

SCHEMAS

Contract: N° IST-1999-10100

Forum for Metadata Schema Implementers

SCHEMAS: Best practice guidelines for managing a registry

Document number: SCHEMAS-UKOLN-WP5-D52-Final-20020207

General Information

Title	SCHEMAS: Best practice guidelines for managing a registry
Creator	Rachel Heery
Subject-Keywords	SCHEMAS D52, disclosure of local metadata usage, metadata registry, management
Description	This document suggests high level guidelines on best practice for managing a registry. The guidelines are based on discussion arising from the SCHEMAS workshops, from experience gained in the wider project., and from related work elsewhere.
Publisher	UKOLN
Contributors	PwC, GMD
Date	7 February 2002
Type	Text Manuscript
Format	application/msword
Identifier-URL	
Identifier-Document Number	SCHEMAS-UKOLN-WP5-D52-Final-20020207
Language	English
Rights	European Commission

Dublin Core Metadata in HTML

```
<META NAME="DC.Title" CONTENT=" SCHEMAS: Best practice guidelines for managing a
registry">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">

<META NAME="DC.Creator" CONTENT="Rachel aHeery">
<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 D52">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Subject" CONTENT="WP5">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Subject" CONTENT="schemas registry management">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Description" CONTENT="This document proposes guidelines for good practice in
managing a schemas registry">
<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) 2002-01-26">
<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">
```

1. INTRODUCTION	5
2. CONTEXT.....	6
2.1 METADATA REGISTRY INITIATIVES.....	6
2.2 DEVELOPMENT OF SCHEMA AND ONTOLOGY LANGUAGES.....	7
2.3 UNDERLYING TECHNOLOGIES.....	7
2.4 QUALITY FRAMEWORKS	8
3. PROPOSED BEST PRACTICE GUIDELINES.....	9
3.1 DEFINE AND PUBLISH THE SCOPE OF THE REGISTRY.....	9
3.2 DEFINE AND PUBLISH LANGUAGE POLICY	10
3.3 DECLARE POLICY ON QUALITY CONTROL	10
3.4 DEFINE APPROPRIATE DATA MODEL	11
3.5 DECLARE TECHNOLOGY AND STANDARDS IN USE.....	12
3.6 FACILITATE DISTRIBUTED CREATION AND UPDATE OF SCHEMAS	12
3.7 ENSURE EFFECTIVE USER INTERFACE	13
4. TABULAR SUMMARY.....	14
5. REFERENCES.....	15

1. Introduction

This document suggests guidelines on best practice for managing a schemas registry. The aim is to propose a small set of principles which might be readily accepted as good practice by registry implementors. The measure by which we recognise 'good management practice' is that of *delivering to the user an effective and positive experience of the registry*. The guidelines therefore focus on key areas which affect the user's perspective of the registry service.

The SCHEMAS Glossary outlines the overall scope of registry activity within the project, and in particular explains our approach to building a metadata registry:

.... As used in the SCHEMAS Project, the term "registry" refers, ideally, to a database that harvests various types of metadata vocabularies from their maintainers over the Web. In response to queries, such a registry should provide term-level documentation of definitions and usage along with contextual annotations. It should in effect function as an indexing engine for dynamically updating, merging, and serving up a large corpus of "dictionary" entries for metadata terms. The context for such a registry is the notion of a Semantic Web where anybody or any organisation can declare a metadata vocabulary and assert a relationship between that vocabulary and any other vocabulary on the Web.
[SCHEMAS GLOSSARY]

The guidelines presented here emerge from experience of building such a registry, and in addition, they draw on good practice in other existing registry services. Previous work undertaken as part of the DESIRE project concerning quality control in subject gateways has also proved valuable. The guidelines are of a general, high-level nature and are intended to be relevant to the variety of new metadata registry services now emerging. Indeed we hope deployment of the guidelines will assist users of registries to understand the purpose and characteristics of whichever registry service they might use.

We would like the guidelines to encourage implementors (including ourselves in our various registry activities) to express information about the construction of their registry and the way it is maintained. The guidelines are not intended in any way to deprecate experimental work, rather to consider how, within this relatively new area of activity, users might be best informed. This should help the user to discover whether a registry meets their requirements. In addition, by articulating suggested 'good practice', we hope to encourage exchange of experience amongst registry implementors.

The guidelines are presented in summary form appropriate for use in a future workshop and for publication on the SCHEMAS web site. These guidelines will also be put to use within the CORES registry, as CORES will continue the registry activity started in SCHEMAS. We hope that validation of the guidelines within CORES will give a sound basis for further discussion of the guidelines, with a view to their promotion and dissemination.

2. Context

Although there is little uniformity as to the precise nature of the services offered, the commonality between metadata registries is that they 'add value' in some way to the distributed, but isolated, listings of schema on the Web. Such added value might be

- navigation of schemas used in a particular domain
- translation of definitions and descriptions
- annotation of schemas giving usage guidance and links to related information
- machine readable access to enable interaction with software agents

The good practice guidelines will facilitate identification of the 'added value' that is being offered by a particular registry. Given the immature nature of registry services it is helpful for services to inform the user as to how a service is 'located' amongst the various initiatives. Many users will not have time to track the different strands of activity, and will rely on the service itself to inform them as to its purpose and motivation.

2.1 Metadata registry initiatives

Metadata registry services on the Web can trace an historical line back to shared data dictionaries, and to a number of registries of data elements encouraged by the ISO/IEC 11179 community. New impetus for the development of registries has come with the Semantic Web activity. The motivation for establishing registries arises from different groupings: domain communities, standardisation communities, corporate knowledge management. Examples include:

- Agencies maintaining directories of data elements in a domain area in accordance with ISO/IEC 11179. This standard specifies data element definitions and registration process for managing metadata to ensure consistency and accuracy. Example implementations are the National Health Information Knowledgebase hosted by the Australian Institute of Health and Welfare [NHIK]; and the Environmental Data Registry hosted by the US Environmental Protection Agency [EDR]
- The xml.org directory of XML document specifications facilitating reuse of DTD's hosted by the Organization for the Advancement of Structured Information Standards [XML.ORG]
- The MetaForm database of Dublin Core usage and mappings maintained at the State and University Library in Goettingen [METAFORM]
- The Semantic Web Agreement Group Dictionary, a database of terms for the semantic web that can be referred to by humans and software agents [SWAG].
- LEXML is hosting the open source development of a multi-lingual and multi-jurisdictional RDF Dictionary for the legal world [LEXML]

- The Microsoft Schema Registry supporting the Microsoft corporate intranet [BLISS]

Activities which project partners have been closely involved with include

- The Dublin Core Metadata Initiative Registry is in prototype exploring the best methods to provide an authoritative source of information regarding DCMI terms, both for humans and software [DCMI REGISTRY].
- The Metadata for Education Group (MEG) Registry brings together various metadata schemas for educational materials as part of the process of developing consensus on the description of learning resources [MEG].

More examples exist and others are in preparation. The guidelines presented here have been informed by reviewing this range of services and identifying the various aspects of good management presented. Although the aims and technologies differ, there seems sufficient commonality in such initiatives to reach a shared view on good practice for such registries.

2.2 Development of schema and ontology languages

During the period of the SCHEMAS project, the W3C has been active in advancing technologies to enable the structured expression of schemas. The Resource Description Framework has produced RDF specifications providing a basis for sharing metadata and schemas to support the exchange of knowledge on the Web. Related activity from the Knowledge Representation community has been focused in the W3C Web Ontology Working Group and this is beginning to impact the work of the RDF Interest Group. We are now seeing the Semantic Web Co-ordination Group bringing this related activity together, and may see in future a gradual convergence of work related to RDF with that of OIL+DAML into the Ontology Web Language. [SEMANTIC WEB]

The richness of ontology languages will affect the expressiveness of schemas. Whilst the user will typically want to be screened from the details of schema languages, it is useful for users to be aware of how a registry is located within current approaches.

2.3 Underlying technologies

There are a variety of technologies underlying present day registry services. Whilst many services use traditional relational database technology, there are experimental implementations where the RDF data model forms the basis for both the schemas and database.

The SCHEMAS project explored a variety of technologies for its own registry. We looked at three different technological approaches for implementing the registry: a simple HTML based 'list of links', an RDF based approach, and a 'traditional' relational database approach. Three different prototypes were implemented, each had its strengths and weaknesses.

The simple HTML based 'list of links', pointing to existing schema definitions and related initiatives, whilst easy to set up provided limited functionality.

The RDF approach offered the potential of a scaleable system based on a common data model (RDF) both for the schema and for the database. The project was looking towards implementation of a database which would be populated with schemas harvested directly from their maintainers in an open Web environment. We built this prototype with the Extensible Open RDF (EOR) Toolkit, an open-source software development project at the Online Computer Library Center (OCLC) [EOR].

However software tools for such a solution proved immature and required a level of development effort beyond that available to the project. In addition the chosen standard for schema specification (RDFS) was itself still under development, and conventions for expressing metadata schemas, in particular application profiles, were still to emerge [GLOSSARY].

For the final implementation of the Registry the project used a traditional database approach developed for an earlier project, DESIRE [DESIRE]. We entered metadata vocabularies into this registry directly, rather than harvesting them from the Web. It was accepted that this approach would be problematic if the registry were to draw in increasingly diverse schemas. Also the issue remained that simple tools and templates were not available for maintainers of schemas to declare their own vocabularies.

This state of flux , with sporadic development of somewhat competing technologies, with prototyping based on immature specifications, is commonplace in the context of the emergence of new services on the Web. Registry implementors will no doubt themselves be aware of the strengths and weaknesses of their chosen software solutions. However it is helpful for services to declare to users their location within the 'spectrum of innovation'. Users can then paint their own picture of the benefits offered by new technologies.

2.4 Quality frameworks

It is worthwhile relating good practice regarding registry services to activities elsewhere concerned with quality issues for web sites and services.

A number of subject gateways (subject based resource discovery services) have produced quality control guidelines and identified detailed criteria for ensuring quality control. The DESIRE project developed the basis of a quality framework for subject gateways, considering ways in which the gateways themselves could 'declare' information about their service, as well as considering the criteria gateways might use for selecting web resources for their catalogues. [HOFMAN]

The DESIRE quality framework is comprised of a comprehensive list of 'quality criteria'. The criteria were categorised as follows:

- Scope Policy: Considering your Users
- Content Criteria: Evaluating the Information
- Form Criteria: Evaluating the Medium
- Process Criteria: Evaluating the System
- Collection Management Policy: Considering your Service

Each of these categories was broken down into further detail. The comprehensiveness of the listed criteria in this report allow it to be viewed as a reference tool for building policy frameworks for particular services.

There is potential for applying this quality framework to a wider range of services than gateways alone [HEERY]. A collaborative effort is underway at present, informed by this framework, to develop quality standards for sites delivering cultural content via the Web [BRUSSELS]. Consideration of this ongoing work on quality frameworks has contributed significantly to the formation of guidelines for registries.

Investigation of the quality criteria emerging from this work reveals common themes. Information services on the Web face many similar issues, indeed both Web sites and metadata registries can be viewed as 'services' and it can be argued that all Web services, however large or small, need to meet similar quality criteria.

3. Proposed best practice guidelines

3.1 Define and publish the scope of the registry

Users are faced with a plethora of different offerings on the Web and they need a clear statement of the objectives of services they encounter. This is as true for registry services as for any other Web based information service. A statement of scope should be easily accessible from the home page of the registry. It should be worded to enable the user to quickly understand the overall purpose of the registry service on offer, and allow the user to judge if this service meets their requirements.

A statement on the scope of the registry might be expected to include the following aspects:

- ***Mission statement***
 - *What is overall objective of the service?*
- ***Motivation and benefits***
 - *How will the registry benefit users? Is the registry addressing particular research, technology or schema language issues?*

- **Target users**
 - *Who are seen as the primary users of the registry? Are they human and/or software agents? What are the categories of users e.g. metadata creators, schema designers, information professionals?*
- **Content 'collection policy'**
 - *What are the criteria for inclusion of schemas in the registry*
- **Areas out of scope**
 - *Are there significant areas that are 'out of scope'?*

3.2 Define and publish language policy

A statement of policy should be readily accessible to the user outlining the multilingual nature of the registry in regard to the user interface and to the schemas themselves.

In particular it should be made clear which parts of the schemas are translated, and who is responsible for that translation, whether this is the registry manager, or the owners of the schema, or third parties.

A declaration of language policy might be expected to cover::

- **Language of user interface**
 - *What languages are available?*
- **Language of schemas**
 - *Does the registry focus on schemas in a particular language? Are translations of appropriate parts of schema available? What is the policy for translating schemas? Are schemas translated by the registry manager, or the owners of the schema, or third parties?*

3.3 Declare policy on quality control

The user needs information on quality assurance procedures in order to evaluate the usefulness of the registry. Some registries may intend to include new, emerging and draft schemas, and it is valid to provide such a service. On the other hand some registries may aim to provide access only to those schemas approved by an authoritative source, or perhaps those schemas with successful deployment within a community. Declaring a policy regarding quality control makes it possible for the user to evaluate the registry.

A policy on quality control might be expected to include statements on:

- **Accuracy**
 - *Are the schemas included checked for accuracy? What means are used to ensure accuracy?*
- **Completeness**
 - *Are the schemas complete or do they represent a sub-set of other schema? How comprehensive is the information about the schema? Is additional information available elsewhere?*

- **Currency**
 - *How often are the schemas checked for currency? What is the registry policy on versioning of schemas? Are draft schemas included in the registry, if so are they clearly indicated? Are superseded versions of schemas included in the registry, if so are they clearly indicated?*
- **Registration authority**
 - *Is it clear on what authority schemas are entered into the registry? Are schemas created by the schema maintainer? Or are they created by the registry manager?*
- **Filtering**
 - *Are the schemas quality assured? Is the information in the registry peer-reviewed or refereed in any way?*

3.4 Define appropriate data model

The purpose of a registry is to 'add value' over and above that given by individual, distributed listings of schema. One can assume many registries will add value by indexing the distributed schema, and providing navigational aids to a number of related schema. It is important that by doing this a registry does not significantly distort the semantics of the schema or the structural relationships between terms within the schema. Having identified the scope of a registry, the registry manager needs to define the data model to be used within their registry. It is necessary to ensure the data model in the registry will express in as comprehensive and accurate way as possible the semantics and structure of the metadata vocabularies to which it gives access. In other words the view given by the registry must have a close affinity to the 'native data models' of the particular metadata vocabularies it navigates.

This is a fundamental challenge for a 'registry service', one that may only be partially achievable at present given that there is no consensus on the structure of metadata vocabularies and given that ontology languages for expressing schema on the Web are immature. Further discussion of the issues involved is available in the SCHEMAS Glossary [SCHEMAS GLOSSARY].

Nevertheless for the stakeholders in the registry service (both the user and the schema maintainer) it is important that there is no significant distortion of the schema semantics or structure introduced by the registry. This means it may be necessary to limit the content of a particular registry to schema with a similar data model, or to ensure that the 'presentation layer' of the registry is sufficiently sophisticated to deliver a correct interpretation of the schema to the user.

In summary a level of 'lossiness' may be inevitable in providing a merged view of several schemas, but good practice dictates that this should be at an acceptable level. Stakeholders in the registry service should judge that the 'partial understanding' on offer is sufficiently true to the original schema to be useful.

Appropriate information regarding the data model might include:

- *Structural representation and explanation of the data model used within the registry*
 - *What are the components of the data model? Why is this model in use?*
- *Impact of data model on user's view of schemas*
 - *Does the data model in the registry lead to lossiness? Is this level of lossiness acceptable in order to achieve the objectives of the registry? Have schema maintainers approved the view of schemas presented by the registry?*

3.5 Declare technology and standards in use

It is clear that good management of a registry requires that the technology in use will deliver the objectives of the registry. For a registry service intended to be experimental and technically exciting, then use of immature software is acceptable. A proof of concept registry may need to use unproven, innovative technology. Alternatively, if reliability and low maintenance are priorities for the service, then good management may well dictate that the technology chosen is tried and tested.

Appropriate standards should be used to take advantage of shared effort, to enable exchange of schemas produced to a common format, and as the basis for interworking between registries.

In order to inform the user a statement on appropriate use of technology might be expected to include:

- *Standards*
 - *Which standards are used? Why have these standards been chosen as appropriate?*
- *Technical architecture*
 - *What technology underlies the system? What is the level of innovation in the system? What is the development path for the underlying software?*

3.6 Facilitate distributed creation and update of schemas

In order to create an infrastructure whereby schemas can be 'shared' an essential step is to support implementors in the creation and declaration of their schema. The deployment of simple tools to enable local implementors to construct and declare their schemas in a standard way would provide registries with the potential offered by standardised input. If the schema could be stored locally to the implementor, then several registries could then index and link to the schema (or indeed harvest them) in an automated way.

Registries and local implementors would benefit by collaborating on the deployment of tools for the local creation of schemas in a standard format.

3.7 Ensure effective user interface

It is critical that registries gather user feedback, consider usability issues, and thereby discover whether the services they offer fulfil user requirements. By undertaking such formative evaluation registries will ensure they develop over time into useful and effective services.

The user's experience of the registry will be dependent very much on the design of the user interface. Good web design principles should be followed. It is not intended to try to replicate or summarise those principles here in relation to a registry service, as that would be a major piece of work in itself.

Registries should declare a policy to address user interface issues which might be expected to include the following aspects:

- *Evaluation*
 - *What is the evaluation policy? How have usability issues been addressed? What are the usage patterns of the registry?*
- *Feedback mechanisms*
 - *Are contact details of registry manager readily available for users to feedback comments? Are contact details given for schema maintainers?*
- *Accessibility*
 - *Have accesssibility guidelines been applied?*
- *Navigability*
 - *Is the composition and organisation of information helpful to the user?*

4. Tabular summary

Scope	<ul style="list-style-type: none">• Mission statement• Motivation and benefits• Target users• Content collection policy• Highlight significant out of scope areas
Language policy	<ul style="list-style-type: none">• Language of user interface• Language of schemas
Quality control	<ul style="list-style-type: none">• Accuracy• Completeness• Currency• Registration authority• Filtering
Data model	<ul style="list-style-type: none">• Structural representation and explanation of the data model• Impact of data model on user's view of schema
Use of technology	<ul style="list-style-type: none">• Standards• Technical architecture
Distributed creation of schemas	<ul style="list-style-type: none">• Collaboration on deployment of tools for local creation of schemas in a standard format.
User interface	<ul style="list-style-type: none">• Evaluation• Feedback mechanisms• Accessibility• Navigability

5. References

[BLISS] Vivian Bliss, Metadata Registry supporting a corporate intranet. Microsoft Corp. Presentation at Schemas Project in Europe Workshop 3, Budapest, Hungary May 2001

www.schemas-forum.org/workshops/ws3/presentations/VBliss.ppt

[BRUSSELS] Coordination of National Digitisation Policies & Programmes

<http://www.cordis.lu/ist/ka3/digicult/en/eeurope.html>

[DESIRE] The DESIRE registry hosted at UKOLN.

<http://desire.ukoln.ac.uk/registry/>

[DCMI REGISTRY] DCMI registry prototypes. Hosted at OCLC.

<http://wip.dublincore.org:8080/registry/Registry>

[EOR] The EOR toolkit.

<http://eor.dublincore.org>

[HEERY] Rachel Heery, Quality issues for cultural web sites: experience from DESIRE and Renardus. Experts meeting on co-ordination of digitisation policies and programmes, Centre Albert Borschette, Brussels, 17 July 2001.

http://www.renardus.org/news/position_summ.htm

[HOFMAN] Paul Hofman, Emma Worsfold et al. Specification for resource description methods Part 2. Selection criteria for quality controlled information gateways. DESIRE, 1997. <http://www.ukoln.ac.uk/metadata/desire/quality/>

[LEXML] LEXML: Open Source Development of an RDF Dictionary

<http://home.snafu.de/mmuller/lexmlde/rdf.htm>

[MEG] Registry of MEG-related schemas. Hosted at UKOLN

<http://www.ukoln.ac.uk/metadata/education/registry/contents.html>

[METAFORM] MetaForm: Database containing Dublin Core manifestations and other metadata formats. Hosted at the State and University Library in Goettingen

<http://www2.sub.uni-goettingen.de/metaform/>

[NHIK] National Health Information Knowledgebase. Hosted by the Australian Institute of Health and Welfare

<http://www.aihw.gov.au/knowledgebase/index.html>

[SCHEMAS GLOSSARY] Thomas Baker and Gauri Salokhe, The SCHEMAS Forum – a Retrospective Glossary.

<http://www.schemas-forum.org/info-services/d74.htm>

[SEMANTIC WEB] Semantic Web Activity Statement
<http://www.w3c.org/2001/sw/Activity#intro>

[SWAG] SWAG Dictionary
<http://webns.net/>

[XML.ORG] The XML Registry hosted by OASIS.
<http://www.xml.org/xml/registry.jsp>