TopQuadrant

helping enterprises envision, architect and plan knowledge-based solutions

The Business Value of Semantic Technologies

A TopQuadrant Special Report

September, 2004

TopQuadrant, Inc.
141 Howard Drive
Beaver Falls, PA 15010

Mills Davis
Managing Director
mdavis@topquadrant.com

724.846.9300
www.topquadrant.com

This special report has been sponsored in part by
Digital Harbor
www.dharbor.com

Contents

Figures

1. Executive Summary

What is the value of semantic technologies for business?

- This whitepaper explores the business value of semantic technologies. It is written for decision-makers in corporations, governments, and non-profit organizations who are examining potential technology investments to improve enterprise performance.
- Semantics are shared meanings, associations, and know-how about the uses of things. In a computer, what has meaning is what we can represent. Semantic technologies represent meanings separately from data, content or application code, using the open standards for the semantic web.
- Semantic technologies are tools. They are used to create, discover, represent, organize, process, manage, reason with, present, share, and utilize meanings to accomplish business purposes. More than 50 companies are currently providing semantic technologies and solutions.
- What makes semantic technologies different from previous information technologies is that they represent meanings separately from data, content and application code so that machines as well as people can understand, share and work with them. The basic building blocks for digital semantics are data, metadata, context, and ontology.
- Semantic technologies are "meaning-centered" rather than data or document-centered. They include tools for auto-recognition of topics and concepts, information and meaning extraction, and categorization. Given a question, semantic technologies can directly search topics, concepts, associations that span a vast number of sources. The results are fast, relevant, and comprehensive. Plus, semantic technologies can deliver answers, not just lists of sources.
- Semantic technologies organize meanings using taxonomies, ontologies and knowledgebases. These are relatively easy to modify for new concepts, relationships, properties, constraints and instances. Semantic technologies integrate data, content, applications, and processes via a shared ontology, which minimizes costs and effort to develop and maintain.
- Semantic technologies reason via associations, logic, constraints, rules, conditions, and axioms that are represented in the ontology. This declarative structure allows reasoning in multiple directions. For example the same knowledgebase can be used to answer questions about how, why, and what-if as well as give factual responses. Also, semantic technologies allow development of programs that can "learn" (infer and create new knowledge), simulate and test, and adapt behavior based on experience.
- By using ontologies, semantic technologies can auto-discover and provision web services and functionality. Ontologies can link applications into composites that deliver a comprehensive view of situations with all data and information in context. Also, by representing meanings in language and media neutral forms, semantic technologies can auto-generate text, graphics, drawings, documents, and natural language

dialogs. Similarly, they can auto-personalize, customize, and generate multiple versions of communications from the same knowledgebase automatically.

- This report describes applications of semantic technologies across industry and government in a wide-range of business situations. For example, they are used to integrate data and content resources. Also, they integrate business processes and workflows — both internally as well as with external entities. They integrate multiple applications together into composites so that humans and applications have a full context within which to reason and make decisions.
- The true test of whether semantic technologies deliver business value is if the benefits exceed the cost and risks. This report documents 2-10 times improvements in key measures of performance for applications addressing core business processes. This sort of value proposition is compelling for any technology.
- Case examples reported in this study document 20-90% reductions in labor hours, cycle time, inventory levels, operating cost, development time, and development cost. Similarly, these cases show 2-50X improvements in quality, service levels, productivity, and return on assets. Strategic benefits cited include 2-30X revenue growth; 20-80% reduction in total cost of ownership; 3-12 month positive return on investment; 3-300X positive ROI over 3-years; mitigated risk and reduced vulnerability to fraud, liability, or litigation; and improved odds of achieving or exceeding mission objectives.
- From a lifecycle perspective, semantic integration greatly speeds up the process and drives down the cost to develop, deploy, maintain, and evolve complex (integrated, distributed) applications, which are fundamental building blocks of network-centric era computing.
- Given the compelling nature of the business value proposition as demonstrated by the case examples, TopQuadrant expects the market for semantic technologies to develop rapidly during the next few years. By 2010 we estimate the market for semantic solutions, services, and software will exceed $60B.
- Our recommendation to businesses and government agencies examining potential technology investments is not to wait. The value of semantic technologies is established. Any enterprise can prove the value of these technologies in as little as 100-days. This special report outlines a methodology for doing exactly this.

How is this report organized?

This special report is divided into the following topics:

1. *Executive summary* — Highlights important ideas and take-away points for executives and decision-makers.
2. *Semantic technologies* — Introduces key concepts and terminology you need to know in order to talk about semantic technologies. Discusses key functions and capabilities of semantic technology, how they differ from information technologies, and identifies some companies providing semantic solutions.
3. *Appreciating the value of meaning* — Discusses the key elements of business value and how semantic business drive these measures across the lifecycle of the investment.
4. *Case examples showing semantic technology ROI* — Shows where semantic technologies can and are being applied in business and presents 35 case examples reporting significant return on investment.
5. *Proving the value of semantic technologies* — Outlines a 100-day methodology and executive action plan for harnessing the value of semantic technologies for the business.
6. *About TopQuadrant* — Describes the background and capabilities of the report authors.

2. Semantic Technologies

Key questions:

- What are semantics in a computer?
- What are the functions and capabilities of semantic technologies?
- How does semantic technology differ from conventional IT?
- Who are some of the providers of semantic solutions?
- How is semantic technology developing?

Semantics are shared meanings, associations, and know-how about the uses of things.

Semantic technologies are tools. They are used to create, discover, represent, organize, process, manage, reason with, present, share, and utilize meanings to accomplish business purposes. What makes semantic technologies different from previous information technologies is that they represent meanings separately from data, content and application code so that machines as well as people can share, understand and work with them.

Figure-01 describes four elements of meaning of meaning that semantic technologies deal with: instances, concepts, relationships, and constraints.

Figure-01 Components of meaning

Component	Description	Example
Instance	▪ Identity, datum, fact, resource, specific values ▪ Instances are atomic, unique, singular occurrences	▪ "Admiral Jack Johnson" (this person) ▪ Bruno (my dog) ▪ $150 (the price of this report) ▪ La plume de ma tante (object this phrase refers to)
Concept	▪ Idea, class, sub-class, attribute, property ▪ Anything applicable to more than one instance ▪ Anything that has ever been thought	▪ Cat ▪ Dog ▪ Process ▪ Length ▪ Age
Relationship	▪ Link, association, or connection between things	▪ LocatedIn ▪ RequiredBy ▪ AffectedBy
Constraint	▪ Condition ▪ Logical axiom or theory ▪ Conditional or unconditional assertion	▪ Cardinality is at least 1 ▪ Maximum value is 300 ▪ If A, then B precedes C ▪ Person must be over 21 to buy liquor ▪ Admiral RanksHigherThan lieutenant

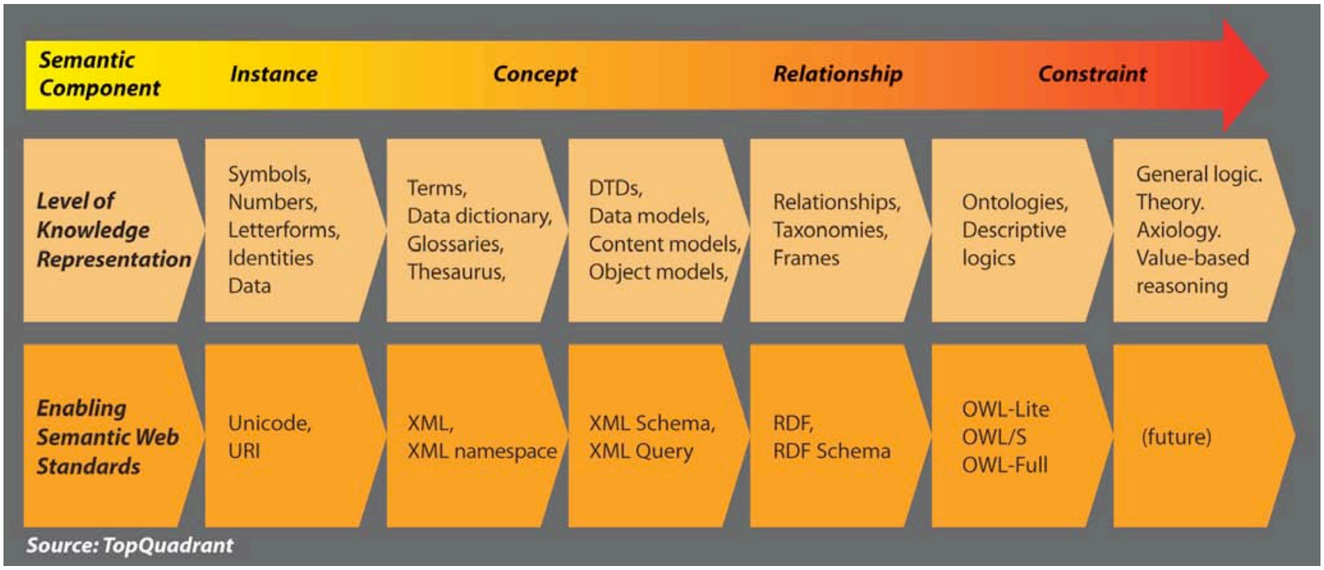
Source: TopQuadrant

Semantics is the idea of getting knowledge within context. For example, the simple phrase "Admiral Jack Johnson" connects to concepts we English speakers have about persons, male gender, armed forces, high-ranking, and so forth. We derive information and meaning based on how these concepts relate to one another. Figure-01 singles out four components of this mesh of associations we call meaning.

- An *instance* is something singular, unique and individual. It is a datum, a fact or identity. An instance is atomic. The name "Admiral Jack Johnson" refers to a unique individual.
- A *concept* is an idea, a class, and abstraction that is applicable to more than on instance, even if it's just the same idea perceived as persisting through different moments of time. "Bruno" is a dog. And dogs have attributes and properties such as legs, color, age, etc.

  Ideas and concepts are not the same thing as the language we use to express them. "La plume de ma tante" and "my aunt's pen" both talk about a writing instrument belonging to a sister of a parent. In digital semantics a goal is to represent meanings in a language-neutral way.
- *Relationships* are the essence of meaning. A relationship is a link or connection between things. "Jack Johnson" is an "admiral." A "salute" *is RequiredBy* the navy (which *is a* "military organization") for persons of higher rank. The words in italics denote relationships. In a dictionary, the meanings of all words are defined in relationship to each other.
- *Constraints* are rules, logical axioms, or theories that assert knowledge about something. For example, "Admiral RanksHigherThan lieutenant." Constraints can be conditional (e.g., if A, then B precedes C) or unconditional (the sun always shines). Knowledge enables reasoning. For example, if "Jack Johnson" is an admiral in the navy and I am a lieutenant; and a salute is required by the navy for persons of higher rank; and an admiral ranks higher than a lieutenant. Then, by inference, I salute Jack.

Figure-02 Representing meaning in a computer



What are semantics in a computer?

Figure-02 overviews different ways that semantic technology represents meaning. The layers of the diagram (from top to bottom) depict: (1) semantic components, (2) levels of knowledge representation, and (3) enabling semantic web standards.

The general progression from left to right is from simpler or more foundational aspects of semantics to more complex and expressive constructs: that is, from lower to higher knowledge value.

In a computer, what's real is what we can represent. Semantic technologies represent meanings separately from data, content or application code using the open standards for the semantic web.

- At the instance level, universal character sets (Unicode) and universal resource identifiers (URI) provide the basis for representing identities, symbols, numbers and words that make up data.
- At the concept level, extensible markup language (XML) standards provide a language for defining metadata (data about data) tags, name spaces, and data structures.
- At the relationship level, an extension of XML called resource definition framework (RDF) provides a way to define semantically meaningful connections between resources and concepts.
- At the level of constraints, the web ontology language (OWL) and forthcoming logic standards provide a way to model and describe the ideas and relationships that exist in an area of knowledge as well as rules, conditions, and logical axioms that apply when reasoning across this knowledgebase.

Figure-03 Key building blocks of digital semantics

Component	Description	Example
Data	▪ Data are facts: atomic, unique, singular occurrences, like individual letters of the alphabet	▪ 8811343 ▪ "Bill"
Metadata	▪ Metadata are data about data, or information about information. Metadata are the building blocks of context and meaning in computers.	▪ <order number> ▪ <Customer name> ▪ <Address>
	▪ ***Syntactic metadata*** provides labels, like a post-it note, attached to data, and formatting ▪ The combination of metadata tag + data is like individual words in a language.	▪ 8811343 = <order number>, number, 7-digits ▪ "Bill" = <Customer>, character string ▪ "Mini-iPod" = <Product ID> ▪ "123 4th St." = <Address>
	▪ ***Structural metadata*** tell us how data are grouped and put in ordered arrangements with other data. This is analogous to a phrase in a language. Customer has a name and address. Order has a number, product ID, and a customer.	▪ Customer record: Customer name, Address ▪ Order record: order number, product ID, customer name, ship to date
	▪ ***Semantic metadata*** adds relationships, rules, and constraints to connect the dots between pieces information and show how they are linked.	▪ If we know that a vendor shipped a product to a customer, and the customer has a location, then by inference, we know that the vendor shipped to a location.
Context	▪ Context is a set of relationships surrounding something that help determine its meaning.	▪ Bank can be a financial institution or the edge of a river, depending on context. ▪ Dictionary (all words are defined in relationship to other words.)
	▪ ***Information*** is data in the context of other data — any ordered arrangement of data that is useful for human understanding such as a list, report or document.	▪ Ordered collection of data fields in a form, or rows and columns of a table, or a database schema, or an object model.
	▪ ***Knowledge*** is a network of concepts that provides context for both data and information.	▪ Theory + rules for visual arrangement of symbols in a diagram ▪ Rules, prescriptives, and rationale of a policy provision
Ontology	▪ An ontology specifies ideas that exist in an area of knowledge as well as the relations between these concepts. This includes attributes, properties, concepts and rules, and how they relate one to another. ▪ Ontology provides a reference model for data and process integration as well as knowledge sharing between systems and people in ways that facilitate machine reasoning and inference ▪ A knowledgebase is an ontology plus instances (or data).	▪ A taxonomy (such as the Dewey decimal system) is a vocabulary or set of terms denoting concepts plus structure describing class/sub-class relationships. ▪ Ontology = taxonomy + relationships, constraints and rules ▪ Knowledgebase = ontololgy + instances.

Source: TopQuadrant

Figure-03 describes the basic building blocks of digital semantics. These include data, metadata, context, and ontology.

Figure-03 shows that:

- *Data* are the most basic building blocks. Data are facts, instances. For example, the number 8811343. By itself, it doesn't mean much.
- *Metadata* add meaning to data through concept labels, data structure, and semantic relationships. When we add the tag <order number> to the number 8811343, we've linked a concept to an instance and created a building block of meaning.
- *Context* denotes patterns of relationships that are particularly useful for understanding, and action. Data in context is information. When we group all the tagged data belonging to a customer, and all the tagged data belonging to an order, we create two ordered arrangements of data (contexts) which we call records.
- An *Ontology* is a network of concepts, relationships and constraints that provides context for data and information as well as processes. When we add relationships between the records and the concepts they contain as well as add constraints such as rules that tell us how to reason about these relationship, then we've created an ontology that we can use to correlate information and make inferences. For example, if we know a vendor shipped a product to a customer, and we know the customer has a location, then by inference we know that the vendor shipped to a location.

Business ontology

A business ontology is a formal specification of how to represent the concepts of the business and their interrelationships in ways that facilitate machine reasoning and inference.

A "business ontology" ties systems together, much as a database ties together discrete pieces of data. It is a "schema across systems" that logically maps information sources and describes the interaction of data, processes, rules, and messages across systems. In a relational database, we can only define "that" a relationship exists, then must write code to specify how it fits together with processes, rules, events, or applications. In a business ontology, on the other hand, semantic definitions allow us to specify "what kind" of relationship exists, "how" it is related to other elements of the architecture (e.g. a web service). We don't write code to do this. Rather, we exploit relationships we've specified to save development time and derive more business value from our data. For example, an item may not have a location, but is contained in a warehouse that has a location or is supplied by a supplier that has a location. The data in these scenarios may be different, but when the relationships are mapped, these semantics (semantic definitions) allow users to capture meaning from multiple data sources without writing code. Benefits from ontology-based technologies include:

- Precise, fast knowledge generation, navigation, and retrieval
- Flexible integration through shared metadata layers
- Multi-device capability through unambiguous definition and specification of any web resource
- Seamless end-to-end security through shared understanding of user roles and context

Figure-04 Functions of semantic technologies

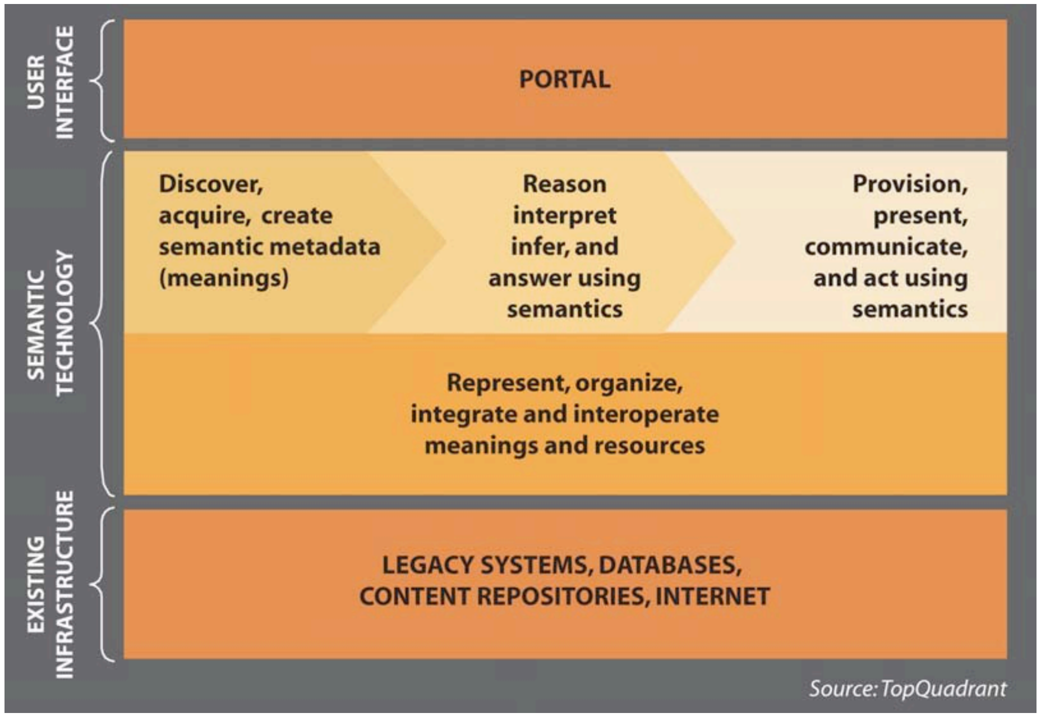


Figure-04 identifies high-level functions of semantic technologies. The diagram has three layers. The top-most layer, user interface, and the bottom-most layer, existing infrastructure, refer to existing information technologies (IT/COM) for portals, legacy systems, databases, content repositories, and the Internet. The middle two layers refer to semantic technologies for discovering, managing, reasoning with, and utilizing meanings in ways that differ from conventional information technologies:

Semantic technologies provide an abstraction layer above existing IT technologies that enables bridging and interconnection of data, content, and processes. Second, from the portal perspective, semantic technologies can be thought of as a new level of depth that provides far more intelligent, capable, relevant, and responsive interaction than with information technologies alone.

What are the functions and capabilities of semantic technologies?

Figure-05, which follows this page, highlights some of the capabilities of semantic technology. The table is divided into three columns. Functions and related activities are to the left. Semantic capabilities are listed to the right. Capabilities package semantic technology functionality to meet business needs. For example:

- *To discover, acquire and create semantic metadata*, semantic technologies sense, discover, recognize, extract information, encode metadata, and tag content. Semantic capabilities that provide these functions include recognizers, auto-taggers, content annotators, automatic categorizers, and terminology brokers.
- *To represent, organize, integrate and interoperate meanings and resources,* semantic technologies model, classify, categorize, index, cross-reference, interconnect, map, bridge, federate, store, and manage. Semantic capabilities that provide these functions include semantic web service discovery, composition, choreography, registry, semantic web service advisor, data integrator, application integrator, process integrator, ontology builder, and semantic broker.
- *To reason, interpret, infer, and answer using semantics*, semantic technologies query, search, meta-search, retrieve, match, mediate, simulate, analyze, answer, explain, plan, schedule, and optimize. Semantic capabilities that provide these functions include assistants, agent, pattern and connection explorer, match-maker, mediator, design advisor, simulator, configurator, trade-off analyzer, planner, scheduler, optimizer, and autonomous learner.
- *To provision, present, communicate, and act using semantics,* semantic technologies create composite applications, browse, navigate, author text, draw graphics, create languages, and orchestrate services. Semantic capabilities that provide these functions include semantic portals, semantic browsers, application composer, concept-based searcher, context-aware retriever, expert locator, enhanced search query, social network navigator, interest-based deliverer, natural language generator, forms generator, document generator, engineering drawing generator, display and media producer.

Figure-05 Semantic technology capabilities

Function	Activity	Semantic Capabilities
Discover, acquire, & create semantic metadata	Sense Discover Recognize Extract Encode Tag Represent	Recognizer Automated tagger Content annotator Auto-categorizer Terminology broker
Represent, organize, integrate, and interoperate meanings & resources	Model Classify Categorize Index Cross-reference Interconnect Map Bridge Federate Store Manage	Semantic service discovery Semantic web service composition Semantic service choreography Semantic content registry/advisor Semantic data integrator Semantic content integrator Semantic application integrator Semantic process integrator Ontology builder Semantic broker
Reason, interpret, infer, & answer with semantics	Query Search Meta-search Retrieve Match Mediate Simulate Analyze Answer Explain Plan Schedule Optimize	Assistant Agent Pattern & connection explorer Match maker Mediator Design advisor Simulator Configurator Trade-off analyzer Planner Scheduler Optimizer Autonomous learner
Provision, present, communicate, and act using semantics	Composite apps Browse Navigate Author text Draw graphics Create docu-ments Orchestrate services	Semantic portal Semantic browser Application compositor Concept-based search Context-aware retrieval Expert locator Enhanced search query Social network navigator Interest-based delivery Natural language generator Forms generator Document generator Information display generator Engineering drawing generator Media generator

Source: TopQuadrant

Figure-06 Differences between semantic and information technologies

Category	Semantic Technology	Information Technology
Represent meaning	▪ Semantic web standards ▪ Language-neutral ▪ Machine interpretable ▪ Semantic metadata ▪ External to system (sharable) ▪ Knowable at run time	▪ Data format; program language ▪ Language-based ▪ Human knowledge required ▪ Syntax/structure metadata ▪ Internal to system ▪ Predefined at design time
Discover, acquire, create, access	▪ Meaning-centered. ▪ Auto-recognition, information extraction, and categorization ▪ Topic and concept-based search with enhanced recall and precision (relevance)	▪ Message, file, page, and document-centered ▪ Manual, full-text & statistical categorization ▪ Data, word, & document search with limited recall and precision
Organize, integrate, interoperate	▪ Taxonomy, ontology, knowledge base ▪ Knowledgebase easily modified for new concepts, relationships, properties, constraints and instances. ▪ Integrate data, content, applications, and processes via shared ontology	▪ RDBMS, OODBMS, flat file. ▪ Database structure difficult to modify to change /add new relationships (e.g., requires coding, reorganization) ▪ Integrate data, content, applications, and processes via point-to-point interfaces
Reason, interpret, infer, answer	▪ Application reasons via logic constraints, rules, axioms separate from cod ▪ Knowledge answers questions such as how, why, and what if, (combining theory & data) ▪ Machine can learn (infer new knowledge), simulate, test and adapt based on experience.	▪ Application reasons via fixed algorithm that is embedded in the application code. ▪ Information processing provides situation awareness, e.g.,: what, where, when, and how much ▪ Preprogrammed responses are like instinct. Logic updated off-line (new version) Program does not learn or adapt
Provision, present, communicate, act	▪ Auto-discover, provision, semantic web services ▪ Ontology-linked composite applications give 360-view of relevant data in context. ▪ Auto-generate text, documents, graphics, drawings & dialogs ▪ Auto-personalize, customize, version (e.g. languages) ▪ Autonomics: systems with self-knowledge can self-configure, self-optimize, self-protect, self-heal, and self-manage. ▪ Services & products that know, learn, & reason as humans do	▪ Manually discover and implement data and application connections & interfaces ▪ Human search to find data & information and then put it into useful context for decisions ▪ Computers as “electronic pencils” for humans to author and develop content, visuals, and media formats

Source: TopQuadrant

Figure 06 summaries ways that semantic technologies differ from information technology. The table compares how meanings are: (1) represented; (2) discovered, acquired, and created; (3) organized, managed, and utilized to integrate and interoperate resources; (4) used to reason, infer, interpret and answer; and (5) used to provision services, communicate content and media, and manage actions.

How does semantic technology differ from conventional IT?

All software technologies, to some extent, start with meanings. Where they differ begins with how the represent meanings.

Semantic technologies encode meanings separately from data and content files and separately from application code. This enables machines as well as people to understand, share and reason with them at execution time. With semantic technologies, adding, changing and implementing new relationships or interconnecting programs in a different way can be just as simple as changing the external model that these programs share.

With information technologies, on the other hand, meanings and relationships must be predefined and "hard wired" into data formats and the application program code at design time. This means that when something changes, or we want exchange information we hadn't previously, or two programs need to interoperate in a new way, the humans must get involved. Off-line, the parties must define and communicate between them the knowledge needed to make the change, then recode the data structures and program logic to accommodate it, then apply these changes to the database and the application. Then, and only then, can they implement the changes.

Semantic technologies are "meaning-centered." They include tools for auto-recognition of topics and concepts, information and meaning extraction, and categorization. Given a question, semantic technologies can directly search topics, concepts, associations that span a vast number of sources. The results are fast, relevant, and comprehensive. Plus, semantic technologies can deliver answers, not just lists of sources.

Information technologies are data, page, and document centered. They can only directly search these primary sources, by browsing, by word or number indices, or with statistical categorization. Precision and recall is more limited, plus information technologies only return lists of pages, documents, and files to consult.

Semantic technologies organize meanings using taxonomies, ontologies and knowledgebases. These are relatively easy to modify for new concepts, relationships, properties, constraints and instances. Semantic technologies integrate data, content, applications, and processes via a shared ontology which minimizes costs and effort to develop and maintain.

Information technologies organize meanings using flat files (simple schemas), relational data models (RDBMS), and object-oriented models (OODBMS). Database structure is relatively rigid, and difficult to modify for new concepts and relationships. Integration of data and processes typically requires point-to-point interfaces and connectors that are costly to develop and maintain since the knowledge required must be hard coded in each connection rather than shared via a common metamodel.

Semantic technologies reason via associations, logic, constraints, rules, conditions, and axioms that are represented in the ontology separately from application code. This declarative structure allows reasoning in multiple directions. For example the same knowledgebase can be used to answer questions about how, why, and what-if as well as give factual responses. Also, semantic technologies allow development of programs that can "learn" (infer and create new knowledge) simulate and test, and adapt behavior based on experience.

Information technologies reason via fixed algorithms that are embedded in application code. Information technologies give us situation awareness. For example, they answer questions about what, where, when, and how much. Algorithms are preprogrammed behaviors, like instinct. They perform a rote task. If anything is learned, people must update the logic off-line to create a new version of the program.

Semantic technologies use ontologies to auto-discover and provision services and functionality (e.g., semantic web services, semantic grid services, etc.). They use ontologies to link applications into composites that deliver a comprehensive (e.g., virtual, 360 degree) view of situations with all data and information in context. By representing meanings in a language and media neutral form, semantic technologies can auto-generate text, graphics, drawings, documents, and natural language dialogs. Similarly, they can auto-personalize, customize, and generate multiple versions of communications from the same knowledgebase automatically. Semantic technologies enable "autonomics:" systems with self-knowledge that can self-configure, self-optimize, self-protect, self-heal, and self-manage. They provide the foundation for developing new categories of services and products that can know, learn, and reason as humans do.

Information technologies require humans to manually discover and implement data and application connections and interfaces. Alternatively, humans must search to find data and information, then put it into the right context for decision-making. Information technologies use computers as "electronic pencils" for humans author and develop content, visuals, and media formats.

Figure-07 Representative companies providing semantic technologies

Function	Sample Solution Providers		
Discover, acquire, & create semantic metadata	Autonomy Captiva ClearForest Convera Copernic	Endeca Entrieva Factiva FAST H5 Technology	Interwoven Inxight Primus Stratify Verity
Represent, organize, integrate, and interoperate meanings & resources	Celcorp Cogito CognIT Connecterra Contivo Digital Harbor Empolis Grand Central	Enigmatec Intellidimension KFI L&C Metamatrix Miosoft Modulant Network Inference	OntologyWorks Ontoprise SchemaLogic Semagix Triple Hop Tucana Unicorn Vitria
Reason, interpret, infer, & answer using semantics	HP IBM Network Inference	NuTech OntologyWorks Ontoprise Open Cyc	Protogé Unicorn
Provision, present, & communicate, and act using semantics	Cogito Digital Harbor Endeca Inxight	Merant Oracle PeopleSoft Pinnacor	SAP Sybase Vignette

Source: TopQuadrant

Who are some of the providers of semantic solutions?

Figure-07 identifies companies that provide semantic technologies. This list includes approximately 50 firms.

This list is representative, and is by no means exhaustive. This listing is organized by major function. Vendors list alphabetically. Most vendors have multiple capabilities. We have listed vendors in the category that correlates most closely with the case examples we review later in this report.

Also, the degree to which these vendors comply with standards for the semantic web varies. Most companies that developed proprietary technologies and capabilities before semantic web standards were formally disseminated are in process of providing support for them now.

3. Appreciating the Value of Meaning

Key questions:

- What elements of business value matter?
- How do semantic technologies drive these measures?
- How will the value of semantic technology play our in the marketplace?

In the previous section we explained what semantic technologies are, the functions they perform, how they differ from conventional information technologies, and who are some of the companies that deliver semantic solutions. In this section we ask: why should you care?

Business and government executives should only care about semantic technologies if they deliver compelling business value that far-and-away exceeds the cost and risk associated with investing in these solutions.

What we have found is that semantic technologies do indeed impact the measures of business performance that matter. The value proposition for semantic technologies is strong in all of the following categories: efficiency, effectiveness, business edge, and lifecycle economics.

In this section we introduce the elements of business value that matter and will set expectations for how much semantic technologies affect performance. In the next section we will present case examples that demonstrate quantum improvements in every measure of performance we discuss here.

What elements of business value matter?

The classic motivations for investing in new technologies are basically three:

Efficiency gain — Doing the same job faster, cheaper, or with fewer resources than it was done before. The key measurement is cost savings.

Effectiveness gain — Doing a better job than the one you did before, making other resources more productive and improving attainment of mission. The key measurement is return on assets.

Business edge — Changing some aspect of what the business does, resulting in growth, new value capture, mitigation of business risk, or other strategic advantage. The key measurement is return on investment.

The acid test for any new technology or solution investment is this: do the benefits represented by these “3-Es” outweigh the costs and risks associated with making the change. If the answer is “yes”, we have something to consider. If not, the matter ends.

Figure-08 Business value and the impact of semantic technologies

EFFICIENCY	EFFECTIVENESS	EDGE
Cost savings. Doing the same job faster, cheaper, or with fewer resources than it was done before.	*Return on assets*. Doing a better job than the one you did before, making other resource more productive and increasing their return on assets and attainment of mission.	*Return on investment.* Changing some aspect of what the business does, resulting in growth, new value capture, mitigation of business risk, or other strategic advantage.
IMPACT OF SEMANTIC TECHNOLOGIES*		
20-80% less labor hours 20-90% less cycle time 30-60% less inventory levels 20-75% less operating cost 25-80% less set-up & development time 20-85% less development cost	50-500% quality gain 2-50X productivity gain 2-10X greater number or complexity of concurrent projects, product releases, & units of work handled 2-25X increased return on assets.	2-30X revenue growth 20-80% reduction in total cost of ownership 3-12 month positive return on investment 3-300X positive ROI over 3-years

**Source: TopQuadrant*

Figure-08 identifies three primary categories of business value and summarizes the potential impact that semantic technologies can have in each area. The metrics listed come from the case examples reviewed in section 5 of this report.

How do semantic technologies drive these measures?

Efficiency gains

Semantic technologies can have a dramatic impact on labor hours, cycle time, inventory levels, operating cost, development time, and development cost. Case examples show 20-90% reductions in these measures.

Semantic technologies drive efficiency gains in multiple ways, for example:

- *Semantic information access* — applied to content and data sources, semantic automation allows people to cope with information overload. Semantic solutions allow them to access, tag, extract, classify and make sense of vastly more and varied sources than ever previously possible. This matters, for example in research, intelligence, and publishing. But it impacts any area requiring search.

- *Enterprise and supply chain integration* — Semantic integration changes the economics of bridging disparate and redundant data sources, and linking legacy systems together. Huge savings are possible. Today's motivation for using web services is to take cost and time and effort out of the process of integrating applications and maintaining interoperability. Exposing functionality as web services can reduce TCO by significant percentages. However, it still requires people to make the interfaces, even when they can find out about the service in a directory. Moving to semantic web services enhances the value proposition. Service ontologies provide a way to take people out of the loop of configuring and managing the integration. The idea here is akin to "autonomics": software with self-knowledge that can auto-configure, self-maintain, and self-manage — resulting in even better TCO.
- *Knowledge-centered manufacturing* — Semantic virtualization means driving a product life cycle (from research, design, engineering, manufacturing, through in-service maintenance and sunset) using knowledge-based models. Together with simulation and auto generation of drawings and technical document, semantic virtualization dramatically increases productivity, and improves the quality, speed, and economics of design and manufacturing.
- *Generative communications* — Another area where semantic technologies impact efficiency is generating the right content, in the right media format, delivered to the right audience, through the right media channel, with the right economics. A knowledgebase that represents the meaning of content separately from the language forms, as well as the attributes of audience, language, media, and delivery channel allows semantic technologies to automatically generate audience-specific text, drawings, documents, animations, and media displays directly from the knowledgebase. Huge savings in time, labor, and cost are possible.

Effectiveness gains

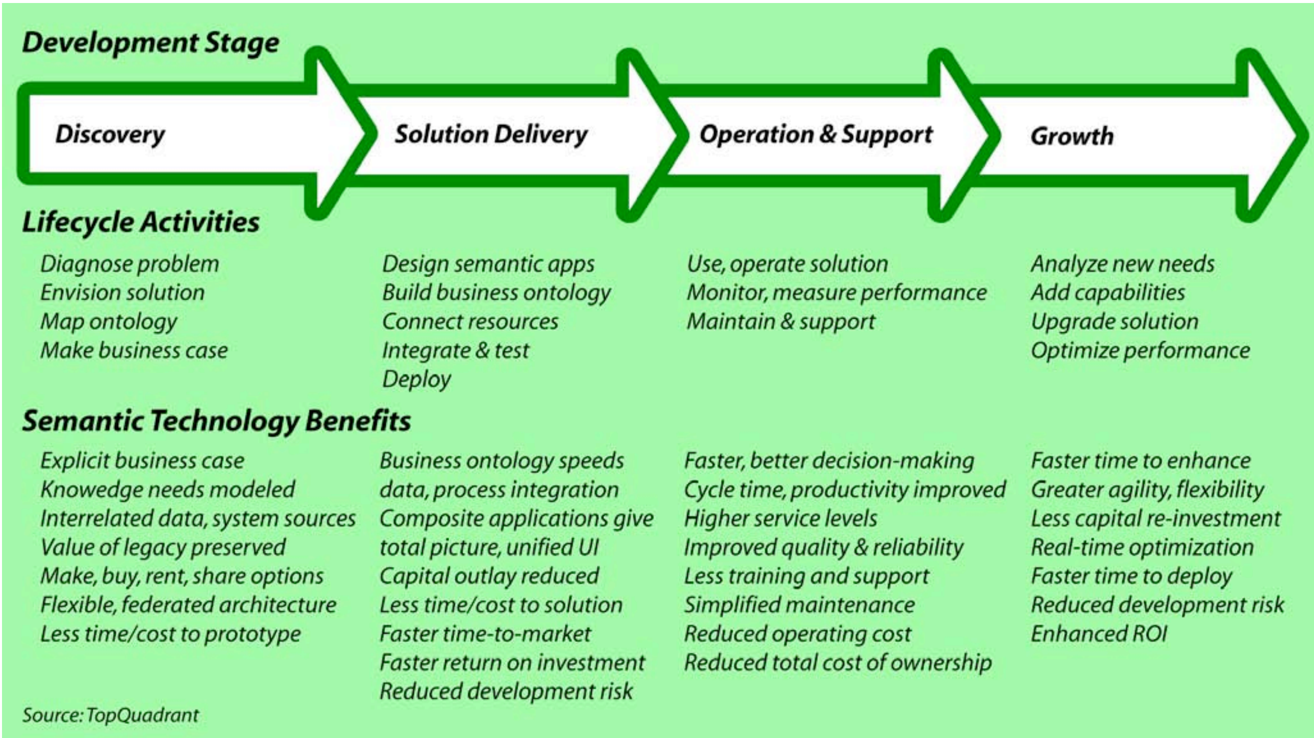
Semantic technologies can drive dramatic improvements in quality, service levels, and productivity. Combined with process improvements, these can allow existing staff to handle a greater number (or complexity) of concurrent projects, product releases, and units of work handled. Case examples show increases in effectiveness and return on assets from 2-50 times.

Business edge

The strategic value of semantic technologies is the ability to create new advantages. Case examples evidenced this in several ways, for example: 2-30X revenue growth; 20-80% reduction in total cost of ownership; 3-12 month positive return on investment; 3-300X positive ROI over 3-years; mitigated risk and reduced vulnerability to fraud, liability, or litigation; and improved odds of achieving or exceeding mission objectives.

The conclusion we reach is that semantic technologies pass the acid test for business value. 2-10 times improvement in performance for applications addressing core business processes is a compelling value proposition for any technology.

Figure-09 Semantic technologies maximize lifecycle value and returns



Since the business benefits that semantic technologies promise are so compelling, it might be well to look into the economics of the path that a business or government agency takes to get there. *Figure 09* portrays the life cycle of a system investment. The diagram is divided into three levels. Development stages list across the top. Lifecycle activities in each stage appear across the middle of the chart. Semantic technology benefits during each stage are summarized across the bottom of the diagram.

The semantic solution life cycle spans four stages: discovery, solution delivery, operation and support, and growth (secondary and tertiary development). In this example, we focus on a project whose objective is semantic integration. What we see is that semantic technologies positively impact total cost of ownership and return on investment across all stages of the lifecycle, while delivering a better way to integrate information and functionality.

Discovery

During the discovery stage, the key activities are to diagnose the problem (or opportunity), envision the solution, map the knowledge required, and make the business case. Semantic integration provides more choice and control over investment options and ROI. By focusing on questions needing answers, knowledge integration requirements become clear. By interrelating data and functionality across systems, business ontologies preserve the value of legacy applications while supporting federated deployment. Mapping connector technology to ontology reduces switching costs and gives multiple options for acquiring new functionality (make, buy, rent, share, etc.). Representing knowledge in high-level models vs. coding preserves flexibility and results in less time and cost to prototype.

Solution

The design-build process can be up to 3-5 times faster for A2A and B2B integration. Semantic integration requires knowledge engineering skills (to develop and maintain the ontology), but less overall IT resources to implement equivalent functionality. This means less time and cost to solution with reduced developmental risk. Since the "smarts" are in the ontology, not the code, test-debug-revision cycles are faster. Semantic integration makes solutions involving multiple systems and data sources much simpler to deploy. Federated deployment is practical either across multiple locations, or locally, via web services. Dependency on proprietary solutions and vendors is reduced, since ontologies are based on public standards. Performance of semantic integration engines can be up to 20-30 times faster than relational database technology.

Operation and support

With semantic integration, operating costs can be reduced as much as 50-75 percent, especially for training, technical support, and maintenance. Knowledge applications provide real-time information in context, improving productivity, and reducing decision cycle time. A common user interface across systems and data sources is simpler to use and reduces training and support costs. Ontology-based middleware simplifies maintenance and deployment of changes.

Growth

Semantic integration builds in flexibility and adaptability. It is easier to add, subtract, and change components because changes impact the ontology layer, with far less coding involved. It's is like a virtual data warehouse, a virtual process schema. With semantic integration, it is much easier, faster, less expensive to modify to add new capabilities, relationships, etc. vs. stand-alone applications or relational data base technology. High-level modeling enables many new capabilities, displays, etc. to be added by business personnel who are non-programmers.

How will the value of semantic technologies play out in the marketplace?

We have seen that the business value and lifecycle economics of semantic technologies are favorable. Next, we look at how the value of semantic technology is likely to play out in the marketplace through the end of the decade.

Semantic technologies have been characterized as an early market. Indeed the majority of current expenditures are more for R&D than for operational deployments. What this study reveals is that semantic technologies are now poised and ready for mainstream deployments. Of course, the technology is still evolving. But, as the case examples will bear out, the benefits are already here.

Whenever an industry, or in this case and entire economy, is about to retool to make a transition from one category of technology to another, there has got to be a compelling economic driving force. It's a bit like lighting a match. If there's enough heat and light is released the reaction goes forward and is self-sustaining. If not, the match goes out and no fire is lit. In the case of semantics, it appears that we are just at that point where the next bonfire is being lit.

Figure-10 Growth of semantic and information technology to 2010

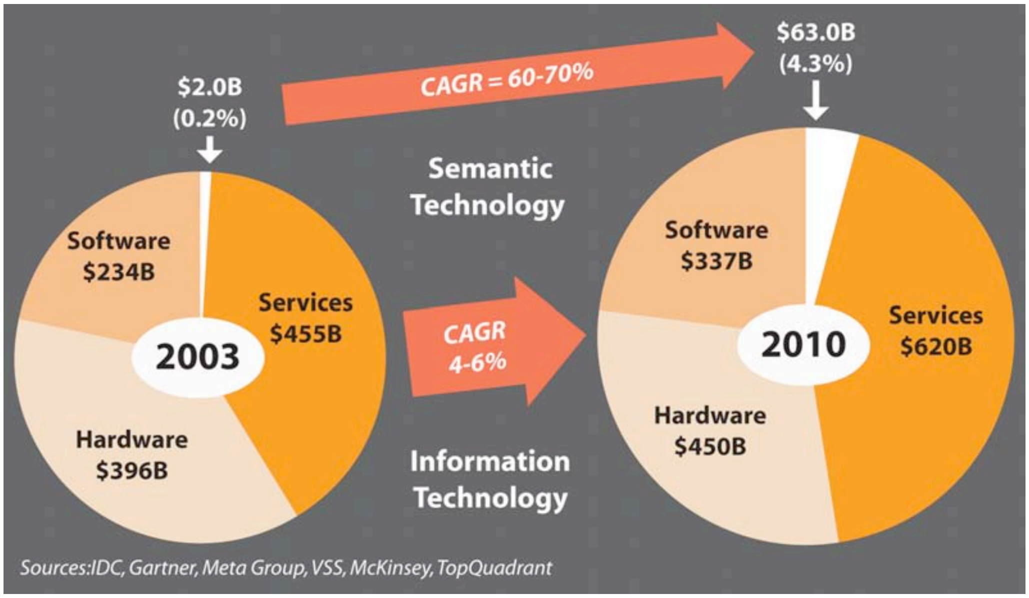


Figure-10 projects the overall growth of semantic technologies and information technologies to 2010. The baseline year show to the left. The forecast year shows to the right. And the compound annual growth rates (CAGR) appear in the arrows connecting the two pie charts.

The numbers for information technology represent total sales of software, services, and hardware. The numbers for semantic technology represent a composite of services and software.

Information technology is projected to grow for the rest of this decade at an overall growth rate of 4-6%, with services growing fastest, and hardware the slowest.

Semantic technology during this period is projected to grow from less than $2.0 billion to around $63.0 billion with a compound annual growth rate approaching 70%. Given the significant ROIs being realized by early adopters of semantic solutions reported here, we believe that level of growth is both reasonable and sustainable.

Figure-11 Semantic technology market in 2010

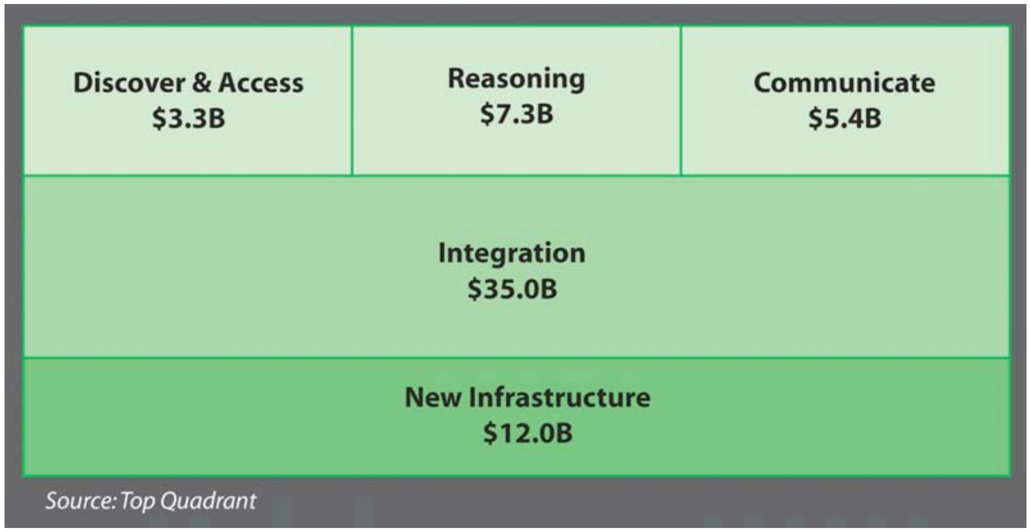


Figure-11 estimates the overall size of the semantic technologies market in 2010 by functional capabilities. This horizontal segmentation generally follows our earlier description of semantic technologies. We want to emphasize that these market projections are preliminary estimates, based on best available data that future research will refine. The diagram depicts five segments as follows:

- *Discovery and access* — Sense-making, data and content mining, moving from documents to knowledge-centered processes, intelligent information access, and social networking will be growth areas. The estimate for this segment is $3.3B.
- *Reasoning* — Semantic technologies enable intelligence, real-time auditing and compliance, simulation-based "virtual" product design, engineering and manufacturing, virtual data centers, adaptive logistics, and supply chain optimization. New application categories will have huge economic benefits. The estimate for this segment is $7.3B
- *Communication* — Representing the knowledge about things separately from content and media files will spawn new categories of enterprise publishing, especially relating to product lifecycle management, professional publishing, and business information services. The estimate for this segment is $5.4B.
- *Integration* — By far, semantic integration will be the largest category of service and software during this decade. The estimate for this segment is $35.0B
- *Semantic infrastructure* — The emergence of semantic web services, context and situational computing, semantic grid, pervasive computing, and large-ontology reasoning engines will include new operating systems and hardware categories. The estimate for this segment is $12.0B.

Another useful way to segment the market for semantic technologies is by industry. Some significant industry verticals for semantic technologies include:

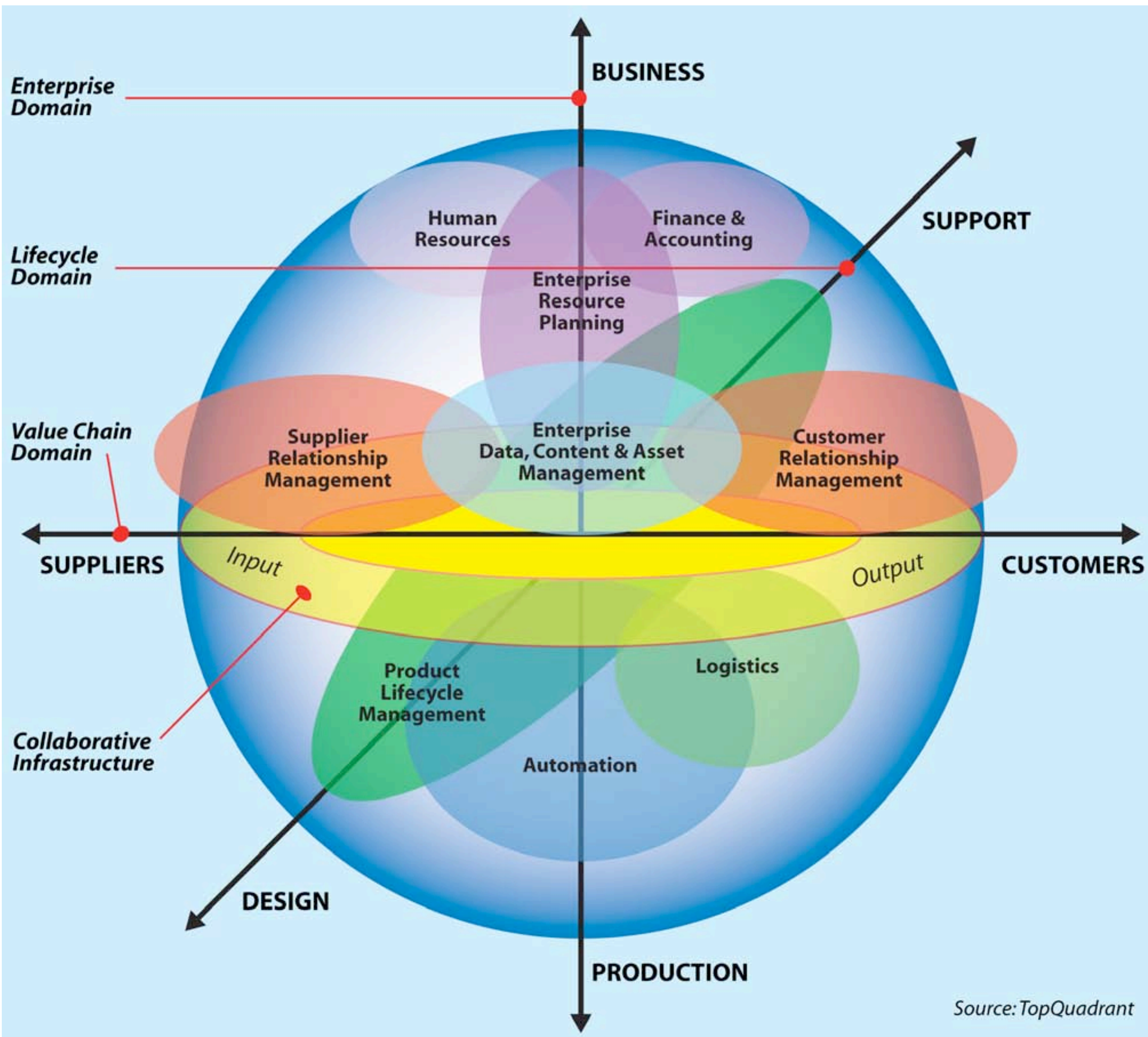
- *Life sciences, health care and medical* — Bioinformatics, drug discovery, document sharing, process outsourcing including steamlining of research, development, and approval processes, integrated patient care delivery.
- *Infrastructure and geographic and spatial information intensive industries* — including energy, telecommunications, logistics, and distribution networks. Gas and oil exploration, network design, semantic service grids, data-set visualization, network design, routing and optimization.
- *Advanced manufacturing* — Simulation-based acquisition, engineering, and test and evaluation for technology development for chip design, aerospace, automotive, discrete and continuous process manufacturing as well as IT hardware, software, and telecommunications engineering.
- *Financial services* — Regulatory compliance and auditing, straight through processing, portfolio risk management, integration of customer services, and enhanced performance for compute-intensive tasks, e.g.:
- *Government, defense, Intelligence and R&D* — Semantic integration, simulation and design, distributed database coordination, service utilities, net-centric computing, etc.
- *Publishing and media communications* — reference knowledge, professional information services, games, versioning, cross-media and integrated communications.

4. Case examples showing semantic technology ROI

Key questions:

- Where in the business landscape can semantic technologies be applied?
- What situations call for semantic solutions?
- How are semantic technologies being applied today?
- What do case examples show?

Figure-12 Business domains where semantic technologies can deliver value



Where in business can semantic technologies be applied?

Figure-12 depicts areas in business and government where semantic technologies can be applied and deliver value. The diagram is divided into several business domains:

- *Enterprise domain* is the business—production axis. The business processes include enterprise resource planning, human resources, finance and accounting, and business intelligence. Production processes include execution and automation systems.
- *Lifecycle domain* is the design—support axis. The key process is product lifecycle management, which includes research and development, design, engineering, manufacturing, distribution, and support.
- *Value chain* domain is the customer—supplier axis. The key processes are customer relationship management involving marketing, sales, and customer service; and supplier relationship management, which involves planning, sourcing, making, and delivery.
- *Collaborative infrastructure* is the integrative platform that interlinks and supports these domains. The processes involve information input, enterprise data, content and asset management, and information output.

Figure 13 Enterprise situations that call-out for semantic solutions

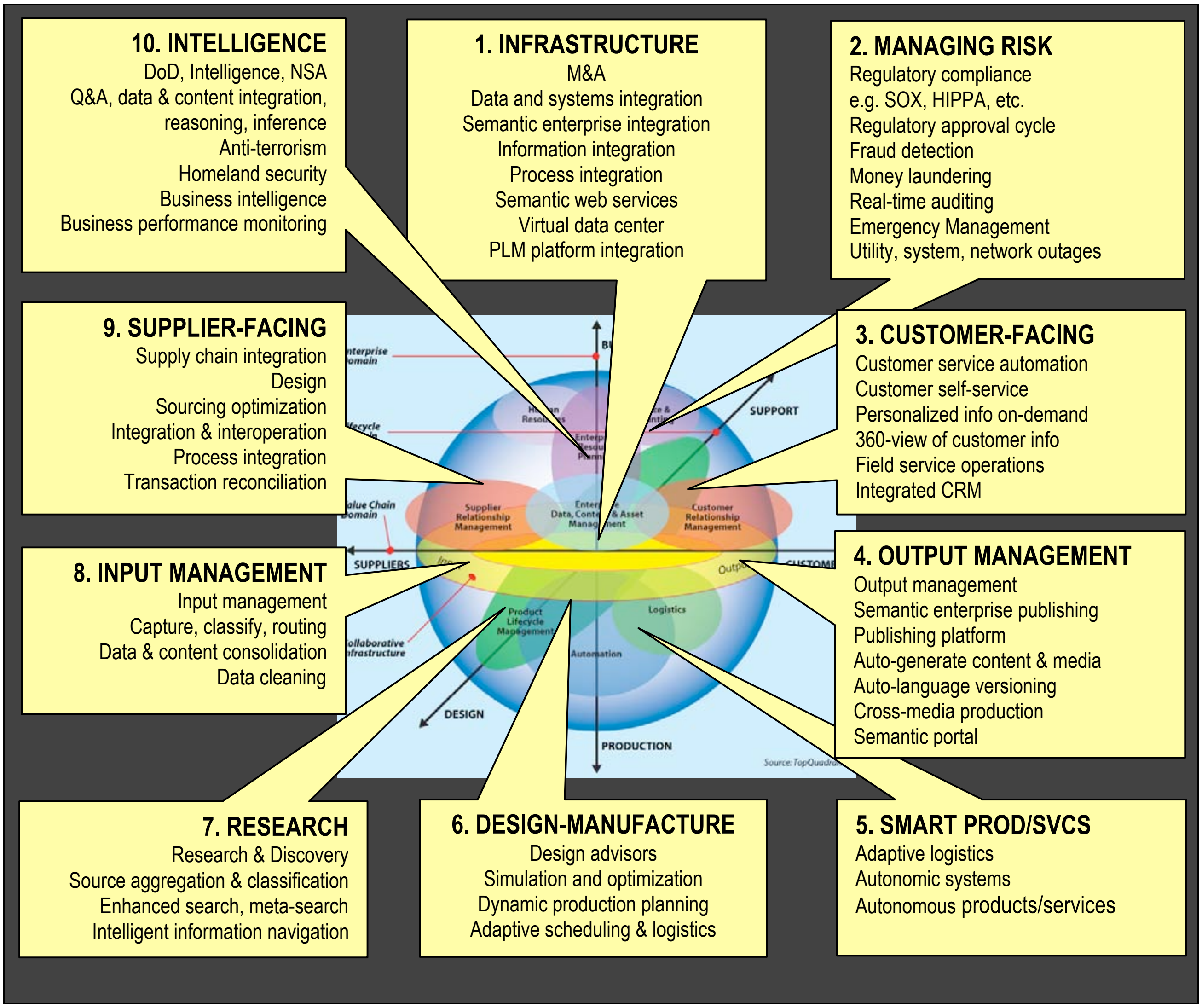
Situation	Characteristics
Significant legacy systems that need to connect and evolve	Disconnected business processes. Redundant and fragmented databases and content repositories. Maintenance time bomb is ticking. High cost of ownership is even though a lot of time and money spent to support information on legacy systems and develop point-to-point API and transaction interfaces.
Business situation, processes, or IT infrastructure frequently change	Current IT systems and integration solutions are rigid, do not adapt easily to changes and cannot be easily migrated. RDBMS and OODBMS deployments are costly and time consuming to upgrade with new/changed entities and relationships. In net-centric world, lack capability to adapt systems quickly or maintain context for situation awareness. Organization reaction time (e.g., to observe, orient, decide, and act) lags.
Mergers, acquisitions, and divestitures	Global companies grow by acquisition and divestiture. Excessive cost and time to specify, build and implement solutions in order to quickly access information systems and ensure quality customer service while removing redundancies. Cross-business unit integration is monolithic, inflexible, and difficult to decouple during divestiture.
Coordinated access to external systems required	B2B interface standards absent, only standardized formats for certain types of data Access external systems manually or by developing custom interfaces. Hoping they can convince their partners, customers, and suppliers to make similar expenditures. Lack capability to rapidly reconfigure and optimize operations across the supply chain.
Need to use regulated systems	In order to make certain changes to their systems, these companies must navigate federal, state, or local bureaucracies in order to secure approval. To avoid this, often use manual "work-arounds."
Need to make sense of large volume of external information	Information overload. Difficulty coping with quantity of information; finding relevant content in a timely fashion; ensuring that relevant sources have not been overlooked. Difficult to access and aggregate sources. Manual classification and indexing. Poor search precision, recall, and relevance.
Access information sources in context	Access to needed information in systems within the company, but managed by other business units or divisions, is limited. Separate sources not in context. Manual search, correlation, and integration. Difficult to create total picture, 360-view.
Need to deliver right content across media, languages, and audiences economically	Content and media delivery process that spans product/service lifecycle, geographic regions, languages, and media formats is driven manual authoring, design, graphics, versioning, and media production. Relatively slow and labor intensive. High costs. Problems with quality, consistency, timeliness and fit to audience needs.

Source: TopQuadrant

What situations call for semantic solutions?

Figure-13 illustrates eight business situations with symptoms and characteristics that call for semantic technology solutions. This list is not intended to be exhaustive. However, in each case, application of semantic technologies can address the core problem and deliver strong benefits.

Figure-14 Areas where semantic technologies are being applied in business and government today



How are semantic technologies being applied today?

Figure-14 calls out ten specific areas where semantic technologies are being applied today in business and government. Listed clockwise, these areas are (1) infrastructure and integration, (2) managing risk, (3) customer-facing services, (4) output management, (5) smart products and services, (6) design and manufacture, (7) research, (8) input management, (9) supplier facing processes, and (10) intelligence.

What do case examples show?

In the following topics, we present 35 case examples that highlight the benefits being obtained from today from semantic technologies. Examples are drawn from industry and government and are grouped by call-out area. Each case demonstrates significant gains in the measures of business value that truly matter: efficiency, effectiveness, and return on investment.

1. Infrastructure & integration

1.1 Enterprise publishing platform

Professional publisher needed flexible, integrated business, supply chain, and production platform to support multiple customer segments with product lines that span customer lifecycles, processes, and interests, utilizing multiple media channels. Having grown by acquisition, business units have disparate legacy systems and repositories. Goal: to rationalize systems and infrastructure; take cost and time out of operations; and facilitate business growth through development of new product and service extensions.

- *Semantic solution* — Semantic M&A, BPM, and publishing: Ontology-integrated federated publishing platform linking CRM, PLM, SCM, and ERM functions with value chain deployed content origination, asset management, media production, and cross-media delivery capabilities.
- *Efficiency gain* — Reduced cost and time to market to develop new products. Reduced cost and time and effort to publish. Less labor to develop, operate.
- *Effectiveness gain* — Cost and time to market to develop new products reduced by 1/3. Reduced cost and time and effort to publish.
- *Edge* — Accelerate new revenue opportunities. Reduced total cost of ownership through vendor rationalization and simpler development and maintenance.

1.2 Enterprise application integration

Global 2000 corporation needed to speed the process and reduce the cost and effort required to integrate enterprise business processes and applications across multiple locations. Maintaining point-to-point data transformation was becoming unsustainable. It was becoming cost-prohibitive to make changes in underlying data sources, message formats, and business rules since critical business logic and metadata was locked into proprietary applications and middleware.

- *Semantic solution* — Ontology-based semantic information model providing leverage for integrating enterprise applications and data.
- *Efficiency gain* — High-level ontology-mapping reduces time and effort to integrate. 2-5X faster solution delivery. Reduced training and support and operating costs. Faster time to upgrade and enhance.
- *Effectiveness gain* — Enterprise processes and data sources map to each other through a common meta-model, simplifying application integration, maintenance, and upgrades. Semantic development environment accelerates new & composite application deployment. Semantic portal puts information in context of total process, other applications, and all data sources.
- *Edge* — Reduce TCO by 20-65%. Financial exposure and developmental risk mitigated.

1.3 Enterprise search

Manufacturer needed search capability that would scale to massive numbers of records but whose performance (numbers of queries per second) would not degrade as with RDBMS or OODBMS indexing.

- *Semantic solution* — Ontology-enhanced graph database search.
- *Efficiency gain* — RDBMS and OODBMS search required indexing at each step; thus performance degraded as L log (N) where L is the path length and N is the number of records. Semantic graph database performed at scale because it required no indexing, eliminating the log (N) from the performance equation.
- *Effectiveness gain* — Graph-based search proved >10X faster than traditional query and delivered relatively constant performance regardless of number of records being searched.
- *Edge* — Scalable search.

1.4 IT infrastructure rationalization

A global financial services provider needed to overcome shackles of its client-server architecture. It needed only 6 databases to operate, but found it had more than 80 copies of some of these. New infrastructure solution needed to decouple applications from data, eliminate redundancies, and provide higher quality data.

- *Semantic solution* — Ontology meta-modeled information integration mapping data sources and interrelationships.
- *Efficiency gain* — Operations, maintenance, and future development costs greatly reduced. Savings over 5 years in $10s of millions.
- *Effectiveness gain* — Ontology decoupled applications from data. Eliminated 1/2 of redundant databases. Ontology permits creation of data transformations and "virtual databases" and 'virtual data warehouses" providing real-time integrated queries across federated sources, with improved data control and quality.
- *Edge* — Faster time to deployment than conventional approaches. Substantially reduced TCO.

1.5 PLM platform integration

To remain competitive, a supplier of industrial automation to manufacturers needed to compress time and cost to handle orders, plan production, and handle change orders. Solution must enable "DABASA" (design anywhere, build anywhere, support anywhere).

- *Semantic solution* — Ontology-driven federated product lifecycle and integrated supply chain management.
- *Efficiency gain* — Reduce order processing from 4 days to 1 hour. Reduce production planning from 4 days to 5 hours. Cut engineering change notice cycle time in half, and cost by $200
- *Effectiveness gain* — Totally connected supply chain as well as internal locations accelerates demand-fulfillment cycle time and reduces costs.
- *Edge* — Strategic cycle-time improvement.

1.6 Virtual data centers

A financial services firm needed scaleable architecture for provisioning and managing IT across 1000s of databases, applications, and locations. Current fragmented situation is an operational cost time bomb.

- *Semantic solution* — Ontology-based semantic service grid; virtual data centers; autonomics
- *Efficiency gain* — Development cost 25%. Operational savings > $4M per year. Running 24 hours/day for several years.
- *Effectiveness gain* — a comparison
  Traditional: Bankers Trust
  - MVS - IBM mainframe
  - US based solution
  - 110 operators & programmers

  Service grid: Deutsche Bank
  - Win NT cluster (Oracle on Sun Solaris
  - World-wide solution
  - 16 operators & programmers
  - 1/4 as expensive to develop
- *Edge* — More sustainable IT infrastructure.

1.7 Supply chain platform integration via services

Fortune 1000 company needed to integrate data and processes internally and with supply chain partners, while minimizing capital investment, time-to-solution, and total cost of ownership.

- *Semantic solution* — Semantic web service based shared resource platform for EAI and B2B.
- *Efficiency gain* — No hardware, software, staffing. No maintenance or upgrade fees. TCO reduced up to 70%.
- *Effectiveness gain* — Fast partner on-boarding. Simple, self-service provisioning. Flexible change management.
- *Edge* — Service-oriented shared resource architecture enables faster ROI. No up-front investment. No firewall exposure. Readily scaleable, subscription based.

2. Managing Risk

2.1 Emergency management

Organizations that must manage emergencies (e.g. outages, breaches, service disruptions, etc.), need to make time-critical decisions that require total access to information in real-time, and in a context that supports its effective use. Solution must integrate disparate data, content and applications, and be deliverable within reasonable cost, time, effort, and risk.

- *Semantic solution* — Business ontology that connects data and processes providing real-time comprehensive integrated situation awareness. Semantic development environment for building composite applications and portal UIs.
- *Efficiency gain* — Semantic solution development is 2-5X faster and less costly. Productivity gain form having information in context: e.g, eliminates searching, aggregating sources, and accelerates time to value.
- *Effectiveness gain* — Ontology-based integration delivers real-time, 360 view from all relevant sources giving total picture for sense-making and decision support. Information in context enables faster, better decision-making.
- *Edge* — Reduced TCO. Business risk mitigated.

2.2 Fraud detection

Financial services need high performance fraud detection solutions that can process massive data from a wide variety of sources, provide highly accurate detection and classification of suspected fraudulent transactions; reduce "false positives"; identify and adapt to new schemes; and deliver fast reporting.

- *Semantic solution* — Ontology and rule-based expert systems, neural networks, statistical methods, rough sets techniques, enhanced classifier systems, and fuzzy logic rule induction systems.
- *Efficiency gain* — Since this application runs in addition to current fraud detection software, the resources required for this activity remain essentially the same. The gain comes from not having to deal with the consequences of failing to detect fraud.
- *Effectiveness gain* — Accuracy in fraud detection increased by 10-20% vs. current systems. Reducing losses translates into a measurable gain in return on assets.
- *Edge* — Measurable reductions in exposure.

2.3 Intellectual asset management

A chemical company needed to improve yield on intellectual property. Solution must provide for knowledge asset classification, strategy, valuation, competitive assessment, and investment.

- *Semantic solution* — Ontology-enhanced text mining, auto categorization, intelligent tagging, and document screening.
- *Efficiency gain* — Tax liabilities maintenance costs for non-productive IP decreased by $50M. Competitive intelligence and material research productivity increase of 30-50X. Isolating valuable documents in weeks rather than years.
- *Effectiveness gain* — Re-engineered system for managing portfolio of 30,000 patents increased annual licensing revenue from $25 M to $125M over 5 year period. Speeded time-to-market for new IP.
- *Edge* — Increased return on assets.

2.4 New drug approval cycle

A global pharmaceutical firm needed to accelerate the approval cycle for sale of new drugs.

- *Semantic solution* — Knowledge chain modeling and social networking applied to approval process
- *Efficiency gain* — Reduced time to file with FDA by 6X from 18 months to 3 months. Reduced time to approval by 4X from 36 months to 9 months
- *Effectiveness gain* — Accelerating time to market increases lifecycle revenue and profit by 5-10%.
- *Edge* — First phase of program saved $94 million.

2.5 Regulatory compliance

Public corporations need to integrate policies, information, and processes into one view that provides legally defensible evidence of compliance with regulations such as Sarbanes-Oxley, HIPPA, Gram-Leech-Bliley.

- *Semantic solution* — Ontology-based regulatory and standards models, semantic information and process models create "virtual databases" that provide the metaview needed for compliance reporting and auditing.
- *Efficiency gain* — Reduced cost to establish compliance. Reduced cost to comply. Reduced cost to adapt as regulatory requirements, and internal systems change.
- *Effectiveness gain* — Ontologies map relationships between data sources and processes. Provide a unified view across all compliance-affected operations. Facilitates near real-time regulatory reporting and compliance audits. Provides foundation for cost-effective integration of process & data as well as process upgrades.
- *Edge* — Reduced TCO frees resources. Litigation risk mitigated.

3. Customer-facing Services

3.1 Customer access to technical publications

A $400 million company, which delivers technical standards, safety publications, design guidelines, vendor and logistics information, and software tools through portals, CD-ROM subscriptions, and printed documents, needed to improve the speed, convenience, and effectiveness of customer access to its library of over 1.5 million documents and product catalogs.

- *Semantic solution* — Concept-based search and guided navigation.
- *Efficiency gain* — Indexing time accelerated by 4X. Reduced customer support costs because of ease-of-use and lowered training requirements
- *Effectiveness gain* — High customer satisfaction, reflected in positive feedback and 300% faster response times.
- *Edge* — Time to deploy new applications is nearly 2X faster, plus staff is able to handle 5-10 new products concurrently.

3.2 Customer self-service call center

A manufacturer needed to improve quality and reduce the cost of customer service. Complex products cause increased need for customer service, which is costly to provision, even with outsourcing. Cross-industries 40-80% of customers say they are dissatisfied with customer support. Also 2/3 change contractor after unsatisfactory service.

- *Semantic solution* — Ontology-based self-service access to integrated content combined with case-based reasoning across similar problems to provide customer self-service.
- *Efficiency gain* — Electronic self-service reduces costs by more than 1/2. Cost savings through call avoidance was $3M in first year. Maintenance of knowledgebase at 1/5 person-year.
- *Effectiveness gain* — 3/4 of the customers and 2/3 of the employees rate intelligent customer self-service as “good” or “very good.”
- *Edge* — Positive ROI in less than 12 months. Risk of customer defections mitigated.

3.3 Health care delivery

Healthcare providers need to link disjointed data from heterogeneous sources in order to improve data quality, reduce errors and omissions, streamline operations, comply with federal mandates, and provide effective patient care. Current systems are fragmentary, disconnected, and capture overlapping and redundant information.

- *Semantic solution* — Business ontology-enabled 360 view of all patient data, treatment status, financial records, and history.
- *Efficiency gain* — Ontology-based semantic solution is faster, less labor-intensive to develop. Risk of litigation mitigated.
- *Effectiveness gain* — Better patient care decisions and diagnoses resulting from access to integrated total patient information.
- *Edge* — Better patient care. Reduced exposure to litigation.

4. Managing Output

4.1 Enterprise publishing

A property and casualty insurance firm needed to accelerate process and reduce the effort required to compile claim summary reports (required for litigation) from disparate data maintained on five separate legacy systems. Solution requires a common user "front-end" across all legacy systems.

- *Semantic solution* — Semantic information mapping to bridge disparate systems and provide unified portal access to integrated information.
- *Efficiency gain* — Time required to generate a claim summary reduced from 8-hours to 25 seconds. Annual labor savings exceeds 250,000 person-hours. Annual process savings in the millions.
- *Effectiveness gain* —
- *Edge* — 4-month deployment. Positive ROI in 6 months with "very substantial" return on investment within the first year. Risk of late, missing, or incorrect summary information for litigation reduced.

4.2 Enterprise publishing

Global corporations need to improve the effectiveness of lifecycle product communication while taking cost, time, and effort out of the process. Technical knowledge management spans content creation, content management, localization, cross-media publishing, and project and process management across geographic regions, business units, and supply-chain relationships. In support of PLM and global CRM, the strategy is to create once, localize once, store once, and deliver in multiple ways including web, CD, email, and print.

- *Semantic solution* — Ontology-integrated platform and content for PLM and CRM technical knowledge creation versioning, and cross-media delivery. Automated content metatagging. Semantic technology based generation of multi-lingual text, graphics, documents, web pages, and interactive media. Knowledgebase and semantic web service-based integrated collaborative supply chain for print and cross-media content production.
- *Efficiency gain* — Save 1/4 to 1/2 of enterprise's media communication cost through sourcing and procurement standardization, integrated communications management, and semantic technology enabled process improvements. Labor savings of 1/4 to 1/2 affects authoring, graphics and illustration, production, and administration.
- *Effectiveness gain* — Time-to-market accelerated by 2-to-10 times. Enables concurrent support for multiple product launches in multiple geographic regions using multiple media channels concurrently.
- *Edge* — ROI of semantic technology-based solution is 2-5X faster.

4.3 Knowledge-centered engineering

Across the engineering lifecycle, a part design can translate into hundreds of drawings, schematics, and documents prepared for different disciplines, or usages at different stages. Currently, the workflow is document-centric, utilizing CAD and CAE tools as electronic pencils for creating and recreating documents. As project size and complexity grows, internal document maintenance and management consumes 80-90% of resources.

- *Semantic solution* — Ontology-based engineering captures, represents, and maintains total product knowledge in a language-neutral, federated repository. Semantic applications generate all categories of engineering drawings, specifications, project documents, and technical literature as needed.
- *Efficiency gain* — Up to 5-10X faster design, build cycle. Up to 5-10X reduction in project costs. Up to 5-10X fewer engineering resources.
- *Effectiveness gain* — High-performance. Knowledge-centered engineering can deliver the same result in less time using significantly fewer engineering resources. Knowledge-centered engineering enables effective control of larger and more complex than can be handled with conventional methods.
- *Edge* — Return on investment during each stage of the engineering lifecycle. ROI comes from taking huge amounts of labor, cost, and time out of the process. Lifecycle knolwedge-base removes errors and inconsistencies; gives visibility to all parts and phases of the project; and stops knowledge erosion due to personnel changes.

4.4 Research publishing

Brokerage firm needed to integrate a range of proprietary information and 3rd-party commercial financial information services to produce a query-driven, near real-time consolidated research report utilizing rich media technology.

- *Semantic solution* — Meta-modeled information integration (virtual database) for research content access and aggregation.
- *Efficiency gain* — Less cost and labor develop. E.g. common API for all developers, vs. one-offs.
- *Effectiveness gain* — Faster time to market.
- *Edge* — TCO reduced by 50% vs. conventional development. Positive ROI in 6 months.

5. Smart products and services

Examples we reviewed in this category are in various stages of research and development. Not deployed operationally, data on business value was not available.

6. Design & Manufacture

6.1 Network design optimization

A telecommunications company needed to reduce the cost, time, and risk to design "surviable" network additions.

- *Semantic solution* — Agent-based network modeling and genetic algorithms for evolutionary design optimization.
- *Efficiency gain* — Average network design time decreases from 2 months to 2 days. Cost savings (vs. manual design) averages $1-10M per network.
- *Effectiveness gain* — Results equal or exceed expert designers within a fraction of time.
- *Edge* — Strong ROI. Evaluating more complete spectrum of solution alternatives mitigates risk of sub-optimum design.

6.2 Product development

Auto manufacturers need to reduce the cycle time, cost, and labor required to develop new parts and product designs.

- *Semantic solution* — Ontology, rule, and parametric based design advisors
- *Efficiency gain* — 20-40% gain in productivity. 25-95% savings in total cost of design.
- *Effectiveness gain* — Design advisors in use and proven effective for transmissions, crankshafts, powertrain components, drive line layouts, rack density, hood and decklid, stamping dies, direct field vision, tool design, injection molding, and many other applications.
- *Edge* — 50-75% gain in quality attaining 6-sigma certification.

6.3 Virtual product development

R&D unit of first-tier aerospace manufacturer needed to reduce product development cycle time and cost, while improving quality.

- *Semantic solution* — Ontology-integrated collaborative infrastructure, 3-D modeling, simulation and virtual reality tools
- *Efficiency gain* — Product cycle time and costs reduced by more than 50%.
- *Effectiveness gain* — Eliminate the need to build costly prototype hardware. Produce more efficient, supportable, higher performance systems with first-time quality.
- *Edge* — Customer and stakeholder access to virtual prototypes improves product quality and mitigates development and business risk.

7. Research

7.1 Intellectual property R&D

Manufacturer needed to improve the process of research and development of patentable inventions.

- *Semantic solution* — Knowledge-base enhanced guided innovation discovery and evaluation process; facilitated web and patent search; document summarization.
- *Efficiency gain* — Productivity gains through ready knowledge access and process improvements.
- *Effectiveness gain* — Enhanced rate of conversion of ideas into patents into products (time-to-market). Integrated access to 2000 scientific web sites and 9000 scientific effects.
- *Edge* — Improved ROI on R&D. Comprehensive market and competitive research mitigates new product development risk.

7.2 Investment information access

Investment decision-makers need tools that enable fast, comprehensive access to all relevant research. If source remain dispersed across multiple search services with inadequate aggregation, classification, search, and filtering capabilities, the process of finding and utilizing the right information remains slow and tedious, placing the investment decision-maker at a competitive disadvantage.

- *Semantic solution* — Semantic information integration, industry taxonomies, source aggregation, auto- categorization, agent-based filtering, and concept-based search
- *Efficiency gain* — Reduced labor and time to assemble the relevant research needed to support investment decisons
- *Effectiveness gain* — Research productivity gain through improved recall (integrated access to the right sources) and precision (relevant results)
- *Edge* — Accelerated time to deal. Better investment decisions faster.

7.3 Litigation discovery

The discovery phase of the litigation process is critical for preparing a winning argument. Litigation teams must examine volumes of documents in a short period of time in order to identify all that are relevant to their case. Failure to identify and examine all relevant documents can incur significant risks to firm and its client.

- *Semantic solution* — Ontology-based directed discovery applies a knowledgebase of legal expertise together with case-specific criteria to automate scanning, evaluation, and identification of all documents relevant to the case out of the total collection. Benchmarking to establish accuracy follows set-up.
- *Efficiency gain* — Up to 2-3X faster document review. Up to 2-3X more accurate, comprehensive, and consistent review process across all stages of litigation.
- *Effectiveness gain* — Accelerates document review. Increases review accuracy -- both precision and recall. Results of benchmarks vs legal and paralegal researchers are as good as humans in precision (i.e.: when system says a document is relevant, it is.) but much better in recall Semantic/ AI-based system misses between 80% and 95% fewer actually relevant documents than humans typically do. Humans are good when they tell you: these 10 documents are relevant; but, they are terribly bad at finding all of the relevant documents.
- *Edge* — Return on investment comes from acceleration of discovery process, reduced cost to litigate, and improved odds (competitive advantage.) Mitigates legal and financial risks. Using more complex metrics that calculate relative odds, the knowledge-based process is between 10 and 20 times more effective than manual attorney review.

8. Input Management

8.1 Source data automation

A health insurance firm receiving more than 400,000 claims documents daily needed a better way to manage document processing in order to meet customer service commitments. Data was manually keyed from paper in a process that was labor-intensive, expensive, prone to errors and exception, and resulted in slow payments and increased CRM burden. Solution required automation of capture, classification, and delivery.

- *Semantic solution* — Document scanning with OCR, auto-categorization, automated information extraction, data validation, and rules-based routing.
- *Efficiency gain* — Labor reduction by nearly 1/2. Claims processing time reduced from 9-10 days to 1-2 days.
- *Effectiveness gain* — Eliminates manual sorting and keying of data from documents, improving data accuracy to 99%. Automated capture, information extraction and indexing increases productivity by 5X.
- *Edge* — Positive ROI in ≤ 12 months. Risk of mis-handling claims reduced through improved data accuracy.

8.2 Publishing repository auto-tagging

Scholarly publisher needed faster, more consistent tagging of content in order to leverage content assets and come to market faster with new, related products

- *Semantic solution* — Text mining, and automatic tagging and classifying facts and events using parts-of-speech tagger, noun phrase and verb phrase grouper, verb and noun pattern extractor, named entity recognizer, template and rule-based relationship extractor.
- *Efficiency gain* — Automated 2/3 of tagging while improving accuracy and consistency.
- *Effectiveness gain* —Integration with corporate workflow increased utilization and value by 20%. Increased scalability of content assets through efficient acquisition /integration process.
- *Edge* — Improved return on content assets.

8.3 Repository access, search and retrieval

University press needed to reduce cost to index and categorize >12M articles and 300 online journals while providing better online discovery-search and retrieval capabilities.

- *Semantic solution* — Ontology-enhanced scaleable content tagging, and taxonomy management, and discovery-search.
- *Efficiency gain* — Reduced labor required to index assets.
- *Effectiveness gain* — Automate 22,000 categories and 300,000 concepts with quarterly refresh of taxonomies.
- *Edge* — Improved search efficiency. Greater customer satisfaction.

9. Supplier-facing Processes

9.1 Invoice reconciliation

A $4.3B business unit needed to reduce the cost, time and effort required to reconcile disputed invoices with its supply chain partners. Current process was labor intensive with 95% of time consumed with information collection and aggregation. They had a growing backlog of past-due disputes, and were only managing >60 past due. Solution required fast, flexible integration with multiple external systems.

- *Semantic solution* — Semantically modeled integration of information and processes, linking internal and external systems.
- *Efficiency gain* — Labor reduction from 20 agents to 15, to accomplish more. 8-day reconciliation cycle reduced to <1 day. Automation delivers >$1million in cost savings per year.
- *Effectiveness gain* — Managing >15 day past due. Much higher levels of customer service. Collectors now dedicated to dispute resolution.
- *Edge* — 6-month deployment. Positive ROI in less than 9 months with a +500% return on investment within 1 year. Bad debt risk and capital costs mitigated by more timely collection.

9.2 Freight routing and schedule optimization

Airfreight company needed to optimize scheduling of freight movement to improve delivery, minimize freight handling, and increase revenues.

- *Semantic solution* — Agent-based modeling, simulation, and adaptive network routing.
- *Efficiency gain* — Amount of freight handling on ground reduced by 20%. Amount of freight handled multiple times reduced by 75%.
- *Effectiveness gain* — Increased cargo revenue by $10 M, without incurring additional labor costs.
- *Edge* — Positive ROI in less than 12 months, with substantial return on investment.

9.3 Supply chain inventory and cycle time optimization

A major consumer goods company needed to reduce cycle time and inventory levels across its supply chain from factory to consumer. Solution must optimize performance across an increasingly dense and interconnected (non-linear) supply web.

- *Semantic solution* — Agent-based supply chain modeling, adaptive simulation, and machine learning to determine and implement optimal supply chain policies.
- *Efficiency gain* — Company saves $300M per year.
- *Effectiveness gain* — Results reduce both cycle time and inventory levels by 50%.
- *Edge* — Strong ROI. Investment to research, prototype, test, and deploy is around $3M for a return of 100X in the first year of operation. Simulation and ongoing fact-based evaluation of alternatives validates strategy and manages risks

10. Intelligence

10.1 Analyst and decision-maker Q&A

Agencies concerned with security and intelligence need technology that can analyze and integrate massive data and content from disparate sources to deliver near-real time, reliable, scaleable and actionable information analysis and decision-making.

- *Semantic solution* — Integrate, manage, interpret, Q&A reasoning across massive amounts of data and content. Knowledgebase combining ontology (reference knowledge, domain models, concepts, relationships) and entities or instances; automatic metadata enhancement; machine learning and knowledge discovery; semantic portal.
- *Efficiency gain* — Human-supervised auto-categorization of web pages is 10-20 times faster (>1000 pages per day with higher quality metadata generation. Auto-metatagging of news feeds is 100-1000 times faster + no supervision needed. Metatagging of internal/enterprise research is 10 times faster. Content editors can supervise auto-tagging of 5-50 times more sources per day.
- *Effectiveness gain* — Staff able to manage 100s to 1000s more sources, with storage and search scaleable to millions of document. Auto-categorization accuracy around 90 percent. Relevance: near 100%, as good as human reviewers, but with vastly greater coverage, plus personalization and semantic associations. Near real-time search/ personalization of new content and breaking news. Inclusion of new content sources within 2-8 hours. Development of new domain models in 2-8 weeks leading to useful results (approximately 1-million entities and relationships. Q&A at 1M queries per hour per server, with extraction processing/delivery time at several documents per second.
- *Edge* — Vulnerability from information omission and overload reduced.

10.2 Business intelligence

Having implemented ERM, SCM, and CRM systems, organizations are seeking to monitor and manage their business processes in order to better understand status, identify bottlenecks, reallocate resources and improve performance.

- *Semantic solution* — Business ontology integrates process and data to create "virtual data warehouses" and provide rich multi-dimensional linkages for business intelligence and business activity management applications.
- *Efficiency gain* — Reduced development cost and time. Easier and less expensive to operate and maintain.
- *Effectiveness gain* — Integrates legacy data input to business intelligence packages. Allows for a richer set of questions to be asked and answered and actions taken than are possible with most BAM packages.
- *Edge* — Ability to explore tough questions of about how, why, and what-if leading to better business decisions faster, based on comprehensive information in context.

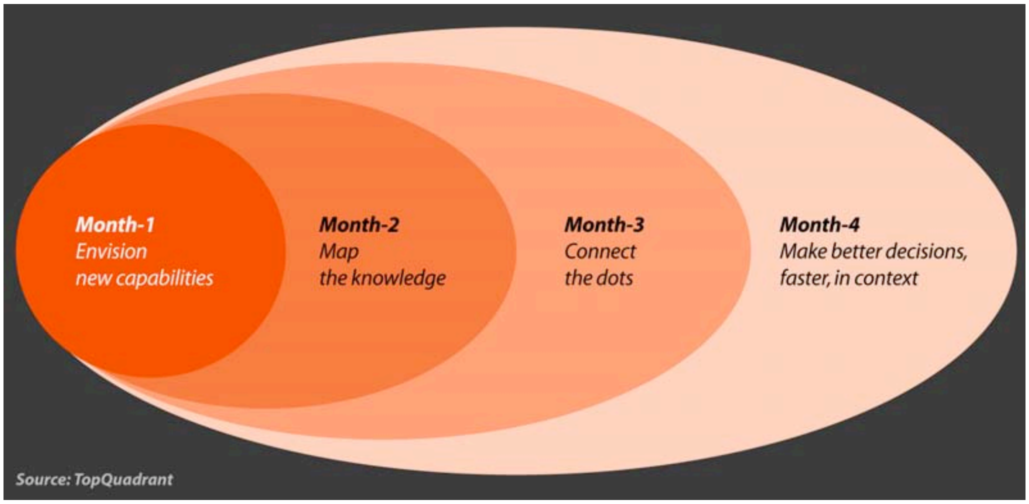
10.3 Business performance management

Asset management firm needed unified reporting and business analytics to manage performance of restructured business lines. Legacy systems were fragmented across business units. Solution must be economically sustainable and quick to implement and evolve.

- *Semantic solution* — Ontology-based information integration for business intelligence and program management.
- *Efficiency gain* — 2X faster development time than conventional EAI middleware approaches with much.
- *Effectiveness gain* — Bridging data and system sources through an information integration server provides common real-time connectivity from legacy to business intelligence solutions, enabling fast, complete reporting and analytics.
- *Edge* — Reduced TCO.

5. Proving the Value of Semantic Technologies

Figure-14: Proving the value of semantic technologies in 100 Days



Start with a 100-day roadmap

Figure-14 provides a simple roadmap to follow to prove the value of semantic technologies in as little as 100 days. This roadmap has the following steps:

- *Month-1: Envision new capabilities*. Learn about semantic integration. Embrace open standards — semantic web stack. Train in knowledge engineering.
- *Month-2: Map the knowledge*. Represent knowledge about data, applications, and processes in a business ontology. Establish context, a 360-degree view, the total picture. Achieve semantic interoperability by integrating the knowledge, not the code, to bridge information sources and link processes and functionality.
- *Month-3: Connect the dots*. Link data sources, applications, and processes to the business ontology using web services and connector technology. Create composite applications with portal building blocks linking them to the business ontology.
- *Month-4: Make better decisions, faster, in context*. Build composite applications with one interface that interconnects and interacts with multiple applications and data sources.

Tips for travelers

Semantic solutions are never an “all or nothing” proposition. There is virtue in starting small, solving a problem with immediate payback, then moving on to tackle the next problem. As the business moves forward, its business ontology is based on a technology designed to adapt and grow as needs change.

Select the right problem

Avoid the "big bang." The tendency of many program managers is to outline a definitive blueprint that is too broad and involves integrating too many business processes, systems, content and data sources, and requirements that are in constant flux. Stay pragmatic and avoid the big bang. Here's why. A scope that is too broad is difficult to manage and tends to lower the impact and value of individual requirements. Also, the probability of successful project execution decreases with the number of integration points and decision-makers involved. Big projects tend require significant expenditure over a long period of time before yielding positive returns. This increases risk. For example, business requirements change and are virtually impossible to predict accurately outside of a 10-to-12-month window.

Beware of long-term commitments. There is a natural tendency on the part of software vendors and service companies to foster large-scale implementations. Manage your vendor relationships responsibly. Large-scale deals tend to favor software vendors. Long-term implementations tend to favor service vendors. Long-term project engagements are risky to manage and do not provide the proper incentives for timely project completion.

Instead, tackle smaller, yet practical, projects with well-defined process objectives that produce measurable financial impact. Limiting scope in terms of time and budget has the following benefits:

- Tactical goals are attainable
- Shorter time frames stabilize requirements and assumptions
- Single-budget-period projects force completion before allocating new funds
- Tight scopes improve training and change management for all involved
- Iterative cycles mitigate project risk by providing a clear exit point

Focus on the value of semantic technologies

Understand the business value you are seeking before you commit to implementation. Identify the questions to be answered and decisions to be made that demand knowledge discovery, integration of resources, knowledge-based reasoning, or semantics-based communication. Build the business case. Test before you invest.

Put the knowledge in the data. If you can describe a problem as facts, you don't have to do it in code. Your ontology is what matters. Two rules are worth remembering about semantic integration:

Rule #1 is "Whatever it is you want to integrate, first integrate the knowledge about it."

Rule #2 is "Represent the knowledge about what your integrating separately from what you're integrating."

When it comes to your business ontology, there is no "one size fits all." Tools and resources exist. Some of these are better and more useful than others. But, currently there is no ready-made whole solution off the rack.

Apply the Right Skills

Get the training you need. Knowledge engineering is a skill that is in somewhat short supply. Your people will need to understand the vision of the semantic web and its emerging standards. They will need training in semantic technologies, especially how to build and manage ontololgies. Interestingly, the best people to do your knowledge engineering may not be IT professionals, but rather, people with business domain expertise.

Should you bring in experts to jump-start the process? In several areas you should consider doing this, for example:

- Briefings to build awareness and educate senior management.
- Training for internal personnel in semantic technologies.
- Envisioning solutions — working with internal personnel to visualize capabilities and solutions to problems. Often business users don't know what they want until they see what they can get. Outside expertise can be helpful to identify the right solution and build the business case for it.
- Picking the right platform and tools for the job.
- Prototyping solutions — working with internal staff to quickly demonstrate business value.
- Solution delivery — outsourcing for resources to engineer solutions may be cost-effective.

Summary

This completes TopQuadrant's special report on the business value of semantic technologies. We began by explaining what semantic technologies are, their functions and capabilities, who makes them, and how semantic technology differs from more conventional information technology. Next, we explored how semantic technologies drive key elements of business value, and concluded that the benefits are compelling and will power significant market growth this decade. We supported this position with case examples that show where semantic technologies are being applied today and are delivering significant ROI. Based on these findings, our recommendation to businesses and government agencies examining potential technology investments is not to wait. Then, we concluded by mapping out a simple roadmap that any enterprise can follow to quickly prove the value of semantic technologies for their business in as little as 100-days.

About TopQuadrant, Inc.

Mills Davis

The author of this report is Mills Davis, managing director with TopQuadrant. Mills specializes in strategic applications of next-wave semantic, information, content, and media technologies. His clients include leading technology manufacturers, global 2000 corporations, and government agencies. As a researcher and industry analyst, Mills has authored more than 100 reports, whitepapers, articles, and industry studies. He is a frequent speaker at industry events, and his work has appeared in more than 30 trade magazines.

TopQuadrant

TopQuadrant, Inc. is a leading consultancy focused on the emergence of semantic web technologies and solutions. We provide knowledge, resources, and services for semantic solutions from research to implementation. As business consultants and technologists, TopQuadrant acts as a trusted intermediary to help enterprises envision, architect, plan and realize knowledge-based solutions that deliver significant ROI. Our professionals bring expertise in artificial intelligence, object technology, adaptive systems, ontology engineering, knowledge and content management, publishing and media, semantic web and grid computing, and methodologies for knowledge, software and systems engineering. TopQuadrant has developed a set of unique tools, methodologies, and services that jump-start successful building of semantic solutions, and help clients transition to next generation, semantically integrated systems while sustaining and optimizing their investments in current technologies. These include:

- **TopConnexion™**, a multi-company knowledge service that conducts research; publishes case studies, technical assessments, and whitepapers;and produces workshops and conferences.
- **TopDrawer™**, a comprehensive knowledgebase of semantic technology capability cases — solution patterns for ontology-based applications.
- **TopMind™**, executive briefings on semantic technology; hands-on trainings in semantic web standards, languages, tools and ontology development.
- **Solution Envisioning**, scenario-driven workshops to explore system options and design future solutions through analogies and examples using capability Cases.
- **Semantic Solution Development Services**, including optimal technology and vendor selection, ontology development, and full implementation of semantic solutions.

Digital Harbor

www.dharbor.com

Digital Harbor has in part sponsored this TopQuadrant special report. Digital Harbor develops and markets PiiE™, a leading platform for composite applications. PiiE™ fuses semantic, portal, process, and integration technologies to enable enterprises to build, deploy and use inter-networked applications in a faction of the time and cost of other approaches. Developed for major intelligence agencies and now, commercial enterprises, PiiE™ provides superior information integration and management by extending existing applications and data sources into dynamic portal-accessed composite applications.