Recipes for Semantic Web Dog Food — The ESWC and ISWC Metadata Projects

Knud Möller[1], Tom Heath[2], Siegfried Handschuh[1], and John Domingue[2]

[1] Digital Enterprise Research Institute, National University of Ireland, Galway
knud.moeller@deri.org, siegfried.handschuh@deri.org
[2] Knowledge Media Institute and Centre for Research in Computing, The Open University, Milton Keynes, UK
{t.heath, j.b.domingue}@open.ac.uk

Abstract. Semantic Web conferences such as ESWC and ISWC offer prime opportunities to test and showcase semantic technologies. Conference metadata about people, papers and talks is diverse in nature and neither too small to be uninteresting or too big to be unmanageable. Many metadata-related challenges that may arise in the Semantic Web at large are also present here. Metadata must be generated from sources which are often unstructured and hard to process, and may originate from many different players, therefore suitable workflows must be established. Moreover, the generated metadata must use appropriate formats and vocabularies, and be served in a way that is consistent with the principles of linked data. This paper reports on the metadata efforts from ESWC and ISWC, identifies specific issues and barriers encountered during the projects, and discusses how these were approached. Recommendations are made as to how these may be addressed in the future, and we discuss how these solutions may generalize to metadata production for the Semantic Web at large.

1 Introduction

This paper reports on the efforts made to generate, maintain and deploy structured metadata for both the European and the International Semantic Web Conferences (ESWC and ISWC). In particular, we discuss the experiences gained and lessons learned during ESWC2006 and ISWC2006. However, the work done for those conferences is continued and further refined in the 2007 ESWC in Innsbruck, Austria and the 2007 ISWC in Busan, South Korea. The main contributions of this paper are the in-depth reporting of real deployments of Semantic Web technologies, and analysis of the technical lessons learned. Through analyzing our own experiences we provide a number of recommendations to others pursuing similar projects. Whilst the deployments were carried out by Semantic Web enthusiasts, the lessons learned apply to a wide range of potential deployment scenarios. As the Semantic Web moves from research and development into deployment and adoption, understanding these experiences and their implications becomes increasingly important. Crucially, challenges encountered by

those knowledgeable about the Semantic Web will almost certainly be encountered by those with less experience.

The following section discusses the aims of the projects, and related work. Section 3 characterizes the source data from which the metadata was generated, and discusses the output requirements. Sections 4.1 and 4.2 discuss the approaches taken at ESWC2006 and ISWC2006 respectively, the problems encountered and the solutions applied. In Section 5, we make recommendations for those undertaking such efforts at future conferences. These may also be relevant to any generation of metadata on a larger than personal scale.

2 Background and Aims of the Metadata Projects

The ESWC2006 and ISWC2006 metadata projects addressed a number of related aims. Firstly, to generate data for use with existing Semantic Web applications (thereby providing a showcase for such applications) and to further Semantic Web research. Secondly, to evaluate the practicality and feasibility of producing, managing, and deploying Semantic Web data for events such as ESWC2006 and ISWC2006. Only by undertaking such projects can we fully understand these processes. By reporting our experiences we believe we can inform similar projects in the future, whilst also highlighting the challenges of producing, managing and deploying Semantic Web data. Finally we argue that the Semantic Web community has an obligation to carry out such activities, if we are to better comprehend the challenges faced by others who may wish to adopt Semantic Web technologies. "Eating our own dog food" [4] is an essential mechanism by which to gain the appropriate insights.

2.1 Related Work

The use of technologies to support technology-related events is not novel. For example, [8] reports on the deployment of IRC backchannels at the ACM 2004 Conference on Computer Supported Cooperative Work (CSCW 2004), as a means to complement existing communication channels at the event.

Given the specific metadata requirements of the Semantic Web compared to other technological advances, the production of structured metadata describing events has greatest prevalence at conferences in the Semantic Web and related fields. For example, [3] reports on the W3Photos project, an initiative to enable the addition of semantic descriptions to photos taken at the Worldwide Web Conferences and related events. At the Worldwide Web Conference in 2006, a version of the conference programme was produced in RDF/XML[3] and deployed together with related services aimed at end users, such as the mSpace Schedule Explorer[4]. Organizers of the 4th International Semantic Web Conference in 2005 made the PiggyBank semantic web browser extension[6] available to delegates,

[3] `http://www2006.org/programme/dynamic`
[4] `http://www06.mspace.fm/`

as a means to annotate items related to the conference such as papers presented or local restaurants.

Whilst each of these initiatives is commendable, they share a common limitation due to the restricted scope of the metadata produced. To the best of our knowledge, the metadata projects at ESWC2006 and ISWC2006 represent the first instances of conferences attempting to offer comprehensive semantic descriptions of the event. Furthermore, as a result of this comprehensiveness, we believe that these events have moved significantly closer to integrating metadata production with broader conference workflows than previous events. These integration attempts raised a number of issues and challenges, which will be examined in the remainder of this paper.

3 Problem Description

In the course of the metadata efforts for both ESWC2006 and ISWC2006 we faced a number of specific problems and challenges, which will be discussed in this section. These challenges can be divided into two main areas: those arising from the specific characteristics of the input source data, and those arising from the requirements on the output data (the RDF metadata). If we view the metadata efforts as a process, then those challenges define its input and output, respectively. We will describe both in turn in the following sections.

3.1 Data Characteristics

When describing an academic conference, the three object types of greatest interest are people, events, and publications. People may take the role of paper authors, delegates at the conference, and committee members. Events may consist of talks (e.g. paper or poster presentations), conference sessions in which several papers are presented, or entire tracks. Various kinds of non-academic events also occur, such as meals, social events, and even coffee breaks. Publications can consist of full papers and poster/demo papers, plus a bound or electronic volume of the entire conference proceedings. In addition, artefacts such as sets of slides can be of great value, whilst not being formally published. Apart from this core set of data, other kinds of information, such as rooms within the conference venue, or sponsoring organizations, can be relevant.

In the course of producing metadata for both conferences, it became apparent that the source data from which both corpuses were produced had some novel characteristics relative to other Semantic Web data sets we had previously encountered. We predict that these characteristics are common to academic conferences in general, and found that they raised a number of challenges in producing the RDF descriptions of the conferences.

Firstly, data sets related to academic conferences are typically small in volume, compared to existing, established data sets available as RDF, such as DBLP[5]. Secondly they are heterogenous in nature, covering concepts as diverse

[5] `http://www4.wiwiss.fu-berlin.de/dblp/`

as people, places, artefacts, and events. Consequently there are few economies of scale in producing this data, as many different export or conversion tools may be required, relative to large, homogenous datasets from one central source.

Thirdly the input data set typically originates from many different sources, such as conference submission and registration systems, email messages, or text documents, and is traditionally managed by many different people, and each may use different methods to manage the data. In our experience these ranged from spreadsheets, to lists in documents, and HTML pages. This presented a number of challenges, as very few of these systems were already web-based, or designed to publish to public web sites.

3.2 Output Requirements

Just as the input data influenced the metadata efforts, so did the requirements on the output data. The metadata was to be used in a Semantic Web context, and so needed to adhere to the principles of linked, Semantic Web data. This made RDF a natural choice of data model. It was also important to use established vocabularies and ontologies. This was done to an increasing extent, starting with the integration of FOAF and SWRC at ESWC2006 and continuing with the iCalendar and BibTEX standards at ISWC2006. Where existing URIs could not be used to identify resources (as was often the case) new URIs were necessary. These were minted following consistent patterns and based on domains within our control. This allow for the configuration of properly de-referenceable URIs.

Many tools that used the conference metadata were targeted at visualisation, and so needed to traverse the RDF graph in all ways and access resources from all kinds of angles. This made the use of inverse properties very useful, if not necessary. Also, inverse functional properties are a good way of getting around the URI problem, and transitive properties can simplify ontology modeling significantly. OWL lite provides these features, and was therefore chosen to model the ontologies.

4 Approaches Taken

4.1 3rd European Semantic Web Conference (ESWC2006)

The ESWC2006 Semantic Web Technologies project[6] combined the creation and publishing of metadata describing the conference, with deployment of a range of applications (such as a semantic Wiki, photo annotation tool, and semantic search engine) that were intended to enhance the conference for delegates by making use of the ESWC2006 Conference Ontology and associated RDF/XML dataset. Fuller descriptions of these applications are given in [5] alongside the results of an evaluation into how they were received by conference delegates.

[6] `http://www.eswc2006.org/technologies`

Ontology Existing event and conference ontologies, such as the AKT Reference Ontology[7], the Conference ontology by Jen Golbeck[8], and the eBiquity Conference Ontology[9] were initially surveyed to assess their suitability for use in the ESWC2006 Technologies project. These ontologies were found to lack the expressivity required for the project. Consequently an exercise was carried out to model the Conference domain as the basis for a new ESWC2006 Conference Ontology. The ESWC2006 Conference Ontology has the following top-level classes: Artefact, Call, Event, Place, Role, Sponsorship, all of which (except Sponsorship) are extensively sub-classed to provide a high degree of expressivity. In contrast to other ontologies, the ESWC2006 Conference Ontology explicitly models relationships between people, roles, and events. So for example, the act of *giving a paper at a conference* is modeled in terms of a person holding a role of presenter at a specific *talk event*, with which there may be one or more associated *artefacts*, such as a paper or a slide set.

Wherever possible the ontology sought to reuse existing classes from widely deployed ontologies, instead of starting from scratch. Consequently, the ontology makes use of the Person and ResearchTopic classes from the FOAF[10] and SWRC [12] ontologies, respectively. For example, the `eswc:heldBy` property has a domain of `eswc:Role` and a range of `foaf:Person`. It is hosted at[11], according to the "Best Practice Recipes for Hosting RDF Vocabularies" [10].

Metadata Creation RDF descriptions[12] were made available of the ESWC2006 Organizing Committee, the tracks, sessions, talks, and roles that existed at the conference, the papers, posters, and demos presented, rooms in the conference venue, and delegates who opted into the public Semantic Delegates List. Where an Organizing Committee member was responsible for a particular area (such as posters and demos, or workshops), that individual provided the source data for the corresponding RDF descriptions. These data sets were generally small in volume, and came in a range of different formats, such as Excel spreadsheets, tables in documents, and HTML, as this was generally how members of the organizing committee managed the data for which they were responsible. These characteristics influenced how the RDF descriptions were then produced. It was not deemed efficient to automate the production of RDF descriptions where the source data was not already well structured, or in areas such as workshops, tutorials, or demos, where there were relatively few instances. Consequently, the majority of descriptions were produced manually using a generic XML editor. Aside from being very resource intensive this created issues with maintenance, for example when sessions changed venue.

[7] `http://www.aktors.org/publications/ontology/`
[8] `http://www.mindswap.org/~golbeck/web/www04photo.owl`
[9] `http://ebiquity.umbc.edu/ontology/conference.owl`
[10] `http://xmlns.com/foaf/0.1`
[11] `http://www.eswc2006.org/technologies/ontology`
[12] `http://www.eswc2006.org/rdf`

The one area where automation was clearly beneficial was in producing the Semantic Delegates List. On an opt-in basis, additional data was collected from delegates at registration via the conference registration system. Those who gave their consent were featured in the Semantic Delegates List, an RDF representation of people present at the conference. This gave basic information such as the delegate's name, and a hash of their mailbox URI, in addition to any further information they had supplied such as their homepage URI, the URI of their FOAF file, their workplace homepage, and their areas of interest. The information provided by delegates was exported from the conference registration system as an Excel spreadsheet, and processed by a PHP script to generate RDF according to the FOAF ontology. This strategy proved very effective where specific pieces of information has been given dedicated fields in the registration system. However, due to limitations in the registration system, information about delegates' areas of interest had to be provided into one text field, despite the interests coming from a fixed vocabulary. This presented numerous challenges when parsing the data to produce RDF.

An additional challenge concerning topics of interest centered around how the Areas of Interest for a specific conference are managed and integrated with other data sources. For ESWC2006, the existing topic hierarchy from the SWRC ontology [12] was used when collecting delegates' interests. A preferable approach would be for conference topics to be marked up using SKOS[13] from an early stage, as this topic listing could then be reused within other systems deployed for the conference. Such an approach would also facilitate the creation of mappings between topics occurring at related conferences whilst still allowing local flexibility in how topics are defined and how they evolve.

One objective of the ESWC2006 Technologies project was to bring the same degree of semantic descriptions to workshops, as to the main conference track. This presented significant challenges, in that all workshops maintained separate web sites maintained by different people, and all structured differently. The intended approach was to use GRDDL[14] to generate RDF descriptions of workshop proceedings, participants, and committees. This objective was not met at ESWC2006 due to time constraints, however, we believe it would be a valuable objective for future conferences.

Finally, creation of the metadata for ESWC2006 required the minting of new URIs to identify all relevant entities. A lack of tools to support this process made it difficult to ensure they were being used consistently across all RDF descriptions (for example in the RDF describing full papers, and in the Semantic Delegates List). The most pragmatic solution to this problem involved simply using a consistent syntax for minted URIs.

Metadata Deployment In deploying the RDF descriptions of the conference, close integration of the RDF with the conventional Web site was desired. Consequently, and to aid human management of the data, many separate RDF files

[13] `http://www.w3.org/2004/02/skos/`
[14] `http://www.w3.org/TR/grddl/`

were created and deployed on the conference web site[15]. The *FOAF Autodiscovery* technique[16] was used to link HTML pages to their corresponding RDF files. The Apache web server running the conference web site was also configured to ensure URIs could be de-referenced, using approaches similar to those outlined in [1].

4.2 5th International Semantic Web Conference (ISWC2006)

The approach taken and decisions made during the planning of the ISWC2006 metadata efforts[17] were largley influenced by the metadata efforts of ESWC2006. In this section we will therefore concentrate on the aspects where changes or a development between the two conferences took place. Just like for for ESWC2006, the data corpus was made available prior to the conference, and a number of applications who used the data were featured on the metadata website.

Ontology Since the ESWC schema had already proved useful and adequate, its adoption for ISWC2006 was clearly desireable. By using the same schema, both datasets become interoperable, can be queried and visualized in an integrated fashion and will together have greater impact on the community. However, a number of issues were identified which needed to be addressed. The main points were: (i) increased integration of existing standards and vocabularies, (ii) rearrangement of the document concept space, and (iii) fully utilising the possibilities OWL lite offers, such as inverse properties, which aid in data visualization.

Addressing those points did not mean literally changing the existing ESWC ontology, but instead establish a new ISWC ontology. However, instead of starting afresh, this new ontology simply imports the old one, adds a number of new concepts and properties, makes some additional statements about existing concepts and properties and deprecates others. In effect, a dataset conforming to the ISWC ontology is still largely compatible with the ESWC ontology. In the following paragraphs we will discuss some of the changes made in moving from one conference ontology to the next. We think this discussion illustrates very well some typical issues for ontology evolution.

Even though the ESWC ontology integrated FOAF for people-related metadata, it still used custom classes and properties to represent events and publications. For ISWC, we decided to integrate more established standards for those kinds of entities. For event data, the iCalendar format [2] was adopted. Since iCalendar itself does not have an RDF schema, we decided to use the schema suggested in a W3C interest group note[18]. For representing publications, we adopted the widely used BibTeX schema [11]. A number of implementations in RDF exist, but we decided to use the BibTeX-related classes from the SWRC (Semantic Web for Research Communities) ontology[19], due to reasonably wide

[15] `http://www.eswc2006.org`
[16] `http://rdfweb.org/topic/Autodiscovery`
[17] `http://iswc2006.semanticweb.org/program/tech_links.php`
[18] `http://www.w3.org/TR/rdfcal`
[19] `http://ontoware.org/projects/swrc/`

usage (outside SWRC-related projects, the ontology is also used by Flink [9] and openacademia [7]) and tool support.

The integration with both iCalendar and BibTEX was performed in a straightforward fashion. We chose an appropriate concept from the ESWC ontology as the entry point and established a subclass relationship to a matching concept in the external ontology. Figure 1 gives an example of how this worked with the iCalendar integration. The `eswc:OrganizedEvent` concept, which is the superconcept of all event types in the ESWC ontology was made a subconcept of `ical:Vevent`, thereby allowing all ESWC event types to be treated as iCalendar events.

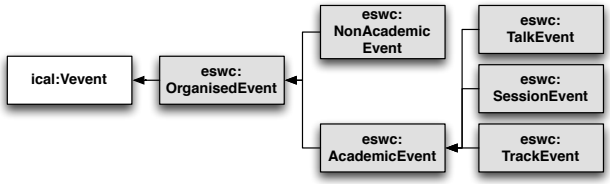
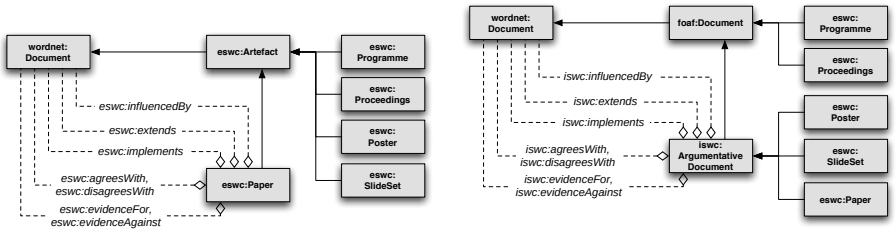


Fig. 1: Integrating the iCalendar Ontology

The document concept space in the ESWC ontology is homogeneous — papers, posters, slide sets, proceedings and programmes are all grouped under a common superconcept `eswc:Artefact`. Properties regarding scientific discourse such as `eswc:influencedBy` or `eswc:agreesWith` are defined to apply to instances of `eswc:Paper` only (see Fig. 2a). For ISWC, we decided to split the document concept space into *argumentative* documents (papers, posters and slide sets) and *non-argumentative* documents (proceedings and programme). This allowed us to redefine the properties listed above so that they apply to all kinds of argumentative documents. Also, we wanted to further strengthen the integration with FOAF and establish a relation to FOAF's `Document` class. In order to achieve this, we introduced a new class `iswc:ArgumentativeDocument`, which is the superclass of all relevant document classes in the ESWC ontology, and a subclass of `foaf:Document`[20]. All document classes which are deemed non-argumentative are direct subclasses of `foaf:Document`. In addition, we had to introduce new properties such as `iswc:agreesWith` and at the same time deprecate the ones like `eswc:agreesWith` (see Fig. 2b).

To utilize the expressiveness of OWL lite to a greater degree, the ISWC ontology defined `owl:inverseOf` relationships for a number of properties from the ESWC ontology. Finally, transitivity was used in defining properties that express sub- and super-event relationships (e.g. a talk is a sub-event of a session, is a sub-event of a track).

[20] Note that, since existing statements from the ESWC ontology are not changed, all document classes are still also subclasses of `eswc:Artefact`!



(a) ESWC (b) ISWC

Fig. 2: The Evolution of the Document Concept Space from ESWC to ISWC

Metadata Creation The process of creating the metadata for ISWC2006 was affected by the same issues that had already surfaced during ESWC2006. Again, the source data came in a variety of different formats, none of which were particulary suitable for automatic transformation. This reflects the fact that, even though ESWC and ISWC are Semantic Web conferences, data integration wasn't yet a critical point to consider in organizing them.

The main portion of the generated data was papers, authors and events such as talks and panels. In addition to that, conference and session chairs were represented, as well as those conference attendees who volunteered to provide a FOAF profile of themselves. The source data for both papers and authors was an Excel table provided by the proceedings chair, as well as the actual PDF documents of the papers. Automated generation of structured metadata was partially possible. However, author email addresses and affiliations, as well as paper abstracts still had to be extracted by hand. The source data for the events list was a timetable in HTML, which was sufficiently unstructured to make automatic conversion to RDF difficult.

To ease the creation of RDF, we therefore decided to add an intermediate step. We first generated ordinary, non-RDF BibTEX and iCalendar documents. The benefit was that we were now able to edit and maintain the data with dedicated tools that are tailored towards dealing with such data and make this task much less frustrating and error-prone that maintaining the data in a general-purpose text editor. Also, it was now possible to use other tools to automatically generate the desired target RDF (we used a modified version of the customizable bibtex2rdf converter[21] and Python scripts made available by the W3C Semantic Web Interest Group[22]).

Not all data could be generated and maintained in this way. Email addresses and affiliations of authors cannot be represented in ordinary BibTEX, and so had to be added by hand later. Similarly, links between papers and their talks, links

[21] http://www.l3s.de/~siberski/bibtex2rdf/
[22] http://www.w3.org/2002/12/cal/

between events (e.g. between individual talks and their session), as well as the various conference and session chairs were added manually.

Table 1 provides an overview of the two datasets that were generated for both conferences. Even though almost identical ontologies were used, some obvious differences can be observed. The ISWC dataset did not contain any workshops, and was thus less comprehensive than that of ESWC. As a result, much fewer instances of `eswc:Role` and `eswc:Artefact` were defined. On the other hand, it is striking that the ISWC dataset has a much larger number of triples (before inferencing). This is mainly due to a small change in the ontology, which was introduced to improve the immediate usefulness of the data for consuming SW tools: the authors of publications were modelled as complex `foaf:Person` objects with various assertions about them, whereas at ESWC they were represented by a URI that was not further defined. This also explains the rise in person instances from ESWC to ISWC.

Table 1: Overview of the ESWC2006 and ISWC2006 datasets

Top Level Class	**ESWC2006**	**ISWC2006**
`eswc:OrganisedEvent`	129	130
`eswc:Role`	130	48
`eswc:Artefact`	100	68
`foaf:Person`	189	301
`swrc:Topic`	59	0
`eswc:Place`	8	0
Total Number of Triples	2939	5902

Hosting/Serving the Metadata In a move from ESWC2006, we decided to use a database solution for hosting, instead of individual files. The complete dataset first created in the form of RDF documents and then loaded into a Jena RDF store[23] and made available through a Joseki[24] SPARQL server at a public URL[25]. Using a hosting setup based on an RDF store made reasoning capabilities available to all external consumers of the conference data (mainly the tool providers), and, more importantly, made it possible to perform queries over the complete dataset. Again, the webserver was configured to allow dereferencing of resource URIs.

[23] `http://jena.sourceforge.net/`
[24] `http://www.joseki.org/`
[25] `http://128.192.251.191:8080/joseki/iswc`

5 Recommendations for Semantic Web Dog Food

Having encountered the challenges reported above, work has been undertaken towards their resolution; both in the evolutions seen between ESWC2006 and ISWC2006, and in further changes being made in how the metadata will be produced for ESWC2007 and ISWC2007. However, many issues remain. Consequently we offer the follow recommendations to others undertaking similar projects in the future, and to those planning any deployment of metadata for the Semantic Web.

5.1 Process Recommendations

- **Release metadata early**, thereby allowing tools to be developed using real data, and deployed in advance of the conference. This may be of particular significance where tools are designed to support delegates in planning their travel to and schedule at the conference. An example is the conference scheduler that was deployed during ISWC2006[26].
- **Provide sample data** with which developers can work in the period before the final conference data is available. Datasets from previous conferences may be sufficient for this purpose.

Addressing these recommendations requires more integrated and efficient workflows across the entire conference planning process.

5.2 Workflow Recommendations

At ESWC2006 and ISWC2006 metadata was produced centrally by one person, but from heterogenous sources. This differs from a number of existing metadata production workflows. For example, Semantic Web-compatible versions of databases such as DBLP (see Sect. 3.1) are produced by exposing existing databases as RDF, enabling a simple workflow and single source to yield large amounts of data. In contrast, production of personal FOAF files has traditionally been carried out by one individual creating RDF by hand, or semi-manually. We argue that production of conference metadata to date has followed a third workflow, of production by one individual with relatively low degrees of automation, but from many sources. This may provide an interesting case study relevant to Semantic Web deployments in general, as not all metadata will be produced on a very large or very small scale, as in the DBLP and FOAF examples. However, we believe that future conference workflows should attempt to distribute production of metadata across many individuals, ensuring production is as automated and integrated with wider conference organization as possible. In addition to increasing efficiency, automating production may also help ensure comprehensiveness of the resulting dataset — something that can be hard to achieve with more manual production. Automation may be aided by the use of commmon

[26] `http://schedule.semanticweb.org`

platforms for managing conference information that are capable of publishing data directly to the Semantic Web, or through an intermediate conversion step. On this basis we offer the following recommendations:

- **Reuse** wherever possible. Ontologies, and tools and methodologies for producing data may all be sufficiently generic to be reused. Naturally, the SWC conferences all build on the work that has been done at the preceeding conferences, and all reuse vocabularies such as FOAF, iCal or BibTEX.
- **Integrate metadata production into conference workflows** at all possible stages: in the work of the organizing committee members, and in the submission and registration systems. For ISWC2007, this has been applied by using the same submission system throughout (for both conference and all workshops) and ensuring that the system provides all necessary data.
- **Devolve responsibility for metadata production** to authors, delegates, and members of the organizing committee. For example, for ISWC2007, all authors of accepted papers are required to add additional data to be used in the metadata corpus.

5.3 Technical Recommendations

- **Define clear formats for minted URIs**. This will help ensure consistency in the absence of URI management tools. E.g., starting with ESWC2007, all conference metadata will be hosted at `http://data.semanticweb.org`, and the URIs for all entities adhere to strict format within this domain, such as `http://data.semanticweb.org/conference/eswc/2007/person-59437`.
- **Support the de-referencing of URIs** by configuring web servers appropriately. Servers should ideally support HTTP 303 redirects and Content Negotiation.
- **Provide crawlable and browsable metadata** on the conference web site that is well integrated with conventional HTML content.
- **Provide a queryable repository of all metadata** that supports the SPARQL query language and protocol.
- **Provide mechanisms to update metadata** that require minimum manual intervention.
- **Maintain a central repository of data** in order to deliver the features recommended above. For the SWC conferences starting with ESWC2007, this is `http://data.semanticweb.org`.
- **Move towards a layered architecture** that takes a service-oriented approach, adding a Services layer on top of Data and Query layers.

5.4 Functionality Recommendations

- **Define and enable, at the Services layer, a range of common functions** associated with a conference, or shown to be beneficial to delegates.
- **Support conference organization workflows** by exploiting Semantic Web technologies.

Common functions or services might include a session attendance suggestion service, such as that proposed in the ESWC2006 Design Challenge[27]. Prior to the conference a travel arrangement service may also be beneficial. Different services are likely to be of value before, during, and after the conference. Understanding which provides maximum value at each point in time requires further investigation.

There are also many areas in which Semantic Web technologies may assist with organization of a conference itself, and these deserve further consideration. For example, Semantic Web technologies may be able to assist with Programme Committee selection, ensuring that a suitable range of location, affiliation, and expertise is achieved. Research into scheduling on the Semantic Web may be able to inform the development of systems able to plan transfers between airports and the conference venue. Using Semantic Web technologies in planning a conference adds further motivation for using structured data at all stages of the planning process.

6 Conclusions

We argue that the metadata efforts at ESWC2006 and ISWC2006, and the ongoing efforts of 2007, have brought numerous benefits to the Semantic Web community, such as the creation and deployment of significant datasets which can be used in future research, opportunities for the deployment of applications, and demonstrations of the community's commitment to eating its own dog food. That these efforts are seen as important is supported by the ongoing efforts to generate and deploy metadata at ESWC2007 and ISWC2007. The ontology used for the metadata efforts which was started with ESWC2006, was refined during the following conferences and has now reached a relatively stable version[28].

Furthermore, and perhaps most significantly, the projects represent a form of action research by which the community can identify issues for ongoing research, and barriers to wider adoption of the Semantic Web. We believe that the value of this paper lies in the fact that we report on real world use cases of Semantic Web technology, albeit in the context of research-focussed Semantic Web conferences, identify issues encountered, and make recommendations about how they may be overcome. These are important lessons to be learned for the application of Semantic Web technology in general; in broader settings and by other groups outside the core Semantic Web community. Despite being Semantic Web enthusiasts, in coordinating the metadata projects at both conferences we found the experience of eating our own dog food just as challenging as the idiom implies. We hope that by sharing our recipes and ideas for improvements, eating one's own dog food can become significantly more appetizing.

[27] `http://www.eswc2006.org/technologies/designchallenge/semantic-conference-program.pdf`

[28] `http://data.semanticweb.org/ns/swc/ontology`

Acknowledgements

The work presented in this paper was supported (in part) by the Líon project supported by Science Foundation Ireland under Grant No. SFI/02/CE1/I131 and (in part) by the European project NEPOMUK No FP6-027705. Furthermore, it was supported (in part) by the Advanced Knowledge Technologies (AKT) project, an Interdisciplinary Research Collaboration (IRC), which is sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/N15764/01. The AKT IRC comprises the Universities of Aberdeen, Edinburgh, Sheffield, Southampton and the Open University.

References

1. C. Bizer, R. Cyganiak, and T. Heath. How to publish linked data on the Web, 2007. `http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/`.
2. F. Dawson and D. Stenerson. Internet calendaring and scheduling core object specification (iCalendar), 1998. RFC 2445: `http://www.ietf.org/rfc/rfc2445.txt`.
3. G. Elin. Is a picture worth a thousand clicks? Challenges of adding semantic data to images. In *WWW2004 Workshop on Interaction Design and the Semantic Web (IDSW2004)*, New York, NY, USA, May 2004.
4. W. Harisson. Eating your own dog food. *IEEE Software*, 23(3):5–7, 2006.
5. T. Heath, J. Domingue, and P. Shabajee. User interaction and uptake challenges to successfully deploying Semantic Web technologies. In *The 3rd International Semantic Web User Interaction Workshop (SWUI2006) at ISWC2006*, Athens, GA, USA, November 2006.
6. D. Huynh, S. Mazzocchi, and D. Karger. Piggy Bank: Experience the Semantic Web inside your Web browser. In *4th International Semantic Web Conference (ISWC 2005)*, Galway, Ireland, November 2005.
7. M. Klein, P. Mika, and R. Serban. Semantics-based publication management using RSS and FOAF. In *Poster Track, 4th International Semantic Web Conference (ISWC2005)*, Galway, Ireland, November 2005.
8. J. F. McCarthy and D. M. Boyd. Digital backchannels in shared physical spaces: Experiences at an academic conference. In *CHI '05 extended abstracts on Human factors in computing systems*, Portland, OR, USA, April 2005.
9. P. Mika. Flink: Semantic Web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3, 2005.
10. A. Miles, T. Baker, and R. Swick. Best practice recipes for publishing RDF vocabularies, March 14 2006. `http://www.w3.org/TR/swbp-vocab-pub/`.
11. O. Patashnik. BibTexIng, February 8 1988. BibTeX Documentation.
12. Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, and D. Oberle. The SWRC ontology - Semantic Web for research communities. In *12th Portuguese Conference on Artificial Intelligence (EPIA 2005)*, Covilha, Portugal, December 2005. Springer.