(DA3)

Metadata Schema Registry

Version 1.0, MAY 2007

Lead Investigator: Suzanne Little – email N/A

Chief Investigator: Jane Hunter – jane@itee.uq.edu.au

Prepared by: Andrew Newman – anewman@itee.uq.edu.au

Executive Summary

This is a report for the DA3 – "Centralized repository of metadata schemas and application profiles". The work package is part of the DART (Dataset Acquisition, Accessibility and Annotation e-Research Technologies) project.

This work package includes the design and implementation of a metadata schema registry. This is defined as an online electronic repository of schema or application profiles that describe the structure and/or semantics of scientific data for recording and exchange. These registries can be used by researchers to format scientific data sets to improve interoperability and ensure that data can be easily disseminated to researchers within their discipline.

This report covers:

- A brief description of the background, available research, the motivation behind this work package and how it relates to other DART packages,
- The milestones delivered in order to achieve the aims of this work package including discussions of any issues discovered, and
- A description of the architecture of the current system delivered and the interaction with external systems.

Table of Contents

1 Introduction

The purpose of this document is to outline the objectives of the work package SI3, the work that was performed and the outcomes and conclusions from the work.

The DA3 Metadata Schema Registry work package investigates the use of metadata schema registries within eScience. The package objective is to reduce wasted effort on creating metadata schemas and improve interoperability of metadata schemas. This work package will develop and provide access to a centralized repository/registry of metadata schemas, building on the open source software tools being developed within the JISC IE Metadata Schema Registry Project (IEMSR) by UKOLN and ILRT (http://www.ukoln.ac.uk/projects/iemsr/).

The work package aim is to develop software (or use existing open source software) to enable users to create new schemas, submit schemas to the registry and search and browse the registry. This involves:

- Understanding the approach, technologies being developed by JISC IEMSR;
- Working on extensions, refinements, gaps;
- Building prototype tools and a preliminary registry.

2 Project Milestones

The initial milestones for DA3 included:

- Investigation of search mechanisms for schemas,
- Development of Proof of Concept demonstration, and
- Development of Prototype.

The development of this work package was developed by Suzanne Little.

3 Project Outcomes

The outcomes of the project include:

- A storage and retrieval mechanism for metadata schemas,
- A user interface to submit, query, and restrict access to the stored metadata schemas.

3.1 Q1 March 2006 – Investigation of Metadata Search

3.1.1 Summary of Work

Investigated and setup an installation of the JISC IE Metdata Schema Registry (IEMSR) as a prototype for the proposed system. This included:

- A summary of existing schema registries was been written,
- A small collection of example schemas were gathered for use in development,
- A number of usage scenarios and a metadata schema have been written, and
- The architecture and requirements for the DART metadata schema registry (DMSR) was created (see Appendix A).

The key functionality for a prototype implementation was highlighted including:

- Basic user authentication (to act as a placeholder prior to integration with other DART projects),
- Schema validation,
- Submission,
- Metadata capture,
- Storage,
- Browsing, and
- Basic metadata search.

3.1.2 Issues Encountered

There remained questions around around the domains of the schemas to include for example, scientific, cultural or education and which formats to support such as RDF and XML.

3.2 Q2 June 2006 – Development of Proof of Concept

3.2.1 Summary of Work

To articulate the purposes and functionality of the DART metadata registry prototype a usage scenario document was produced (Appendix B). The key functionality required for the DA3 prototype was:

1. Submission (aka: registration, uploading, entering) ;

2. Metadata capture;
3. Verification and validation;
4. Search or query of the schema and metadata;
5. Browse of the schema or metadata;
6. Retrieval (aka: downloading);
7. Updating (aka: editing, (metadata) maintenance);

A number of example scenarios were written to elaborate on the types of uses for the registry and to elucidate the purpose of the online repository in the key DART application domain of climatology.

3.2.2 Issues Encountered

There were no issues highlighted during this milestone.

3.3 Q3 September 2006 – Development of Prototype

3.3.1 Summary of Work

A prototype metadata schema registry was implemented. The prototype was written in PHP, Python and XSLT. It used MySQL, Redland and eXist as backend storage components. Figure 1 shows the architecture diagram for the implementation.

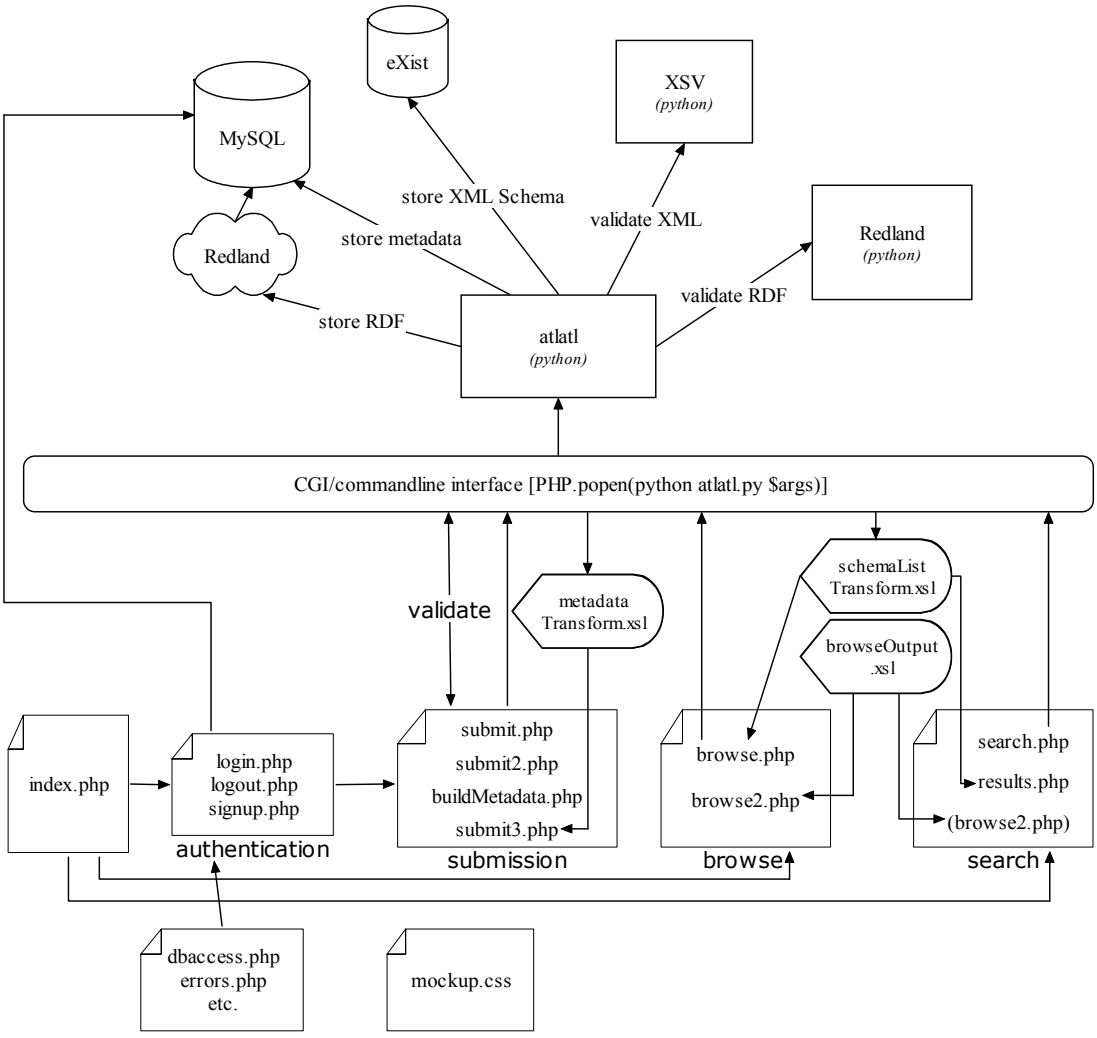

Figure 1 DMSR Prototype Architecture

The prototype demonstrates the functionality, procedures and potential application of metadata schema registries both within the DART project and in the wider field of Australian eScience domains. Users are able to submit schema that describe the format of their scientific data or alternatively they can access the repository to discover authoritative or suggestive standards for structuring scientific data for exchange and reuse. This will improve the ability of researchers to exchange, publish and disseminate scientific research data and therefore increase their capability for distributed collaboration both within their field and across domains.

3.3.2 Issues Encountered

The DMSR prototype requires further work to become a stable, public tool. Evaluation with a set of users is needed to establish the interface requirements and the level of functionality required. Preliminary work on a public version of the prototype designed to be more stable but with limited functionality has been conducted. This implementation could be used to collect example schema for use in evaluation and to gather user feedback on the interface and desired functionality.

4 Technical Requirements

The system comprises of a single component that relies on various systems:

- Apache Web server with mod_python and PHP 5,
- MySQL, and
- eXist XML Database

4.1.1 Apache 2.2.x, mod_python and PHP

In both Windows and Linux deployment an Apache 2.2.x server is required to be installed.

The Apache module, mod_python, is required as is a Python runtime. The atlatl service has been tested with both Python 2.4 and Python 2.5. MySQL for Python is also required in order for the service to retrieve and update the database. XML parsing and validation is performed by the Python library, minixsv.

PHP version 5 was is required to use the web based user interface. Also required is to enable the PHP XSLT extension.

4.1.2 MySQL

The version of MySQL required is version 5.0. This has been installed on both a Windows and Linux based server.

4.1.3 eXist XML Database

The eXist database is used to store and query XML schemas. The version used was eXist 1.1.1 and requires a Java 1.4 environment in which to run.

5 Archival Storage of Project Deliverables

The software and installation instruction for the DA3 work package is available from the eResearch DA3 home page:

http://www.itee.uq.edu.au/~eresearch/projects/dart/outcomes/metadataschemareg.php

This links to the download page for the work package which is at:

http://www.itee.uq.edu.au/~eresearch/projects/dart/outcomes/dmsrDownload.php

It is also stored on DVD for archival purposes.

6 Recommendations

Overall, the DA3 project has summarised the state of existing metadata schema repositories and established the characteristics of a schema repository for the DART project. This project has shown how a functional metadata schema repository can be used within scientific domains to assist researchers in formatting interoperable data for exchange, reuse and dissemination. Three usage scenarios have been written to demonstrate possible applications of the metadata schema registry and illustrate required functionality. A prototype demonstrator has been built, modelled upon the architectural structure of the JISC IEMSR and designed to allow integration with future DART portals or access control mechanisms.

Other future work includes:

- The implementation of alternative or extra functionally including APIs for agent interaction such as SOAP, WSDL or REST.
- More sophisticated browsing interfaces that exploit RDF graphing tools and visualisation of schema relationships.
- Versioning and updating of schema, particularly how to determine if a schema has changed and explicit unique identifiers for different versions of a schema.
- The integration of previous work on harmonising and applying domain ontologies to provide advanced semantic search capabilities.
- The integration of the registry prototype with the proposed DART portal, the existing work using PLONE or related work on authentication and access rights.

A key remaining question is the scope and authority of the schema. This includes determining where the registry would be hosted. Possible suggestions have included: CSIRO, other government authorities, a university or universities or the National Library of Australia. The storage capabilities are not expected to be onerous – although currently the schema are recorded locally to enable efficient searching of the schema content. How to determine if a schema is suitable for a 'scientific' schema repository and establishing the authority of the submitter are also issues.

Also, more work remains to be done on the prototype DMSR:

- Search on schema contents across both XML and RDF Schema,
- Transforming schema and schema fragments for display to and interaction with users (e.g. using XSLT to build a (SVG) graph of the relationships, extending existing RDF graph browsing tools; using GraphVis to visualise RDF relationships),
- Update functionality (verifying user rights, managing versions),
- Bug testing and stabilising the prototype implementation,

- Stabilising the public version on maenad and soliciting user feedback, and
- User testing and collecting user feedback

7 Publications

No publications have so far been created from this work.

8 Terms of Reference

8.1 Glossary

Acronym	Definition
DMSR	DART Metadata Schema Registry
IEMSR	JISC IE Metadata Schema Registry
OWL	Web Ontology Language
RDF	Resource Description Framework
XML	eXtensible Markup Language

9 Report Signoff

It is agreed between

Suzanne Little

and

Prof Jane Hunter

and

Andrew Treloar

That the **Final Report Document** for the DA3 – Metadata Schema Registry gives a full account of the work undertaken for the DART Project.

- has been read and reviewed by all parties,
- shows that the work package DA3 has been completed satisfactorily,
- clearly outlines the functionality that was delivered.

Dated this 1st day of June 2007

Signed by Prof Jane Hunter for and on behalf of the Chief Investigator	Signed for and on behalf of DART by the Project Director Andrew Treloar

10 Appendix A – Survey of Metadata Registries

10.1 Characteristics of Metadata Registries

A *metadata registry* can be defined as a central location where metadata definitions are stored and maintained. Metadata registries are used by a variety of organisations to structure their metadata and enable interoperability, sharing and re-use of data. The registries act as a central source of authorative schemas or vocabularies for use within a domain.

Metadata registries differ in a number of key ways depending on their purpose and application domain.

1. schemas vs vocabularies (and the formats supported)
2. common model vs repository
3. breadth of domain
4. levels of access
5. human and software agents (and the APIs supported)

Metadata registries can be used to store either schemas or vocabularies. Schemas describe the structure of data or metadata. For example, XML Schemas describe the structure of XML documents and database schemas describe the table structure of relational databases. A central resource which contains specifications of schemas in formats such as SQL, UML, XML Schema or as application profiles, enables data creators or recorders within a domain to produce metadata which adheres to a common structure. This data is then more easily integrated.

Metadata registries that store domain vocabularies enable users to produce both structurally standard metadata and use common or standard terminology for metadata fields. Ontologies describe the semantics of a domain and can be used to connect or relate terms. These vocabularies, common terminologies or ontologies can be recorded in formats such as RDF Schema, OWL, DAML+OIL or SKOS.

If a metadata registry is used to store schemas or application profiles, then it is often desirable to relate these schemas through a common data model such as Dublin Core or IEEE LOM. This enables users to search for related terms and to produce metadata that can be integrated through this common model even though the schemas used may be different. Other metadata registries act more as repositories and store schemas or vocabularies without relating them to a common model.

Registries relating schema to a common data model are generally quite specific and restricted to a particular domain. Repositories, which do not use a common model, can be very broad and may store schema from a wide variety of domains. The interfaces to both these styles of registry need to provide users with convenient methods for searching, browsing and querying the stored schema or vocabularies.

Varying levels of access are also generally required by metadata registries. Some registries act as a service provider and do not support public registration of schema

while others allow anybody to submit a schema for registration. Security issues can also be a concern particularly in sensitive or commercial applications.

Registries provide many different services and interfaces. Some of these interfaces are used by human agents to discover relevant schema or vocabularies. These interfaces may be built on advanced query languages such as RDFQL or SPARQL or use syndication/notification services based on RSS. Registries are also accessed by software agents to automatically define terms and discover relationships between metadata objects defined using schema in the registry. APIs (application programming interfaces) can be used for providing web services using standards such as SOAP, REST and WSDL.

This document describes a number of existing metadata registries, the approaches they use and standards they apply. Each registry is summarised according to its administrator, domain, type (schema or vocabulary), APIs used and availability.

10.2 Standards and APIs

10.2.1 Registry Standards

The ISO/IEC 11179 Metadata Registry (MDR) standard is a multipart standard that defines the structure and behaviour of an enterprise metadata registry. The standard describes a hierarchy of "concepts" with associated properties. Each concept and property has a precise "Data element definition". Many existing metadata registries or data models, chiefly from government organisations, have been based on ISO/IEC 11179. Examples include:

- Australian Institute of Health and Welfare - Metadata Online Registry (METeOR)
- Global Justice XML Data Model (JXDM)
- Environmental Data Registry (EDR)
- Cancer Data Standards Repository (caDSR)
- National Information Exchange Model (NIEM)

While the ISO/IEC 11179 standard is large and very comprehensive, there are still groups who are proposing extensions and modifications. One example is the Extended Metadata Registry which is proposed by a consortium of scientific organisations. Their aim is to extend ISO/IEC 11179 "to support more diverse types of metadata and enhanced capabilities for semantics specification and queries."

10.2.2 Schema Languages and Formats

XML: eXtensible Markup Language. A W3C standard, basic data storage format.

XML Schema. XML format for defining the structure (syntax) of XML documents.

RDF: Resource Description Framework. A W3C standard for describing triples - object-property-subject relationships.

RDF Schema. XML format for defining relationships (semantics) in RDF.

DAML+OIL: DArpa Markup Language + Ontology Inference Language. Used to define semantics in ontologies. Richer than RDF Schema.

OWL: Web Ontology Language. A W3C standard XML format for defining ontologies.

SKOS: Simple Knowledge Organization System. A set of three RDF vocabularies for describing "thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary, and perhaps also terminologies and glossaries, within the framework of the Semantic Web."

10.2.3 Access Standards and APIs

RDFQL: RDF Query Language. An earlier query language for RDF.

SPARQL: SPARQL Protocol And RDF Query Language. A W3C standard query language and protocol for querying and access RDF documents.

These are the RDF query languages most often referred to by metadata registries. Further information about RDF query languages can be found in [KAR99].

RSS: Really Simple Syndication. An XML format for describing news feeds. This is useful for providing simple, automatic updates as the contents of metadata registries change.

REST: Representational State Transfer. An architectural style proposed by Roy Fielding for use in web service based applications.

SOAP A W3C protocol for exchanging XML-based messages over a computer network, normally using HTTP. Used as the basis for web services.

10.3 Registries

The section briefly describes some existing metadata registries. Each registry is summarised according to its domain, type, the interfaces it provides and its availability.

10.3.1 Information Environment Metadata Schema Registry (IEMSR)

Administrator: Joint Information Systems Committee (JISC)
Web site: http://www.ukoln.ac.uk/projects/iemsr/
Domain: Education, Digital Libraries
Type: application profiles based on the common model of Dublin Core (DC) or IEEE LOM
APIs/Standards: web search interface,
Availability/Access: open source software; public demo site, submission not available

The IEMSR is preceded by a number of registry projects - DESIRE, SCHEMAS, CORES and MEG. It is based on software from the MEG registry project and aims to extend the work done there to include the IEEE Learning Objects Metadata (LOM) as

a common model for application profiles. A document describing the predicted use cases [IEMSR-usage] is useful for understanding the scope of IEMSR.

10.3.2 SchemaWeb

Administrator: VicSoft (Victor Lindesay)
Web site: http://www.schemaweb.info/
Domain: any
Type: RDF Schemas (RDFS, OWL, DAML+OIL)
APIs/Standards: web search and query interfaces; query via REST, SOAP; RSS feeds
Availability/Access: public registrations, source not available(?)

SchemaWeb was announced on the w3c-rdf-interest list on the 19th of November 2003. It is a large and active repository containing a wide variety of schemas including standards such as OWL and RDF as well as ontologies for beer and music. A variety of interfaces are provided for both human and software agents.

10.3.3 Dublin Core Metadata Initiative (DCMI) Registry

Administrator: Dublin Core Metadata Initiative
Web site: http://dublincore.org/dcregistry/
Domain: Dublin Core
Type: Dublin Core element set and related terms
APIs/Standards: web browse and search interfaces
Availability/Access: just for DC so no public registration, source not available(?)

Provided by the Dublin Core Metadata Initiative solely for use as a reference to the Dublin Core metadata model. It provides usage examples for elements in the models as well as definitions, descriptions, possible refinements and related schema. Multiple views of the underlying model are available in formats such as RDF/XML, N-triple and N3.

10.3.4 Metadata Online Registry (METeOR)

Administrator: Australian Institute of Health and Welfare
Web site: http://meteor.aihw.gov.au/content/index.phtml/itemId/181162
Domain: health, community services and housing assistance
Type: schemas, data standards, data dictionaries. Also support for building new schemas/standards.
APIs/Standards: based on ISO/IEC 11179, download in word or pdf format.
Availability/Access: freely available to search/browse, login to create/submit. Source appears to be unavailable.

Developed to address the need in Australia for a formal approach for communicating national health data for statistical purposes. It was preceded by Knowledgebase and the National Health Data Dictionary.

10.3.5 NSDL Metadata Registry

Administrator: National Science Digital Library (National Science Foundation - US)
Web site: http://eg2.ischool.washington.edu/registry
Domain: digital libraries, science education

Type: vocabularies (SKOS), schemas (XML, RDF), application profiles
APIs/Standards: unclear - still in development
Availability/Access: under-development, grant awarded Oct 2005

Built on both the DCMI registry and CORES, the NSDL Metadata Registry intends to complement the existing NSDL Metadata Repository by providing the means to publish schemas and vocabularies. A useful list of terminology used by the project can be found in [NSDL-terms].

10.3.6 XML.org Registry

Administrator: OASIS
Web site: http://www.xml.org/xml/registry.jsp
Domain: broad range listed inc. finance, eGovernment, health, tax/accounting
Type: XML Schemas and vocabularies
APIs/Standards: web search interface
Availability/Access: registration upon submission, source not available(?)

Appears to be currently inactive as the last submission was in 2001.

10.3.7 DoD Metadata Registry and Clearinghouse

Administrator: US Department of Defence
Web site: http://diides.ncr.disa.mil/taxgal/user/index.cfm
Domain: Military, geographic systems (Global Information Grid)
Type: XML Schemas, taxonomies (OWL)
APIs/Standards: unknown
Availability/Access: restricted access, closed source

Access to parts of the site containing content are restricted to users with an account.

10.3.8 Other Registries

- Distributed Interoperable Metadata Registry - proposes an architecture but no online implementation or demonstrator
- ROADS (Resource Organisation and Discovery in Subject-based services) Metadata Registry - circa 1997

10.4 Commercial Tools

- Oracle Enterprise Metadata Manager
- SAS Metadata Repository
- Data Foundations (consulting service)

10.5 References

[wikiMDR] Wikipedia - Metadata Registry

[KAR99] "RDF Query Languages: A state-of-the-art" Greg Karvounarakis.

[IEMSR-usage] Usage Scenarios for the IE Metadata Schema Registry Joint Information Systems Committee.

[NSDL-terms] Metadata Policy for NSDL

11 Appendix B – Usage Scenarios

11.1 Participants and Terminology

SchemaDocument: the document that describes the schema in XML or RDF schema formats.

User: human user of the system. Has two main tasks: submission of the the SchemaDocument to the Registry or retrieval of information from the Registry.

Registry: the application that manages, verifies and stores the SchemaDocuments and their metadata. Also provides APIs for search, browse, submission and retrieval.

RegistryInterface: web-based interface for uploading and downloading SchemaDocuments, capturing metadata from the User about the SchemaDocument, providing the search/browse interface to the Registry.

Registry components:

- **storage:** for storing the SchemaDocuments and related metadata with search and query capabilities. E.g. IEMSR RegistryServer; eXist; MySQL
- **validation:** a method or application for validating the SchemaDocument. E.g. possibly part of IEMSR RegistryServer capabilities; rosco; W3C's RDF validator; W3C's XML Schema Validator
- **presentation:** (part of the RegistryInterface) methods for transforming the SchemaDocument, its relationships and metadata into formats viewable and browsable by the User. E.g. IEMSR webclient interface; SchemaWeb interface; OWL2HTML

11.2 Metadata

(*Initial thoughts on types of metadata that might be required. Possible sources for metadata terms include: IEMSR, ISO/IEC 11179, SchemaWeb, Dublin Core.*)

Metadata Field	Description	Example
Title	The title of the schema	Crystallography Schema (CIF)
Description	A (short) human readable description of the schema contents and purpose	This XML Schema describes the syntax of CIF files for describing crystallographic structures.
Identifier	A uri of the schema	http://dart.edu.au/dmsr/cif/200606012/

	for the registry	
Format	Currently either XML Schema or RDF Schema	XML Schema
Administrator	The organisation or authority who created and manages the schema. Also record address, phone, email and website details for this organisation.	Name: ACME Co; Address: PO Box 1111, Brisbane 4001; Phone: (07) 3333 4444; Website: http://www.acme.com; email: admin@acme.com
Contact	A nominated contact within the administrating organisation and their contact details	Name: John Smith; email: jsmith@acme.com
Author (alt)	The name and contact details of the schema's creator/administrator where a schema is submitted by an independent individual	Name: John Smith, Organisation: ITEE, UQ; Phone: (07) 3365 4444; email: jsmith@itee.uq.edu.au
Publisher	The organisation which makes the schema available for use	The DART Metadata Schema Registry
Date.Created	The date the schema was created	2006-03-06
Date.Submitted	The date the schema was submitted to the registry	2006-03-15
Date.Updated	The last date any changes were made to the schema or its metadata where no version change was made	2006-05-05
Version	Version number for the schema	1.3
Replaces/ IsReplacedBy	Identifier of schema which this schema replaces or is replaced by	Replaces: http://dart.edu.au/dmsr/20050208/

Coverage	The subject, area or domain to which the schema applies (value is possibly from a controlled vocabulary)	CIF (crystallography) files
ExampleInstance	A url to an example instance of the schema	http://dart.edu.au/dmsr/examples/cif/20060612.xml
ReferenceDocument	One or more urls to descriptive reference documents explaining or providing human-readable definitions of the fields in the schema (in txt, html or pdf format)	http://www.acme.com/cif/schema-description.pdf

11.2.1 Metadata Fields Used by Other Registries

The IEMSR online demo provides the following metadata fields -- Description; Format; Date created; Date modified; Publisher -- when browsing the schemas.

SchemaWeb provides the following metadata fields -- Name; Description; Namespace; Location; Website; Contact Name; Contact Email; Local Version -- when browsing the ontologies/schemas.

METeOR (Aust. Institute of Health and Welfare) provides the following metadata fields -- Metadata item type; Synonymous names; METeOR identifier; Registration status; Definition; Classification structure; Guide for use; Steward; Origin; Reference Documents; Revision Status; Value domains based on this classificiation scheme; Dataset specification type; Submitting organisation -- when browsing classification schemes or data set specifications.

11.3 Tasks

These are the tasks that need to be supported by the MSR and the participants in each one:

1. submission (aka: registration, uploading, entering) [User, Registry]
2. metadata capture [User, Registry]
3. verification and validation [Registry]
4. search or query [User, Registry]
5. browse [User, Registry]
6. retrieval (aka: downloading) [User, Registry]
7. updating (aka: editing, (metadata) maintenance) [User, Registry]

11.4 Example Scenarios

A metadata schema registry for the climatology field has been developed and implemented. It contains a dozen schemas in both XML and RDF schema formats. These schemas are used to structure data records from a variety of specific domains related to the study of climatology including water quality testing; meteorology reports; satellite imagery of vegetation density and distribution etc. The registry is available on the web and is used by researchers and administrators from organisations such as: CSIRO; University of Queensland; the Queensland Department for Environment; the Environmental Protection Agency; the CRC for Greenhouse Accounting etc.

11.4.1 Scenario 1

A team working for the CRC for Greenhouse Accounting has developed a schema for recording the level of air-born pollutants in an atmospheric sample. They wish to submit this schema, which is in XML Schema format, to the climatology metadata registry for use by other researchers in the field.

Possible Process:

1. User goes to submission page on RegistryInterface website
2. User fills in form to provide metadata related to the SchemaDocument
3. User selects SchemaDocument location from local machine (or supplies a URL for SchemaDocument location)
4. User pushes submit button
5. RegistryInterface checks that all compulsory elements of the form have been completed (Error is returned to the User if the form is incomplete)
6. RegistryInterface formats the metadata and passes the SchemaDocument (or its URL) and the metadata document to the Registry
7. Registry validates the SchemaDocument (Error is returned to the User if it's invalid)
8. Registry saves the SchemaDocument and metadata document and associates metadata document with SchemaDocument
9. Registry passes success message to RegistryInterface
10. RegistryInterface informs the User that submission has been successful and provides website address for browsing the submitted SchemaDocument

11.4.2 Scenario 2

A researcher from the University of Queensland wishes to make their research data from water testing at the Indooroopilly stretch of the Brisbane River available to other researchers in the climatology field.

They need to find the appropriate schema for recording and structuring water analysis results (e.g. water temperature, oxygen levels, salt levels, pollutants etc.)

Possible Process:

1. User goes to search page of the RegistryInterface website
2. User selects to search for "water" in the domains field
3. RegistryInterface constructs query and passes it to the Registry
4. Registry returns list of matching SchemaDocuments
5. RegistryInterface retrieves metadata about SchemaDocuments from Registry
6. RegistryInterface provides list of SchemaDocuments and selected metadata to the User
7. User browses SchemaDocuments until finding the appropriate one for their requirements
8. User downloads the SchemaDocument

11.4.3 Scenario 3

The researcher from scenario 2 finds the schema for water testing results but it doesn't have enough detail on recording the level of pollutants. Further searches discover the air pollutant schema registered by the research team from scenario 1. The researcher is able to develop an updated version of the water quality schema which includes fields for recording pollutant levels based on the structure used for air pollutants. The researcher registers this new schema as an updated, alternative version of the original water quality schema.

Possible Process:

1. *continuing from scenario 2's possible process step 7*
2. User examines SchemaDocument but is dissatisfied with fields for recording pollutant levels
3. User returns to search pages of the RegistryInterface website
4. User searches for all SchemaDocuments which have the field 'pollutant'
5. RegistryInterface constructs query and passes it to the Registry
6. Registry returns list of matching SchemaDocuments
7. RegistryInterface retrieves metadata about SchemaDocuments from Registry
8. RegistryInterface provides list of SchemaDocuments and selected metadata to the User
9. User finds the schema submitted in scenario 1 and decides that the data recorded about pollutants is better in that schema
10. User downloads both the original water schema and the new air pollutant schema
11. User incorporates sections of the air pollutant schema into the original water schema

12. User follows process described for scenario 1 to submit the new schema and includes that this is an extension of the original water schema