SWAD−Europe Deliverable 3.4: Initial Workshop Report

Project name:

Semantic Web Advanced Development for Europe (☞SWAD-Europe)

Project Number:

IST-2001-34732

Workpackage name:

3 Dissemination and Implementation

Workpackage description:

☞http://www.w3.org/2001/sw/Europe/plan/workpackages/live/esw-wp-3

Deliverable title:

3.4 Initial Workshop Report

URI:

☞http://www.w3.org/2001/sw/Europe/reports/initial_workshop_report

Author:

Charles McCathieNevile

Abstract:

This report summarizes the SWAD-E Initial Workshop, held in Florence on 16 and 17 October 2002 in conjunction with the DC2002 conference. The workshop explored the use of RDF to support resource discovery, and the integration of RDF technology for discovery with more traditional technology for improved results.

STATUS:

First version published 2002-08-31. This is the completed report dated 2003-06-11; however this report may be further updated over the life of the SWAD-Europe Project to link to new work emerging on the topics of the workshop.

Contents

Executive Summary

This workshop was originally scheduled to be held in the first four months of the project. It was held over until October 2002 (month 6) in order that it be hosted in conjunction with the Dublin Core 2002 conference, which brought together an international audience with related interests and expertise.

This workshop was divided into two parts. the first was a short general introduction to the project, its goals and methodology.

The second part consisted of a specific technical developers workshop, focusing on the issues of combining free text searching and metadata-based discovery. The major outcome was gathering, and where possible answering, ☞frequently asked questions about best practices for using RDF.

The workshops were open to all, through attending the Dublin Core 2002 Conference.

1 Introduction

This report is part of the ☞SWAD-Europe project ☞Work package 3: Dissemination and

<u>Implementation</u>. It serves as the report on the Initial Workshop deliverable.

Standards - The areas investigated in this workshop are "pre-standardization". In other words, although there is some sentiment for using existing standards and for more standards in the relevant areas, there is not yet standardization of the schemes and vocabularies used. Each topic builds on existing standards, most particularly RDF.

Some of the time that had been allocated to the workshop was in fact used for Dublin Core technical work.

2 Background

Introductory workshop - The project, as a major European Initiative, needed to be explained in a presentation style, as well as the material available which currently focuses on the administrative detail or work to be completed and is designed primarily for the participants and the Commission as a reference. In addition to a physical presentation this would provide an archived slide presentation of the project itself.

Technical Workshop - There have been a number of searching methods provided for the Web - from manually creating links of content, or automatically providing systems to ease the process, through plain text searching. Dublin Core metadata was designed to assist in searching and finding resources, particularly via the Web.

There have been different approaches to encoding information for enhancing searches, and one of the goals of the technical phase of the workshop was to draw some conclusions about which of these approaches represents best practices. It is assumed that best practices might include techniques drawn from the use of RDF, from plain text searching technology, from the experience of Dublin Core users, and perhaps other sources.

3 Workshop

The workshop was held over two days, in two parts, as a series of sessions at the DC2002 Conference.

The first was a general introduction to the goals, methodology and work plan of the project. This workshop sought to foster awareness of the SWAD-Europe project internationally, and will seek feedback from an experienced international development community on the general directions and work areas of the project.

☞*An Introduction to SWAD-E*

Chairs: Dan Brickley, Libby Miller, Charles McCathieNevile
This session will present the main work areas of the project and the partners directly involved, with discussion of the overall goals, methods of work, tools available and needed, and strategies for development and deployment.

The second was a developers workshop to discuss the issues arising from combining free-text searching methods with metadata-based discovery in order to provide more focussed information.

☞*Data interoperability - tools, techniques and use cases*

Chairs: Dan Brickley, Dave Beckett, Eric Miller
This session will investigate the issues of interoperability between data based on Dublin Core, but extended using various different schemas. Using a set of data made available that includes Dublin Core with different extensions it will look at the combination of four strategies

- *Using hierarchical classification schemes, for example qualified Dublin Core, or the RDF Schema property* subProperty*.*
- *Full text indexing of element content (such as RDF literals) including searching for phrases and substrings in plain text.*
- *Tracking provenance of statements - who said that resource X is of type Y*
- *Enhanced aggregation - using metadata to support more powerful queries and give more accurate results.*

An important question is how to enhance people's ability to find resources on the Web without requiring them to become experts in Dublin Core and RDF.

As the SWAD-Europe workshop was hosted within the Dublin Core conference, a number of joint sessions were held to tak advantage of the opportunity. Rather than holding parallel sessions the organisers agreed to merge two parallel sessions - one discussing Dublin Core Architecture and the other Schema Registries.

4. Outcomes

The introductory session provided an important group of information managers with an introduction to the goals of the SWAD-Europe project. It provided the opportunity to develop relationships between the project and metadata-related activity (most particularly in the field of library management) in Europe, and the opportunity to explore possibilities for furthering the outreach goals of the project.

The time used for working on issues specific to Dublin Core was used for discussion of best practice in encoding Dublin Core in RDF, within the Dublin Core Architecture group. A one hour session discussed strategy and priorities for the ongoing work of the Dubiln Core Architecture Working Group, and how that work relates to the Semantic Web Initiative. A second session on RDF schema registries allowed the SWAD workshop attendees and Dublin Core Registry Group members to join in shared discussion of tools and techniques for managing RDF schema registries.

The time spent on technical issues led to ☞a rough set of notes from the discussion which include some ☞Frequently Asked Questions, and some proposed answers. There were also several sets of slides presented, and some tools (see the next section for more details).

5. Slidesets and tools

Several sets of slides were presented during the workshops. In addition, several tools were presented during the technical workshop

Introductory workshop -

Dan Brickley, ☞Introduction to SWAD-E
>   An overview of the project, its goals and methodology.

Charles McCathieNevile, ☞SWAD-E and Europe
>   Looking at the community focus of the project and in particular its focus on the various developer communities within Europe.

Dave Beckett, ☞Reports and Development
>   Highlighting the ☞Scalability report already available and some development work in progress both within the project and in the broader community but related to the goals of the project.

Libby Miller, Semantic Web tools
>   Looking at Semantic Web tools being used to support the project and being further developed as part of the project. No slides were presented, but tools covered included the ☞SWAD-Europe event viewer, ☞RDF calendaring work, and the use of IRC bots such as ☞chump.

Technical workshop -

Dan Brickley, ☞Technical workshop
>   Outlining the goals of the technical sessions.

Eric Miller
>   Some tools for browsing RDF information were presented

Tom Baker - ☞Some SWAD-E Issues
>   A discussion of ways to manage information about the Dublin Core Schemas

6. Frequently Asked Questions

The following were questions raised in the Technical sessions with answers suggested:

☞When to use XML, when to use RDF?
>   XML is a hierarchical structure, which implies knowing the structure of the information you want to encode before you start. Querying XML is fast, because of the regular structure imposed. It is

easy to do syntax-based validation on XML. But it is important to know in advance what the XML schema you are using means.

RDF allows for easy encoding of diverse information and easy merging with other sources. It is simple to devolve the creation of vocabularies to other groups, and then mix them. Querying is slower, but more flexible, and allows for easy inferences based on subtype or equivalence relationships.

☞How to link to metadata / where to put it?

There are different answers for different types of problem. In many formats it is possible to put a link (e.g. a `link` element in HTML) or include the data directly (e.g. with the `metadata` element in SVG). In other cases the person creating the metadata does not have control over the original resource, but can associate it via third-party query services (e.g. Annotea)

☞What is the best practice for multilingual XML/RDF?

This question was the focus of ☞discussion based around ☞Tom Baker's presentation

Some questions raised are addressed by work done elsewhere, in some cases within the SWAD-E project:

How scalable is RDF and how well can RDF systems perform?
What are the advantages of different sorts of databases?
How good is 'native' RDF database performance?

The ☞Tools for Semantic Web scalability and storage: Survey of Free Software / Open Source RDF storage systems is a report produced for the SWAD-E project. It provides answers to these questions for Open Source systems available in mid-2002.

What tools and applications are available for end users?
What tools are available to manipulate the metadata in its various locations?

As part of the ☞RDF Resource Guide there are lists of ☞RDF Editors and tools.

Easy introduction documents, tutorials?

The ☞RDF Resource Guide also lists ☞Tutorials and examples
The ☞RDF Primer, produced as part of the Semantic Web Activity at W3C

is designed to provide the reader with the basic knowledge required to effectively use RDF. It introduces the basic concepts of RDF and describes its XML syntax. It describes how to define RDF vocabularies using the RDF Vocabulary Description Language, and gives an overview of some deployed RDF applications. It also describes the content and purpose of other RDF specification documents.

The following questions were also raised, but with no answers recorded from the discussion. Through the life of the project it is planned to gather answers to these questions, and this report wil be updated accordingly.

- How are RDF and XML related?
- How do topic maps relate to RDF?
- How do we merge and share data when we don't always identify things in the same way - in practice, uris often don't exist or people don't know what they are. (handling anonymous resources and resources with multiple identifiers)
- Access control - how do you protect information? (we are exposing our information to others' systems - people in want to have security - don't want to share, or don't want to share wrong thing.) (danbri: close to creative commons - give people access but say what they can use it for.)
- Who should we contact to find solutions to problems like: - annotating content (e.g. annotating aboriginal work) - user end tools - statements in RDF about accessibility - can this user agent access this content as it is, or not until its been transformed in some way (e.g. annotating the stylesheet).
  (Liddy)
- How can we handle provenance tracking (e.g. 5 people making 5 different claims about a page's accessibility)
  (Charles)
- Do you have any advice about distinguishing between the object and a manifestation of the object? This keeps cropping up (Simon)
- How can you extract RDF from legacy systems?
- How do I/should I use XML databases for XML/RDF content?
- What is the appropriate level of detail of description for an object? (Danbri)
- At what level of granularity should I describe my resources? e.g. how much do you encode the data? date, gregorian month, gregorian day etc (Charles McCathieNevile)
- If people use different levels of granularity, how can we link them together? (danbri)
- Is there a role for XML and RDF in bringing ERP systems, documentation management systems, desktop together - getting at the information through one interface? 'enterprise data integration'

(Sandy)
- How should you handle versioning for XML schemas and DTDs and RDF schemas? What are the pros, cons and implications of different approaches? (Eric Miller)

7 References

[RDF-GUIDE]

The ☞RDF Resource Guide, maintained by Dave Beckett, provides a list of briefly annotated links to RDf projects, implementations, articles, specifications and other documents. It is available at http://www.ilrt.bris.ac.uk/discovery/rdf/resources/

[RDF-PRIMER]

The ☞RDF Primer is designed to provide the beginner with the information they need to understand and use RDF. It is produced as part of the W3C's Semantic Web Activity, and is available at http://www.w3.org/TR/rdf-primer/

[RDF-STORES]

The ☞Tools for Semantic Web scalability and storage: Survey of Free Software / Open Source RDF storage systems is a report produced for the SWAD-E project. It is available at http://www.w3.org/2001/sw/Europe/reports/rdf_scalable_storage_report/