Adoption 16

5.1 XML-based Knowledge (Ontology) Modeling Languages **16**

6. Recommendations for Getting Started: *19*

6.1 About Vendor Selection **19**

6.2 Tools Update **19**

About the Authors *20*

Companies interviewed for this report: *21*

Additional TopQuadrant Technology Briefings are Available *21*

About TopQuadrant *22*

1. Executive Summary

A growing number of vendors are responding to the critical need to manage and integrate large numbers of disparate applications and data sources present in today's enterprise. This briefing is focused on the use of semantic technology to integrate structured data and applications. We have analyzed offerings from 9 leading vendors and summarized information on the emerging ontology standards from W3C.

The most common current solution to integration and translation is field to field mapping. Schemas from two data sources are imported and fields are mapped to each other. This approach doesn't scale well as the number of maps grows exponentially with each new data source. Enterprises working with this technology often discover that creating correct maps is a challenge because it requires that the person doing each mapping has an in depth knowledge of both data sources, which is rarely possible.

Semantic technologies offer a new way to integrate data and applications. Before making mappings, a model (or an ontology) of a given business domain is defined. The model is expressed in a knowledge representation language and it contains business concepts, relationships between them and a set of rules. By organizing knowledge in a discrete layer for use by information systems, ontologies enable communication between computer systems in a way that is independent of the individual system technologies, information architectures and applications.

Compared to one-to-one mappings, mapping data sources to a common semantic model offer a much more scaleable and maintainable way to manage and integrate enterprise data. The "common business model" terminology used here may remind readers of the enterprise data and process modeling initiatives. These initiatives have proven to be long on cost and resources and short on ROI. Does the use of semantic integration solutions depend on an enterprise-wide modeling effort? We don't believe so. In fact, we recommend a targeted way to start by situating your first semantic integration solution within a specific project, as opposed to having it as a separate initiative. The model has to be large enough to provide value – sufficient to integrate specific data or applications. It doesn't need to be enterprise-wide. Using knowledge representation approaches based on W3C standards ensures open, future proof implementations where models can be expanded, interlinked, merged and federated.

Semantic technologies are proving to offer enterprises competitive advantage. With the growing adoption of XML and the attendant need to reconcile meanings across different vocabularies, these technologies are becoming increasingly important. Beyond managing and connecting disparate enterprise data, key future capabilities include intelligent web services discovery and orchestration.

Now is the right time to begin developing the expertise in modeling and learning more about semantic technologies. As forecast by Gartner: "By 2005, lightweight ontologies will be part of 75 percent of application integration projects. The relative scarcity of skills in semantic modeling and the unification of information models may be the greatest challenge. Beyond initial development, the need for ongoing information-management processes at the enterprise level will severely tax most enterprises"[1].

To begin understanding and responding to these challengers, learning more about RDF/S and OWL is an important suggested step. Likewise, acquiring methodologies for modeling and information management is recommended.

[1] Gartner, "Semantic Web Technologies Take Middleware to the Next Level", 8/2002

2. A Need to Integrate and a Need to Manage

Integration is arguably the most pressing and expensive IT problem faced by companies today. A typical enterprise has a multitude of legacy databases and corresponding applications. The disconnected systems problem is the result of mergers, acquisitions, abundance of "departmental" solutions and simply implementation of many silo applications created for a specific purpose.

We know of a bank with over 40 different call center systems, a financial services company with more than 1,000 databases and a manufacturing company with over 2,000 CAD/CAM systems. These systems contain valuable information and often are still good for supporting specific tasks they were intended for. Unfortunately, the information they contain can not be leveraged by other systems without a considerable effort. When the changes in business needs or available technology require modifications to these applications to provide additional capabilities and to streamline workflows, integration and extension become a very expensive undertaking. Simply tracking all the enterprise data sources and their relationship to each other is proving to be a challenge. In fact, many IT organizations spend up to 80% of their budgets maintaining the legacy systems leaving limited funds to support new business opportunities or to satisfy new regulatory requirements.

Many companies have been moving to XML to take advantage of standards based integration. However, XML doesn't capture the contextual meaning (or semantics) of the data. And a growing number of "standard" XML dialects (currently over 400) intended to standardize business vocabularies make the need for a semantic translation layer even more apparent.

2.1 The Most Common Solution Strategy

The most common solution to data integration and translation is field to field mapping. Schemas from two data sources are imported and fields are mapped to each other. Rules can be defined to split or concatenate fields or to perform other simple transformations. Once this is done the tool can do data translations either directly at run time or by generating code that will perform the transformations. There are a number of tools on the market that support this approach. Vendors include IBM and Microsoft. Some of the tools have been available for nearly a decade, but the adoption has been slow for a number of reasons:

- Field to field mapping works on a small scale. However, the number of maps grows exponentially with each new data source. Maintenance and evolution become a problem since any change in the schema of one data source will require you to redo multiple maps.
- Enterprises working with this technology often discover that creating correct maps is a challenge. It requires that the person responsible for each mapping has an in depth knowledge of both data sources, which is rarely possible. As a consequence, mapping mistakes are quite common.
- Mapping and translating between two schemas that are using a different design paradigm (i.e., different degree of normalization or nesting) can be very difficult. There is more then one way to design a schema. Performance considerations may result in de-normalized database schemas. When schemas are expected to change, designer may opt for a reflective design. Some XML schemas are deeply nested, others are shallow. Mapping between relational (RDBMS) and hierarchical (XML) stores can suffer from significant impedance mismatch of the models.

- Direct mapping may fail in the situations requiring more conceptual and conditional transformations.

Is there a better solution?

2.2 Semantic Solutions

Semantic technologies offer a new way to integrate data and applications. Before making mappings, an ontology (or a model) of a given business domain is defined. It can be “jump started” by importing data schemas. The model is expressed in a knowledge representation language and it contains business concepts, relationships between them and a set of rules. This is the knowledge that the users of the systems want to store and access, rather than the data that implements that knowledge. The knowledge model is then mapped to fields in databases, XML Schema elements, or operations, such as SQL queries or sets of screen interactions. This approach solves many maintenance, evolution and schema compatibility problems.

The key ingredients that make up an ontology are a vocabulary of basic terms, a precise specification of what those terms mean and how they relate to each other. The term 'ontology' has been used in this way for a number of years by the artificial intelligence and knowledge representation community, but is now becoming part of the standard terminology of a much broader community including object modelers and XML users. By organizing knowledge in a discrete layer for use by information systems, ontologies enable communication between computer systems in a way that is independent of the individual system technologies, information architectures and applications. As a common model an ontology helps in the management of enterprise data sources.

Once the data sources are mapped to the model it can be used as an enterprise data management tool and to transform and validate data at design or run time. We can also envision future applications composed of very thin components that dynamically change their behavior based on the interactions with the business knowledge embedded in the model.

The distinct advantage of knowledge representation languages as ways to express the model is that they are optimized for capturing relationships between concepts and for defining generic and specific rules (assertions) that logical reasoning can be based on. Some examples of such rules are:

- If A is a part of B and B is a part of C then A is a part of C
- If a person has blood-contact with someone at risk of an HIV infection risk, then they are potentially at risk of an HIV infection
- If John wrote a paper on semantic integration, he knows about semantic integration

The attraction of logic as a technology for supporting semantic integration stems from the capability of logical languages to express relationships in generic ways, and the availability of sophisticated automated systems for finding combinations of related items that satisfy certain constraints. The variants of logic used for semantic integration (including Horn logic (prolog), frame logic, and description logic) differ primarily in the expressiveness of the logic and the tractability of the reasoning system. Another technology that provides similar capabilities is "means-ends analysis", which grew out of a different research background. Some vendors (Celcorp) base their integration products on this technology. Using models of knowledge, semantic engines can make inferences and create dynamic (on the fly) relationships between different concepts.

The model in the Figure 1 shows a unified view of billing and contractual databases. The blue arrows indicate explicitly defined relationships, while yellow arrows indicate derived ones. The derived relationships were established by the system based on the defined rules some of which are also shown in the figure below. For example:

- The rule "If customer is subject to a contract and invoice is billed to the customer then invoice is subject to a contract" has resulted in establishing a dynamic runtime connection between an invoice and a customer
- The rule "If contract has terms and invoice is subject to the contract then invoice is subject to each term" has built on the connection inferred by applying the previous rule and established connections between an invoice and the specific terms of the contract.

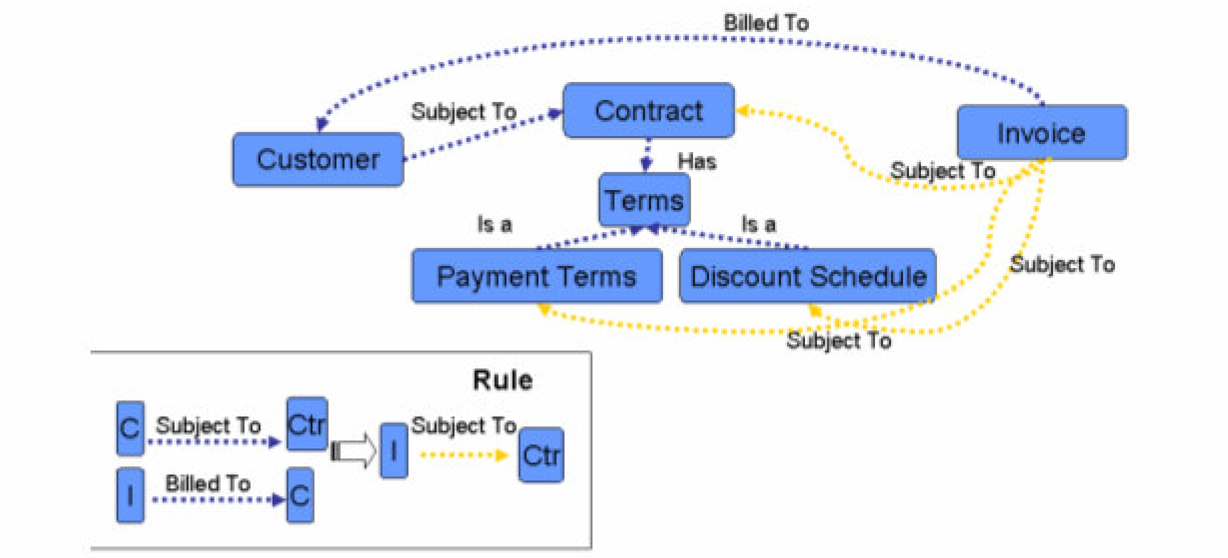

Figure 1: Illustration of a Unified View of Billing and Contractual Databases

Some ideas behind semantic models or ontologies for integration may remind you of metadata management. It is, in fact, based on the similar concepts. However, proponents of semantic integration argue that the use of W3C standard knowledge representation languages gives them distinct advantages:

- ***Open Standards.*** Using knowledge representation approaches based on W3C standards ensures open, future proof implementations where models can be expanded, interlinked, merged and federated.
- ***Rich Semantics.*** Knowledge representation languages offer support for richer and more precise semantics then UML, a standard language behind meta-data repositories. W3C languages like RDF (resource description framework), RDF Schema and the new Web Ontology Language (OWL) have been specifically designed to capture relationships between concepts and to define generic and specific rules (assertions) with the precision that logical reasoning needs.
- ***Native to the Web.*** RDF and OWL are serialized in XML and are, therefore, native to the Web. W3C sees semantic standards as a fundamental enabler for the next phase of web solutions.

3. Semantic Integration Vendors

Table 1 lists companies offering semantic integration solutions. Most of the vendors in this emerging technology field are relatively young (less then 5 years old), privately held companies. Many are capitalizing on the research work that started in early 1990s.

Vendor Name	Product Name	Description	Year Founded	Company
Celcorp	Celware	Engine: Server and Real-time Planner integrate applications streamlining users' workflow where multiple systems must be accessed in order to perform a task. The software uses intelligent agent technology based on proprietary extensions to the "Plan Domain Model and the Graph Plan Algorithm." Modeling: Models are automatically generated by running Celware Recorder, a design time tool.	1990	Celcorp is privately held and based in Santa Monica, California. The company was originally established in Canada and has been offering business integration software for sometime. It has a number of reference clients.
Contivo	Enterprise Integration Modeling (EIM) Server	Engine: Server includes a Semantic Dictionary containing enterprise vocabularies, such as various XML, EDI, and ERP standards; a Thesaurus with synonyms that match business concepts; and a Rules Dictionary that governs the field level data transformation. Modeling: Modeling (mapping) is done using Contivo Analyst tool. Some pre-built maps are available.	1998	Contivo is a privately held company with offices in Palo Alto, California. Contivo's corporate investors include industry leaders BEA Systems, TIBCO Software and webMethods. Venture capital investors include BA Venture Partners, Voyager Capital and MSD Capital LP. It has received a 3rd round of funding in January 2003.
enLeague	Semantic Broker	Engine: Semantic Broker's goal is to overcome differences in business vocabularies, data definitions and Web services by translating and mapping internally and externally developed data and services to business concepts. Services Manager acts as the "traffic cop" that monitors the interaction of services based on the rules and flows defined with the Modeler and Services Flow Manager. Product focus is on creating a scaleable run time environment integrated with popular web application servers. Modeling: Business and Service Flow Modeler uses ontologies to rapidly describe and model critical business processes, goals, and objectives. The Modeler also enables companies to use existing business models (e.g. database	2000	Partially owned by Coca-Cola and located in Atlanta, Georgia (on Coca-Cola campus) enLeague was formed in September 2000. The company has recently acquired Killdara's XML Integration Platform bringing a total number of employees to 16.

		schemas), industry standards, and information from legacy systems by importing and integrating them.		
Modulant	Contextia Product Suite	Engine: Contextia Dynamic Mediation uses a central description of enterprise data called Abstract Conceptual Model (ACM) to enable disparate applications exchange information by transforming messages at runtime. It reconciles semantic conflicts among disparate applications and data sources. Modeling: Modeling is done using Contextia™ Interoperability Workbench capturing the meaning, relationships, and context of data elements of all source and target applications, and mapping them to ACM. The mapping specifications and ACM are then used by the Modulant Contextia Dynamic Mediation to transform data from source to target at runtime. The Interoperability Workbench accepts a variety of inputs for mapping and modeling, including XML schemas, native schemas, database tables, and delimited files.	2000	Modulant was founded in 2000, and subsequently merged with Product Data Integration Technologies (founded in 1989) in order to develop commercially-deployable software based on PDIT's proprietary technology and patent-pending methodology. Modulant is a private, venture-backed company whose existing investors include Sandler Capital Management, Guardian Partners and First Lexington Capital. Modulant's world-wide headquarters is in Charleston, SC, with additional offices in Long Beach and San Francisco, CA, Chicago, IL, Dallas, TX, Washington, DC, London, England and Stockholm, Sweden.
Network Inference	Cerebra Platform	Engine: Cerebra Inference Engine creates dynamic connections between different ontologies using reasoning based on description logic. While Cerebra can work with the central model its value proposition is based on the assertion that only a few key connections between disparate schemas are needed. Cerebra can dynamically infer the rest of the connections thereby minimizing mapping efforts. Modeling: Modeling is done using Cerebra Construct, a MS Visio based graphical modeling tool.	2000	Founded in late 2000 to commercialize a description logic reasoner from the University of Manchester. The company is headquartered in London, UK with plans to open US offices. Network Inference is backed by Nokia Ventures.
Ontology Works	IODE	Engine: IODE utilizes a central description of enterprise data to determine answers to complex queries. Each link in the enterprise ontology is mapped to a query in the "ontology database"; this can either be a warehoused database created as part of the ontology engineering process, or a mediated connection to a legacy database. Solutions to queries in the ontology are build using the rules and relations in the ontology,	1998	The company is privately held and has offices in Maryland and Arkansas. In the first quarter of 2000, it completed development of an initial version (V 1.0) of its tool set and secured its first customer.

		so that the "proof" of the result can be translated in a simple fashion into a program that runs over the databases, to determine the correct answer. Modeling: Modeling can be done using UML tools, translated into a proprietary Ontology Works Language.		
Ontoprise	OntoBroker	Engine: Data integration is done via a several step process that includes importing data schemas from existing databases, and using OntoMap to map concepts and relations from one ontology to the next. These mappings are translated into F-Logic statements, so that Ontobroker can reason over the combined ontology results in data references in the original data sources. Modeling: Modeling is done using OntoEdit and OntoMap. Two more tools are needed to complete this picture, which are a rule editor and a rule debugger, both of which are currently in the proposal stage. The rules state the actual connections between the newly merged concepts, and are susceptible to bugs; hence they must be viewable and debuggable.	1999	The Ontoprise® GmbH is venture capital backed; it achieved a break even point in 2002. The company is headquartered in Germany. Ontoprise was founded as a spin off of the University of Karlsruhe which implemented the first version of technology in 1992.
SchemaLogic	SchemaServer	SchemaServer captures and communicates data definitions (enterprise schema) used across all applications and languages. To help create the active repository of schema and metadata, SchemaServer imports existing schema, taxonomy and classification criteria from databases, applications or content management systems. It supports distributed, collaborative management of enterprise taxonomy. SchemaServer manages the associations and links among the separate schemas by providing the tools necessary to model, map, and describe the multiple relationships.	2001	Privately held company founded by ex-Microsoft employees. Located in Redmond, WA.
Unicorn Solutions	Unicorn System	The Unicorn is a design time tool that imports schemas from multiple data sources including XML, RDBMS, COBOL, IMS, and EDI.	2001	The company is privately held. It is headquartered in New York City with R&D in Israel. Unicorn's investors include:

		They are then mapped to a central enterprise model (ontology). Mapping supports creation of data transformation rules. Unicorn can generate transformation scripts as executable SQL, XSLT, Java Bean code.		Jerusalem Global Ventures, Bank of America Equity Partners, Intel Capital, Israel Seed Partners, Tecc-IS and Apropos.

Table 1: Overview of Semantic Integration Vendors

Other companies worth mentioning in this category include IGS (www.igs.com) and MetaMatrix (www.metamatrix.com) that have UML and MOF based approaches to integration, as well as Vitria (www.vitria.com), an EAI vendor that incorporates business vocabularies.

The market for semantic integration is expected to grow fairly quickly fueled by the needs of enterprises and by the growing maturity of the AI (Artificial Intelligent) technologies that underlie many of these solutions. According to a report released by Business Communications Company, Inc., (*www.bccresearch.com*) they expect AI technologies that assist existing applications handle more complex data analysis, addressing the potential variability in a situation via a set of rules, will see strong growth and implementation across sectors. Their estimate is that this technology will reach $4.8 billion in sales and an AAGR of 14.5% through 2007.

4. Capabilities of Semantic Integration Platforms:

We have identified the following as key capabilities offered by semantic integration solutions:

Enterprise Data Management

- Creating and publishing shared vocabularies of business concepts
- Cataloging data assets, including their schemas and other metadata.
- Formally capturing the semantics of corporate data by mapping database and message schemas to the ontology
- Importing a variety of standard data definition formats
- Supporting model management and evolution

Data Transformation

- Generating scripts and transformations to copy or move the data from one data source to another

Dynamic Code Generation

- Generating executable code such as SQL, XSLT and Java
- Generating "wrappers" for data sources
- Embedding of business rules in models
- Automatic updates after change in the model and schemas

Semantic Data Validation

- Using inference rules to validate integrity of the data based on a set of restrictions. The inference rules will automatically identify inconsistencies when querying for information.

Run-time Support

- Scaleable semantic engine that supports high volume of real time queries

Orchestration of Web Services

- Integration broker
- Intelligent discovery and orchestration (composition and chaining) of web services

Table 2 compares capabilities currently offered by each of the vendors.

	Enterprise Data Managmnt	Data Transformation	Dynamic Code Generation	Semantic Data Validation	Run-time Support	Web Services Orchestration
Celcorp Celware	-	-	Yes	-	Yes	-
Contivo EIM Server	Yes	Yes	Yes	-	-	-
enLeague Semantic Broker	Yes	-	-	-	Yes	Yes
Modulant Contextia Product Suite	-	Yes	-	-	Yes	-

Network Inference Cerebra Platform	-	-	-	Yes	Yes	-
Ontology Works IODE	-	Yes	Yes	Yes	Yes	-
Ontoprise Ontobroker	-	-	-	Yes	Yes	-
SchemaLogic SchemaServer	Yes	Yes	-	-	-	-
Unicorn System	Yes	Yes	-	-	-	-

Table 2: Comparison of Capabilities Offered by Vendors

Table 3 provides a detailed look at each product and its support for open standards.

Product	Product Adoption and Usage	Knowledge Representation	Reasoning Capabilities	Interfaces	Support for Web Services Standards
Celcorp Celware	Mature product, offers a unique approach to application integration. Have a number of reference customers in the financial services industry.	Proprietary, planning to go to RDF in 2003.	Based on proprietary extensions to the "Plan Domain Model and the Graph Plan Algorithm."	Import: Screen scraping, SQL statements	
Contivo Enterprise Integration Modeling (EIM) Server	Relatively mature, has a number of reference customers. Focused on complementing webMethods and Tibco.	Proprietary on top of relational database, evaluating RDF	None evident, integration with a reasoning engine would be hard to implement until support for RDF is offered	Import: XML Schema, RDB (Oracle only), flat files Export: XML Schema (XSLT), EAI (WebMethods, TIBCO), Java	XML, SOAP, WSDL
enLeague Semantic Broker	New, currently in beta	RDF, DAML+OIL	Is designed to Incorporate 3rd party inference engines	Import: XML Schema, RDB (JDBC), RDF/S, DAML+OIL Export: XML Schema (XSLT), RDF/S, DAML+OIL	XML, SOAP, WSDL, UDDI
Modulant Contextia Product Suite	Relatively mature, has a number of reference customers. Focused on government, STEP customers.	XML, proprietary, evaluating RDF	None evident	Import: XML, RDB, flat files, STEP 21 files Export: XML	XML, SOAP

Network Inference Cerebra Platform	New, currently in beta. Initial focus on biotechnology.	RDF, DAML+OIL, OWL	Description Logic	Import: XML Schema, RDB (JDBC), RDF/S, DAML+OIL Export: XML Schema (XSLT), RDF/S, DAML+OIL	XML, SOAP, WSDL
Ontology Works IODE	Relatively mature, has a number of reference customers in government.	Proprietary	Robust, based on a proprietary Ontology language OWL (a variant of KIF, not related to w3c standard by the same name)	Import: UML, RDF/S Export: RDB (Oracle, DB2), DDB, RDF/S, XML	XML
Ontoprise Ontobroker	Relatively mature semantic engine has a number of reference customers. New to the integration market.	RDF, DAML+OIL, OWL support planned	F-Logic	Import: RDB, RDF/S, DAML+OIL, XML Schema Export: RDF/S, DAML+OIL	XML
SchemaLogic SchemaServer	New. The product can unify structured and unstructured data management. Focuses on helping existing customers of Portal and Content Management products.	XML, Proprietary	No	Import: RDF, XML Schema Export: ?	XML, SOAP
Unicorn System	Relatively new, focused on enterprise data management. First customer implementations are in progress.	RDF, DAML+OIL, OWL support planned	A third party reasoning engine could be integrated with this standards-based tool	Import: RDB (Oracle 7i/8i/9i, MS SQL Server 7/2000, DB2), XML Schema, UML (via adopter), ERWin, RDF/S, DAML+OIL Export: RDF/S, DAML+OIL, SQL Transformation Scripts, XSLT	XML

Table 3: Support for Open Standards of Semantic Integration Products

Figure 2 compares how these solutions are positioned within the semantic integration space. The vertical axis represents a vendor's ability to integrate disparate information based on semantics. The horizontal positioning represents a vendor's solution focus. The vertical axis represents a progression – the higher positioning indicates more powerful semantic capabilities. The horizontal line doesn't end with an arrow because, unlike the vertical axis, it is not intended to represent a progression of capabilities. The right most position of a vendor indicates that its major strength is in "Integration and Orchestration". The vendor may also offer some support, but not the full functionality, in the areas of "Model Management", "Validation" or "Run-time".

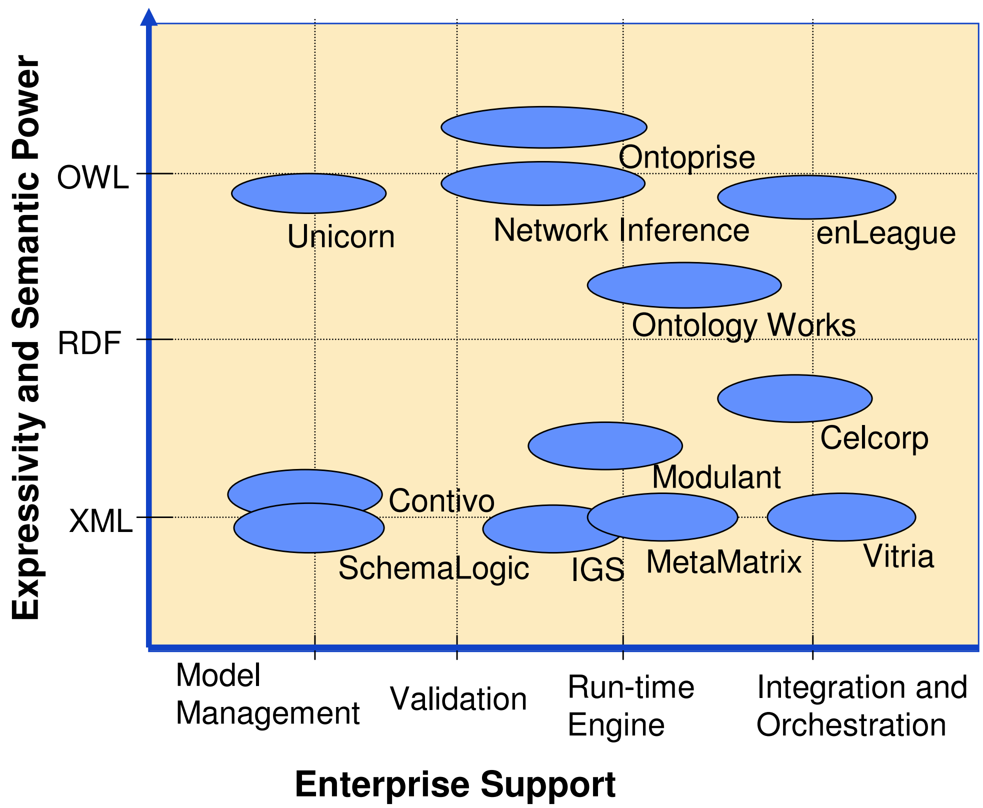

Figure 2: Positioning of Vendor's Solutions within the Semantic Integration Space

5. Standard Languages for Knowledge Modeling

By now we all have heard of HTML and XML. A few important developments preceded HTML, but many have occurred since XML became popular. What we are witnessing today is the emergence of standards for the semantic WEB. These and other important influences from AI, Software Engineering and Process Modeling make up what we are illustrating in Figure 3 as "The Tree of Knowledge Technologies"

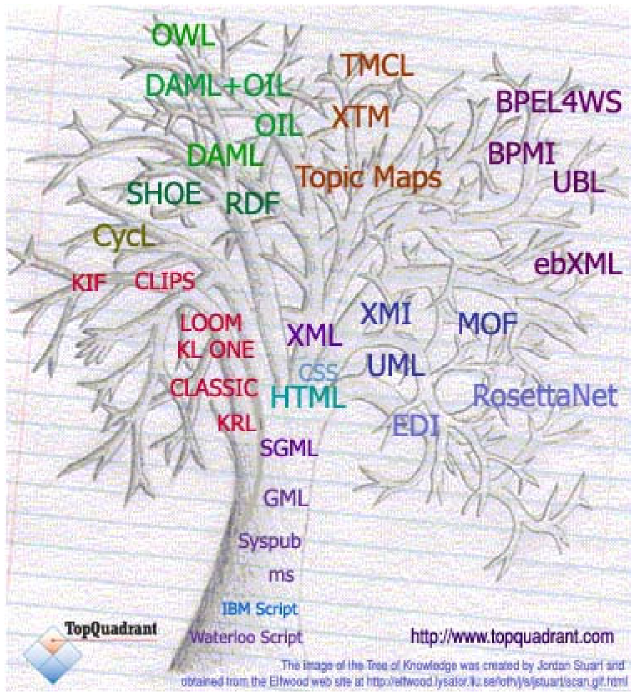

Figure 2: Tree of Knowledge Technologies

The current state of the art on representing and using ontologies has grown out of several efforts that started in the 1980s. Back then, KL-ONE was the most influential of the frame-based representation languages; it allowed for the representation of categories and instances, with inheritance of category properties, and a formal logic for expressing the meaning of properties and categories. At about the same time, rule-based systems were a promising technology. The NASA-sponsored C-Language Integrated Production System (CLIPS) became a de-facto standard for building and deploying rule-based systems.

The Knowledge Interchange Format (KIF), and it accompanying translation tool Ontolingua, were developed to allow knowledge to be shared among these different efforts, and provided the capability to translate knowledge bases in one representation language to another.

These languages were ahead of their time. As a result, they have remained largely within purvey of academia, gaining little commercial support.

With the advent of the World Wide Web, and the acceptance of XML as a de-facto standard for representation of information on the web, ontology efforts joined in. An early project at the University of Maryland produced SHOE, a system for expressing ontologies in XML, and marking up web pages with ontology-based annotations. Many of the ideas from this work made it into the World Wide Web Consortium (W3C) proposal for the Resource Description Framework (RDF) Language.

The DARPA Agent Markup Language (DAML) is built on RDF providing particular logical relationships that standardize the semantics of inferences that can be made over the information in a resource description. The DAML effort drew much of the formal semantics for its logical approach from a parallel effort called OIL (Ontology Inference Layer), which encoded the semantics of Description Logic into an XML-based language. The joining of the two efforts resulted in DAML+OIL language. It allows for a strict interpretation of the statements, so that reasoning agents can collaborate in their use of ontologies. DAML+OIL became a foundation for W3C Web Ontology Language (OWL).

While we have seen some use of UML as a knowledge language and a few MOF (Meta Object Framework) based integration solutions, RDF-based languages have the most potential for success. Table 4 provides a high level view of standards and an indication of the marketplace adoption.

	KIF/OKBC/ CG/CycL	UML	Topic Maps/XTM	RDF(S)	DAML+OIL	OWL
Description	*Legacy KR Languages*	*Universal Modeling Language*	*Topic Maps/XML Topic Maps*	*Resource Description Framework*	*DARPA ML + Ontology Inference*	*Web Ontology Language*
Governance	ANSI American Natio... and others	OMG	ISO Internat... Organi... Standar...	W3C	DARPA	W3C
Years since proposed	>10	>5	>5	>3	>2	>1
Commercial Support (KL*)	□	□	□□	□□□□	□□□	□
Open Source Support	Yes	Yes	Yes	Yes	Yes	

□ 2 or less vendors
□□ 5 or less vendors
□□□ 10 or less vendors
□□□□ > 10 vendors
* - knowledge language

Table 4: View of Knowledge Modeling Standards and Marketplace Adoption

5.1 XML-based Knowledge (Ontology) Modeling Languages

The standards below represent convergence of conceptual modeling and mark up languages:

ISO/IEC 13250 Topic Maps

Topic Maps defines a method of using SGML to represent networks of concepts to be superimposed on content resources (documents of various types), providing a means to represent, navigate, and query the network itself, rather than the full text of a document collection. ISO Topic Maps is an approach for representing topics, their occurrences in documents, and the associations between topics.

XTM is an XML representation of Topic Maps.

Standard Status = Released

There are 3 commercial vendors that offer Topic Maps tools. The Topic Maps standard has been developed in an effort parallel to RDF-based ontology languages. Convergence is not likely, but interoperability is possible. Several approaches for mapping between Topic Maps and RDF have been published. We do not recommend using Topic Maps standards for semantic integration.

RDF/S

The **Resource Description Framework** [W3C-RDF] defines a model and XML syntax to represent and transport metadata. RDF integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web.

Standard Status = Released

The Resource Description Framework (RDF) is a foundation for representing and processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web.

RDF Schema, RDF's vocabulary description language, is an extension of RDF. It provides mechanisms for describing groups of related resources and the relationships between these resources. RDF Schema does the same thing for RDF that DTD and XML Schema do for XML.

Standard Status = Draft

RDF is making good inroads in terms of vendor support. Commercially available tools range from development environments to RDF databases to semantic integration and search/categorization solutions.

DAML+OIL and OWL

DAML + OIL is a semantic markup language for Web resources. It builds on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modeling primitives. DAML+OIL was built from the original DAML ontology language DAML-ONT (October 2000) in an effort to combine many of the language components of OIL.

A DAML+OIL knowledge base is a collection of RDF triples. DAML+OIL prescribes a specific meaning for triples that use the DAML+OIL vocabulary.

The W3C Web Ontology Working Group (WebOnt) has been tasked with producing a web ontology language extending the reach of XML, RDF, and RDF Schema. This language, called OWL, is based on the DAML+OIL web ontology language. The only substantive changes from DAML+OIL are the removal of qualified number restrictions, the ability to directly state that properties can be symmetric; and the removal of some unusual DAML+OIL constructs, particularly restrictions with extra components. There are also a number of minor differences, including a number of changes to the names of the various constructs.

There are three levels of OWL defined (OWL Lite, OWL Full and OWL DL) with progressively more expressiveness and inferencing power. These levels were created to make it easier for tool vendors to support a specified level of OWL.

Standard Status = Draft released in August, 2002.

DAML+OIL and OWL both depend on RDF/S semantics. Thus, the development of these standards is presently a fairly interlocking sequence. Today a number of vendors offer DAML+OIL support. As OWL matures we expect to see them moving from DAML+OIL to OWL.

6. Recommendations for Getting Started:

The "common business model" terminology used by some vendors may remind readers of this report of the enterprise data and process modeling initiatives. These initiatives have proven to be long on cost and resources and short on ROI. Does the use of semantic integration solutions depend on an enterprise-wide modeling effort? We don't believe so. In fact, we recommend a targeted start by situating your first semantic integration solution within a specific project, as opposed to having it as a separate initiative. The model has to be large enough to provide value – sufficient to integrate specific data or applications. It doesn't need to be enterprise-wide. Using knowledge representation approaches based on W3C standards ensures open, future proof implementations where models can be expanded, interlinked, merged and federated.

You may be implementing or enhancing a CRM, portal or a supply chain solution. Any of these projects can be a good starting ground for the semantic integration. It could be used to help you with the data migration or to actually serve as an integration broker. Start with a limited model necessary to support your project. Grow it as needed. Using open standards based technology will enable you to leverage this model with other tools and projects.

Now is the right time to begin developing the expertise in modeling and learning more about semantic technologies. As forecast by Gartner: "By 2005, lightweight ontologies will be part of 75 percent of application integration projects. The relative scarcity of skills in semantic modeling and the unification of information models may be the greatest challenge. Beyond initial development, the need for ongoing information-management processes at the enterprise level will severely tax most enterprises"[2].

To begin understanding and responding to these challengers, learning more about RDF/S and OWL is an important suggested step. Likewise, acquiring methodologies for modeling and information management is recommended.

6.1 About Vendor Selection

Vendors covered in this issue have different strengths as well as different industry and problem focus areas. Choosing the right product will depend on:

- How well it integrates with your data and content sources, infrastructure and applications
- The degree to which you need run time support
- Product's support for the industry specific XML schemas and vocabularies
- Vendor's flexibility and interest in evolving the product to support your requirements

6.2 Tools Update

- Network Inference has unveiled its Cerebra Construct tool for building ontology models to work with its server-based inference engine. Construct is developed on top of MS Visio. Trial version of software is expected to be available for download on the Network Inference web site at www.networkinference.com
- enLeague's Semantic Broker has entered beta testing. The company is looking for two additional beta sites.

---

[2] Gartner, "Semantic Web Technologies Take Middleware to the Next Level", 8/2002

About the Authors

Irene Polikoff is an Executive Partner with TopQuadrant. She is an editor and the main writer of TopQuadrant's quarterly Technology Briefings.

Irene has over 15 years of experience in business application development and deployment, consulting, software development and strategic planning. Irene has held a number of executive positions at IBM. She was Senior Development Manager and Project Executive for worldwide consultant's tooling and methods. Most recently she was a Principal in the national Knowledge, Content Management and Portals Practice in IBM Global Services. Ms. Polikoff was part of the team that developed and deployed a world-wide project management method for IBM Global Services.

Prior to IBM, Ms. Polikoff held IT management positions at Fortune 500 companies where she was responsible for development and deployment of enterprise-wide mission critical information systems. Irene has a background in Operations Research and a strong interest in technologies for software innovation.

Dean Allemang is a Senior Consultant with TopQuadrant who has contributed content to sections of this report.

Dr. Allemang specializes in innovative applications of knowledge technology, and brings to TopQuadrant over 15 years experience in research, deployment and development of knowledge-based systems. Prior to joining Top Quadrant, Dr. Allemang was the Vice-President of Customer Applications at Synquiry Technologies, where he helped Synquiry's customers to understand how the use of semantic technologies could provide measurable benefit in their business processes.

Dr. Allemang has filed two patents on the application of graph matching algorithms to the problems of semantic information interchange. In the Technology Transfer group at Swiss Telecom, he co-invented patented technology for high-level analysis of network switching failures. He is a co-author of the Organization Domain Modeling method, which addresses cultural and social obstacles to semantic modeling, as well as technological ones.

Dr. Allemang combines a strong formal background (MSc in Mathematics, University of Cambridge, PhD in Computer Science, Ohio State University) with years of experience applying knowledge-based technologies to real business problems. Dr. Allemang is a lecturer in the Computer Science Department of Boston University Metropolitan College.

Companies interviewed for this report:

Celcorp - www.celcorp.com

Contivo - www.contivo.com

enLeague Systems – www.enleague.com

Network Inference – www.networkinference.com

MetaMatrix - www.metamatrix.com

Ontology Works – www.ontologyworks.com

Ontoprise – www.ontoprise.com

Unicorn – www.unicorn.com

Additional TopQuadrant Technology Briefings are Available

Current:

- Dictionary of Search
- Ontology Development Lifecycle and Tool Survey

Planned:

- Modeling Techniques
- Semantic Solutions for Search and Self Service

To access these papers, please visit our web site at www.topquadrant.com

About TopQuadrant

TopQuadrant is a trusted intermediary for the intelligent application of knowledge technologies. As knowledge system architects, we are assisting leading enterprises to envision, architect, plan and realize knowledge-based solutions. Our consultants have many years of experience in large consulting organizations, for example IBM Global Services, and have a background in AI, Object Technology, Knowledge Management and Methodologies for Knowledge, Software and Systems Engineering.

Using the following unique tools, we address major obstacles to success in building knowledge solutions:

- ***Solution Envisioning***, a scenario-driven approach to experiencing a future system through analogies and examples using a Database of Capability Cases.
- ***Capability Cases***, application solution patterns (e.g., for ontology-based knowledge applications) expressed in a business context with examples of known uses, applicable technologies and leading practices.
- ***TopDrawer***™, a comprehensive knowledge base for storing and dynamically working with Capability Cases.

With a proven track record in the practical application of knowledge technologies, **TopQuadrant** helps clients transition to next generation, semantically integrated systems, while sustaining and optimizing their investments in current systems.