

# SCHEMAS

Contract: N° IST-1999-10100

## Forum for Metadata Schema Implementers

D43: SCHEMAS THIRD WORKSHOP REPORT

Document number: SCHEMAS-UKOLN-WP4-D43-Final-20010615

### General Information

**Title** SCHEMAS: Third Workshop Report

**Creator** Pete Johnston

**Subject-Keywords** Deliverable D43; third workshop, managing schemas in a multilingual Semantic Web

**Description** This document describes the third SCHEMAS workshop, held 10-11<sup>th</sup> May 2001 in Budapest, Hungary

**Publisher** UKOLN

**Contributor** PwC, GMD

**Date** 15<sup>th</sup> June 2001

**Type** Text Manuscript

**Format** application/msword

**Identifier-URL**

**Identifier-**

**Document Number** SCHEMAS-UKOLN-WP4-D43-Final-20010615

**Language** English

**Rights** European Commission; Internal circulation within project

<META NAME="DC.Title" CONTENT="SCHEMAS: Third Workshop Report">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#title">

<META NAME="DC.Creator" CONTENT="Pete Johnston">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#creator">

<META NAME="DC.Creator.Address" CONTENT="UKOLN, University of Bath, UK">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#creator">

<META NAME="DC.Subject" CONTENT="Deliverable D43">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#subject">

<META NAME="DC.Subject" CONTENT="WP4">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#subject">

<META NAME="DC.Subject" CONTENT="Report resulting from the third SCHEMAS Workshop">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#subject">

<META NAME="DC.Description" CONTENT="This document reports on the third SCHEMAS workshop which took place in Budapest, 10-11<sup>th</sup> May 2001">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#description">

<META NAME="DC.Publisher" CONTENT="UKOLN">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#publisher">

<META NAME="DC.Date" CONTENT="(SCHEME=ISO8601) 2001-06-15">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#date">

<META NAME="DC.Type" CONTENT="Text.Manuscript">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#type">

<META NAME="DC.Format" CONTENT="(SCHEME=IMT) application/msword">

<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin\_core\_elements#format">

## Contents

- [1. Introduction](#)
- [2. Aims of the Workshop](#)
- [3. Programme](#)
- [4. Summary of Discussions](#)
  - [4a. Semantic Web](#)
  - [4b. Multilingual schemas](#)
  - [4c. Controlled vocabularies](#)
  - [4d. SCHEMAS Forum Registry](#)
- [5. Training Materials](#)

- [6. Costs Incurred](#)
- [7. Conclusions from the Workshop](#)
- [8. Concluding Remarks](#)
- [9. References](#)

## Introduction

This document forms the report on the third workshop held by the SCHEMAS project. The workshop was held at the Computer and Automation Research Institute of the Hungarian Academy of Sciences (MTA SZTAKI) in Budapest, Hungary from Thursday 10 May to Friday 11 May 2001. The title of the workshop was "Managing schemas in a multilingual Semantic Web". The number of delegates registered for the workshop (including partners in the SCHEMAS project) was 32.

## Aims of the Workshop

The third workshop was planned to build on the work of the previous two events on the construction, publication and management of metadata schemas, and how metadata schema registries can support this. The publication of metadata schemas in a standards-based form, usable both by human readers and by software agents, will be a key component of the emerging "Semantic Web" [1]. At this time, it appears likely that the mechanisms of publication to provide this functionality in the Web environment will make use of models based on the Resource Description Framework (RDF) [2] and the syntax of the Extensible Markup Language (XML) [3].

The SCHEMAS project recognises that the Web is an inherently multicultural and multilingual arena. If they are to enable information exchange and semantic interoperability in a global environment, metadata registries must address issues of multilinguality and the internationalisation of schemas.

While previous workshops have focused on the construction and publication of metadata element sets, this workshop extended its scope to consider the nature of "controlled vocabularies" (from simple "flat" lists to complex hierarchical thesauri).

In addition to these specific themes, the workshop presented an opportunity for the project to describe progress on the development of the SCHEMAS Forum Registry and for those developing and managing schemas in a range of domains to share experiences, identify common problems and explore solutions.

## Programme

The programme for the workshop, with hypertext links to the speakers' presentations and related materials, can be found on the SCHEMAS website [4]. The original call for participation is also available [5].

The workshop began with an introductory presentation setting this workshop in the context of the other activities of the SCHEMAS project. This was followed by an introduction to the idea of the "Semantic Web" which defined some of the key concepts, introduced some of the enabling technologies and outlined the relationship of the SCHEMAS project to these broader activities.

The remainder of the first day concentrated on **multilinguality** issues, with presentations on:

- the construction of a multilingual registry for the Dublin Core Metadata Element Set at the University of Library and Information Sciences (ULIS), Tsukuba, Japan [6];
- the critical role of a framework to support multilingual controlled vocabularies for Information/Knowledge Management at the World Agricultural Information Centre (WAICENT) (FAO of the UN) [7];

- the SALT Project's development of XML- and RDF- based formats and tools for modelling terminologies [8];

The presentations were followed by a "breakout" session in which participants divided into two groups to discuss whether the requirements for multilingual support could be accommodated within the simple typology of schemas envisaged by the SCHEMAS project. The two groups presented a summary of their discussions and conclusions.

The second day focused on practical experiences of **managing metadata schemas** (and other related objects). Presentations covered:

- the development of the SCHEMAS Forum Registry [9] (including a demonstration), using the EOR Toolkit [10];
- a brief review of the application profile concept, in the context of registry functionality;
- the use of a metadata registry (for both schemas and controlled vocabularies) to support information retrieval on the Microsoft corporate intranet;
- the role of a semantic registry in supporting distributed, multi-schema repositories of learning objects within the EASEL project [11]

This was followed by a session in which workshop participants were given the opportunity to make short presentations. The first presentation described the architecture of the Knowledge on Demand (KOD) toolkit for publishing and delivering packages of learning objects. There was a more informal discussion on the role of a registry (or registries) within the Gateway to Educational Materials (GEM) initiative and some of the more general requirements which this highlighted.

The breakout session on the second day examined the questions of whether a controlled vocabulary or thesaurus could be described within the same conceptual framework as a namespace schema and whether vocabularies should be stored in the same registries as schemas.

## Summary of Discussions

Extensive discussions took place within the scheduled breakout groups and the associated plenaries, and in question-and-answer sessions following the formal presentations. The following represents an attempt to bring together points made throughout the workshop under broad thematic headings. One recurring issue, particularly in the breakout groups, was the need for careful use of terminology and the potential for misunderstanding caused by a failure to define terms clearly, to differentiate specific contextual uses of the same term (like "vocabulary") or otherwise to make explicit any assumptions being made by the speaker.

## Semantic Web

The consensus regarding the "Semantic Web" activity was that the emphasis should be on establishing basic principles and enabling simple applications (the lower layers of Berners-Lee's model [12]) now, while planning for gradual increases in complexity in the future. The functionality provided by metadata schemas registries can contribute to the basic infrastructure. The limitations of this web of relations should be acknowledged: semantic interoperability would develop on the basis of a "partial understanding".

An "open" Semantic Web would also require the layer which enables verification of who (really) makes assertions, but in the short term it is possible that progress can be made in the context of "pockets" or "communities" of trust within which the quality of statements is accepted.

## Multilingual schemas

Neither discussion group focused closely on the use of the suggested "typology of schemas" (Namespace Schemas, Translated Namespace Schemas, Application Profiles) in a multilingual context. This perhaps reflected a greater interest on the part of workshop participants in multilingual issues in the context of controlled vocabularies rather than in the

context of namespace schemas.

There is a general sense in which an element set and the set of terms in a vocabulary are both instances of "value lists" and "translation" becomes analogous to cross-walking.

It was suggested that problems of translation for some metadata element sets may be more complex than for others. The case of Dublin Core, as a simple element set, should perhaps not be taken as representative. Some metadata element sets, which would certainly be used in multi-language, multi-cultural contexts, employ element names and definitions of semantics that are both precise and highly culturally dependent. The domains of educational resources and rights management were presented as examples of this complexity. Although Dublin Core had abandoned the notion of employing a different namespace for each language, it was not clear whether this convention should be applied to more complex element sets.

In some cases, a simple translation of an element set may be insufficient, and a "mapping" between language/culture-specific versions may be required. The internal structure of an "address", for example, varies between languages.

Some consideration was given to the problems highlighted by Shigeo Sugimoto's presentation i.e.

a translated version of a "standard" namespace schema might evolve at different rates from the standard source version,

a translated version might be amended/extended to incorporate "local" properties specific to a language or culture, in which case it might strictly be described as an application profile.

The case of such "local" extensions raised the questions of whether they might be re-translated to the "original" language, and if so whether that should change the original namespace schema or be accommodated in the form of an application profile.

The choices which might be made to address these problems raise the more general question of whether a registry based on such choices can be culturally neutral.

## Controlled vocabularies

There was general agreement that the different types of controlled vocabulary could be accommodated within the same model/framework as that used to describe Namespace Schemas and Application Profiles.

However, although schemas and vocabularies may fit within the same general descriptive framework, their characteristics may be quite different. A schema and a vocabulary are both finite sets of terms (with associated properties), which have an "authoritative" form. At least some instances of vocabularies are much larger than a typical element set. Vocabularies are typically more volatile than element sets and the management of change may be complex, particularly in a distributed environment.

There is a clear requirement to standardise conventions for describing vocabularies in a machine-readable manner so that they can be shared effectively. Such standardisation of descriptive practice will require a clear typology of vocabularies, to encompass both simple "flat" lists and more complex hierarchical and networked vocabularies. There was some doubt expressed about whether it was possible to standardise all the types of associative relationships which one may wish to describe between the terms of a vocabulary, because the semantics of such relationships tend to be highly domain-specific (and language/culture-specific).

There was no clear consensus on the question of whether entire vocabularies (i.e. the sets of terms which make up the vocabularies) should be stored in the same registries as schemas. The shared conceptual model would permit this, but the difference in use (both end use and administration/management) may require quite different interfaces. There was greater agreement it was useful for a registry of schemas to contain high level descriptions of the vocabularies referenced by those schemas, and that it was desirable to standardise practice for that description.

The common practice of specifying the use of subsets of a "standard" vocabulary raised the question of whether the

concept of the "application profile" could be used to describe this. Concern was expressed that such descriptions of "reuse" might fail to capture the context of relationships in which a term is embedded in that "standard" vocabulary. The elements of a namespace schema are not embedded in a network of relationships in the same way. Furthermore, language-specific and culture-specific dependencies become most apparent in the relationships between terms of a vocabulary. While the application profile concept does appear to be useful in this context, its deployment requires further careful consideration.

## SCHEMAS Forum Registry

The SCHEMAS Forum registry is still under development. In particular, work is required to build more user-friendly interfaces for the creation of entries to describe namespace schemas and application profiles.

It was emphasised that the SCHEMAS project was not seeking to create one central "authoritative" registry database. It was rather seeking to test the usefulness of constructs like the "application profile", to explore conventions for describing them using RDF/XML, and to develop a common model for a registry of metadata schemas. That model could be implemented using many different software tools and in the form of many different registry database instances. There was considerable interest expressed in obtaining the EOR toolkit that the SCHEMAS Forum Registry is using.

Building on work by Eric Miller, the project has defined its own vocabulary to describe application profiles in RDF/XML, and this vocabulary is defined in a namespace schema local to the SCHEMAS project. There was some discussion on whether, if the model was shown to be generally useful, it might be appropriate for this vocabulary to be included within a namespace which was sanctioned by another body.

It was noted above that standards are required for the machine-readable representation of controlled vocabularies of various types, as well as for namespace schemas and application profiles. If registries are to function in a distributed networked environment, then standardisation is also required in the area of interactions between individual registries and between registries and other software tools/agents.

## Training Materials

The materials collected during the workshop, including presentations and notes from the breakout sessions, form the initial set of training materials from the third workshop. These may be enhanced later under work package 7 which is concerned with the provision of information materials.

## Costs Incurred

The costs of the workshop were within the allocated budget.

## Conclusions from the Workshop

The need for a network of semantic relations is now widely recognised, and in this sense the "Semantic Web" is already emerging. The precise role of RDF/RDFS in this project is still under debate, but it appears that the RDF model and associated XML syntax at least provide a basis to experiment.

It is clear that one-to-one mapping between the elements of an ever-expanding set of schemas is not a scalable solution to the expression of semantic relationships. However, there is at present no adequate "semantic layer" to provide a basis for an alternative approach.

The discussions of multilinguality highlighted the similarity of languages and domain-specific jargons. The discussions of controlled vocabularies confirmed that they could be encompassed within the conceptual model used to describe schemas. However, there are significant differences in the use made of these different classes of object by humans and by software agents, and consequently in the requirements for their maintenance and administration.

The efforts to develop standards in this area face the competing demands of keeping up with the speed of change and at the same time meeting requirements for quality and trust. There is a great demand for tools, and a real threat that the independent development of different approaches and systems will create barriers to interoperability in the future.

With these factors in mind, the SCHEMAS project should

- allocate more effort to describing tools, and disseminating that information;
- explore best practice for the description/publication of controlled vocabularies, beginning with a convention for the top-level description of a vocabulary
- explore best practice for representing mappings between vocabularies, which will require co-operation on semantics between and across domains
- continue to provide a platform for sharing experiences, with the aim of enhancing "imperfect understanding"

## Concluding Remarks

The feedback obtained informally during the workshop and via the evaluation forms was generally positive, and indicated that the workshop had been successful in providing a forum in which delegates could explore the problems and learn from one another's experiences.

The aspects most often cited as most valuable were the combination of presentations and discussions, and the focus on controlled vocabularies and thesauri.

Areas which might be improved included the need for speakers to clarify their use of terminology; a tendency by some speakers to focus on their own projects rather than on the themes of the workshop; and the lack of structure in some of the breakout discussions. The latter might be improved by a focus on specific problems. It was also suggested that presenting the demonstration of the Schemas Forum registry earlier in the programme might have focused discussion better. Also, the workshop may have benefited from some more critical points of view.

For future events, participants suggested a greater emphasis on technical experience, demonstrations and tools.

Overall, the content of the programme, the organisation and the venue were rated as good to excellent on the evaluation forms.

## References

[1] W3C Semantic Web Activity  
<http://www.w3.org/2001/sw/>

[2] W3C Resource Description Framework (RDF)  
<http://www.w3.org/RDF/>

[3] Extensible Markup Language (XML)  
<http://www.w3.org/TR/REC-xml>

[4] Third SCHEMAS Workshop programme  
<http://www.schemas-forum.org/workshops/ws3/programme.html>

[5] Third SCHEMAS Workshop call for participation  
<http://www.schemas-forum.org/workshops/ws3/index.html>

[6] University of Library and Information Science, Tsukuba, Japan  
<http://www.ulis.ac.jp/>

[7] World Agricultural Information Centre

<http://www.fao.org/waicent/>

[8] SALT project

<http://www.loria.fr/projets/SALT/>

[9] SCHEMAS Forum Registry

[10] EOR Toolkit

<http://eor.dublincore.org/>

[11] EASEL project

<http://www.fdgrouop.co.uk/easel/>

[12] Berners-Lee's diagram is reproduced in Tom Baker's workshop presentation. See

<http://www.schemas-forum.org/workshops/ws3/presentations/baker/sld017.htm>