Registry for Educational Metadata Schemas

Final Project Report and Recommendations

Author	Rachel Heery
Date	7 April 2003
Version	Public circulation – financial details removed
Document Name	regfinalreport

1.   Executive Summary

As the number and range of Web based services increase there is a need to ensure that processes for managing the metadata workflow are effective. Metadata registries have emerged which address, in particular, the publication, navigation and sharing of schemas, schemas which express both 'standard element sets' and 'customised application profiles'. A managed schema registry becomes an authoritative source for information about metadata schemas, and provides an access point to those schemas for re-use.

The benefits of schema registries are in summary

- Increased interoperability between schemas as a result of re-use across many applications

- Less duplication of effort amongst implementors

- Promotion of existing solutions

- Harmonisation between 'competing' standards

The MEG Registry project has delivered a prototype registry and schema creation tools. There has been a limited evaluation of these tools by practitioners and implementers. UKOLN staff have seeded the registry with Dublin Core and IEEE/LOM schemas. The software is now available on an open source basis for installation by third parties, downloadable from ILRT.

Recommendation: Evaluation of the registry and schema creation tool software, and interest shown in these tools, lead us to recommend that JISC funds the project for a further period to establish a framework for use of these tools within the JISC Information Environment community.

Recommendation: Initially to provide a pilot registry service within the JISC Information Environment, and, when experience has been gained, to encourage take-up of the software by other interested parties within the Common Information Environment (noting that interest in using the tools has been shown by the British Library and the eEnvoy's office).

Recommendation: UKOLN and CETIS to collaborate on modelling to enhance support of the IEEE/LOM and to link to XML schemas.

Recommendation: ILRT to enhance the creation tool and registry software to incorporate enhanced data model and support for IEEE/LOM

Recommendation: Further evaluation based on user trials and usability testing is required

Recommendation: Improve presentation and usability of tools based on evaluation

Recommendation: Establish a shared open source development model (such as SourceForge) to facilitate co-ordination of development work, packaging of releases, software testing and quality assurance

2.   Purpose of the document

This document is the final deliverable of the MEG Registry project and is intended to provide a summary of work achieved. It includes recommendations for the future role of schema registries in the context of the developing JISC Information Environment.

Partners in the project were the University of Bath, UKOLN (project lead, data modelling, user requirements) and the University of Bristol, ILRT (software design and development).

3.   Background to Metadata Schema Registries

Structured information (metadata) is the basis for an ever increasing range of Web based services, services where metadata is being used for resource discovery, personalisation, rights negotiation, preservation and more. Many different element sets have been defined to fulfil these various requirements, resulting in domain specific and functional groupings of metadata elements. Some of these element sets are the result of formal standardisation processes, others are de facto standards emerging as new services are developed. Inevitably local extensions and variants (application profiles) are appearing to accommodate the specific requirements of particular applications. Although unnecessary proliferation of element sets is unwelcome, attempts to impose uniformity are unlikely to succeed and indeed are counter-productive, as innovations may well emerge from differentiation between services (and the metadata on which they are based).

As the number and range of Web based services increase there is a need to ensure that processes for managing the metadata workflow are effective. There have been a number of initiatives exploring provision of information about metadata formats whether at the level of data elements, standardised element sets or, at the implementation level, application profiles. The adoption of more structured forms for expressing information about data elements and element sets, in particular the Resource Discovery Framework (RDF) Schema Specification[1], is an enabling technology that has added impetus to these developments. The emergence of Semantic Web and Web Services technologies has encouraged the structured representation of element sets as machine readable schemas that can be manipulated by software tools. Metadata registries have emerged which address, in particular, the publication, navigation and sharing of schemas, schemas which express both 'standard element sets' and 'customised application profiles'. The overall aim of such registries is to provide users, whether human or software, with the means to navigate and access existing schemas, and register new schemas. A registry enables an implementor to 'publish', navigate and obtain information about metadata schemas and to re-use existing schemas. A managed schema registry becomes an authoritative source for information about metadata schemas, and provides an access point to those schemas for re-use.

The benefits of schema registries are in summary

- Increased interoperability between schemas as a result of re-use across many applications

- Less duplication of effort amongst implementors

- Promotion of existing solutions

- Harmonisation between 'competing' standards

Registries might take different forms whether providing

---

[1] Dan Brickley and R.V. Guha . RDF Vocabulary Description Language 1.0: RDF Schema W3C Working Draft 23 January 2003 <http://www.w3.org/TR/rdf-schema/>

- An authoritative source for a particular standard such as the DCMI Registry

- A repository of schemas (both element sets and application profiles) relevant to a particular sector or domain such as the MEG registry itself which focuses on the requirements of UK educational services

- A mapping and crosswalk registry as explored within the DESIRE project.

Over the last few years, research on schema registries within the DESIRE and SCHEMAS projects established a data model for declaring standard element sets and local usages (application profiles). The DESIRE and SCHEMAS demonstrator registries offered a Web interface for humans to use, enabling navigation by browsing and searching registered schemas. At that stage the RDF specification for declaring schemas was in its early days, and there were few tools available to provide support for applications based on RDF. The projects recognised the potential that a machine-readable standard syntax such as RDF Schema would bring, both in terms of scaleabilty and sustainability. Schemas encoded in RDF would be re-usable by other applications outwith the registry, both by applications such as metadata editors, as well as by other registry services. RDF also provides a framework for m2m access to a registry. However these projects did not have sufficient effort available to implement an RDF based solution, particularly given the immaturity of the RDF standards at that time.

The MEG Registry project provided an opportunity to implement a prototype schema registry based on RDF specifications and tools that had by this time become more stabilised. The aim was to seek a more sustainable and scaleable solution than had been available to previous projects.

Scaleability also needed to be addressed regarding creation and registration of schemas. Within the initial DESIRE and SCHEMAS projects, UKOLN staff created entries themselves for schemas in the registry. This was not a feasible solution in the longer term, and also meant that implementers were not gaining experience of structuring and developing their own schemas. The MEG project aimed to develop a schema creation tool, suitable for users with no knowledge of RDF, to be used in an interactive way with the MEG Registry.

4.   MEG Registry project objectives

The MEG group consists of over sixty members from a range of UK educational agencies whose interests cover the spectrum of primary and secondary education, life long learning, HE and FE. Approximately twenty members are known to be active in creating schemas to describe educational resources. UKOLN acts as co-ordinator for the group. The existence of such a focused group offers great potential for sharing and collaboration regarding the design and re-use of schemas.

Prior to the start of the project, consensus had been reached within MEG that a registry was required to:

- Provide easy access to standard educational metadata schemas through a Web interface which allows the user to navigate relationships between the components of those schemas

- Provide information about extensions to standard formats used by MEG members

- Give authoritative definitions for local elements, and permitted schemes for elements

Users of the MEG registry would typically come under the following categories:

- Publishers of standards

- Implementers seeking appropriate schemas

- Developers comparing schemas

- Metadata creators seeking assistance in using particular schemas correctly

The MEG registry aims to provide a mechanism for various learning initiatives to manage interoperability between their metadata implementations by publishing their schemas in a registry,

making them available to others. This provides a concrete way of encouraging sensible uniformity alongside necessary divergence. It helps avoid unnecessary duplication of effort, and supports sharing of common approaches.

As a first step towards an operational service the MEG registry project intended to provide enhanced software tools. In particular the focus was on

- ensuring the data model is extensible by using RDF to accommodate a variety of schemas

- improving ease of data input by enabling distributed input of schemas by MEG members, as input centrally by UKOLN staff was not feasible in the long term

- use of RDF schema language for input and output of schemas in order to facilitate interoperable exchange of schema data beyond the MEG system

In order to achieve these goals the MEG Registry project undertook development of a complementary schema creation tool and schema registry in order to support interactive creation of schemas. Schema creators would have search and browse access to schemas already registered in the registry, and could re-use data elements by incorporating them into their own application profiles.

5.   Methodology

The MEG Registry data model underlies the schema creation tool and the registry. The project undertook detailed modelling of the various classes of entity manipulated by the tools, and the relationships between instances of those classes: element sets, elements, element usages, encoding schemes, values within encoding schemes, responsible agencies.

Building on previous work within DESIRE and SCHEMAS, and informed by the RDFS specification, this data model is designed to provide flexibility for creation and registration of both metadata schemas and application profiles. The aim of such a modelling exercise is to ensure that schemas can be represented accurately in the MEG tools, and that the tools are able to deliver the searching and navigation functionality required by end-users.

Functional requirements were drawn up for a registry system made up of two components:

- The **registry** proper, which reads schemas, stores the aggregated data, and presents an interface to that data which allows a human reader to browse and search it.

- A **schema creation and registration tool**, which will allows a human user to create and edit schemas, and to submit/re-submit them to the registry.

The functional requirements were informed by a number of Use Cases

Use Case 1: Publishing a description of a Element Set

Use Case 2: Publishing a description of an Application Profile

Use Case 3: Indexing a standard schema for a Element Set

Use Case 4: Exploring Element Usage

The project reviewed existing software on which to base the schema creation and registration tool. Although various RDF schema creation tools existed, and some of these might have been modified to interact with the Registry, none was available which would sufficiently hide the underlying RDF and data model from a non-expert user. A schema creation and registration tool was therefore custom built for the project.

The Registry was re-engineered using the Redland toolkit, replicating the core functionality of the previous DESIRE prototype in an RDF environment.

In order to populate the registry, the project decided to seed the registry with well known standard schemas relevant to the target audience. Standards makers are now beginning to publish their element sets using RDF, and the project was able to use RDF schemas produced by DCMI as input

to the registry, and more recently has been able to use copies of the IEEE/LOM draft RDF schema for this purpose. Given the hierarchical nature of the IEEE/LOM significant effort was involved in deciding the best way map the IEE/LOM RDF schema onto the registry model.

It should be noted that the registry gives information on entities within its data model (such as application profiles, element sets) that are not known by the RDF model. This means that the DCMI and IEEE/LOM schemas produced by the standards bodies need to be supplemented with additional metadata about the terms described by the registry. The project has attempted to minimise the application specific extensions to RDF. In future there may be potential to upgrade to OWL (the Web Ontology language) to further minimise local RDF extensions.

Initial alpha testing of the tools was undertaken at UKOLN which resulted in a number of software fixes and minor enhancements being carried out by the software developers.

A workshop was held at the end of the project to support MEG members inputting their schemas in a structured format using the new tools, partly as a means of populating the registry and partly to evaluate the tools.

6. Project Outputs

Note that all outputs are available from the project web pages at
<http://www.ukoln.ac.uk/metadata/education/regproj>

6.1 Review of existing software solutions

A review of existing software was undertaken by ILRT, informed by their involvement in emerging Semantic Web activities and the RDF Core activity at W3C. A summary report of findings is available linked from the MEG registry project web site.

6.2 Functional specification

A functional specification was drawn up by UKOLN in collaboration with MEG users. This outlined the required functionality both for the schema creation tool and registry. A significant part of this specification was a revised data model for the MEG Registry that modelled in a formal and systematic way the various entities to be manipulated by the creation tool and registry.

6.3 New implementation of registry based on sustainable software

A new implementation of schema registry software was developed by ILRT according to the functional specification produced by UKOLN. The new lightweight MEG registry was based on Redland Toolkit developed by Dave Beckett at ILRT. This software is made available as open source software and can be downloaded from the ILRT CVS repository.

6.4 Schema registration tool

A schema creation and registration tool was developed by ILRT according to the functional specification produced by UKOLN. The tool enables distributed creation of RDF schemas for element sets and application profiles either in a standalone fashion or in interactive mode with the MEG registry. The RDF schema can then be registered and input into the MEG registry automatically. This software is made available as open source software and can be downloaded from the ILRT CVS repository.

6.5 Schema creation workshop

A MEG workshop was run by UKOLN where the new registration tool was introduced to MEG members. Attendees were given the opportunity to use the tool to create their own schemas in a RDF

format so that they could be registered in the MEG registry. This was a hands-on workshop held in January 2003, and was attended by nine members of the MEG group, including two members of CETIS, and in addition two UKOLN RDNC staff.

7.   Dissemination and collaboration

MEG project partners are involved in various activities related to schema creation and registration:

Dave Beckett, ILRT, who developed the registry software is a member of the W3C RDF Core working group. Damian Steer, who developed the schema creation tool, went on from ILRT to a placement at HP Labs where he followed up his interests in RDF tools, producing a schema browsing tool BrownSauce. Pete Johnston, UKOLN, has contributed to working groups formulating RDF and XML expressions of DCMI element sets, work which has been informed by experience with modelling and design of the MEG Registry tools. Dave and Pete have been active within the DCMI working group looking at RDF bindings.Rachel Heery, UKOLN, has continued her involvement with the DCMI Registry development.

A paper outlining the MEG Registry and Schema Creation and Registration Tool (SCART) was presented at the DCMI 2002 Conference in Florence[2], followed by a demonstration. Positive feedback was received from information service providers and implementers from both the corporate and education sector. This paper formed part of other registry related activities at the conference including presentation of the DCMI Registry, and a meeting of the DCMI Registry working group.

The MEG Registry project has also collaborated closely with the EC funded CORES project (UKOLN being partners in both projects). The CORES project is funded as a supporting measure under the EC 5th Framework Semantic Web activity, and has an overall aim of moving forward interoperability between metadata schemas. CORES main objectives are to bring standards makers together to agree on a common approach to declaring schemas, and to enable implementers to share schemas by means of a registry and creation tool. CORES is a relatively short term 15 month project with an end date of June 2003. It has a small amount of software development effort located at SZTAKI, the Computer and Automation Research Institute of the Hungarian Academy of Sciences. The timing of the two projects has been fortuitous, enabling SZTAKI to build on the MEG open source software. SZTAKI have enhanced the MEG tools with authentication and annotation functionality. Potentially annotation is a powerful facility allowing 'expert comment' (whether domain, technical, implementation or other expertise) on element sets, application profiles, and encoding schemes within the registry.

8.   Evaluation

There were a number of themes that emerged from evaluation by various parties: at the workshop from the MEG group including CETIS staff, peer review by RDNC staff, and feedback from dissemination at DC2002.

There was overall agreement on the benefits of the MEG Registry and schema creation tool:

- providing sustainable and extensible tools for creation, navigation and re-use of DCMI-like schema

- enabling sharing of information about schemas

- as a basis for schema registry Web Services

---

[2] Rachel Heery, Pete Johnston, Dave Beckett, Damian Steer. The MEG Registry and SCART: complementary tools for creation, discovery and re-use of metadata schemas. October 2002.
In: *Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002.*
Florence: Firenze University Press, 2002, pp. 125-132.
<http://www.bncf.net/dc2002/program/ft/paper14.pdf>

- as a focus for alignment on use of standards within communities of use

It was recognised that some areas needed to be addressed:

- *How well the data model accommodated non-DCMI schema, in particular the IEEE/LOM element set and application profiles derived from that schema.*

This was the main topic for discussion at the MEG Registry workshop, where most of the participants were IEEE/LOM users. The constraints of the RDF binding and the registry data model mean that much of the hierarchical aspects of the IEEE/LOM element set are stripped away and represented within the registry by a 'flattened' model. The difficulty in expressing IEEE/LOM in RDF has been acknowledged by Mikael Nilsson in his account of issues associated with this binding:

> Thus, in order to encode the LOM 1.0 data model in RDF in a manner compatible with Dublin Core, we have had to do some re-modelling of LOM, trying to interpret the element hierarchy in terms of "properties" and "values".[3]

It is as a consequence of this re-modelling that the representations of the IEEE/LOM provided by the registry appear as "flat" listings of properties, rather than the tree-like structures perhaps more familiar to IEEE/LOM implementers.

Following the workshop discussion within the project and with other interested parties has focused on possibilities for extending the registry data model to better accommodate the IEEE/LOM. Some adaptations within the registry would enable pointers to existing external XML schemas, and there may be possibilities to view elements in a 'native' format. However the basis of the tools assumes that the RDF expression of the schema is 'primary' and there may need to be some compromise solution.

The motivation for developing the registry tools has been very much based on supporting interoperability, and it would be disappointing not to pursue a common approach to digital library and eLearning metadata formats within these tools. For this reason the alternative option of developing an IEEE/LOM specific registry is not immediately attractive.

- *Improvements in usability of tools*

It was acknowledged that more work needs to be done on usability of tools, introducing for example context specific help for novice users, better sizing of windows, more intuitive labelling of input fields. Most significant was a requirement for schema creators to build on existing application profiles.

- *Policy framework regarding persistence of the registry service and schema content, collection scope and quality assurance*

There was discussion at the workshop regarding the need for policies to ensure good management of the registry. While acknowledging that a number of policy issues are more relevant to a 'production' registry, concerns about persistence are particularly significant as the funding for the MEG Registry project comes to an end. Workshop attendees did not feel able to commit any significant efforts to creating and registering schemas unless they could have confidence in continuance of the registry. In order to commit time to gaining familiarity with the tools they need some assurance regarding future maintenance of the underlying software, and future management of the content of the registry.

- *Need for a number of relatively minor enhancements and bug fixes*

A number of fixes and enhancements were identified during user testing. These need to be addressed prior to more widespread use of the tools.

---

[3] Nilsson, Mikael. Semantic issues with the LOM RDF binding, 2003-01-15.

<http://kmr.nada.kth.se/el/ims/md-lom-semantics.html>

9.   Recommendations for the future

9.1   Establish pilot service

The MEG Registry project has delivered a prototype registry and schema creation tools. There has been a limited evaluation of these tools by practitioners and implementers. UKOLN staff have seeded the registry with DC and IEEE/LOM schemas. The software is now available on an open source basis for installation, downloadable from ILRT.

Recommendation: Evaluation of the registry and schema creation tool software, and interest shown in these tools, lead us to recommend that JISC funds the project for a further period to establish a framework for a pilot service within the JISC Information Environment community.

9.2   Business model

With the MEG Registry project funding coming to an end, there is no longer funded effort for managing and maintaining the MEG Registry nor for software maintenance and support. Although there is some effort within the UKOLN Interoperability Focus, only limited time can be dedicated to the registry activity. With no direction, after a limited time, the registry content will become confused. Also with no further funding, the open source software on which the registry tools are based can only be maintained on a best efforts basis, and there will be no effort for enhancement or trouble-shooting.

There is now an opportunity to consider options for the provision of such a service. Options for a 'business model' can be summarised as

- Service for JISC IE – funded by JISC as a service to the JISC IE community

    o   Supporting sharing and re-use of schema amongst implementers

    o   Providing a repository of authoritative schemas in use within JISC IE applications and services

    o   Providing focus for joint digital library and eLearning metadata activity (e.g. UKOLN and CETIS working together to provide solution for their communities)

- Service for MEG community – funded by JISC and Becta using the same model as the MEG Registry project

    o   Wider target audience than JISC IE,

    o   Challenging organisational and funding model

- Service for Common Information Environment – joint funding by Common Information Environment lead players

    o   Is provision of enabling infrastructure an attractive short term goal for the CIE?

    o   Challenging organisational and funding model

Recommendation: Initially to provide a pilot registry service within the JISC Information Environment, and, when experience has been gained, to encourage take-up of the software by other interested parties within the Common Information Environment (noting that interest in using the tools has been shown by the British Library and the eEnvoy's office).

9.3   Collaboration between digital library and eLearning communities

A key issue for the project, and indeed for the wider provision of schema registries, is how 'accommodating' a common data model can be to the variety of schemas that might be registered. Much schema registry activity over the past few years has focused on Dublin Core element sets and application profiles. Although the MEG tools are not overtly 'Dublin Core centric', the data model is

hospitable to simple 'Dublin-Core-like' schema. In order to facilitate creation and registration of more hierarchical structured schema such as the IEEE/LOM more work needs to be done particularly on the data model, and the schema creation tool.

Recommendation: UKOLN and CETIS to collaborate on data modelling to enhance support of the IEEE/LOM and to link to XML schemas.

Recommendation: ILRT to enhance the creation tool and registry software to incorporate enhanced data model and support for IEEE/LOM.

9.4   Usability

The project has undertaken preliminary testing and evaluation of the tools, and as part of this process has identified the need for a number of minor enhancements and fixes. Further user trials are required to refine the interface and ensure usability issues are addressed.

Recommendation: Improve presentation and usability of tools based on evaluation.

9.5   Open source software maintenance

Over the last year SZTAKI, within the CORES project, have collaborated with ILRT adding enhancements to the registry and schema creation tool software. This has resulted in separate versions of the software available from the respective CVS repositories of the two institutions. There are other potential collaborators. Interest in downloading and further enhancing the software has been shown by the GEM project in the context of a CORES workshop. There are opportunities to collaborate with OCLC's registry development work. Others may be interested after further dissemination is carried out.

This is a good opportunity to leverage such contributions by managing the software in a centralised open source software environment such as SourceForge. SourceForge provides support management tools (Tracker) and release management using a Web-based file release system. Release testing and integration of fixes could be co-ordinated.

The lead developer (Dave Beckett, ILRT) has expressed an interest in leading this work in collaboration with Andras Micsik at SZTAKI.

Recommendation: Establish a shared open source development model (such as SourceForge) to facilitate co-ordination of development work, packaging of releases, software testing and quality assurance.

10.   Useful Links

MEG Registry project <http://www.ukoln.ac.uk/metadata/education/regproj/>

GEM (Gateway to Educational Material) R&D Server <http://gemstar.ischool.washington.edu/>

SourceForge <http://sourceforge.net/>

CORES <http://www.cores-eu.net/>

DCMI Registry <http://dublincore.org/dcregistry/index.html>

DESIRE <http://desire.ukoln.ac.uk/registry/index.php3>

SCHEMAS <http://www.schemas-forum.org/>