

SWAD-Europe: Semantic Web Advanced Development in Europe A Position Paper

D.Brickley¹, S.Buswell², B.M.Matthews³, L.Miller⁴, D. Reynolds⁵,
and M.D.Wilson³

¹ World-Wide Web Consortium, INRIA, France

² Stilo Ltd., UK

³ CLRC Rutherford Appleton Laboratory, UK

⁴ ILRT, University of Bristol, UK

⁵ Hewlett-Packard Laboratories, UK

Abstract. For the Web to reach its full potential, it must evolve into a Semantic Web, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people. The 'Semantic Web' is a recent initiative of the World Wide Web Consortium (W3C), with the goal of extending the current Web to facilitate Web automation, universally accessible content, and the 'Web of Trust'. However, if the semantic web is going to be adopted and assimilated a clear migration path from present technologies to new ones is required. The SWAD-Europe project aims to support the W3C's Semantic Web initiative in Europe, providing targeted research, demonstrations and outreach to ensure Semantic Web technologies move into the mainstream of networked computing. The project aims to support the development and deployment of W3C Semantic Web specifications through implementation, research and testing activities.

1 Introduction

Tim Berners-Lee has provided a roadmap for developing the Semantic Web [TBL98] with a "pyramid of technologies", with XML at the base, through RDF, RDF Schemas, ontologies, queries and rules, logic and proof to arrive, with Digital signatures, at the Web Of Trust. This scheme has inspired a large amount of activity at various levels of the pyramid, including RDF [RDF] and RDF Schema [RDFS], the DAML+OIL initiative and a variety of other approaches which come under the W3C's Semantic Web Activity Domain. This work has been taken up in some practical applications, notably CC/PP for customising content to different devices, P3P for privacy preferences [P3P], and RSS [RSS] for site syndication. However, this ambitious body of work is in danger of passing the average web developer by. So far, the work undertaken has concentrated on solving the important technical questions underlying the semantic web, with not enough effort on providing complete and readily comprehensible tools and examples for the practical user. If semantic web technologies are going to be adopted as a normal and natural part of the next

generation of the web, then an effort has to be made to provide such accessible material. A new European project, SWAD-Europe, seeks to fill this gap.

2 A Scenario-Led Perspective

As longstanding participants in the Semantic Web, XML and Web developer communities, the project team are familiar with a number of ‘*frequently asked questions*’ that arise when considering Semantic Web technology. These come from technical, consumer, content creator and business perspectives, but a common theme recurs: *technology integration*.

“Which standard should I use?”

“How do use RDF with XML Schemas?”

“...or Web Services with Web Ontologies? ...MathML with RDF-rules?”

Such questions are often themselves a means to an end. The goal is typically not to integrate two different W3C data formats, but to complete some more specific task. Technology-oriented questions often mask an application-oriented need.

We often hear questions such as:

“I am re-engineering our Intranet and want a standards-based way of exchanging (amongst other things) ‘organisational chart’ information about departments and groups. The XML Schema Specification seems relevant, since we keep much of this data in relational databases, and tools exist to export data using XML Schema. Everybody recommends the use of XML, but there seem to be so many different ways of using it. The XML specification provides DTDs; there is also now an XML Namespaces specification, and a number of alternative XML Schema languages.”

“Added to this, articles I read about the Semantic Web suggest I should be using an ‘Ontology language’ (instead?), based on RDF Schema and(/or) DAML+OIL. Since the information we are trying to represent is an organisational chart, we are also considering the use of SVG to create, exchange and edit this information. It is not clear which, if any, of these technologies are most appropriate to use, nor what the relationship between them is.”

From the concerns of managers and technologists, content creators and policy makers, we note this same need. Web technology, and now Semantic Web technology in particular, presents a daunting array of tools, specifications and techniques. The full range of relevant technology, while in principle extremely powerful, also risks stifling or delaying innovation through providing *too much* to choose from.

3 The SWAD Project

While exaggerated here for effect, these concerns are real, current and addressable. To answer this, we need a combination of advanced technology development and a programme of documentation, demonstration, education and outreach.

Semantic Web Advanced Development for Europe (SWAD-Europe) aims to play a key role in the evolution of the Semantic Web, through education and outreach to developers, organisations and content creators; through Open Source implementation

and testing, and through pre-consensus technology development to drive and inform the creation of new Semantic Web standards.

The overarching aim of the project is thus to provide, through all appropriate means, a body of answers to questions that have to date gone unanswered, and to foster grassroots communities within which such concerns are addressed. It is more important to offer clear answers to these questions than it is for us to write software or complex technical reports. The technical research and advanced development activities are a means to an end: facilitating wide-scale Semantic Web deployment. The project will therefore remain responsive to external developments (such as the appearance of unanticipated third-party work, software libraries etc.), refining the technical focus of the research to track the current state of the art, and to respond to the concerns of stakeholder communities.

The period 2002-2004 will see the first wave of mainstream Semantic Web applications. SWAD-Europe's role will be to ensure that the critical technology components required for widespread Semantic Web adoption are readily accessible to European industry, consumers, and developers. This involves finding and maintaining a balance between "in-house" Open Source tool development, community building, outreach and evangelism, combined with more technologically advanced research and analysis to support and field-test Semantic Web standards.

Goals of SWAD-Europe

- To implement scenario-led examples showing the integration of multiple Semantic Web technologies drawing practical use cases from industry, consumer, and developer perspectives.
- To develop a Semantic Web technology integration strategy that emphasises the utility of XML languages (such as SVG, HTML, MathML, XLink) as complementary rather than competing components of the Web.
- To ensure that the European developers, citizens and content creators are kept aware of Semantic Web technology for supporting universal accessibility, device independence and internationalisation.
- To ensure that European Community is kept aware of international best practice, and that best practice within Europe is recognised internationally.
- To undertake targeted research and development in support of these objectives, and in collaboration with the wider European developer community, W3C Member organisations, and related Open Source initiatives.

The SWAD-Europe project is designed to further the evolution of the Semantic Web through a combination of targeted research and community outreach, taking a set of use-case driven scenarios and illustrating how Semantic Web technologies relate to the practical needs of European consumers, business and content producers.

The most innovative aspect of this project lies in the combination of strategies that will be focussed on the problem. By ensuring that our research and development work is driven by scenario-led case studies, SWAD-Europe will provide the European community with a 'big picture' understanding of the Semantic Web, as well as with detailed knowledge of the tools and techniques that make it practically deployable.

4 Practical Steps towards the Semantic Web in Europe

The project will concentrate on the following areas.

Semantic Web Services. Several aspects of the project address the need for convergence between the Semantic Web and Web Service perspectives. This is addressed at three levels: through demonstrating specific worked examples of Semantic Web Services such as those relating to Annotations and to Trust; through exploring options for technical convergence between RDF and Web Service specifications (e.g. W3C RDF and SOAP serialisation syntaxes); and through showing new functionality gained by the application of a Semantic Web approach to the creation, discovery and exploitation of Web Services.

XML / Semantic Web Integration. The project seeks to show significant progress towards the integration of Semantic Web technology (specifically RDF and Web Ontology languages) with the broader family of XML specifications. This includes the creation of strategies that allow knowledge represented diagrammatically (in SVG) to be made accessible to RDF tools and to non-visual user environments such as speech browsers and mobile phones. Similarly, combining MathML with RDF logic will progress the integration of mathematical information into the Semantic Web. Considering the amount of information currently represented in diagrams (e.g. PowerPoint) and spreadsheets (e.g. Excel), advances in these two content areas alone could expose vast amounts of structured information to Semantic Web tools.

APIs and Query Languages. The SWAD-Europe work on API and Query language convergence is intended to combine advances in technical knowledge with the rationalisation and documentation of deployed best practice. If the Semantic Web is to provide a stable platform for advanced information services, it is important to seek consensus amongst developers and application creators on basic functionality for query and data access. While this work will necessitate a certain amount of technical innovation, the emphasis here is on consolidating and stabilising the interface conventions that have emerged over the last 2-3 years amongst RDF developers.

Trust. SWAD-Europe includes research work focussed on issues of trust management for the Semantic Web. Building on earlier RDF-based work at W3C/MIT and elsewhere, this will include an analysis of the *Capabilities* and *Proof Carrying Authentication* approaches to trust, to digitally signed Semantic Web content, and to the creation of 'Web of Trust' applications. The related work on annotations provides a practical, consumer-oriented test-bed for exploring the deployment of this technology.

Annotations and information quality. The developer community has for many years been exploring a variety of strategies for building Annotation systems for Web content. Recent RDF-based Annotation systems have shown the potential offered by Semantic Web technology in this area. The technical innovation targeted by SWAD-Europe is, again, centred around convergence, integration and scalability. Web Annotation systems will only realise their true potential when there is a large body of data accessible for applications to use. We therefore focus our efforts on the identification of common ground amongst current annotation systems, and on demonstrating systems that combine the best aspects of each. This is complemented by the workpackage on scalability, which addresses a need apparent from existing Annotations systems, the storage and query of large bodies of RDF data. Annotations also provide one of the project's application scenario areas, and as such will be the

subject of documentation in the Education and Outreach work, and be the basis of worked examples for the work on queries, the API, trust and scalability.

Scalability. Scalability is a common concern amongst Web developers exploring Semantic Web technology. The project will identify and document a number of strategies and techniques for storing and querying large repositories of RDF data. This technical research will also seek to progress the state of the art in this area, through the exploration of schema-aware data storage and indexing systems.

Thesaurus management. Thesaurus and classification systems are a critical technology for the Semantic Web. They provide a bridge between traditional digital library applications and more complex Ontology-based Semantic Web systems. The need for classification and semantic-tagging was one of the prime motivations for the creation of RDF. As with Annotations, there is a significant body of knowledge in the Semantic Web developer community (and the project team) relating to the use of RDF with Thesaurus systems. Rather than seek a 'great leap forward' in the underlying approach, the innovation goal here is to innovate through consolidation. SWAD-Europe will provide a stable, well-documented approach to RDF Thesaurus deployment that will allow application developers and content producers to integrate established traditional digital library systems with other Semantic Web tools. The importance of multilingual thesauri in a European context motivates the focus on developing a language-tagged multilingual thesaurus system in this workpackage. Particularly in the context on non-linguistic content (e.g. Multimedia, especially digital images), the use of a multilingual thesaurus can provide a very cost-effective means to re-purpose content to new audiences across national and linguistic boundaries.

References

- [TBL98] Berners-Lee (1998) *Semantic Web Road map*
<http://www.w3.org/DesignIssues/Semantic.html>
- [RDF] Lassila O & Swick R (1999) *Resource Description Framework (RDF) Model and Syntax*. <http://www.w3.org/RDF/Group/WD-rdf-syntax/>
- [RDFS] Guha L. & Brickley D. (2000). Resource Description Framework (RDF) Schema Specification 1.0 W3C Candidate Recommendation 27 March 2000
<http://www.w3.org/TR/rdf-schema>
- [RSS] RDF Site Summary (RSS) 1.0
<http://groups.yahoo.com/group/rss-dev/files/specification.html>
- [P3P] Brian McBride, Rigo Wenning, [Lorrie Cranor](#), *An RDF Schema for P3P* W3C Note 25 January 2002. <http://www.w3.org/TR/p3p-rdfschema/>