Business Tools: *The Next* QUANTUM LEAP

By William M. Baker, CMA, and F. Douglas Roberts, CPA

Joanna Forrest has just found an IMA conference she'd like to attend. Using the Internet, she instructs the software "agent" on her PC to gather information about the event. In seconds, the agent provides a detailed meeting agenda, a choice of flights to and from the conference city, and hotel and car rental information, including Joanna's preference for a room facing east. The automated agent also recommends three restaurants for dinner. It displays the conflict between the meeting dates and Joanna's dentist appointment and even searches the schedules of all of her colleagues to determine who else might be attending the conference. The final screen Joanna sees is a prompt asking her whether she wants to register for the conference. She clicks "Yes," and, after negotiating the best rates, the agent books the flights and reserves the hotel room and car. The agent contacts Joanna's dentist and reschedules her appointment, then e-mails her colleagues who are attending to tell them to look for Joanna when they arrive. Last, Joanna receives a message telling her that her calendar has been updated and that confirmation numbers and turn-by-turn driving directions for the trip have been sent to her computer.

If you think this scenario sounds far-fetched, activities such as these (and those shown in Table 1) already take place on the Semantic Web. Tim Berners-Lee, director of the World Wide Web Consortium (W3C) and widely credited as the inventor of the Web, sees the Semantic Web—where computers talk to and work with each other—as the next logical step for the Internet. Although only five million to 10 million of the billions of documents on the Web currently enable Semantic Web activities, research and business in this area are growing, and change will come quickly. Here's what management accountants need to know.

The Semantic Web Is a Data Web

While Berners-Lee describes the current Web as a *document* Web, he refers to the Semantic Web as a *data* Web. At first, moving from a document Web to a data Web seems backwards; after all, documents are composed of data. But documents are actually made up of words, files, and text that are labeled by means of HTML tags. Using these tags, a person can create documents that Web browsers (and other people) can read.

But computers can't read documents meaningfully. Computers read data. In the World Wide Web of data, computers can read everything and make decisions in much the same way that people make decisions based on the information in documents. If you've ever searched for a particular document on the Web and instead received links to millions of documents that weren't helpful, you begin to understand that a *data* Web would be far more useful.

This data Web isn't another separate World Wide Web. Instead, think of it as an additional layer on top of (or within) the existing Web. To move from human-only consumption to one where both humans and computers can work with the same information, data will link to computer-readable definitions that provide the meaning of the data along with inference rules for logically reasoning with this information. By linking data with their meanings—what each piece of data is and how pieces relate to one another—automated software agents can solve well-defined problems using systematic operations. This will lead to a variety of uses whereby intelligent computer agents will be able to perform many tasks related to searching for, retrieving, filtering, integrating, and presenting information that humans must today do themselves. Computer agents will even be able to take action in certain situations—or at least provide the user with a list of options. These agents will give people the opportunity to spend more time on those innately human tasks: analysis and judgment.

How It Works

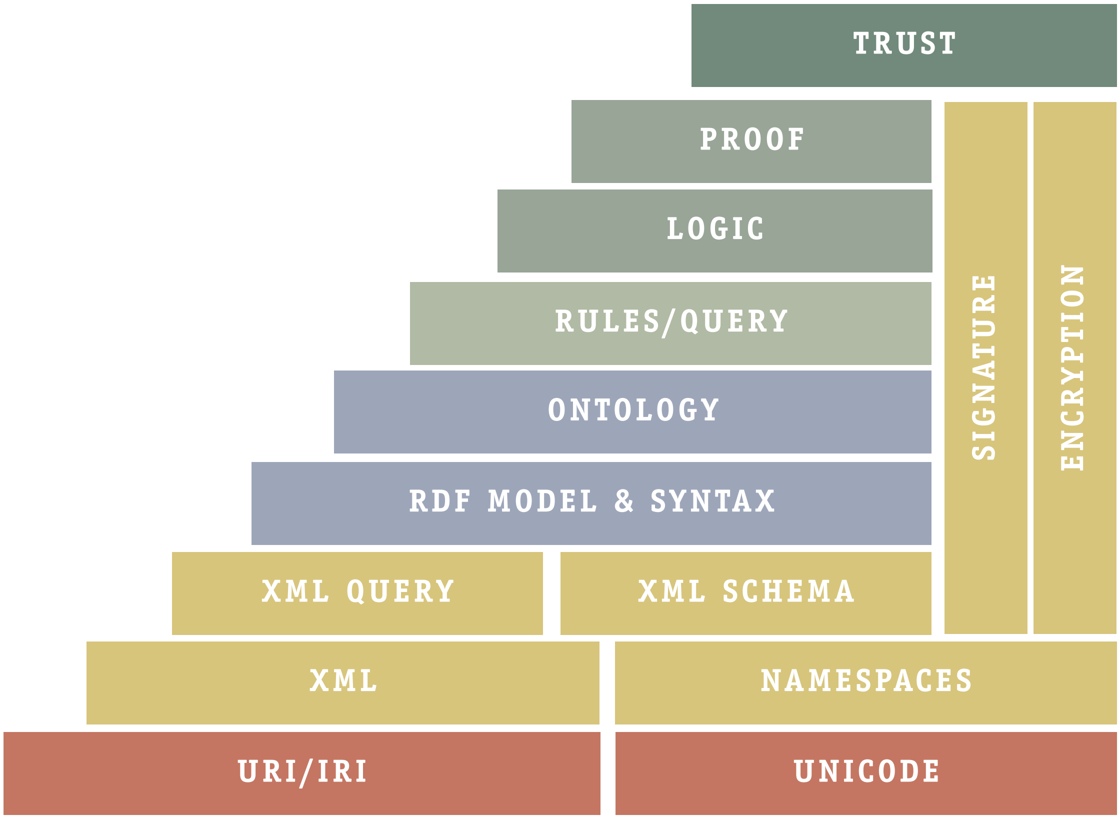
The building blocks of the Semantic Web are shown in Figure 1. To some extent, though sometimes only experimentally, they already exist on the Web today. To enable the Semantic Web, however, these blocks must be able to work together to achieve a state of *interoperability*—where computers can work together without human intervention.

The basic blocks, URI/IRI and Unicode, are already common. Most people are familiar with URIs (Uniform Resource Identifiers, also called URLs, which were cocreated by Berners-Lee) as the "addresses" that identify locations on the Internet. IRIs (International Resource Identifiers) are simply another way to name those same addresses that will allow for more characters than URIs. Unicode refers to the set of universally accepted ways that computers display and represent text in documents on the Internet. The major building blocks for enabling the Semantic Web are XML, RDF, and OWL. Although most people don't yet understand these three blocks, the W3C endorses and supports them. Let's look at each one in more detail.

XML. Many accountants are already familiar with XML (Extensible Markup Language). While HTML uses tags that determine how documents are displayed, XML employs tags to define elements of data. Users can define their own sets of tags, and, if those sets comply with XML rules, they become acceptable as part of XML. Such tag sets use namespaces—collections of unique nametags—to eliminate the possibility of ambiguity when computers read documents. The tag set that accountants use most is XBRL (eXtensible Business Reporting Language), a special tag set that describes financial data commonly used in financial reporting. Since 2005, the Securities & Exchange Commission (SEC) has encouraged publicly traded companies to submit their reports in XBRL format. It voted unanimously in May 2008 to propose a requirement that "approximately the 500 largest" companies begin submitting XBRL-tagged financial statements in 2009, with all publicly traded companies submitting them by 2011. This represents the accounting profession's first step toward enabling the Semantic Web.

RDF. The next big building block is RDF (Resource Description Framework), a standardized, commonly accepted model for metadata (data about data). RDF is much newer than (and not yet as developed as) XML.

Figure 1: Building Blocks for the Semantic Web



Source: Released into the public domain on Wikipedia by the diagram's author, Sebastian Faubel.

While XML identifies elements on the Web, standardized RDF sets (such as Dublin Core) make statements about the traits of those elements in a subject-predicate-object format. For example, an RDF statement would indicate that a particular XML tag "Net Income" has the trait of being "GAAP-based" (based on Generally Accepted Accounting Principles). Additional specific RDF descriptions would also indicate that the "Net Income" tag is based on "FIFO" (First-In, First-Out) inventory calculations, "Zero" extraordinary items, and "Straight-Line" depreciation. The RDF would specify that accounting basis, inventory flow assumption, presence of extraordinary items, depreciation method, and many other elements are used to describe what XML calls net income. Thus, RDF can provide a highly detailed definition of net income that any computer can read.

OWL. The third major building block for enabling the Semantic Web is ontology. XML identifies the elements, RDF describes the traits associated with those elements, and ontology describes how the elements (and their traits) relate to each other. The ontology sets endorsed by the W3C are collectively called OWL (Web Ontology Language). OWL sets such as FOAF (Friend of a Friend) provide standardized, computer-readable representations of relationships. In short, these relationships represent knowledge, and, using them, computers can think and reason to the point where they can completely understand the elements of business transactions and the traits associated with those transactions. This means that if XML, RDF, and OWL are tested logically—and it's proven that they interact accurately—computers can conduct business with other computers without human intervention. Indeed, computers can execute "person-less" financial transactions.

Benefits of a Turbo-Charged Web

Who will benefit from the Semantic Web? Essentially, everyone. For example, in healthcare, most or all of a patient's medical records are stored on computers in various physicians' offices and locations. A patient checking

Table 1: Accounting Examples of Person-less Semantic Web Activities

SEMANTIC WEB AGENT	PERSON-LESS ACTIVITIES
Auditing Agent	Verify all documentation—internally and externally
	Substantive and attribute testing
	Monitor and data mine all documents and data in search of fraud and errors
	Administer internal controls
Communication Coordinator Agent	Enable any two (or more) systems to become a single database management system (DBMS) by reconciling all queries (thus, all data are in one single data store)
	Integrate all computer systems across all platforms and boundaries, regardless of applications
Financial Reporting Agent	Monitor the external environment, searching for enhancements to Management's Discussion and Analysis
	Compile financial reports instantly
	Update/upkeep XBRL
Marketing Agent	Evaluate marketing effectiveness (with both financial and nonfinancial measures) By campaign By product/service By employee
	Redefine customer service
	Recalibrate balanced scorecards
Opportunities Agent	Evaluate current and potential investments
	Facilitate brainstorming
	Provide economic analysis of the potential for outsourcing
	Create forecasts and projections
	Monitor competition
	Revise and update internal controls
	Continuously monitor all risks
Personnel Agent	Hire, manage, and develop staff
	Conduct background checks
	Staff and schedule projects properly
	Merge personal lives with business demands when scheduling
	Evaluate employee performance
Purchasing Agent	Select vendor
	Negotiate purchase
	Schedule delivery
	Facilitate payment
	Optimize inventory
	Develop/enhance supply chain
Tax Agent	Continuously monitor (and determine relevance of) tax codes, cases, rulings, and laws worldwide

into an emergency room can be diagnosed more effectively and more quickly via the Semantic Web because the hospital can issue a query across the Internet and immediately gather all of the patient's electronic medical records to better understand his or her symptoms. In academic arenas, researchers can publish all of their data sets on the Web, and other researchers can issue queries to verify and extend their own research results. CFOs can interweave their financial data with economic data from the same time period to better understand and explain their company's performance.

Accountants also benefit from the Semantic Web. Both internal and external auditors can place software agents on the Web that study the auditing concerns and processes of other companies while continuously auditing all internal corporate transactions. Management accountants not only can compare actual and budgeted results in real time for any account, but they can also query the entire Internet to benchmark the processes of other companies or search for new cost advantages. The variances between actual and budgeted results are always more meaningful when they're enhanced by external data. How can a company truly know it is doing well until it compares its results to the results of other companies?

The requester will search the Internet, as will the provider, and when they discover each other based on specific parameters, the two companies will arrange to conduct business together without human interaction.

All these tasks can be accomplished on the Semantic Web using routine search-engine queries. For example, an accountant can issue a query asking his organization's software agent to present day-by-day cost efficiency variances over a six-month period. The agent will also be instructed to compare the variances to benchmarks that are gleaned from companies throughout the world. The accountant is then free to work on other tasks while the software agent completes the cost-efficiency report.

A Closer Look at Interoperability

As shown in Figure 1, the combination of XML, RDF, and OWL must be logically tested before these queries can function. Then, as more and more companies create *proven* Semantic Web content, they'll be able to conduct business with each other. This interoperability demonstrates the ultimate in Semantic Web activity.

Here's how it works: One company's computer, the **requester**, will "send" software agents onto the Web that describe the goods or services the company needs. Meanwhile, the **provider**—a company that produces the goods or services that the requester needs—will send its software agents onto the Web to "advertise" its goods or services. The requester will search the Internet, as will the provider, and when they discover each other based on specific parameters, the two companies will arrange to conduct business together without human interaction.

Companies that want to take advantage of interoperability must create and demonstrate three elements. Initially, they must demonstrate *capability*. This is achieved by using OWL to formally describe the functionality of the goods or services that they provide or need. Second, interoperating companies must specify *choreography*. Choreography describes the interaction style of a company, specifying the input and output formats that the organization uses (or requires) to execute e-commerce transactions. Finally, the company must specify how the transaction will be orchestrated. *Orchestration* describes the policies and procedures for accepting and completing transactions, such as anticipated delivery times, expected costs, quality specifications, and error-handling processes. The more accurately and specifically companies can demonstrate capability, choreography, and orchestration in their Semantic Web software agents, the more likely they'll be satisfied with their "person-less" business transactions.

Even after companies have created proven semantic agents, they must overcome three additional barriers before they can achieve satisfactory interoperability. The first barrier is *vocabularies*. Though XML, RDF, and OWL use standardized formats, businesses often describe their goods and services using different terminologies. For example, a retailer might send a software agent onto the Web to search for button-down shirts. Naturally, the company would be careful to specify desired colors, quantities, and sizes. A supplier issues its software agent onto the Web. All three required elements—capability, choreography, and orchestration—are present. The supplier has the colors, quantities, and sizes that the retailer needs. But a problem arises: Each of them defines a "button-down" shirt differently. The supplier uses the RDF tag "button-down" to specify shirts that button down the front, using five or six buttons. The retailer uses the button-down tag to specify shirts with button-down *collars*. This difference in vocabularies will impede both companies' satisfaction with interoperability. Thus, another "layer" of the transaction will include very specific definitions.

Second, businesses use different *protocols*. This means that, internally, companies use different software. To be effective, all new software packages must be able to convert (translate) data into data adhering to Semantic Web standards. Microsoft, for instance, has already created software that converts data into XML-tagged data. Competing companies are also racing to create software that easily converts data using RDF and OWL.

Finally, when companies are constructing the orchestration elements of their software agents, they must clearly explain and reconcile their *business processes*, including their credit policies, freight terms, and discount opportunities. Additionally, they must be able to examine the business processes within the software agents of other companies to determine, based on reputations and references, whether the companies (and their software agents) can be trusted.

Levels of Interoperability

Highly effective interoperability occurs on three levels: At the (base) *communication* level, necessary data is defined precisely so that any computer can read it. At the next level, the *syntactic* level, multiple software components interact seamlessly, regardless of differences in application software, operating systems, firewalls, or any other technological concerns. At the third (highest) level, the *semantic* level, companies are able to conduct business together while reconciling all differences in perceptions, assumptions, perspectives, and "hidden" or implied meanings. (The elements of, barriers against, and levels of interoperability are summarized in Table 2).

Table 2: Interoperability

ELEMENTS OF INTEROPERABILITY
Capability
Choreography
Orchestration

BARRIERS AGAINST INTEROPERABILITY
Vocabularies
Protocols
Business Processes

LEVELS OF INTEROPERABILITY
Communication
Syntactic
Semantic

E-Commerce Redefined. XML, RDF, and OWL are sufficiently standardized and sophisticated to render the semantic level available immediately. Once this level is achieved, business will change rapidly and completely. B2B e-commerce, for example, will be redefined. In current B2B e-commerce, business relationships tend to be long term. Companies that conduct B2B transactions together create highly defined business arrangements, and such companies have tightly coupled relationships. On the Semantic Web, transactions will occur quickly, and B2B relationships may be short term. The software agent that searches the Web 24/7 and finds the best business arrangement on one day may find even better ones the next day from two or more providers simultaneously.

Management accountants will face the daunting task of controlling and auditing these complex arrangements while extracting accurate financial, tax, and cost accounting data. Further, when a company's software agent finds the software agent of a potential customer, the company's accountants must devise tests of other agents throughout the Internet that have done business with that customer to verify the consistency and reputation of the customer and its technology. In short, the accountant must create a way for the company's software agents to ask other agents for "references."

Two Immediate Concerns

These complex arrangements need to cross two barriers: language and information security. Many people say that English is the language of the Internet, yet research suggests that some companies and countries avoid using

English. In the new data Web, SQL (Structured Query Language) will become the language of the Internet. Software companies have already begun developing products that translate human-language requests (such as spoken English) into SQL command sets. Unfortunately, these products provide only formal, literal translations; none of them effectively reconciles the cultural and linguistic differences that exist among countries throughout the world.

Information security will be an even greater challenge. For many years, the American Institute of Certified Public Accountants (AICPA) has indicated that information security is the most important technological challenge in accounting. That challenge will be magnified in the Semantic Web. In the current document Web, computers can only *display* data. In the new Web, computers will be able to *read and use* data. Therefore, no software agents or data on the Semantic Web can be trusted unless they're encrypted using the PKI (Public Key Infrastructure) with digital signatures. In the PKI, a company provides a public key (available on the Internet) that other firms can use to encrypt data that they send to the company. At the same time, the company keeps a private key that decrypts the data. This key must be kept secure, never to be shared or made available publicly. As caretakers for the information in their organizations, management accountants will be responsible for encrypting and protecting data on the Semantic Web.

Management accountants will be responsible for encrypting and protecting data on the Semantic Web.

Current State of the Semantic Web

Though only a small portion of the World Wide Web is part of the Semantic Web, every day more websites dedicated to creating and implementing the new data Web appear. Much of the current activity can be tracked at www.semanticweb.org. This wiki demonstrates how XML, RDF, and OWL work and how they function together, and it lists the dates and locations of Semantic Web conferences. The site has also begun soliciting and displaying real-world applications of the Semantic Web, and it has a separate area designed to introduce beginners to the new technology (see the section labeled "Getting started").

Much as earlier technologies such as electronic data interchange (EDI) and basic document-based e-commerce evolved when companies attempted to increase revenue, the Semantic Web will help organizations search for new revenue streams. This is already happening, in fact. For example, individuals who host Facebook pages or Internet blogs often log on to find that new advertisements have been placed on their sites—ads that they knew nothing about and didn't expect to see. How did they get there? Provider software agents search the Web for specific keywords and, when they find them, install the ads. These revenue-driven installations are the first examples of person-less transactions (and they can be alarming to the folks who find them on their sites).

Large companies such as Google and Yahoo! realize the lucrative potential of the Semantic Web. For example, Yahoo! has made all of its search indices and files available publicly as part of its BOSS (Build your Own Search Service) program. Yahoo! believes that the quest for Semantic Web dominance is so important that it's allowing others to conduct free research that they can't conduct alone. Yahoo! will then charge usage fees to the most successful development sites.

Once again, when both the requester and provider agents are in place, change will come very quickly. Not surprisingly, most of the change will be driven by revenue, and it'll likely evolve based on the iron-fisted techniques that drove EDI and early B2B e-commerce in the 1990s. Paperless and person-less transactions are alike in that companies can't execute these transactions by themselves. In early EDI contracts, large retailers would tell suppliers to either "go paperless" with them or they'd find a supplier who would. The suppliers, completely reliant on the large retailers as customers, had no choice but to comply. Usually, they turned to their accountants for help.

Things will be no different as person-less transactions begin their rise to prominence. Management accountants need to be ready. **SF**

William M. Baker, CMA, Ph.D., is professor of accounting at Appalachian State University in Boone, N.C., and is a member of IMA's Catawba Valley Chapter. You can reach him at (828) 262-6200 or bakerwm@appstate.edu.

F. Douglas Roberts, CPA, Ph.D., is associate professor of accounting and director of accounting internships at Appalachian State University. You can reach Doug at (828) 262-6213 or robertsfd@appstate.edu.