The Semantic Web

Kieron O'Hara & Wendy Hall

Version of article forthcoming in Marcia J. Bates & Mary Niles Maack (eds.), *Encyclopedia of Library and Information Sciences*, Taylor & Francis

Intelligence, Agents, Multimedia Group

School of Electronics and Computer Science

University of Southampton

Highfield

Southampton SO17 1BJ

United Kingdom

{wh,kmo}@ecs.soton.ac.uk

KEYWORDS

Semantic Web, web of data, linked data, interoperability, information sharing, ontologies, representation, metadata, information processing

ABSTRACT

The Semantic Web is a vision of a web of linked data, allowing querying, integration and sharing of data from distributed sources in heterogeneous formats, using ontologies to provide an associated and explicit semantic interpretation. The article describes the series of layered formalisms and standards that underlie this vision, and chronicles their historical and ongoing development. A number of applications, scientific and otherwise, academic and commercial, are reviewed. The Semantic Web has often been a controversial enterprise, and some of the controversies are reviewed, and misconceptions defused.

INTRODUCTION

The Semantic Web is an extension, in progress, to the World Wide Web, designed to allow software processes, in particular artificial agents, as well as human readers, to acquire, share and reason about information. Whereas the World Wide Web (WWW) consists largely of documents, which are generally created for human consumption, the Semantic Web (SW) will be a web of data, making them more amenable for computers to process [1]. The data will be processed by computer via semantic theories for interpreting the symbols (hence: *Semantic* Web). In any particular application, the semantic theory will connect terms within a distributed document set logically, and thereby aid interoperability.

For instance, people use a lot of data in daily interactions, viewing bank statements, or digital photographs, or using diaries or calendars. But this does not constitute a web of data, because the data are neither exported from the applications in which they are stored or were created, nor linked to other relevant data. In a genuine web of data, such data could be used seamlessly in a number of applications. For example, one could view one’s photographs (which will contain a time stamp) in one’s calendar, which would then act as a prompt to suggest what one was doing when they were taken. The data which one uses would be to some extent freed from the constraints of particular applications, and instead could be interlinked and reused creatively.

As another example, Web services can currently be accessed and executed via the Web, but because the Web does not provide much information-processing support, services must be specified using semi-formal languages and as with information retrieval humans need to be kept in the loop. Web services described using Semantic Web techniques should provide support for autonomous agents and automatic systems [2].

The world of linked information is a very unstructured, “scruffy” environment. The amounts of information that systems need to deal with are very large indeed. Furthermore, systems must pull together information from distributed sources, where representation schemes can be expected to be highly heterogeneous, information quality variable, and trust in information’s provenance hard to establish. Semantic Web technology needs to be based on standards that can operate in this heterogenous information world.

The SW therefore requires two types of information standard to operate. First, it requires common formats for integrating information from these diverse sources. And second, it needs a language to express the mapping between the data and objects in the real world, in order to allow a seamless understanding of a distributed set of databases. Hence, for instance, we could signal that a database containing a column *zip code*, and another database with a column labelled *ZC*, were actually both referring to the same concept with their different labels, and by creating such a semantic link, we could then start to reason over both databases in an integrated fashion. Such semantic links are often obvious to humans, but not to computers. A key formalism here is the *ontology*, which define the concepts and relationships that we use in particular applications. Ontologies are central to the SW vision, as providing the chief means by which the terms used in data are understood in the wider context [1, 3].

THE AIM OF THE SEMANTIC WEB

The aim of the SW is to shift the emphasis of reasoning from documents to data, for three reasons. First, it will facilitate data reuse, often in new and unexpected contexts. Second, it will help reduce the amount of relatively expensive human information processing. Third, it will release the large quantity of information, not currently accessible, that is stored in relational databases (RDBs) by making it directly machine-processable [4].

This implies that RDB objects must be exported to the Web as first-class objects, which in practice entails mapping them onto a consistent system of resource identifiers – called Universal Resource Identifiers (URIs – see below). The SW itself is a suite of languages and formalisms designed to enable the interrogation and manipulation of representations which make use of URIs [1].

It is hoped that the SW will exhibit the same *network effects* that promoted the growth of the WWW. Network effects are positive feedback effects connected with *Metcalfe's Law* that the value of a network is proportional to the square of the number of users/members. The more people share data that can be mapped onto URIs, the more valuable that data is. As value increases, more agents join the network to get the benefits, and include information that they own in the network which further increases its value. This, like the WWW model, is radically different from other models of the value of information, wherein value is dictated by *scarcity* (copyright, intellectual property restrictions, etc). In decentralised networks like the Web the value of information is dictated by *abundance*, so it can be placed in new contexts, and reused in unanticipated ways.

This is the dynamic that enabled the WWW to spread, when the value of Web documents was seen to be greater in information-rich contexts. One initiative to support the development of the SW is the creation of a discipline of *Web Science*, which is intended to exploit study of both technical and social issues to predict such matters with more accuracy [5, 6].

If the SW is to grow in an analogous way, more data has to be exposed to the Web that can be mapped onto URIs. In practice that means that the data must be exposed in the Resource Description Framework (RDF), an agreed international standard whose role in the SW is described below [7]; in particular, it can be used not only to assert a link between two resources, but also to name (and therefore make explicit) the relationship that links them. RDF is the language of choice for reuse, because it is a relatively inexpressive language compared to other formalisms used in the SW (see Figure 1 for a pictorial representation of the layers of formalisms required for the SW vision – expressivity increases as we ascend the diagram). The importance of RDF in this model is dictated by the so-called *principle of least power*, which states that the less expressive the representation language, the more reusable the data [8].

The importance of growth is so that a stage can be reached when reuse of data – one's own or that of other people – is facilitated. There would ideally be so much information exposed in RDF that the contexts into which one's own data can be placed would be rich enough and numerous enough to increase its value significantly. RDF (as described below) represents information as a subject-predicate-object triple each of whose component parts is a URI. If the objects, resources or representations referred to by the URIs are defined in ontologies, then this enables the interoperability at which the SW aims.

Hence another vital component in the SW is the development and maintenance of ontologies. These must be endorsed by the communities that use them, whether they are large-scale, expensive ontologies developed as a result of a major research effort, or relatively *ad hoc* creations intended to support small scale collaboration.

Ontologies can also play an important role in bringing (representatives of) two or more communities together for a common purpose, by expressing a common vocabulary for their collaboration, onto which the terms of each discipline can be mapped. Such collaborative efforts are extremely important for reuse of content [3].

This is not to say that search and retrieval on the current Web is not high quality; the methods pioneered by Google and others work very well. Nevertheless, keyword-based search techniques are vulnerable to a number of well-known flaws. Individual words can be ambiguous. A document can refer to a topic of interest without using the keyword. Keywords are language-dependent. Information distributed across several documents cannot be amalgamated by keyword search. And even though PageRank and related algorithms for search produce impressive results, the user still needs to read manually through the ordered list of retrieved pages, and inspect their content to determine relevance to his/her inquiry. This involvement of the user is a hindrance to scalability.

The SW should make more accurate querying possible, using ontologies to help with problems of ambiguity and unused keywords, and data linking to query across distributed datasets. Furthermore, it should be able to go beyond current search with respect to the three issues of reuse, automation and exploitation of RDBs. And as well as search and retrieval, the addition of information processing support to the Web will help promote other functions such as Web services and knowledge management.

COMPONENTS OF THE SEMANTIC WEB

At one level, the SW is a complex of formalisms and languages each doing a different job in the representation of information, as shown in Figure 1. Each formalism is an internationally-agreed standard (see below), and the composition of the functions these formalisms serve, composed support semantically-enabled reasoning on data.

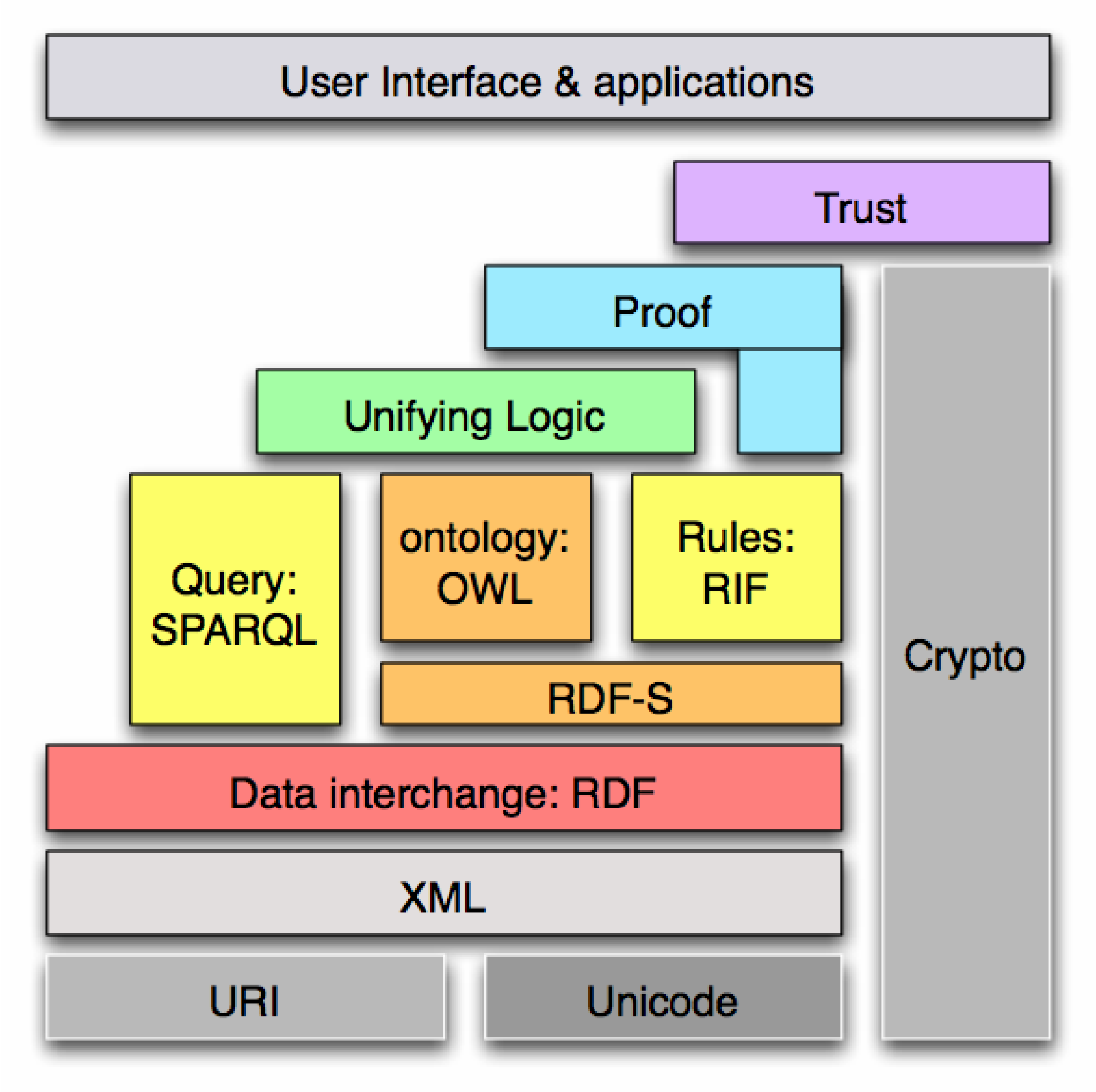

Figure 1: The Layered View of the Semantic Web [6]

At the bottom of this diagram stands the Universal Resource Identifiers (URIs) which identify the resources about which the SW provides reasoning capabilities [9]. The universality of URIs is extremely important – i.e. it is vital that whatever naming convention is used for URIs is adopted globally, so as to create the network effects that allow the SW to add value. Interpretation of URIs must also be consistent across contexts. In other words, when we *dereference* URIs (i.e. when we locate the resource to which the URI refers), we should always get the same object. If these conditions about URI naming schemes are met, then making an association between a URI and a resource means that different people can refer or link to it consistently in their conversations. The other basic formalism, Unicode, is an industry standard that allows computers to represent text in different writing systems.

The next layer up, XML (eXtensible Markup Language), is a language to mark up documents, and a uniform data exchange format between applications [10]. It allows the insertion of user-defined tags into documents that provide information about the role that the content plays. So, for instance, XML allows one to write a document describing a book, and also to *annotate* the document with machine-readable *metadata* to indicate e.g. who the authors of the book are.

The Resource Description Framework (RDF – [7]) is a very minimal knowledge representation framework for the Web, which uses a basic subject-predicate-object structure, with the twist that it assigns specific URIs to its individual fields – including in the predicate position, thereby identifying a relationship between the entities identified by the connected nodes. This use of URIs allows us to reason not only about objects but also about the relationships between them. XML is a metalanguage that provides a uniform framework for markup, but it does not provide any way of getting at the *semantics* of data; RDF is the first step towards semantics.

RDF Schema (RDFS, sometimes known as RDF(S) – [11]) gives greater scope for sharing information about individual domains; whereas RDF is a data interchange language that lets users describe resources using their own vocabularies, and makes no assumptions about the domains in question, RDFS provides a basic set of tools for producing structured vocabularies that allow different users to agree on particular uses of terms. An extension of RDF, it adds a few modelling primitives with a fixed meaning (such as class, subclass and property relations, and domain and range restriction).

A key component for SW applications is the *ontology*. Ontologies [3] are shared conceptualisations of a domain which are intended to facilitate knowledge and information sharing by coordinating vocabulary and allowing basic inference of inheritance and attributes of objects. Several initiatives are developing ontologies, particularly in a number of sciences, which means that the scientists are likely to be among the important early adopters of SW technology (see below). RDFS is an important step towards the SW vision, as the addition of modelling primitives makes it a basic ontology representation language.

However, greater expressivity is likely to be required in the development of more complex ontologies, and the W3C has issued a Web Ontology Language (OWL – [12]) in multiple versions that allows ontologies to be not only represented but also checked for logical properties such as consistency. The three species of OWL are:
(1) OWL Full, containing all the OWL primitives, allowing arbitrary combination of those primitives with RDF and RDFS (allowing changes in meaning even of predefined OWL or RDF primitives), but also providing so much expressive power as to make the language undecidable (i.e. it cannot be guaranteed that a computation using the full expressive power of OWL Full will be completed in a finite time);
(2) OWL DL, which restricts application of OWL's constructors to each other, and corresponds to a decidable *description logic*, but which is not fully compatible with RDF; and (3) OWL Lite, which sacrifices even more expressive power to facilitate implementation and reasoning [12]. This set of relations affects the downward compatibility of the SW layer diagram – the only version of OWL that is downward compatible with RDF and RDFS (i.e. so that any processor for that version of OWL will also provide correct interpretations of RDFS) is OWL Full, which is undecidable [13, pp.113-115, 14].

All varieties of OWL use RDF for their syntax, and use the linking capabilities of RDF to allow ontologies to be distributed – ontologies can refer to terms in other ontologies. Such distributivity is a key property for an ontology language designed for the SW [15].

OWL supports some kinds of inference, such as subsumption and classification, but a greater variety of rules and inference is needed. Hence, work is currently ongoing on the Rule Interchange Format (RIF), which is intended to allow a variety of rule-based formalisms, including Horn-clause logics, higher order logics and production systems, to be used [16]. Various insights from Artificial Intelligence (AI) have also been adapted for use for the SW, including temporal (time-based) logic, causal logic and probabilistic logics [1].

Having represented data using RDF and ontologies, and provided for inference, it is also important to provide reliable, standardised access to data held in RDF. To that end, a special query language SPARQL (pronounced 'sparkle'), which became a W3C recommendation in January 2008, has been designed [17]. Logic and proof systems are envisaged to sit on top of these formalisms, to manipulate the information in deployed systems [1].

A very important layer is that of *trust* [18]. If information is being gathered from heterogeneous sources and inferred over, then it is important that users are able to trust such sources. The extent of trust will of course depend on the criticality of the inferences – trust entails risk, and a risk-averse user will naturally trust fewer sources [19, 20]. Measuring trust, however, is a complex issue [21]. A key parameter is that of provenance, a statement of (a) the conditions under which, (b) the methods with which, and (c) the organisation by which, data were produced. Methods are appearing to enable provenance to be established, but relatively little is known about how information spreads across the Web [22].

Related issues include respect for intellectual property, and the privacy of data subjects. In each case the reasoning abilities of the SW can be of value, and initiatives are currently under way to try to exploit them [23]. Creative Commons [24] is a way of representing copyright policies and preferences based on RDF to promote reuse where possible (current standard copyright assumptions are more restrictive with respect to reuse). And research into the Policy Aware Web is attempting to develop protocols to allow users to express their own privacy policies, and to enable those who wish to use information to reason about those policies [25]. Cryptography protocols to protect information will also play an important role, as shown in Figure 1.

ADDITIONAL FACTORS IN SEMANTIC WEB DEVELOPMENT

INFRASTRUCTURE

Another important part of SW development is the infrastructure that supports it. In particular, if data is to be routinely published to the Web in RDF format, there must be information repositories that can store RDF and RDFS. These *triple stores* (so-called because they store the RDF triples) must provide reasoning capabilities as well as retrieval mechanisms, but importantly must be *scalable*. Examples of triple stores include JENA [26], 3store [27, 28] and Oracle 11*g* [29]. OWLIM is a repository which works as a storage and inference layer for the Sesame RDF database, providing reasoning support for some of the more expressive languages of the SW, RDFS and a limited version of OWL Lite [30, 31].

REASONERS

As representation in the SW is more complex than in previous technologies, so is reasoning. The area of SW reasoning has been the focus of much research, in order to infer the consequences of a set of assertions interpreted via an ontology. In such a context, inference rules need clear semantics, and need to be able to cope with the diverse and distributed nature of the SW.

There are a number of important issues of relevance in this area. (1) Under what conditions is negation monotonic (i.e. the addition of new facts does not change the derivation of not-p), or non-monotonic (including negation as failure, deriving not-p from the failure to prove p)? (2) How should we handle conflicts when merging rule-sets? (3) 'Truth' on the Web is often dependent on context – how should a reasoner represent that dependence? (4) How should scalability be balanced against expressivity? (5) Logic often assumes a static world of given 'facts', but how should it be adapted to the SW, a much more dynamic space where propositions are asserted and withdrawn all the time? (6) The heterogeneous nature of the SW means that data in the SW is of varying trustworthiness; how should a reasoner deal with variable reliability? None of these questions has a 'correct' answer, but any SW reasoning system needs to address them.

There has been a lot of research on SW reasoning, but an important desideratum is that a reasoner should support the W3C recommended formalisms, in particular supporting OWL entailment at as high a level as possible, and SPARQL querying. Examples include: Jena, an open source SW framework for Java, with a rule-based inference engine [32]; Pellet, a sound and complete OWL-DL reasoner [33]; and KAON2, an infrastructure for managing ontologies written in OWL-DL and other SW rule languages [34]. For a short review of the problems and prospects for SW reasoning, see [35].

BOOTSTRAPPING

Bootstrapping content for the SW is one more important issue. Sufficient content is required for the hoped-for network effects to appear. There are initiatives to generate data in RDF and to expose it on the Web as a vital first step. The DBpedia [36] is based on the Web 2.0 community-created encyclopaedia Wikipedia, and is intended to extract structured information from Wikipedia allowing much more sophisticated querying. Sample queries given on the DPpedia website include a list of people influenced by Friedrich Nietzsche, and the set of images of American guitarists. DBpedia uses RDF, and is also interlinked with other data sources on the Web. When accessed in late 2007, the DBpedia dataset consisted of 103 million RDF triples. Other examples of linked data applications include the DBLP bibliography of scientific papers [37], and the GeoNames database which gives descriptions of millions of geographical features in RDF [38].

Even if RDF began to be published routinely, there is still a great deal of legacy content on the Web, and to make this accessible to SW technology some automation of the translation process is required. GRDDL (Gleaning Resource Descriptions from Dialects of Languages) allows the extraction of RDF from XML documents using transformations expressed in XSLT, an extensible stylesheet language based on XML. It is hoped that such extraction could allow bootstrapping of some of the hoped-for SW network effects [39].

Annotating documents and data with metadata about content, provenance and other useful dimensions (even including relevant emotional reactions to content – [40]) is also important for the effort to bring more content into the range of SW technologies

[41]. Multimedia documents, such as images, particularly benefit from such annotation [42]. Again, given the quantities of both legacy data, and new data being created, methods of automating annotation have been investigated by a number of research teams in order to increase the quantity of annotated data available without excessive expenditure of resources [41, 43, 44].

THE SOCIAL CONTEXT: WEB SCIENCE

The SW vision has been delineated with some care by the W3C, and as has been seen involves an intricate set of connections between a number of formalisms, each of which is designed to do a certain job. As we will describe in the next section, that vision has altered and gained complexity over time.

In general, there are severe complications in the mapping between the micro-level engineering of Web protocols, and the macro-level social effects that result from large-scale use of the Web. The combination of scales, effects and phenomena involved is too large to be easily covered by a single discipline, even computer science. The social interactions enabled by the Web place demands on the Web applications underlying them, which in turn put requirements on the Web's infrastructure. However, these multiple requirements are not currently well-understood [45]. Social studies tend to regard the Web as a given, whereas the Web is rather a world changeable by alterations to the protocols underlying it. Furthermore, the Web changes at a rate that is at least equal and may be faster than our ability to observe and analyse it.

The SW is a development bringing the Web vision to a new level of abstraction, yet the current state of our knowledge of the Web and its relation to offline society leaves a number of questions unanswered about how it will impact at a large scale. In particular, it is unknown what social consequences there might be of the greater public exposure and sharing of information that is currently locked in databases. Understanding these consequences is important partly because the developers of the SW want to build a technology that is not harmful to society thanks to emergent social effects, and partly because it is important that the SW goes with the grain of society, in order that it be effective in real-world situations [5].

To this end, in 2006 the Web Science Research Initiative (WSRI) was set up as a joint venture by the Massachusetts Institute of Technology and the University of Southampton to foster the interdisciplinary study of the Web in its social and technical context. WSRI's role includes crafting a curriculum for study across the various relevant disciplines; [6] is a detailed review of the wide range of scientific and social-scientific research that is likely to be relevant, including graph and network theory, computer science, economics, complexity theory, psychology, law etc.

HISTORY AND INTELLECTUAL BACKGROUND

The vision of a web of data was always implicit in the ideas underlying the development of the WWW, and was articulated by Sir Tim Berners-Lee at the first WWW conference in 1994. Berners-Lee is well known as the inventor of the WWW in 1989-91, and has been a leading figure in the development of the SW. As well as holding Chairs at the Massachusetts Institute of Technology, USA, and the University of Southampton, UK, Berners-Lee is the Director of the World Wide Web Consortium (W3C), which he founded in 1994.

A key moment in the development, and public perception, of the SW was an article written for *Scientific American* by Berners-Lee, James A. Hendler and Ora Lassila in 2001 [46]. This paper postulated the next stage of the WWW explicitly as one where data and information, as well as documents, are processed automatically, and envisaged a world where intelligent agents were able to access information (e.g. from calendars, gazetteers and business organisations) in order to undertake tasks and planning for their owners.

This vision of automation of a series of routine information processing tasks has not emerged at the time of writing (2008). The article's agent-oriented vision distracted attention from the main point of the SW, the potential of a web of linked *data* (as opposed to documents) with shared semantics. Hence, in 2006, Berners-Lee, together with Nigel Shadbolt and Wendy Hall, published another article in the IEEE journal *Intelligent Systems*, which made that point explicitly, and argued that the agent-based vision would only flourish with well-established data standards [1].

The *Scientific American* article painted a very enticing picture, but its key message was less to do with the agents and more to do with the semantic information infrastructure that Berners-Lee et al were advocating. Indeed, the infrastructure will be used for many knowledge management purposes, not only in allowing agents to communicate. The agent-focused rhetoric of the article has prompted some to argue that the SW is a restatement of the programme of Artificial Intelligence (AI) in the 1960s and 1970s, and will share its perceived failures. We address this question below, in the section entitled 'Controversies'.

In 2001 (and before), the conceptualisation of the various formal layers of the SW was as shown in Figure 2, with a fairly straightforward cascade up from URIs to XML and namespaces, to RDF and RDFS, through ontologies to rules, logic, proof and trust (the diagram has been widely distributed, but see e.g. [47]). Comparison with Figure 1 shows how the details of the SW layers have had to be amended over time as implementation has continued. The requirements for expression of ontology-related information has led to an extra complexity from that envisaged in 2001, while the criticism of the SW vision based on the *Scientific American* article has led to a realisation that not only to the expressive formalisms need to be in place, but also tools and methods need to be created to allow use of SW technologies to integrate smoothly into organisations' standard information workflows [e.g. 1, 44, 48, 49]. This led to a top layer, User Interface, being added to the Figure 2 structure at a later date.

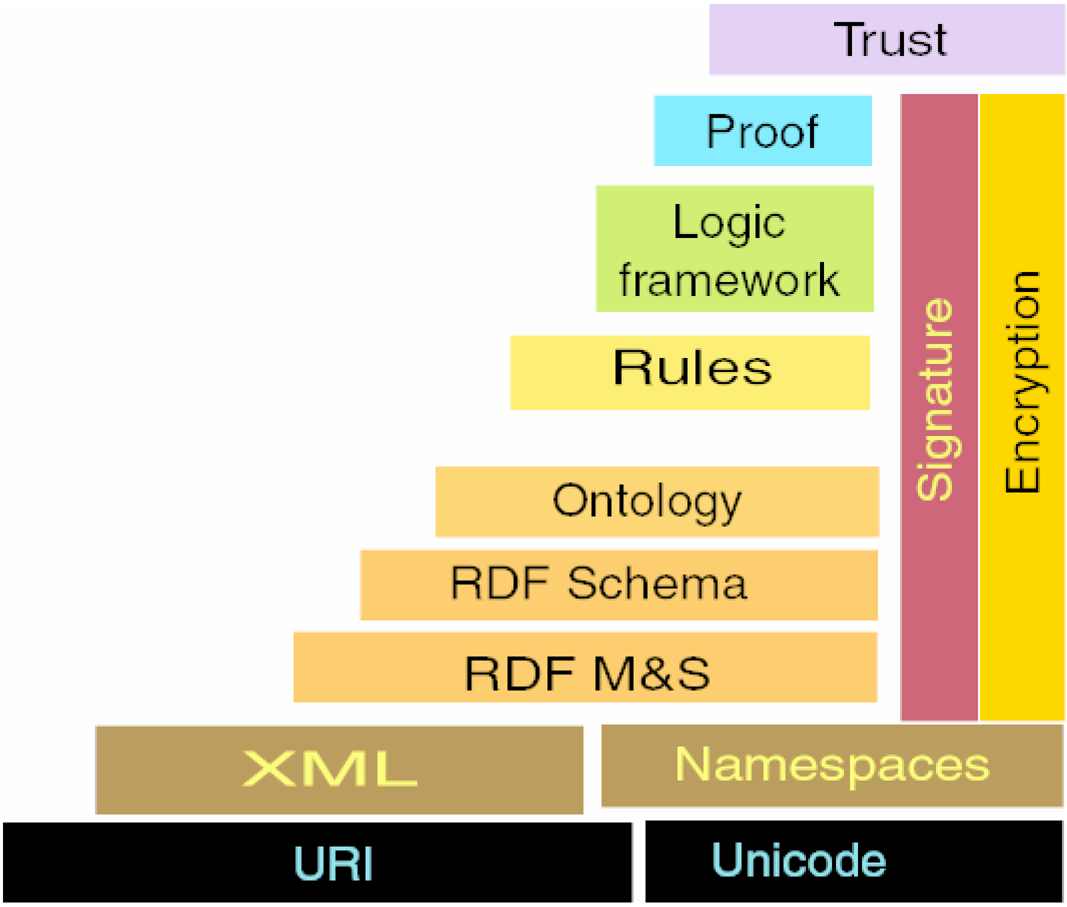

Figure 2: The Early Layered View of the Semantic Web

Where intelligent agency has appeared – and there are currently several applications, including shopbots and auction bots – it has tended to be handcrafted and unable to interact with heterogeneous information types. This is largely because of a lack of well-established scalable standards for information sharing; however, progress is being made towards that goal, especially via the painstaking committee-based standards development processes instituted by the W3C. These standards are crucial for the SW to "take off", and for the hoped-for network effects of a large number of users to emerge [1].

The SW vision has been implemented by standards bodies, such as the Internet Engineering Task Force (IETF) as well as the W3C (the W3C is responsible for standards specific to the WWW), which have orchestrated efforts together with the user community to develop the languages at various levels to share meaning. Once standards are set by the W3C, they are called *recommendations*, acknowledging the reality that with the decentralisation of the Web, and a lack of a central authority, standards cannot be enforced. The first Resource Description Framework (RDF) standard was specified in 1997 and became a W3C recommendation in 1999, thereby providing a minimal knowledge representation language for the Web with the clear backing of the nascent SW community.

Fixed standards for expressing ontologies appeared later in the process, with RDFS and OWL becoming recommendations in 2004. OWL evolved from other ontology language efforts, including OIL [50] and DAML [51], whose merged product, DAML+OIL, was the most important predecessor to OWL [52]. In January 2008, the query language SPARQL became a W3C recommendation, while the Rule Interchange Format RIF was under development in mid-2008.

Figure 3, created in 2003, illustrates the pattern of SW development using the visual metaphor of a tide flowing onto a beach (this diagram is widely available, but see [53]). From top to bottom in the diagram are the various layers of the SW diagram, from trust and proof down to data exchange and markup. From left to right come the various stages in a rough lifecycle from research to deployment: the first stage is a blue-sky research project; the second is the production of a stable system or formalism that is not a standard; the best aspects of these systems are then used as the bases for W3C standards, and the final stage is one of wide deployment. Hence, for instance, early ontology efforts like Cyc and description logics led to efforts such as DAML and OIL, which in turn helped create OWL. Wide deployment of OWL then results in a so-called 'Web of meaning'.

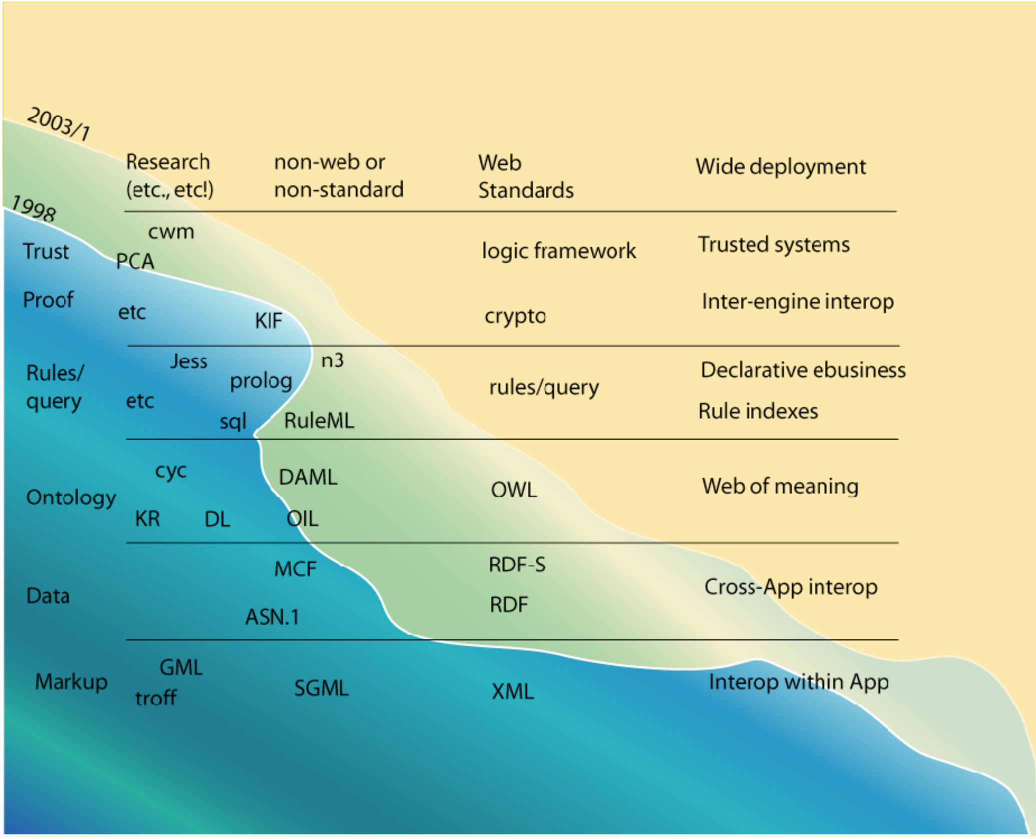

Figure 3: A Representation of the Progress of Semantic Web Development

The 'sea' of research and deployment approaches from the bottom left of Figure 3 to the top right, as the 'tide' comes in. Hence in 1998, various formalisms were in place for all the various levels of representation of the SW, but only XML was a Web standard and beginning to be used widely. By 2003, OWL and RDFS were close to their final forms, and RDF was beginning to be used widely for cross-application interoperability. At the time of writing the 'tide' has advanced further to the right, so work is ongoing on rule language RIF, and query language SPARQL is a candidate W3C standard. Meanwhile OWL is being used more frequently by ontology builders.

The SW's history to date is largely one of standard-setting. However, it has also been argued that, analogous to other systems which have spread quickly and grown exponentially, what is needed is a 'killer app' (i.e. an application that will meet a felt need and create a perception of the technology as 'essential'). Less ambitiously, the SW's spread depends not only on having an impressive set of formalisms, but also software tools to use information represented in those formalisms [49]. The SW is clearly not at the time of writing an information resource in routine use. Nevertheless, there are some applications where SW technologies are serving valuable purposes, and we review some of these in the next section.

APPLICATIONS AND SYSTEMS

PROPERTIES OF SYSTEMS

In general, SW projects tend to exhibit a few constant features. They generate new ontologies for the application domain (art, or computer science), and use them to interrogate large stores of data, which could be legacy data or freshly harvested. Hence a body of evidence is building up that ontologies have an important role in mediating the integration of data from heterogeneous sources.

Furthermore, the results of SW projects are generally presented using custom-built interfaces. This hints at a very important research area, which is the development of scalable visualisers capable of navigating the graph of connected information expressed in RDF. As can be seen, the importance of applications and user interfaces was made clear in the latest version of the layered SW diagram (Figure 1).

In this section we will look at active SW successes, focusing on application areas and types, then commercial/real world systems, before finally looking at some of the more successful academic efforts as judged by the SW development community itself.

APPLICATION AREAS

There are areas where the SW is already an important tool, often in small focused communities with pressing information-processing requirements and various more-or-less common goals. Such communities can function as early adopters of the technology, exactly as the high energy physics discipline played a vital role in the development of the WWW. A series of case studies and use cases is maintained at [54].

The most important application for SW technology is *e-science*, the data-driven, computationally-intensive pursuit of science in highly distributed computational environments [55]. Very large quantities of data are created by analyses and experiments in disciplines such as particle physics, meteorology and the life sciences. Furthermore, in many contexts, different communities of scientists will be working in an interdisciplinary manner, which means that data from various fields (e.g. genomics, clinical drug trials and epidemiology) need to be integrated. Many accounts of distinct and complex systems (e.g. the human body, the environment) consist of data brought from disciplines varying not only in vocabulary, but also in the scale of description; understanding such systems, and the way in which events at the micro-scale affect the macro-scale and *vice versa*, is clearly an important imperative. Many scientific disciplines have devoted resources to the creation of large-scale and robust ontologies for this and other purposes. The most well-known of these is the *Gene Ontology*, a controlled vocabulary to describe gene and gene product attributes in organisms, and related vocabularies developed by Open Biomedical Ontologies [56]. Others include the Protein Ontology, the Cell Cycle Ontology, MeSH (Medical Subject Headings, used to index life science publications), SNOMED (Systematized Nomenclature of Medicine) and AGROVOC (agriculture, forestry, fisheries and food).

E-government is another potentially important application area, where information is deployed widely, and yet is highly heterogeneous. Government information varies in provenance, confidentiality and "shelf life" (some information will be good for decades or even centuries, while other information can be out of date within hours), while it can also have been created by various levels of government (national/federal, regional, state, city, parish). Integrating that information in a timely way is clearly an important challenge (see for instance a pilot study for the United Kingdom's Office of Public Sector Information, exploring the use of SW technologies for disseminating, sharing and reusing data held in the public sector [57]).

COMMERCIAL ACTIVITY

There is an increasing number of applications that allow a deeper querying of linked data. We have already discussed DBpedia [36], DBLP [37] and GeoNames [38]. Commercial applications are also beginning to appear. Garlik [58] is a company seeking to exploit Semantic Web-style technologies to provide individual consumers with more power over their digital data. It reviews what is held about people, harvesting data from the open Web, and represents this in a people-centric structure. Natural Language Processing is used to find occurrences of people's names, sensitive information, and relations to other individuals and organisations.[1] Twine [59] is intended to enable people to share knowledge and information, and to organise that information using various SW technologies (also, like Garlik, using Natural Language

[1] Declaration of interest: Wendy Hall is Chair of the Garlik Advisory Board.

Processing). Twine's developer Nova Spivack has coined the term 'knowledge networking' to describe the process, analogous to the Web 2.0 idea of 'social networking'.

The increasing maturity of SW technology is being shown by the growing number of successful vendors of SW technology. We have already seem OWLIM [31], which was developed by Ontotext, a semantic technology lab focused on technologies to support the SW and Semantic Web services based in Sofia, Bulgaria and Montreal, Canada; Ontotext has been and is a partner in a number of major SW research projects [60]. Ontoprise, based in Karlsruhe, Germany, is a software vendor for implementing SW infrastructure in large, distributed enterprises; its products include OntoBroker, which provides ontology support using the W3C recommended languages OWL, RDFS and SPARQL, and Semantic MediaWiki+, a collaborative knowledge management tool [61]. Asemantics, with offices in Italy, Holland and the United Kingdom, uses a combination of Web 2.0 paradigms with SW technologies such as XML and RDF. The SW technologies are powerful representational tools but often perceived as hard to use and search, so Asemantics attempts to exploit the perceived usability of Web 2.0 to present data in more widely accepted formats [62].

ACADEMIC WORK: THE SEMANTIC WEB CHALLENGE

Much of the major work in the SW has been carried out in the academic sphere, and in funded research projects between academic and commercial partners, and is reported in journals and conferences (see end of article for a list of the more importance conferences). Any review of academic work in this field will inevitably be selective; for the purposes of this article we will focus on a particular effort to nurture applications, the *Semantic Web Challenge*.

The SW Challenge was created in 2003, and associated with the International Semantic Web Conference (ISWC) of that year. Since then it has become an annual competition to create an application that shows SW technology in its best aspects, and which can act as a 'benchmark' application. Hence the SW Challenge gives us a series of illustrative applications thought by researchers' peers to constitute best SW practice [63].

To meet the criteria for the Challenge, a tool or system needs to meet a number of requirements [64], which provide a useful characterisation of the expectations governing an SW system, and are suggestive of the expected properties of SW applications. For instance, it should use information from sources that are distributed and heterogeneous, of real-world complexity and with diverse ownership. It should assume an open world, and that the information is never complete, and it should use some formal description of the meaning of the data. Optional criteria include a use of data in some way other than the creators intended, use of multimedia, use of devices other than a PC. Applications need not be restricted to information retrieval, and ideally the system would be scalable in terms of the amount of data used and the number of distributed components cooperating. All these criteria indicate areas where SW systems would be expected to have an advantage.

The winners of the SW Challenge to date are as follows.

2003: CS AKTive Space (University of Southampton), an integrated application which provides a way to explore the UK Computer Science Research domain across multiple dimensions for multiple stakeholders, from funding agencies to individual researchers, using information harvested from the Web, and mediated through an ontology [65].

2004: Flink (Vrije Universiteit Amsterdam), a 'Who's Who' of the SW which allows the interrogation of information gathered automatically from Web-accessible resources about researchers who have participated in ISWC conferences [66].

2005: CONFOTO (appmosphere web applications, Germany), a browsing and annotation service for conference photographs [67].

2006: MultimediaN E-Culture Demonstrator (Vrije Universiteit Amsterdam, Centre for Mathematics and Computer Science, Universiteit van Amsterdam, Digital Heritage Netherlands and Technical University of Eindhoven), an application to search, navigate and annotate annotated media collections interactively, using collections from several museums and art repositories [68].

2007: Revyu.com (Open University), a reviewing and rating site specifically designed for the SW, allowing reviews to be integrated and interlinked with data from other sources (in particular, other reviews) [69].

CONTROVERSIES

The SW vision has always generated controversy, with a number of commentators being highly sceptical of its prospects. Let us briefly review some of the disputed issues.

THE SEMANTIC WEB AS "GOOD OLD-FASHIONED ARTIFICIAL INTELLIGENCE"

One view holds that the SW is basically a throwback to the project to programme machine intelligence which was jokingly christened by John Haugeland 'GOFAI' (Good Old-Fashioned AI). This proved impossible: so much of human intelligence is implicit and situated that it was too hard a problem to write down everything a computer needed to know to produce output that exhibited human-like intelligence. For instance, if a human is told about a room, further explanations that a room generally has a floor, at least three walls, usually four, and a ceiling, and some method of ingress that is generally but not always a door, are not required. But a computer needs to be told these mundane facts explicitly – and similarly every time it is introduced to a new concept [70].

One attempt to work around this problem is the Cyc project, set up in 1984, which aims to produce a gigantic ontology that will encode all common-sense knowledge of the type about the room given above, in order to support human-like reasoning by machines [71]. The project has always aroused controversy, but it is fair to say that over two decades later, GOFAI is no nearer. The implicit nature of common-sense knowledge arguably makes it impossible to write it all down.

Many commentators have argued that the SW is basically a re-creation of the (misconceived) GOFAI idea, that the aim is to create machine intelligence over the Web, to allow machines to reason about Web content in such a way as to exhibit intelligence [72, 73]. This, however, is a misconception, possibly abetted by the strong focus in the 2001 *Scientific American* article on an agent-based vision of the SW [46]. Like many GOFAI projects, the scenarios in that article have prominent planning components. There is also continuity between the AI tradition of work on formal knowledge representation and the SW project of developing ontologies (see below).

The SW has less to do with GOFAI as with context-based machine reasoning over content (and the provision of machine-readable data on the Web). The aim is not to bring a single ontology, such as Cyc, to bear on all problems (and therefore implicitly to define or anticipate all problems and points of view in the ontology definition), but rather to allow data to be interrogated in ways that were not anticipated by their creators. Different ontologies will be appropriate for different purposes; composite ontologies can be assembled from distributed parts (thanks to the design of OWL); and it is frequently very basic ontologies (defining simple terms such as 'customer', 'account number' or 'account balance') that deliver large amounts of content. It is, after all, a matter of fact that people from different communities and disciplines can and do interact without making any kind of common *global* ontological commitment [1, 6, 74].

Indeed, we can perhaps learn from the experience of hype and reaction that accompanied the development of artificial intelligence (AI). There has been a great deal of criticism of AI, but much has been learned from AI research and some AI methods and systems are now routinely exploited in a number of applications. The same may be expected of the SW. We should not expect to wake up one morning with the SW implemented and ready for use. Rather, a likelier model is that SW technologies will be incorporated into more systems 'behind the scenes' wherever methods are needed to deal with signature SW problems (large quantities of distributed heterogeneous data).

ARGUMENTS FOR AND AGAINST ONTOLOGIES

The importance of ontologies for the SW has been another point of friction with those who believe the programme unrealistic. Ontologies are seen as expensive to develop and hard to maintain. Classification of objects is usually done relative to some task, and as the nature of the task changes, ontologies can become outdated. Classifications are also made relative to some background assumptions, and impose those assumptions onto the resulting ontology. To that extent, the expensive development of ontologies reflects the world view of the ontology builders, not necessarily the users. They are top-down and authoritarian, and therefore opposed to the Web ethos of decentralisation and open conversation. They are fixed in advance, and so they don't work very well to represent knowledge in dynamic, situated contexts [75, 76, 77].

Furthermore, say the critics, the whole point of the Web as a decentralised, linked information structure is that it reflects the needs of its large, heterogeneous user base which includes very many people who are naïve in their interactions. The infrastructure has to be usable by such people, which argues for simplicity. The rich linking structure of the current Web, combined with statistically-based search engines such as Google, is much more responsive to the needs of unsophisticated users. The Semantic Web, in contrast, demands new information markup practices, and corporations and information owners need to invest in new technologies. Not only that, but current statistical methods will scale up as the number of users and interactions grows, whereas logic-based methods such as those advocated by the SW, on the other hand, scale less well [cf. e.g. 78].

Folksonomies

One development as part of the so-called 'Web 2.0' paradigm (of systems, communities and services which facilitate collaboration and information-sharing among users) that has drawn attention in this context is that of the 'folksonomy'. Folksonomies have arisen out of the recent move to allow users to 'tag' content on Web 2.0 sites such as the image-sharing site Flickr, and the video-sharing site YouTube. Having seen content, users are allowed to tag it with key words, which, when the number of users has become large enough, results in a structure of connections and classifications emerging without central control. Their promoters argue that folksonomies 'really' express the needs of their users (since all the structure has arisen out of their use-based classifications), whereas ontologies 'really' express the needs of authorities who can 'impose' their views from the top down [76].

However, folksonomies are much less expressive than ontologies; they are basically variants on keyword searches. A tag 'SF' may refer to a piece of science fiction, or to San Francisco, or something else from the user's private idiolect. Indeed, that ambiguity arises even if we make the unrealistic assumption of a monoglot English user community. Once we realise speakers of other languages will use a system, then there are further possible ambiguities – for instance, in German 'SF' might refer to the Swiss television station Schweizer Fernsehen.

Resolving this controversy

When a community is large enough and the benefits clear, then a large-scale ontology building and maintenance programme is justified. In a recent note, Berners-Lee argues that such conditions will be perhaps more frequently encountered than sceptics believe. On the very broad assumptions that the size of an ontology-building team increases as the order of the log of the size of the ontology's user community, and that the resources needed to build an ontology increase as the order of the square of community size, the cost per individual of ontology building will diminish rapidly as user community size increases. Of course these assumptions are not intended to be deeply realistic, so much as indicative of how the resource implications diminish as the community increases in size. Berners-Lee's moral: "Do your bit. Others will do theirs" [74].

Even so, not all ontologies need to be of great size and expressive depth. Certainly the claim that has been made that the SW requires a single ontology of all discourse on the model of Cyc, but this is not backed up by the SW community. Such an ontology, even if possible, would not scale, and in a decentralised structure like the Web its use could not be enforced. We should rather expect a lot of use of small-scale, *shallow* ontologies defining just a few terms that nevertheless are widely applicable [74]. Experience in building real-world SW systems often shows that expectations about the cost and complexity of the ontologies required are overblown, and the ontology-building process can be relatively straightforward and cheap [79].

For example, the machine-readable Friend-of-a-Friend (FOAF) ontology is intended to describe people, their activities and their relations to other people. It is not massively complex, and indeed publishing a FOAF account of oneself is a fairly simple matter of form-filling (using the FOAF-a-matic tool [80]). But the resulting network of people (showing their connections to other people) has become very large indeed. A survey performed in 2004 discovered over 1.5 million documents using the FOAF ontology [81].

With respect to Folksonomies, it is important to note that ontologies and folksonomies serve different purposes. Folksonomies are based on word tags, whereas the basis for ontology reference is via a URI. One of the main aims of ontology definition is to *remove* ambiguity – not globally, for this may well be impossible, but rather within the particular context envisaged by the developer (see the section on 'Symbol Grounding' below). Folksonomies will necessarily inherit the ambiguity of the natural language upon which they are based. And while folksonomies emerge from data sharing practices, it is not necessarily the case the ontologies are authoritarian; rather, the latter should ideally be *rationalisations* of current sharing practice. This does entail departure from current practice, but not necessarily of great magnitude. Indeed, a strong possibility is to use cheaply-gathered folksonomies as starting points for ontology development, gradually morphing the Web 2.0 structures into something with greater precision and less ambiguity [82].

Symbol grounding

An important aspect of the SW is that URIs must be interpreted consistently. However, terms and symbols are highly variable in their definitions and use through time and space. The SW project ideally needs processes whereby URIs are given to objects, such that the management of these processes is by communities and individuals, endorsed by the user community, who ensure consistency. This URI 'ownership' is a critical to the smooth functioning of the SW [1].

But the process of *symbol grounding* (i.e. ensuring a fixed and known link between a symbol and its referent) is at best hard, and at worst (as argued by Wittgenstein, for instance) impossible [83, 84]. Meanings do not stay fixed, but alter, often imperceptibly. They are delineated not only by traditional methods such as the provision of necessary and sufficient conditions, but also by procedures, technologies and instrumentation, and alter subtly as practice alters.

Any attempt to fix the reference of URIs is a special case of symbol grounding, and is consequently hard to do globally. It is certainly the case that attempting to resist the alteration in community practices and norms, and reformulation of meanings of terms, would be doomed.

Yorick Wilks has argued that since much knowledge is held in unstructured form, in plain text, automatic Natural Language Processing techniques, statistically-based, can be used to 'ground' meanings of terms for the SW [72]. Berners-Lee on the other hand maintains that the SW is necessarily based on logic and firm definitions (even if those definitions were imperfect, or highly situated and task-relative), not words, use patterns and statistics. Wilks' point is that the aim of defining terms in logic is too idealistic, and anyway depends on assumptions about ordinary word meaning. Berners-Lee's counterargument is, in effect, that though meanings are not stable, they can be stable *enough* relative to individual applications and in particular contexts to allow the SW approach to work.

CONCLUSION

The SW has been somewhat misunderstood in some commentaries. Its aim is not to force users to accept large ontologies remote from data-sharing practice imposed by shadowy authorities. Neither is it intended to produce a theory of all discourse, or to reproduce GOFAI. Rather, it is intended to shift the emphasis of the Web from being a web of documents to a web of linked *data*. It is the development of formalisms and technologies facilitating the creation, sharing and querying of linked data using sharable ontologies to establish common interpretations. For this reason, an alternative name for the SW is the *Web of data*.

The SW is a work in progress. As it stands, the ‘buy in’ to the SW has not yet produced the desirable network effects, although several disciplines are enthusiastic early adopters of the technology (e.g. the e-science community). And there are still several important research issues outstanding. It is not yet known how best to: query large numbers of heterogeneous information stores at many different scales; translate between, merge, prune or evaluate ontologies; visualise the SW; establish trust and provenance of the content.

As complex technologies and information infrastructures are developed, there is a dynamic feedback between requirements analysis, engineering solutions and hard-to-predict global behaviour of human, machine and hybrid systems. Understanding how basic engineering protocols governing how computers talk to each other can result in social movements at a very different level of abstraction is very hard, yet essential to realising the SW vision. Indeed, such understanding, the defining purpose of the discipline of *Web Science*, is essential to ensuring that *any* Web-based information structure is beneficial [5].

ACKNOWLEDGEMENTS

The authors would like to thank Tim Berners-Lee, Nigel Shadbolt, James A. Hendler, Daniel J. Weitzner, Harith Alani, Marcia J. Bates and an anonymous referee for helpful comments and discussions.

REFERENCES

1. Shadbolt, N.; Hall, W.; Berners-Lee, T. The Semantic Web revisited. IEEE Intelligent Systems **2006**, *21* (3), 96-101.

2. Fensel, D.; Bussler, C.; Ding, Y.; Kartseva, V.; Klein, M.; Korotkiy, M.; Omelayenko, B.; Siebes, R. Semantic Web application areas. 7th International Workshop on Applications of Natural Language to Information Systems (NLDB 2002), Stockholm, Sweden, June~27-28, 2002, http://www.cs.vu.nl/~ronny/work/NLDB02.pdf, 2002 (accessed July 2008).

3. Fensel, D. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd Ed.; Springer, Berlin, 2004.

4. Berners-Lee, T. Relational databases on the Semantic Web. http://www.w3.org/DesignIssues/RDB-RDF.html, 1998 (accessed December 2007).

5. Berners-Lee, T.; Hall, W.; Hendler, J.; Shadbolt, N.; Weitzner, D. Creating a science of the Web. Science **2006**, *313* (5788), 769-771.

6. Berners-Lee, T.; Hall, W.; Hendler, J.A.; O'Hara, K.; Shadbolt, N.; Weitzner, D.J. A framework for Web Science. Foundations and Trends in Web Science **2006**, *1* (1), 1-134.

7. Klyne, G.; Carroll, J.J.; McBride, B. Resource Description Framework (RDF): concepts and abstract syntax. http://www.w3.org/TR/rdf-concepts/, 2004 (accessed December 2007).

8. Berners-Lee, T. Principles of design. http://www.w3.org/DesignIssues/Principles.html, 1998 (accessed December 2007).

9. Berners-Lee, T.; Fielding, R.; Masinter, L. Uniform Resource Identifier (URI): generic syntax. http://gbiv.com/protocols/uri/rfc/rfc3986.html, 2005 (accessed December 2007).

10. Bray, T.; Paoli, J.; Sperberg-McQueen, C.M.; Maler, E.; Yergeau, F. Extensible Markup Language (XML) 1.0 (Fourth Edition). http://www.w3.org/TR/xml/, 2006 (accessed December 2007).

11. Brickley, D.; Guha, R.V.; McBride, B. RDF vocabulary description language 1.0: RDF Schema. http://www.w3.org/TR/rdf-schema/, 2004 (accessed December 2007).

12. McGuinness, D.L.; van Harmelen, F. OWL Web Ontology Language overview. http://www.w3.org/TR/owl-features/, 2004 (accessed December 2007).

13. Antoniou, G.; van Harmelen, F. *A Semantic Web Primer*; MIT Press, Cambridge MA, 2004.

14. Dean, M.; Schreiber, G.; Bechhofer, S.; van Harmelen, F.; Hendler, J.; Horrocks, I.; McGuinness, D.L.; Patel-Schneider, P.F.; Stein, L.A. OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref/, 2004 (accessed December 2007).

15. Smith, M.K.; Welty, C.; McGuiness, D.L. OWL Web Ontology Language guide. http://www.w3.org/TR/owl-guide/, 2004 (accessed December 2007).

16. Boley, H.; Kifer, M. RIF basic logic dialect. http://www.w3.org/TR/rif-bld/, 2007 (accessed December 2007).

17. Prud'hommeaux, E.; Seaborne, A. SPARQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/, 2007 (accessed December 2007).

18. Golbeck, J.; Trust on the World Wide Web: a survey. Foundations and Trends in Web Science **2006**, *1* (2), 1-72.

19. Bonatti, P.A.; Duma, C.; Fuchs, N.; Nejdl, W.; Olmedilla, D.; Peer, J.; Shahmehri, N. Semantic Web policies – a discussion of requirements and research issues. In *The Semantic Web: Research and Applications,* 3rd European Semantic Web Conference 2006 (ESWC-06), Budva, Montenegro, 2006; Sure, Y., Domingue, J., Eds.; Springer: Berlin, 2006.

20. O'Hara, K.; Alani, H.; Kalfoglou, Y.; Shadbolt, N. Trust strategies for the Semantic Web. In Workshop on Trust, Security and Reputation on the Semantic Web, 3rd International Semantic Web Conference (ISWC 04), Hiroshima, Japan, 2004, http://eprints.ecs.soton.ac.uk/10029/ (accessed December 2007).

21. Golbeck, J.; Hendler, J. Accuracy of metrics for inferring trust and reputation in Semantic Web-based social networks. In *Engineering Knowledge in the Age of the Semantic Web*, Proceedings of 14th International Conference, EKAW 2004, Whittlebury Hall, United Kingdom, 2004; Motta, E., Shadbolt, N., Stutt, A., Gibbins, N., Eds.; Springer: Berlin, 2004, 116-131.

22. Groth, P.; Jiang, S.; Miles, S.; Munroe, S.; Tan, V.; Tsasakou, S.; Moreau, L. An architecture for provenance systems. http://eprints.ecs.soton.ac.uk/13216/1/provenanceArchitecture10.pdf, 2006 (accessed December 2007).

23. O'Hara, K.; Shadbolt, N. *The Spy in the Coffee Machine: The End of Privacy As We Know It*; Oneworld, Oxford, 2008.

24. http://creativecommons.org/about/ (accessed December 2007).

25. Weitzner, D.J.; Hendler, J.; Berners-Lee, T.; Connolly, D. Creating a policy-aware Web: discretionary, rule-based access for the World Wide Web. In *Web and Information Security*; Ferrari, E., Thuraisingham, B., Eds.; Idea Group Inc: Hershey, PA, 2005.

26. http://jena.sourceforge.net/ (accessed December 2007).

27. http://sourceforge.net/projects/threestore (accessed December 2007).

28. Harris, S.; Gibbins, N. 3store: efficient bulk RDF storage. In Proceedings of the 1st International Workshop on Practical and Scalable Systems, Sanibel Island, Florida, 2003; http://km.aifb.uni-karlsruhe.de/ws/psss03/proceedings/harris-et-al.pdf (accessed December 2007).

29. http://www.oracle.com/technology/tech/semantic_technologies/index.html (accessed December 2007).

30. http://www.ontotext.com/owlim/ (accessed July 2008).

31. Kiryakov, A.; Ognyanov, D.; Manov, D. OWLIM: a pragmatic semantic repository for OWL. In *Web Information and Systems Engineering –WISE 2005 Workshops*, Proceedings of the Workshop on Scalable Semantic Web Knowledge Base Systems at WISE 2005, New York, November 2005; Dean, M., Guo, Y., Jun, W., Kaschek, R., Krishnaswamy, S., Pan, Z., Sheng, Q.Z., Eds.; Springer, Berlin, 2005, 182-192, http://www.ontotext.com/publications/ssws_owlim.pdf (accessed July 2008).

32. McBride, B.; Jena: implementing the RDF model and syntax specification. In *Proceedings of the 2nd International Workshop on the Semantic Web: SemWeb 2001*, at World Wide Web Conference 2001, Hong Kong, May, 2001; Decker, S., Fensel, D., Sheth, A., Staab, S.; Eds.; CEUR-WS Vol. 40, 2001, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-40/mcbride.pdf (accessed July 2008).

33. Sirin, E.; Parsia, B.; Cuenca Grau, B.; Kalyanpur, A.; Katz, Y.; Pellet: a practical OWL-DL reasoner. Journal of Web Semantics **2007**, *5* (2), 51-53.

34. http://kaon2.semanticweb.org/ (accessed July 2008).

35. Fensel, D.; Van Harmelen, F.; Unifying reasoning and search to Web scale. IEEE Internet Computing **2007**, *11* (2), 96, 94-95 *(sic)*.

36. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: a nucleus for a Web of open data. In Proceedings of the 6th International Semantic Web Conference 2007, Busan, South Korea, 2007, http://iswc2007.semanticweb.org/papers/715.pdf (accessed December 2007).

37. http://www4.wiwiss.fu-berlin.de/dblp/ (accessed December 2007).

38. http://www.geonames.org/ (accessed December 2007).

39. Connolly, D., Ed.; Gleaning Resource Descriptions from Dialects of Langages (GRDDL). http://www.w3.org/TR/grddl/, 2007 (accessed December 2007).

40. Schröder, M.; Zovato, E.; Pirker, H.; Peter, C.; Burkhardt, F. W3C emotion incubator group report. http://www.w3.org/2005/Incubator/emotion/XGR-emotion/, 2007 (accessed December 2007).

41. Handschuh, S., Staab, S., Eds.; *Annotation for the Semantic Web*; Amsterdam: IOS Press, 2003.

42. Troncy, R.; van Ossenbruggen, J.; Pan, J.Z.; Stamou, G.; Halaschek-Wiener, C.; Simou, N.; Tsouvaras, V. Image annotation on the Semantic Web. http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation/, 2007 (accessed December 2007).

43. Handschuh, S.; Staab, S.; Ciravegna, F. S-CREAM – Semi-automatic CREAtion of Metadata. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Proceedings of 13th International Conference, EKAW 2002, Siguënza, Spain, 2002; Gómez-Pérez, A., Benjamins, V.R., Eds.; Springer: Berlin, 2002, 358-372.

44. Vargas-Vera, M.; Motta, E.; Domingue, J.; Lanzoni, M.; Stutt, A.; Ciravegna, F. MnM: ontology-driven semi-automatic and automatic support for semantic markup. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Proceedings of 13th International Conference, EKAW 2002, Siguënza, Spain, 2002; Gómez-Pérez, A., Benjamins, V.R., Eds.; Springer: Berlin, 2002, 379-391.

45. Hendler, J.; Shadbolt, N.; Hall, W.; Berners-Lee, T; Weitzner, D.; Web Science: an interdisciplinary approach to understanding the World Wide Web. Communications of the ACM **2008**, *51* (7), 60-69.

46. Berners-Lee, T; Hendler, J.; Lassila, O.; The Semantic Web. Scientific American **May 2001**, http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 (accessed December 2007).

47. Berners-Lee, T. Foreword. *Spinning the Semantic Web: Bringing the World Wide Web to its full potential*; Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Eds.; MIT Press: Cambridge, MA, 2003, xi-xxiii.

48. Golbeck, J.; Grove, M.; Parsia, B.; Kalyanpur, A.; Hendler, J. New tools for the Semantic Web. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Proceedings of 13th International Conference, EKAW 2002, Siguënza, Spain, 2002; Gómez-Pérez, A., Benjamins, V.R., Eds.; Springer: Berlin, 2002, 392-400.

49. Alani, H.; Kalfoglou, Y.; O’Hara, K.; Shadbolt, N. Towards a killer app for the Semantic Web. In *The Semantic Web*, Proceedings of the International Semantic Web Conference 2005, Hiroshima, Japan, 2005; Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A., Eds.; Springer, Berlin, 2005, 829-843.

50. Fensel, D.; Horrocks, I.; van Harmelen, F.; Decker, S.; Erdmann, M.; Klein, M. OIL in a nutshell. In *Knowledge Engineering and Knowledge Management: Methods, Models and Tools*, Proceedings of 12th European Knowledge Acquisition Workshop (EKAW 2000), Juan-les-Pins, France, October 2000; Dieng, R., Corby, O., Eds.; Springer, Berlin, 2000, 1-16, http://www.cs.vu.nl/~ontoknow/oil/downl/oilnutshell.pdf (accessed July 2008).

51. http://www.daml.org/about.html (accessed July 2008).

52. Patel-Schneider, P.; Horrocks, I.; van Harmelen, F. Reviewing the design of DAML+OIL: an ontology language for the Semantic Web. In Proceedings of the 18th National Conference on Artificial Intelligence (AAAI02), Edmonton, Canada, 2002; http://www.cs.vu.nl/~frankh/postscript/AAAI02.pdf (accessed December 2007).

53. Connolly, D.; Semantic Web update: OWL and beyond. http://www.w3.org/2003/Talks/1017-swup/all.htm, 2003 (accessed December 2007).

54. http://www.w3.org/2001/sw/sweo/public/UseCases/ (accessed December 2007).

55. Hendler, J.; de Roure, D. E-science: the grid and the Semantic Web. IEEE Intelligent Systems **2004** *19* (1), 65-71.

56. http://www.geneontology.org/ (accessed July 2008).

57. Alani, H.; Dupplaw, D.; Sheridan, J.; O'Hara, K.; Darlington, J.; Shadbolt, N.; Tullo, C. Unlocking the potential of public sector information with Semantic Web technology. In Proceedings of the 6th International Semantic Web Conference 2007, Busan, South Korea, 2007, http://iswc2007.semanticweb.org/papers/701.pdf (accessed December 2007).

58. https://www.garlik.com/index.php (accessed December 2007).

59. http://www.twine.com/ (accessed December 2007).

60. http://www.ontotext.com/index.html (accessed July 2008).

61. http://www.ontoprise.de/index.php?id=134 (accessed July 2008).

62. http://www.asemantics.com/index.html (accessed July 2008).

63. http://www.informatik.uni-bremen.de/agki/www/swc/index.html (accessed December 2007).

64. http://challenge.semanticweb.org/ (accessed December 2007).

65. schraefel, m.m.c.; Shadbolt, N.R.; Gibbins, N.; Glaser, H.; Harris, S. CS AKTive Space: representing computer science on the Semantic Web. In *Proceedings of WWW 2004*, New York, 2004; http://eprints.ecs.soton.ac.uk/9084/ (accessed December 2007).

66. Mika, P. Flink: Semantic Web technology for the extraction and analysis of social networks. *Journal of Web Semantics* **2005** *3* (2). http://www.websemanticsjournal.org/papers/20050719/document7.pdf (accessed December 2007).

67. Nowack, B. CONFOTO: A semantic browsing and annotation service for conference photos. In *The Semantic Web*, Proceedings of the International Semantic Web Conference 2005, Hiroshima, Japan, 2005; Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A., Eds.; Springer, Berlin, 2005, 1067-1070.

68. Schreiber, G.; Amin, A.; van Assem, M.; de Boer, V.; Hardman, L.; Hildebrand, M.; Hollink, L.; Huang, Z.; van Kersen, J.; de Niet, M.; Omelayenko, B.; van Ossenbruggen, J.; Siebes, R.; Taekema, J.; Wielemaker, J.; Wielinga, B. MultimediaN e-culture demonstrator. http://www.cs.vu.nl/~guus/papers/Schreiber06a.pdf, 2006 (accessed December 2007).

69. Heath, T.; Motta, E. Revyu.com: a reviewing and rating site for the Web of data. In Proceedings of the 6th International Semantic Web Conference 2007, Busan, South Korea, 2007, http://iswc2007.semanticweb.org/papers/889.pdf (accessed December 2007).

70. Haugeland, J. Understanding natural language. Journal of Philosophy **1979** *76*, 619-632.

71. Lenat, D.B. Cyc: a large-scale investment in knowledge infrastructure. Communications of the ACM **1995** *38* (11).

72. Jones, K.S.; What's new about the Semantic Web? Some questions. SIGIR Forum **2004**, *38* (2), http://www.sigir.org/forum/2004D/sparck_jones_sigirforum_2004d.pdf (accessed December 2007).

73. Wilks, Y. The Semantic Web: apotheosis of annotation, but what are its semantics? IEEE Intelligent Systems **2008**, *23* (3), 41-49.

74. Berners-Lee, T. The fractal nature of the Web. http://www.w3.org/DesignIssues/Fractal.html, 2007 (accessed December 2007).

75. Pike, W.; Gahegan, M. Beyond ontologies: toward situated representations of scientific knowledge. International Journal of Human-Computer Studies **2007**, *65* (7), 674-688.

76. Shirky, C. Ontology is overrated: categories, links and tags. http://www.shirky.com/writings/ontology_overrated.html, 2005 (accessed December 2007).

77. Stevens, R.; Egaña Aranguren, M.; Wolstencroft, K.; Sattler, U.; Drummond, N.; Horridge, M.; Rector, A. Using OWL to model biological knowledge. International Journal of Human-Computer Studies **2007**, *65* (7), 583-594.

78. Zambonini, D. The 7 (f)laws of the Semantic Web. http://www.oreillynet.com/xml/blog/2006/06/the_7_flaws_of_the_semantic_we.html, 2006 (accessed December 2007).

79. Alani, H.; Chandler, P.; Hall, W.; O'Hara, K.; Shadbolt, N.; Szomsor, M. Building a pragmatic Semantic Web. IEEE Intelligent Systems **2008**, *23* (3), 61-68.

80. http://www.ldodds.com/foaf/foaf-a-matic (accessed December 2007).

81. Ding, L.; Zhou, L.; Finin, T.; Joshi, A. How the Semantic Web is being used: an analysis of FOAF documents. In Proceedings of the 38th International Conference on System Sciences, 2005. http://ebiquity.umbc.edu/_file_directory_/papers/120.pdf (accessed December 2007).

82. Mika, P. Ontologies are us: a unified model of social networks and semantics. Journal of Web Semantics **2007**, *5* (1), 5-15.

83. Harnad, S. The symbol grounding problem. Physica D **1990**, *42*, 335-346; http://users.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad90.sgproblem.html (accessed December 2007).

84. Wittgenstein, L. *Philosophical Investigations*. Basil Blackwell: Oxford, 1953.

FURTHER READING

Antoniou, G.; van Harmelen, F. *A Semantic Web Primer*; MIT Press, Cambridge MA, 2004.

Berners-Lee, T. *Weaving the Web: the past, present and future of the World Wide Web by its inventor*; Texere Publishing, London, 1999.

Berners-Lee, T.; Hall, W.; Hendler, J.A.; O'Hara, K.; Shadbolt, N.; Weitzner, D.J. A framework for Web Science. Foundations and Trends in Web Science **2006**, *1* (1), 1-134.

Berners-Lee, T.; Hall, W.; Hendler, J.; Shadbolt, N.; Weitzner, D. Creating a science of the Web. Science **2006**, *313* (5788), 769-771.

Berners-Lee, T; Hendler, J.; Lassila, O.; The Semantic Web. Scientific American **May 2001**, http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 (accessed December 2007).

Fensel, D. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd Ed.; Springer, Berlin, 2004.

Fensel, D.; Hendler, J.; Lieberman, H.; Wahlster, W. *Spinning the Semantic Web: bringing the World Wide Web to its full potential*; MIT Press, Cambridge, MA, 2003.

Shadbolt, N.; Hall, W.; Berners-Lee, T. The Semantic Web revisited. IEEE Intelligent Systems **2006**, *21* (3), 96-101.

There are several important annual conferences for the SW community, including: the World Wide Web Conference (WWW); the International Semantic Web Conference (ISWC – pronounced Iss-wick); the European Semantic Web Conference. These conferences preserve their proceedings online.

The World Wide Web Consortium's Semantic Web activity page is at http://www.w3.org/2001/sw/, and contains references to interviews, manifestos and statements by key SW developers. It also maintains a useful site of case studies and use cases at http://www.w3.org/2001/sw/sweo/public/UseCases/. For Web Science, see http://webscience.org/.