**Editor: Steffen Staab**
University of Koblenz-Landau
staab@uni-koblenz.de

# The Semantic Web Revisited



**Nigel Shadbolt and Wendy Hall,** *University of Southampton*
**Tim Berners-Lee,** *Massachusetts Institute of Technology*

In the 50 years since the term AI was coined at the Dartmouth Conference, the digital world has evolved at a prodigious rate. It has produced an information infrastructure that few would have anticipated—with the possible exception of Vannevar Bush,[1] although even he might have thought the scale of achievement extraordinary. Today, the World Wide Web links 10 billion pages, and search engines can divine themes embodied in the links to serve useful and relevant content almost instantaneously.

Fifty years ago it might have appeared audacious to build a global web of information, to deploy semantics on such a scale, and to attempt inference over the resulting components. Fifty years ago, even if you could have explained it, a Semantic Web would have seemed as remote as general AI. Yet today we believe that the Semantic Web is attainable. We are seeing its first stirrings, and it will draw on some key insights, tools, and techniques derived from 50 years of AI research.

## From documents to data and information

The original *Scientific American* article on the Semantic Web appeared in 2001.[2] It described the evolution of a Web that consisted largely of documents for humans to read to one that included data and information for computers to manipulate. The Semantic Web is a Web of actionable information—information derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of "meaning" in which the logical connection of terms establishes interoperability between systems. This was not a new vision. Tim Berners-Lee articulated it at the very first World Wide Web Conference in 1994. This simple idea, however, remains largely unrealized.

A Web of data and information would look very different from the Web we experience today. It would routinely let us recruit the right data to a particular use context—for example, opening a calendar and seeing business meetings, travel arrangements, photographs, and financial transactions appropriately placed on a time line. The *Scientific American* article assumed that this would be straightforward, but it's still difficult to achieve in today's Web.

The article included many scenarios in which intelligent agents and bots undertook tasks on behalf of their human or corporate owners. Of course, shopbots and auction bots abound on the Web, but these are essentially handcrafted for particular tasks; they have little ability to interact with heterogeneous data and information types. Because we haven't yet delivered large-scale, agent-based mediation, some commentators argue that the Semantic Web has failed to deliver. We argue that agents can only flourish when standards are well established and that the Web standards for expressing shared meaning have progressed steadily over the past five years. Furthermore, we see the use of ontologies in the e-science community presaging ultimate success for the Semantic Web—just as the use of HTTP within the CERN particle physics community led to the revolutionary success of the original Web.

## A growing need for data integration

Meanwhile, the need has increased for shared semantics and a web of data and information derived from it. One major driver—one that this magazine has reported on extensively—has been e-science (*IEEE Intelligent Systems*, special issue on e-science, Jan. 2004). For example, life sciences research demands the integration of diverse and heterogeneous data sets that originate from distinct communities of scientists in separate subfields. Scientists, researchers, and regulatory authorities in genomics, proteomics, clinical drug trials, and epidemiology all need a way to integrate these components. This is being achieved in large part through the adoption of common conceptualizations referred to as *ontologies*. In the past five years, the argument in favor of using ontologies has been won—numerous initiatives are developing ontologies for biology (for example, see http://obo.sourceforge.net), medicine, genomics, and related fields. These communities are developing language standards that can be deployed on the Web.

Many other disciplines are adopting what began in the life sciences. Environmental science is looking to integrate data from hydrology, climatology, ecology, and oceanogra-



*Published by the IEEE Computer Society*

## Resource Description Framework

RDF assigns specific Universal Resource Identifiers (URIs) to its individual fields. Figure A is an example RDF graph from the W3C RDF Primer (www.w3.org/TR/rdf-primer), showing a representation for a person named Eric Miller. As we create an RDF graph of nodes and arcs, a URI reference used as a graph node identifies what the node represents; a URI used as a predicate identifies a relationship between the things identified by the connected nodes. So, the RDF in Figure A represents

- individuals—such as Eric Miller, identified by http://www.w3.org/People/EM/contact#me;
- kinds of things—such as Person, identified by http://www.w3.org/2000/10/swap/pim/contact#Person;
- properties of those things—such as mailbox, identified by http://www.w3.org/2000/10/swap/pim/contact#mailbox; and
- values of those properties—such as mailto:em@w3.org as the value of the mailbox property (RDF also uses character strings such as "Eric Miller" and values from other data types such as integers and dates as property values).

RDF also provides an XML-based syntax called RDF/XML for recording and exchanging graphs. Figure B shows a small chunk of RDF in RDF/XML corresponding to the graph in Figure A.

The Figure B rendering is actually quite clumsy syntactically, and its lack of transparency and readability might have been a factor inhibiting rapid adoption of RDF. However, there are alternative forms that are easier to interpret; for example, see the N3 notation (www.w3.org/DesignIssues/Notation3.html).

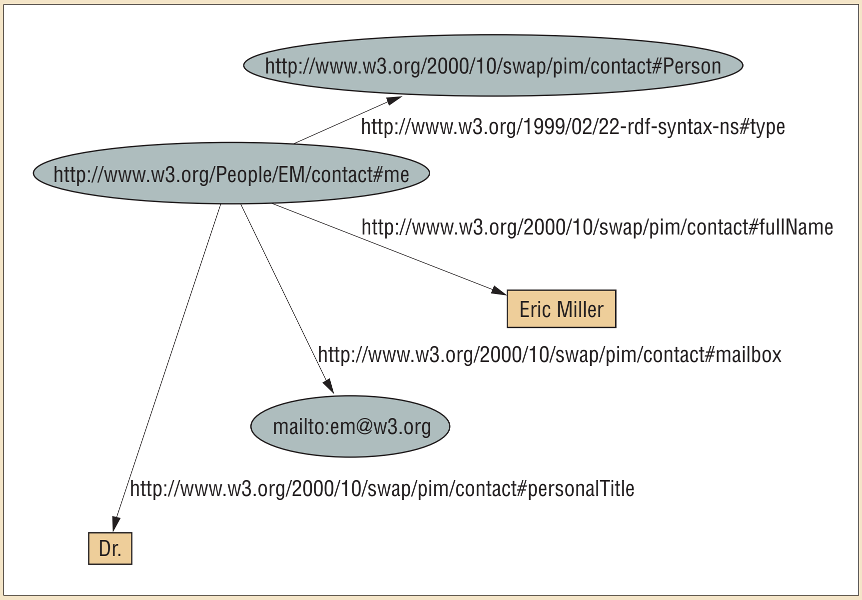



**Figure A. An RDF graph representing Eric Miller.**

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">

  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>

</rdf:RDF>
```

**Figure B. A chunk of RDF in RDF/XML describing Eric Miller and corresponding to the graph in Figure A.**

phy (see http://marinemetadata.org/examples/mmihostedwork/ontologieswork). The need to understand systems across ranges of scale and distribution is evident everywhere in science and presents a pressing requirement for data and information integration.

Various e-government initiatives represent similar efforts. The United Kingdom has developed an Integrated Public Sector Vocabulary (www.esd.org.uk/standards/ipsv). The recently created UK Office of Public Sector Information (www.opsi.gov.uk) is a response to an EU directive (2003/98/EC, http://www.ec-gis.org/document.cfm?id=486&db=document). OPSI aims to exploit the considerable amounts of government data for citizens' benefit. Several EU countries are developing similar programs to implement the EU directive.

Despite these and other significant drivers in defense, business, and commerce, it's still apparent that the Semantic Web isn't yet with us on any scale.

So let's review what progress we've made and consider the various impediments to its global adoption.

## Progress

Consistent with the need for a Web semantics, the user community, including standards organizations like the Internet Engineering Task Force and the World Wide Web Consortium (W3C), has directed major efforts at specifying, developing, and deploying languages for sharing meaning. These languages provide a foundation for semantic interoperability.

In 1997, the W3C defined the first Resource Description Framework specification (see the related sidebar). RDF provided a simple but powerful triple-based representation language for Universal Resource Identifiers (URIs). It became a W3C recommendation by 1999—a crucial step in drawing attention to the specification and promoting its widespread deployment to enhance the Web's functionality and interoperability.

The original Web took hypertext and made it work on a global scale; the vision for RDF was to provide a minimalist knowledge representation for the Web.

### Universal Resource Identifiers

URIs identify resources and so are central to the Semantic Web enterprise.[3] Using a global naming convention (however arbitrary the syntax) provides the global network effects that drive the Web's benefits. URIs have global scope and are interpreted consistently across contexts. Associating a URI with a resource means that anyone can link to it, refer to it, or retrieve a representation of it.

Given the Semantic Web's aims, we want to reason about relationships. URIs provide the grounding for both our objects and relations. They underpin the Semantic Web, allowing machines to process data directly. In this way, the Semantic Web shifts the emphasis from documents to data. Much of the motivation for the Semantic Web comes from the value locked in relational databases. To release this value, database objects must be exported to the Web as first-class objects and therefore must be mapped into a system of URIs.

Languages have evolved to offer greater opportunities for encoding meaning that can support information integration and interoperability. RDF Schema became a recommendation in February 2004. RDFS took the basic RDF specification and extended it to support the expression of structured vocabularies. It has provided a minimal ontology representation language that the research community has adopted fairly widely.

### Triple stores

As RDF and RDFS have gained ground, the need for repositories that can store RDF content has grown. These so-called *triple stores* vary in their capabilities. Some focus on providing a rich means to reason over the triples (for example, see http://jena.sourceforge.net), while others focus on storing large quantities of data (see http://sourceforge.net/projects/threestore for an example from the open source community and www.oracle.com/technology/tech/semantic_technologies/index.html for a commercial example). Some operate as plug-ins to current Web browsers (http://simile.mit.edu/piggy-bank) and others as systems that can operate with a range of existing third-party databases (www.openrdf.org).

As the stores themselves have evolved, the need has arisen for reliable and standardized data access into the RDF they hold. The SPARQL language (www.w3.org/TR/rdf-sparql-query), now in its final review stages for W3C recommendation status, is designed to fulfill this requirement.

### RDF translation

Other significant progress includes GRDDL (Gleaning Resource Descriptions from Dialects of Languages, www.w3.org/2004/01/rdxh/spec), which provides a means to extract RDF from XML and XHTML documents using transformations expressed in XSLT (Extensible Stylesheet Language) and associated with the original content. This capability could potentially overcome the RDF bootstrap problem by generating sufficient RDF for serendipitous reuse to occur. The amount of XML and XHTML data on the Web, especially data generated from back-end databases, is considerable and offers good opportunities for RDF conversion.

URIs have global scope. Associating a URI with a resource means that anyone can link to it, refer to it, or retrieve a representation of it.

### Web Ontology Language

For those who required greater expressivity in their object and relation descriptions, the OWL (Web Ontology Language, www.w3.org/TR/2004/REC-owl-features-20040210) specification integrated several efforts. The W3C recommendation presents three versions of OWL, depending on the degree of expressive power required. OWL's core idea is to enable efficient representation of ontologies that are also amenable to decision procedures. It checks an ontology to see whether it's logically consistent or to determine whether a particular concept falls within the ontology.

OWL uses the linking provided by RDF to allow ontologies to be distributed across systems. Ontologies can become distributed, as OWL allows ontologies to refer to terms in other ontologies. In this way OWL is specifically engineered for the Web and Semantic Web.[4]

OWL is seeing increased adoption but still needs tools and software development environments to support its production and application. These are starting to appear but as yet we have few means to routinely and effortlessly generate Semantic Web annotations using this or other languages at the point of content use or creation.

### Rules and inference

But ontologies are only one part of the representation picture. Rules and inference also need support. The OWL language itself is designed to support various types of inference—typically, subsumption and classification—and a range of automated reasoners are available (for example, see www.cs.man.ac.uk/~sattler/reasoners.html). Because it's difficult to specify a formalism that will capture all the knowledge in a particular domain, there are other approaches to inference on the Web. Work has begun on the Rule Interchange Format (www.w3.org/2005/rules), an attempt to support and interoperate across a variety of rule-based formats. RIF will address the plethora of rule-based formalisms: Horn-clause logics, higher-order logics, production systems, and so on.

Moreover, AI researchers have extended these various logics and modified them to capture causal, temporal, and probabilistic knowledge. Causal logic, such as Glenn Shafer proposed,[5] developed out of action logics in AI, and it's intended to capture an important aspect of commonsense understanding of mechanisms and physical systems. Temporal logic formalizes the rules for reasoning with propositions indexed to particular times; Zhisheng Huang and Heiner Stuckenschmidt suggested temporal-logic approaches for ontology version management.[6] Probabilistic logics are calculi that manipulate conjunctions of probabilities of individual events or states. Perhaps the most well-known of these are Bayesian, which you can use to derive probabilities for events according to prior theories about how probabilities are distributed. Bayesian reasoning is commonplace in search engines. In domains where reasoning under uncertainty is essential, such as bioinformatics, Kenneth Baclawski and Tianhua Niu have suggested using Bayesian ontologies to extend the Web to include such reasoning.[7]

## Data exposure and viral uptake

So far, we've focused on languages, formalisms, standards, and semantics. For this we make no apologies. The Semantic Web can't exist without carefully developed and agreed standards, just as the existing Web couldn't have existed without HTTP, HTML, and XML. But languages and standards are of no consequence without uptake, and uptake requires increasing the amount of data exposed in RDF. (We identify RDF because of the often-encountered *principle of least power*—the less expressive the language, the more reusable the data.)

Uptake is about reaching the point where serendipitous reuse of data, your own and others', becomes possible. We've mentioned the development in the life sciences. Experience suggests that an incubator community with a pressing technology need is an essential prerequisite for success. In the original Web, this community was high-energy physicists who needed to share large document sets. It's easier to mobilize 10 percent of a small but focused community than 10 percent of the general populace—these early adopters are critical.

It's also instructive to consider typical Semantic Web projects of the past five years. They demonstrate a distinctive set of characteristics. Typically, they generate new ontologies for the application domain—whether it's information management in breast diseases[8] or computer science research.[9] They either import legacy data or else harvest and redeposit it into a single, large repository. Then they carry out inference on the RDF graphs held within the repositories and represent the information using a custom-developed interface.

These projects have been important proving grounds for a number of techniques and methods. They show how to facilitate harvesting and semantic integration by using ontologies as mediators. They have served as a development context for RDF stores and a whole range of important Semantic Web middleware. In general, however, they lack real viral uptake. Moreover, in most cases, we aren't able to look up a URI and have the data returned. The data exposure revolution has not yet happened.

URIs provide our symbol grounding in the Web. An RDF triple as a triple of URIs should dereference to terms whose meanings are defined in ontologies. Often, however, the URIs refer to objects that aren't so defined.

Consider a life science example: Uniprot (www.ebi.uniprot.org/index.shtml) is the world's most comprehensive set of databases on proteins, but we can't provide the URI for a Uniprot protein and then simply read off or determine its properties. Rather, the server passes us a zipped bundle of data downloaded as a blob. Moreover, the Life Science Identifier (http://lsid.sourceforge.net) naming-scheme standards that life scientists use aren't HTTP compatible. A process is needed that routinely gives URIs to such objects and entrusts their management to individuals and communities who care about consistent and explicit reference methods.

## Ontology development and management

The challenges here are real. The ontologies that will furnish the semantics for the Semantic Web must be developed, managed, and endorsed by committed practice communities. Whether the subject is meteorology or bank transactions, proteins or engine parts, we need concept definitions that we can use.

> The ontologies that will furnish the semantics for the Semantic Web must be developed, managed, and endorsed by practice communities.

Although some denotations are more persistent than others, we must recognize that they aren't fixed over all time. Even terms used to classify medical diseases change as new procedures and understanding emerges. We need to regard such ontologies as living structures. Some might endure over long periods—for example, terms describing the elements of the periodic table. Others are much more volatile: the 18th-century concept of *phlogiston* doesn't have a place in a modern ontology of chemistry, but it was once thought to be essential to explaining combustion and other chemical reactions. Communities and practice will change norms, conceptualizations, and terminologies in complex and sociologically subtle ways. We shouldn't be surprised or attempt to resist these reformulations. The issue for a Semantic Web built from these conventions is to know when parts need revision.

This brings us to an often quoted concern about the Semantic Web—the cost of ontology development and maintenance. In some areas, the costs—no matter how large—will be easy to recoup. For example, an ontology will be a powerful and essential tool in well-structured areas such as scientific applications. In certain commercial applications, the potential profit and productivity gain from using well-structured and coordinated vocabulary specifications will outweigh the sunk costs of developing an ontology and the marginal costs of maintenance.

In fact, given the Web's fractal nature, those costs might decrease as an ontology's user base increases. If we assume that ontology building costs are spread across user communities, the number of ontology engineers required increases as the log of the user community's size. The amount of building time increases as the square of the number of engineers. These are naïve but reasonable assumptions for a basic model. The consequence is that the effort involved per user in building ontologies for large communities gets very small very quickly.[10]

Not all ontologies have the same characteristics and, in general, we can distinguish *deep* from *shallow* ontologies. Deep ontologies are often those encountered in science and engineering, where considerable efforts go into building and developing the conceptualization. For domains such as proteomics and medicine, the ontology is in a very real sense the data of interest. This becomes apparent when we use an ontology to classify complex sets of properties as constituting certain sorts of object.

Shallow ontologies comprise relatively few unchanging terms that organize very large amounts of data—for example, terms such as *customer*, *account number*, and *overdraft* used in banking and financial contexts or the basic relations that define geospatial information. Some might argue that we've spent rather too much time extolling the virtues of deep ontologies at the expense of the shallow ones that deliver very large amounts of reusable data. Shallow ontologies require effort but over much simpler sets of terms and relations.

## Folksonomies: Web-scale tagging

The complexity of deep ontologies has led some to eschew ontologies altogether in favor of a different approach. *Folksonomies* are a development generating considerable interest at the moment. They represent a structure that emerges organically when individuals manage their own information requirements. Folksonomies arise when a large number of people are interested in particular information and are encouraged to describe it—or *tag* it (they may tag selfishly to organize their own content retrieval or altruistically to help others). Rather than a centralized form of classification, users can assign keywords to documents or other information sources.

Well-known examples of applications that harness and exploit tagging are Flickr (www.flickr.com, a photography publication and sharing site) and del.icio.us (http://del.icio.us, a site for sharing bookmarks). These applications, driven by decentralized communities from the bottom up, are sometimes called Web 2.0 or *social software*.

Tagging on a Web scale is certainly an interesting development. It provides a potential source of metadata. The folksonomies that emerge are a variant on keyword searches. They're an interesting emergent attempt at information retrieval. But folksonomies serve very different purposes from ontologies. Ontologies are attempts to more carefully define parts of the data world and to allow mappings and interactions between data held in different formats. Ontologies refer by virtue of URIs; tags use words. Ontologies are defined through a careful, explicit process that attempts to remove ambiguity. The definition of a tag is a loose and implicit process where ambiguity might well remain. The inferential process applied to ontologies is logic based and uses operations such as *join*. The inferential process used on tags is statistical in nature and employs techniques such as clustering.

This doesn't mean that tags will always replace shallow ontologies. Where a perceived need for ontologies exists, lightweight but powerful ones do emerge and are widely used. Two examples are Friend-of-a-Friend[11] and associated applications such as Flink.[12] This fits in general with calls for the dual and complementary development of Semantic Web technologies and technologies that exploit the Web's self-organization.[13]

Some people perceive ontologies as top-down, somewhat authoritarian constructs—unrelated, or only tenuously related, to people's actual practice, to the variety of potential tasks in a domain, or to the operation of context.[14] This perception might be related to the idea of developing a single consistent Ontology of Everything—like Cyc,[15] for example. Such a wide-ranging and all-encompassing ontology might well have interesting applications, but it clearly won't scale and its use can't be enforced.

If the Semantic Web is seen as requiring widespread buy-in to a particular point of view, then it's understandable that emergent structures like folksonomies begin to seem more attractive.[16] But this isn't a Semantic Web requirement. Ontologies are a rationalization of actual data-sharing practice. We can and do interact, and we do it without achieving or attempting to achieve global consistency and coverage. Ontologies are a means to make an explicit commitment to shared meaning among an interested community, but anyone can use these ontologies to describe their own data. Similarly, anyone can extend or reuse elements of an ontology if they so wish.

Ontologies are attempts to more carefully define parts of the data world and to allow interactions between data held in different formats.

## The next wave

The Semantic Web we aspire to makes substantial reuse of existing ontologies and data. It's a linked information space in which data is being enriched and added. It lets users engage in the sort of serendipitous reuse and discovery of related information that's been a hallmark of viral Web uptake. We already see an increasing need and a rising obligation for people and organizations to make their data available. This is driven by the imperatives of collaborative science, by commercial incentives such as making product details available, and by regulatory requirements. We believe this could bring about a revolution in how, for example, scientific content is managed throughout its life cycle.

This next wave of data ubiquity will present us with substantial research challenges. How do we effectively query huge numbers of decentralized information repositories of varying scales? How do we align and map between ontologies? How do we construct a Semantic Web browser that effectively visualizes and navigates the huge connected RDF graph? How do we establish trust and provenance of the content?

Provenance—that is, the when, where, and conditions under which data originated—has become a key requirement in a range of applications. We might well need the help of researchers in areas as diverse as social network analysis[17,18] and epidemiology to understand how information and concepts spread on the Web and how to establish their provenance and trustworthiness.

We must not lose sight of the fact that the Web, and indeed many of our most important digital environments, depends fundamentally on certain general assumptions about social behavior. The Web relies on people serving useful content; it relies on content generally being on the end of links. We also require that people observe copyright rules. Creative Commons (www.creativecommons.org) is an RDF-based representation of copyright policy to facilitate and maximize appropriate reuse. Policy-aware research takes this further, attempting to express the civic rules of behavior expected in a Semantic Web environment.

The critical factors that led to the Web's success will also be important to the success of our Semantic Web enterprise. As we've seen, some of these factors are social; others have their origin in elementary and fundamental design decisions about the Web's architectural principles. For example, the URL concept embodied the principle that every Web address is equal and all content one jump away. Other critical features included the ability to let links fail (the 404 error).

A great deal of the success relates to what we might call the *ladder of authority*. This is the sequence of specifications (URI, HTTP, RDF, ontology, and so on) and registers (URI scheme, MIME Internet content type, and so on), which provide a means for a construct such as an ontology to derive meaning from a URI. Another example is the construction of a standards body that's been able to promote, develop, and deploy open standards.

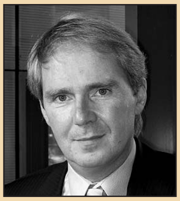
These reflections lead us to ask how we understand the present Web and what developments we anticipate. This is a deep question, and we believe the history of science has something to teach us here. There was a time when our understanding of the world was either a purely philosophical, reflective exercise or else craft-based and rooted in hard-won experience. Empirical methods eventually gave rise to the branches of natural philosophy that became physics, chemistry, and biology. Traces of this legacy can still be found: until quite recently the study of physics at Oxford was termed experimental philosophy. More recently, areas that were once considered amenable only to analytic thought—areas such as epistemology and logic—are to some extent operationalized in computers and computer infrastructures. Knowledge representation and ontology engineering are about trying to capture aspects of shared conceptualizations.

As we build ever more complex computational artifacts and information infrastructures, we observe that large-scale behavior emanates from small-scale and local regularity. We need engineering methods to ensure that our structures conform to reliable and repeatable design requirements. We need scientific analysis to understand and predict the behaviors that result. When we build new opportunities for interaction, we're engaged simultaneously in a synthetic and an analytic project. New rules of interaction such as peer-to-peer protocols result in new macro behaviors—behaviors we can exploit and also analyze. These micro rules can occur at different levels of abstraction—the rules of Wikipedia are beguilingly simple but lead to overall coherence. Local-scale changes in Web architectures and resources can lead to large-scale societal and technical effects. How so?

We expect the developments, methodologies, challenges, and techniques we've discussed here to not only give rise to a Semantic Web but also contribute to a new Web Science—a science that seeks to develop, deploy, and understand distributed information systems, systems of humans and machines, operating on a global scale. AI will be one of the contributing disciplines. AI has already given us functional and logic programming methods, ways to understand distributed systems, pattern detection and data mining tools, approaches to inference, ontological engineering and knowledge representation. All of these are fundamental to pursuing a Web Science agenda and realizing the Semantic Web. ■

## References


1. V. Bush, "As We May Think," *Atlantic Monthly*, vol. 176, no. 1, 1945, pp. 101–108.
2. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific Am.*, May 2001, pp. 34–43.
3. T. Berners-Lee, R.T. Fielding, and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax," IETF RFP 3986 (standards track), Internet Eng. Task Force, Jan. 2005; www.ietf.org/rfc/rfc3986.txt.
4. J.A. Hendler, "Frequently Asked Questions on W3C's Web Ontology Language (OWL)," W3C, 2004; www.w3.org/2003/08/owlfaq.
5. G. Shafer, "Causal Logic," *Proc. 13th European Conf. Artificial Intelligence* (ECAI 98), John Wiley & Sons, 1998, pp. 711–720; www.glennshafer.com/assets/downloads/article62.pdf.
6. Z. Huang and H. Stuckenschmidt, "Reasoning with Multi-Version Ontologies: A Temporal Logic Approach," *The Semantic Web—ISWC 2005: 4th Int'l Semantic Web Conf.*, LNCS 3729, Springer, 2005; www.cs.vu.nl/~heiner/public/ISWC05a.pdf.
7. K. Baclawski and T. Niu, *Ontologies for Bioinformatics*, MIT Press, 2005.
8. S. Dasmahapatra et al., "Facilitating Multi-Disciplinary Knowledge-Based Support for Breast Cancer Screening," *Int'l J. Healthcare Technology and Management*, vol. 7, no. 5, 2006, pp. 403–420; http://eprints.ecs.soton.ac.uk/8955/01/ijhtm03-miakt.pdf.
9. N. Shadbolt et al., "CS AKTive Space, or How We Learned to Stop Worrying and Love the Semantic Web," *IEEE Intelligent Systems*, vol. 19, no. 3, 2004, pp. 41–47.
10. T. Berners-Lee, "The Fractal Nature of the Web," working draft, 1998–2005; www.w3.org/DesignIssues/Fractal.html.
11. D. Brickley and L. Miller, "FOAF Vocabulary Specification," working draft, 2005, http://xmlns.com/foaf/0.1.
12. P. Mika, "Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks," *J. Web Semantics*, vol. 3, no. 2, 2005, pp. 211–223; www.websemanticsjournal.org/ps/pub/2005-20.
13. G.W. Flake, D.M. Pennock, and D.C. Fain, "The Self-Organized Web: The Yin to the Semantic Web's Yang," *IEEE Intelligent Systems*, July/Aug. 2003, pp. 72–86.
14. K. Sparck-Jones, "What's New about the Semantic Web? Some Questions," *SIGIR Forum*, vol. 38, no. 2, 2004; www.acm.org/sigir/forum/2004D/sparck_jones_sigirforum_2004d.pdf.
15. D.B. Lenat, "Cyc: A Large-Scale Investment in Knowledge Infrastructure," *Comm. ACM*, vol. 38, no. 11, 1995, pp. 32–38.
16. C. Shirky, "Ontology Is Overrated: Categories, Links and Tags," 2005; www.shirky.com/writings/ontology_overrated.html.
17. D.J. Watts, P.S. Dodds, and M.E.J. Newman, "Identity and Search in Social Networks," *Science*, vol. 296, 2002, pp. 1302–1305.
18. G. Kossinets and D.J. Watts, "Empirical Analysis of an Evolving Social Network," *Science*, vol. 311, 2006, pp. 88–90.


**Nigel Shadbolt** is a professor of artificial intelligence in the School of Electronics and Computer Science at Southampton University. Contact him at nrs@ecs.soton.ac.uk.



**Tim Berners-Lee** is the director of the World Wide Web Consortium, a senior researcher at the Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Laboratory, and a professor of computer science in the Department of Electronics and Computer Science at Southampton University. Contact him at timbl@w3.org.



**Wendy Hall** is a professor of computer science in the School of Electronics and Computer Science at Southampton University. Contact her at wh@ecs.soton.ac.uk.