

SCHEMAS

Contract: N° IST-1999-10100

Forum for Metadata Schema Implementers

D44: SCHEMAS FOURTH WORKSHOP REPORT

Document number:

SCHEMAS-GMD-WP4-D44-Final-20020110

General Information

Title	SCHEMAS: Fourth Workshop Report
Creator	Gauri Salokhe
Subject-Keywords	Deliverable D44; fourth workshop; Sharing schemas: progress and future plans
Description	This document describes the fourth SCHEMAS workshop which took place in The Hague (Den Haag), The Netherlands on November 30, 2001.
Publisher	GMD
Contributor	Makx Dekkers
Date	2002-01-10
Type	Text Manuscript
Format	application/msword
Identifier-URL	
Identifier-Document Number	SCHEMAS-GMD-WP4-D44-Final-20020110
Language	English
Rights	European Commission

Dublin Core Metadata in HTML

```
<META NAME="DC.Title" CONTENT="SCHEMAS: Fourth Workshop Report">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">

<META NAME="DC.Creator" CONTENT="Gauri Salokhe">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">

<META NAME="DC.Creator.Address" CONTENT="FhG, Sankt Augustin, Germany">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">

<META NAME="DC.Subject" CONTENT="Deliverable D44">
<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="Sharing Schemas: Progress and Future Plans">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Description" CONTENT="This document describes the fourth SCHEMAS
workshop which took place in The Hague (Den Haag), The Netherlands on November 30, 2001.">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#description">

<META NAME="DC.Publisher" CONTENT="GMD">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#publisher">

<META NAME="DC.Contributor" CONTENT="Makx Dekkers">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">

<META NAME="DC.Date" CONTENT="(SCHEME=ISO8601) 2002-01-10">
<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">

<META NAME="DC.Language" CONTENT="English">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">

<META NAME="DC.Rights" CONTENT="European Commission">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#rights">
```

Dublin Core Metadata in RDF

```
<?xml version="1.0"?>
<rdf:RDF xml:lang="en"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:smes="http://www.schemas-forum.org/registry/schemas/SCHEMAS/1.0/smes#"
>

<rdf:RDF rdf:resource = "SCHEMAS-GMD-WP4-D44-Final-20020110">
<dc:title> SCHEMAS: Fourth Workshop Report </dc:title>
<dc:creator> Gauri Salokhe </dc:creator>
<dc:subject> Deliverable D44 </dc:subject>
<dc:subject> Sharing Schemas: Progress and Future Plans </dc:subject>
<dc:subject> WP4 </dc:subject>
<dc:description> This document describes the fourth SCHEMAS workshop which took place in The
Hague (Den Haag), The Netherlands on November 30, 2001. </dc:description>
<dc:publisher> GMD </dc:publisher>
<dc:contributor> Makx Dekkers</dc:contributor>
<dc:date> 2002-01-10 </dc:date>
<dc:type> Text.Manuscript </dc:type>
<dc:format> application/msword </dc:format>
<dc:identifier> SCHEMAS-GMD-WP4-D44-Final-20020110 </dc:identifier>
<dc:language> English </dc:language>
<dc:rights> European Commission <dc:rights>
</rdf:RDF>
```

Table of Contents

1. Introduction	6
2. Aims of the workshop	6
3. Agenda of the workshop	6
4. Summary of the Presentations	6
<i>Review and future of SCHEMAS - Tony Hegarty, PwC</i>	7
<i>Making sense of schemas - Thomas Baker, FhG</i>	7
<i>Schemas in the real world : application profiles - Rachel Heery, UKOLN</i>	7
<i>Using the SCHEMAS Forum Registry - Manjula Patel, UKOLN</i>	7
<i>Knowledge Technologies in the 6th Framework Programme - Pierre-Paul Sondag, EC</i>	8
<i>Interoperability across metadata standards - Makx Dekkers, PwC</i>	8
5. Summary of the Panel Discussion	8
<i>What is the importance of Semantic Web vision – both in general terms and in terms of your domain or own work?</i>	8
<i>What are the main challenges and opportunities in cross-standard co-operation?</i>	8
<i>What are the main technical challenges and what do we need from W3C?</i>	9
6. Information Materials	9
7. Costs Incurred	10
8. Conclusions from the Evaluation Forms	10
9. Concluding Remarks	10
<i>Metadata standards and domain coverage</i>	10
<i>Registries</i>	10
<i>Tools</i>	11
<i>RDF</i>	11
<i>Semantic Web</i>	11
References	11

1. Introduction

This document reports on last of four workshops organised by the SCHEMAS Project. [1] The workshop, with 43 participants, was held at the Koninklijke Bibliotheek (Royal Library) in The Hague, The Netherlands, on Friday November 30, 2001. The theme of the one-day workshop was "*Sharing schemas: progress and future plans*".

2. Aims of the workshop

The purpose of this workshop was to review means by which projects could share information about metadata in use and to consider how collaboration in this area might be made more effective.

The workshop took place in the context of increasing interest in the infrastructure needed for the 'Semantic Web'. The semantic coherence of metadata will be a key issue in the development of the Semantic Web. Understanding metadata schemas in a distributed environment depends not just on sharing vocabularies, but on sharing grammars, or data models. The metadata schemas native to the tremendous diversity of services on the Web are based on local data models that often do not readily make sense in collaborative environments. The SCHEMAS Project has investigated the use of RDF, the Resource Description Framework, for expressing the "profiles" of local metadata models in terms that are mergible across applications. The construction of environments for discovering metadata semantics in use ("registries"), and the standardisation of such models, or lack thereof, is expected to determine the semantic coherence and usefulness of metadata on the Web over the longer term. [2]

The target audience for the workshop were those with a common interest across sectors in learning about recent developments which enable sharing of metadata schemas, and in contributing to future plans. The workshop was intended for those involved in building Web-based systems – whether as metadata creators, system implementers, or software developers within the educational, cultural heritage, government, industrial or other sectors.

The workshop intended to inform participants on recent developments regarding the use of metadata schemas and application profiles. It highlighted the role of registries in providing access to information about schemas complemented by an online demonstration of the SCHEMAS Registry.

A panel discussion addressed issues related to the development and use of metadata standards in various domains.

3. Agenda of the workshop

The programme of the workshop, with links to presentation materials, can be found at: <http://www.schemas-forum.org/workshops/ws4/programme.html> [3]

4. Summary of the Presentations

Review and future of SCHEMAS - Tony Hegarty, PwC

Tony Hegarty of PricewaterhouseCoopers looked back at what SCHEMAS had intended to achieve and how far the project has achieved those objectives. SCHEMAS has achieved a lot in understanding issues related to different types of schemas – for example, the distinction between namespace schemas that 'define' metadata terms and application profile schemas that 'use' them. Creating an RDF-based registry of these schemas, however, turned out to be more difficult to achieve than had been expected. Although SCHEMAS made significant progress in this area, there is a great need for further research.

Making sense of schemas - Thomas Baker, FhG

The presentation introduced metadata as a language based on vocabularies used in the context of particular data models, or grammars. The idea behind metadata registries is comparable to that of a dictionary: dictionaries are reference tools for words in natural languages while registries provide information about the semantics of metadata elements.

The Resource Description Framework (RDF) can be seen as a common grammar for translating, and thereby partially understanding, a vast diversity of native metadata languages on the Web. The W3C Semantic Web activity has defined RDF as a first step to enable integrated access to the Web resources on a grand scale. [4] The success and failures of SCHEMAS experience with representing application profiles in RDF provide input to the longer-term trend towards sharing metadata vocabularies machine-understandably over the Web.

Schemas in the real world : application profiles - Rachel Heery, UKOLN

The experience of implementing resource discovery systems shows that customisation and localisation of standard metadata schemas is widespread. 'Application profile schemas' are designed to express such variants.

The SCHEMAS Project defined "application profile" as a schema which consists of data elements drawn from one or more namespaces, combined together by implementers, and optimised for a particular local application. The concept of application profile contrasts with that of a namespace schema, where metadata terms are defined for reuse by others.. [2,5]

Using the SCHEMAS Forum Registry - Manjula Patel, UKOLN

The SCHEMAS Forum Registry holds several metadata element sets as well as a large number of activity reports which describe and comment on various metadata related activities and initiatives. [6] The current Registry contains a number of namespace schemas and several project application profiles.

The presentation gave an overview of the functionality of the Registry. Two user scenarios presented how the contents of the Registry can be useful to users from different sectors for different purposes.

Knowledge Technologies in the 6th Framework Programme - Pierre-Paul Sondag, EC

An introduction to the 6th Framework Programme (FP6) of Information Society Technologies (IST) was presented in the context of Knowledge Technologies. The field of IST is one of the main thematic priorities of the FP6 proposal and forms a significant part of the eighth priority anticipating science and technology needs. The FP6 aims to provide funding to projects that cater to the issues of usability, representation of content as knowledge, interoperability and standards, and communities and portals.

Interoperability across metadata standards - Makx Dekkers, PwC

In the current Web environment, interoperability across metadata standards is hampered by different metadata standards using different elements, grammars (data models) and encoding formats. The presentation expressed the need for these standards to interoperate to help realise the vision of Semantic Web. The presentation also emphasised the need for future work in shared understanding of objectives as well as in finding appropriate technical mechanisms to underpin this interoperability.

5. Summary of the Panel Discussion

The last part of the workshop consisted of a discussion panel containing experts from different domain areas. Following were the domains represented:

- Rachel Heery (Academic),
- Michael Day (Cultural-Heritage),
- Tony Hegarty (Industrial),
- Tom Baker (Research)

Following sections contain summaries of the questions that were posed to the panel.

What is the importance of Semantic Web vision – both in general terms and in terms of your domain or own work?

It was generally observed that the vision of a Semantic Web (SW) holds out the possibility of a shallow form of interoperability, but domain-specific needs in many sectors imply greater complexity as well as interoperability at a deeper level. A need for interoperability between learning environments and digital libraries was of primary concern to the Academic community.

It appears today that we have parallel universes of discourse with regard to metadata. The worlds of Libraries, of Research, and of Corporations give different names to similar phenomena – for example, "controlled vocabularies" (libraries), "ontologies" (computer science), and "taxonomies" (corporations).

In general, a need for opening access to government information overlaps with needs in the archival and cultural-heritage sectors. Several initiatives in Cultural Heritage and Libraries are based on related Entity-Relationship models, such as the IFLA [RFBF], ABC (Harmony Project), and Cidoc, and there is an evident need for a consistent model to describe them on the Web.

What are the main challenges and opportunities in cross-standard co-operation?

The challenges include:

- The need for recognition, across standardisation activities, of a distinction between "defining" and "reusing" terms. There is also a need to encourage more reuse of vocabularies as opposed to having projects and implementers define their own local standards.
- Although contexts such as the Dublin Core Metadata Initiative (DCMI) have facilitated some collaboration within domains, there is still a lack of openness and collaboration between domains. [7]
- The deployment of metadata needs to be user-driven and needs to be reviewed in terms of its usability. For many generic resource discovery applications, broad-brush categories may be enough, while specialised communities may need more extensive vocabularies.
- It would be useful if standards were to declare their metadata in machine-readable forms which would facilitate cross-referencing, inferencing, and communication among agents.
- In the cultural heritage domain, the problem lies in fragmentation of data as well as the existence of legacy data. The domain itself seems to be fragmented with several different parallel standardisation activities currently taking place. There is a lack of focus on broader Semantic Web objectives.

As for cooperation across standards:

- The use of simple metadata pidgins could facilitate interoperability between domains.
- The Semantic Web initiative and related data models such as RDF will, ideally, provide a focus for efforts to "webify" the diversity of semantic standards in a coherence and machine-usable way.

What are the main technical challenges and what do we need from W3C?

A general consensus was expressed to have a more practical, tool-driven approach to application of the SW concept. The SW group needs to facilitate the creation of simple applications before addressing the grander issues.

There is an obvious lack of tools which would move the vision of SW forward. Tools such as Protégé were deemed too complex. There is a need for tools that can be customised for different types of audience. These tools should help developers declare and manage vocabularies, annotate resources and later edit those annotations, and they should be able to handle the issue of versioning.

It was also proposed that follow-on projects to SCHEMAS could provide a test-bed for SW tools as they become available.

6. Information Materials

All of the workshop will be made available on the SCHEMAS Website.

Further information materials, including guidelines to the registry and writing application profiles, a glossary and a FAQ will be made available over the next two months.

7. Costs Incurred

The costs of the workshop were within the allocated budget.

8. Conclusions from the Evaluation Forms

Feedback from the workshop indicated that the overall content of the workshop proved to be valuable to the delegates in understanding the issues pertaining to schemas, application profiles and metadata in general. The registry demonstration was considered very valuable as it gave an insight into the possible usage of a metadata registry.

Several of the delegates expressed the need for more concrete examples, in standards development and implementation. There was also a need for further demonstration of tools that could be used to build application profiles. There is still a lot of confusion over the use of standards such as XML, RDF, DAML+OIL where more clarification about each of their roles, advantages and disadvantages would help.

One of the delegates observed that most of those present were content providers. Future workshops should try to involve software developers as well.

Overall, the content of the programme, the speakers, the length and timing of the workshop, and the venue were rated between very good to good on the evaluation forms.

9. Concluding Remarks

Metadata standards and domain coverage

SCHEMAS results suggest that there are too many overlapping standards – more needs to be done to achieve rationalisation and, ideally, harmonisation of standards. Greater emphasis should be given to the need for fewer to 'define' and more to 'reuse'. At present, many activities are taking place in research environments. It is necessary to enable vocabulary management for commercial domains - at some point this technology should become a black-box.

Registries

There are quite a few registry activities around, for example SCHEMAS, DCMI, and MetaForm. It is important that networks of registries be established, keeping the information close to where it can be maintained, to be harvested when necessary and reused by multiple registries or portals.

Machine-understandable vocabularies will be a crucial part of a Semantic Web infrastructure. As the growth of standards and schemas gains pace, there is an urgent need to create and maintain distributed registries. This in turn creates a need for a functional registry infrastructure which can process machine-understandable schemas as well as the links and cross-walks between them. With parallel efforts to build registries, the need for encoding vocabularies in a common form has become even more urgent.

Tools

There is a need for simple tools for registries, based on increased understanding and co-operation between tools developers and people who run and maintain registries. For this, funding should be made available. The current model of an RDF registry is not robust enough and an appropriate solution should be found to solve the legacy issue. RDF related tools are also not easy to use and W3C needs to make quick progress in development of easy to use tools.

To facilitate independent maintenance of the information in the registry, tools that will enable applications to create and maintain their own data models that are automatically infusible into a distributed registry are required.

The SCHEMAS Registry, currently a prototype, needs to have more browsing capability and the ability to do free-text search across all types of data.

RDF

For now, RDF seems to surpass XML when dealing with, among other things, vocabularies translated into multiple languages. Just as URIs can be used to merge data from different sources, they can be used to link the names and definitions of metadata terms in different languages.

The RDF model is a good foundation for further progress. It is important to distinguish between internal requirements, which may be complex, and the need for interoperability in the broader Web environment, which calls for simplicity. Deep interoperability among the complex data models of the world is currently not a realistic possibility.

Semantic Web

The Semantic Web looks at a shallow approach to interoperability (linking resources from a wide variety of sources), while most practical work takes place in a more controlled environment, with known cooperation partners and well-defined goals, where deep interoperability needs to be achieved. It is important to keep the approach to broader interoperability simple and to look at it from a practical perspective. Before moving out into unknown territory, intra-domain interoperability needs to be addressed and achieved first.

It was emphasised that the W3C needed to urgently provide tools that are like black-boxes for users, with maximum capabilities for creation and maintenance of schemas. Exemplars of SW applications are also needed from them in the near future to help people realise the concept of SW in the near term.

References

- [1] SCHEMAS Project Website
<http://www.schemas-forum.org/>
- [2] Thomas Baker, Makx Dekkers, Rachel Heery, Manjula Patel and Gauri Salokhe

- What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives* Journal of Digital Information 2 (2), November 2001.
<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/>
- [3] Fourth SCHEMAS workshop programme
<http://www.schemas-forum.org/workshops/ws4/programme.html>
- [4] Resource Description Framework
<http://www.w2.org/TR/REC-rdf-syntax/>
- [5] Rachel Heery and Manjula Patel *Application Profiles: Mixing and Matching Metadata Schemas* Ariadne, Issue 25, September 2000.
<http://www.ariadne.ac.uk/issue25/app-profiles/>
- [6] SCHEMAS Forum Registry
<http://www.schemas-forum.org/registry/>
- [7] Dublin Core Metadata Initiative
<http://www.dublincore.org/>