...a national focus of expertise in digital information management...

Dr. Manjula Patel
Technical Research and Development
m.patel@ukoln.ac.uk
SCHEMAS
Forum for Metadata Schema Implementers

Partners
UKOLN
PricewaterhouseCoopers
Fraunhofer-Gesellschaft
Outline

Terminology
Motivation for SCHEMAS
Schema registries
SCHEMAS Forum Registry
RDF and RDF Schemas
Namespaces and Application Profiles
Example: BIBLINK
Interoperability
Summary
Beyond SCHEMAS
Terminology

Metadata is
- structured data about data
- a form of language (pidgin)

A metadata vocabulary or schema:
- declares a set of concepts or terms and their associated definitions
- the terms are often known as elements
- the definitions provide the semantics, ideally both human and machine readable
- in effect a manifestation of an ontology

A scheme:
- controlled vocabulary or enumerated type
Schemas and ontologies

Conceptual Ontological Model

Metadata Vocabulary

- Human-readable text encoding
- Machine-readable XML encoding
- Machine-readable RDF encoding
- Web readable HTML encoding
SCHEMAS motivation

Help implementers understand the evolving landscape of metadata schemas
  • Overviews of standards development
  • Lists and reviews of schema-creating activities

Standards-based registry (in RDF)
  • Harvest metadata schemas from maintainers
  • Provide integrated access to this distributed corpus of metadata vocabulary terms

Encourage use of existing and emerging standards
The Semantic Web

“The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”


Semantic Web:

• requires machine readability
• highlights importance of metadata schemas in inferencing
The Semantic Web

SCHEMAS Registry

Longer-term goal

Some day... maybe...

Self-desc. doc.

Data

Rules

Proof

Trust

Digital Signature

Ontology vocabulary

XML + NS + xmlschema

RDF + rdfschema

Unicode

URI
Metadata schemas lack a standard, machine-processable publication context for schemas

- Who is defining vocabulary terms?
- Who is using (or adapting) them in projects?
- Who is reviewing, recommending, or certifying those terms?

Comprehensive dictionaries

- Which terms are related to which?
- How are related terms defined?
Schema registries

Add value by providing access to schemas
- search, browse, terms, definitions, usage
- relationships between terms and vocabularies
- crosswalks and mappings
- links to sample metadata instances
- links to annotations
- links to implementations

Support evolution of schemas
- top-down (standards authorities)
- bottom-up (real world usage)

Effective re-use
- inform and promote vocabularies
- provide an authoritative version
- facilitate inter-working
- align with other schemas
Ontology harmonisation

• Knowledge of how other projects are using standards will help avoid duplication of effort
• To help information providers harmonise their schemas for improved interoperation
  • across domains
    e.g. for resource discovery (Dublin Core)
  • in specific domains
    e.g. Education (DC-Ed, IMS), Repositories for preprints (OAi), Theses and Dissertations (NDLTD), Collection Level Descriptions (RSLP), Subject Gateways (Renardus), Mathematics and Physics (MathNat, PhysNet)
• Particular applications or implementations
  e.g. BIBLINK, MathNet, TrialSolution, OAi protocol, EULER, GELOS, GILS, EIONET
SCHEMAS motivation

• Support for schema developers and implementers
• Disclosure or publication environment for vocabularies
• Enable queries across a whole range of schemas
• Clarify relationships between vocabularies
• Encourage sharing of existing vocabularies to help avoid duplication of effort
• Encourage convergence and harmonisation within single domains
• Promotion of standards to improve potential for cross-domain interoperability
Registry architectures

Thick registries (centralised)
- central repository, heavy maintenance burden
- e.g. ISO/IEC 11179 based registries (e.g. Environmental Protection Agency, Australian Health Information Knowledgebase), Dublin Core Metadata Initiative (DCMI), DÉSIRÉ, MetaForm

Thin registries (distributed)
- content and maintenance is distributed, based on a harvesting model
  - push model e.g. European SchoolNet Project (ETB)
  - pull model e.g. SCHEMAS HTML and RDF registries
SCHEMNAS Registry architecture (distributed)
EOR Toolkit

- EOR (Extensible Open RDF) Toolkit: (http://eor.dublincore.org - Eric Miller, OCLC)
- Harvests RDF schemas from distributed servers on the web
- Creates central index for searching
- Schema browser - hyper-links not only between vocabularies, but between related terms
Registry content

- vocabularies (various representations, RDDL)
- metadata watch reports
- standards activity reports
- reviews
- best practice guidelines
- pointers to projects and information about them
- other …guided by user requirements
Resource Description Framework (RDF)

A shared grammar is needed to ensure that humans and software will interpret metadata consistently

- A grammatical **framework** for the **description** of **resources** (Resource Description Framework)
- A language for making statements about properties and relationships of items on the Web
- Basic model for making statements about:
  - **Resource**: anything named with a URI
  - **Description**: stating the properties of the resource using terms named by URIs
  - **Framework**: a common model (grammar) for statements using diverse vocabularies
- Uses XML as serialisation syntax
RDF model & syntax

Expresses the statement:

“The **SCHEMAS Project** is the publisher of the resource which is identified by **http://www.shemas-forum.org/**”
RDF Schemas (RDFS)

- Web-based publication format for declaring semantics
- W3C Recommendation
- Has capability to explicitly declare semantic relations between vocabulary terms
- Machine readable, but also defines properties and classes with human readable labels and comments

Example:

Title - an element from the Dublin Core Element Set

Title: A name given to the resource (defn for humans)

dc:title (unique identifier for machine processing)
Registry Data Model

Schema

- NsSchema
- ApSchema
- WebPage Schema

Document

- MD watch
- Guidelines
Namespaces and Application profiles

Namespace schemas:
Declare a unique set of elements and definitions
- In SCHEMAS, by definition only declare names and definitions of vocabulary terms
- Provide authoritative declaration of definitions
- Indicate semantic relationship between terms
- Support unique identifiers for terms

Application profiles (APs):
Declare which vocabulary terms a particular application or project uses
- In SCHEMAS, a Profile by definition only reuses terms defined somewhere in a namespace
- Data elements drawn from one or more namespaces, optimised for a particular local application
  - may specify dependencies e.g. mandate schemes
  - may adapt existing definitions for local purposes
  - may declare rules for content (usage guidelines)
  - may specify whether an element is mandatory, optional or repeatable
Publication of schemas

• If publish profiles and namespaces machine-processably
  – Usable by search engines and other software agents
  – Annotation capability supports Web of Trust: audit trails about who says what about what

• Use of open standards enhances interoperability (inter-communication between registries, portals, tools etc.,)
**BIBLINK vocabulary**

BIBLINK vocabulary (uses DC and BC)

**DC:** title, contributor, identifier, publisher, format, date, subject, description, language, rights, source

**BC:** creator organisation, contributor organisation, checksum, frequency, edition, place of publication, system requirements

**Qualifies:** title (title alternate), format (extent)

**Schemes:** identifier (URI, URL, DOI, ISBN, ISSN, SICI), subject (LCSH, DDC, UDC, LCC)
# Namespaces

<table>
<thead>
<tr>
<th>Item #</th>
<th>Title</th>
<th>Identifier</th>
<th>Description</th>
<th>Date</th>
<th>Creator</th>
<th>Contributor</th>
<th>Publisher</th>
<th>Type</th>
<th>Language</th>
<th>Relation</th>
<th>Format</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The BIBLINK Vocabulary v1.0</td>
<td><a href="http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc">http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc</a></td>
<td>The BIBLINK vocabulary supports the BIBLINK project which aims to link publishers and National Bibliographic Agencies</td>
<td>2000-08-22</td>
<td></td>
<td></td>
<td></td>
<td>Namespace</td>
<td>en-uk</td>
<td></td>
<td>RDF/XML</td>
<td><a href="http://www.schemas-forum.org/registry/schemas/SHEMAS/1.0/smaes#NsSchema">http://www.schemas-forum.org/registry/schemas/SHEMAS/1.0/smaes#NsSchema</a></td>
</tr>
<tr>
<td>2</td>
<td>The RSLP CLD Vocabulary v1.0</td>
<td><a href="http://www.schemas-forum.org/registry/schemas/RSLP-CLD/1.0/cld">http://www.schemas-forum.org/registry/schemas/RSLP-CLD/1.0/cld</a></td>
<td>The RSLP CLD 1.0 Schema provides a set of metadata attributes for describing collections within the RSLP. It is based on An Analytical Model of Collections and their Catalogues developed previously in the project</td>
<td>2001-03-02</td>
<td></td>
<td></td>
<td></td>
<td>Namespace</td>
<td>en-uk</td>
<td></td>
<td>RDF/XML</td>
<td><a href="http://www.schemas-forum.org/registry/schemas/SHEMAS/1.0/smaes#NsSchema">http://www.schemas-forum.org/registry/schemas/SHEMAS/1.0/smaes#NsSchema</a></td>
</tr>
<tr>
<td>3</td>
<td>The SCHEMAS Vocabulary v1.0</td>
<td><a href="http://www.schemas-forum.org/registry/schemas/SCHEMAS/1.0/smaes">http://www.schemas-forum.org/registry/schemas/SCHEMAS/1.0/smaes</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
XML Namespace mechanism

...can be used to pull in multiple namespaces
BIBLINK namespace

define a new term and associated semantics ...

BIBLINK vocabulary:

   Edition: A statement indicating the version or edition of the resource

   <rdf:Property ID = "edition" >
   <rdfs:label> Edition </rdfs:label>
   <rdfs:comment>
       A statement indicating the version or edition of the resource
   </rdfs:comment>
   <rdfs:isDefinedBy
       rdf:resource =
           "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc"/>
</rdf:Property>
BIBLINK namespace

```xml
<rdf:Property ID="extent">
  <rdfs:label>Extent</rdfs:label>
  <rdfs:comment>The size of the resource in bytes, number of files or CD-ROMs</rdfs:comment>
  <rdfs:subPropertyOf "http://purl.org/dc/elements/1.1/format"/>
  <rdfs:isDefinedBy "http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc"/>
</rdf:Property>

...declare a local qualifier for dc:format

Dublin Core: Format: The format of the resource
BIBLINK: Extent: The size of the resource in bytes, no. files, or CDROMS
BIBLINK namespace define a scheme for identifier …

BIBLINK vocabulary: identifier scheme: URI, URL, DOI, ISBN, ISSN, SIC

```xml
<rdfs:Class rdf:ID="IdentifierScheme">
  <rdfs:label> Identifier Schemes </rdfs:label>
  <rdfs:comment> A set of identifier encoding schemes and/or formats </rdfs:comment>
  <rdfs:isDefinedBy rdf:resource="http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc"/>
</rdfs:Class>

<rdfs:Class rdf:ID="URI">
  <rdfs:label> URI </rdfs:label>
  <rdfs:comment> Uniform Resource Locator </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#IdentifierScheme"/>
  <rdfs:isDefinedBy rdf:resource="http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc"/>
</rdfs:Class>
```
## Application profiles

<table>
<thead>
<tr>
<th>Item #</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>The BIBLINK Application Profile v1.0</td>
</tr>
<tr>
<td>Identifier</td>
<td><a href="http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be-ap">http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be-ap</a></td>
</tr>
<tr>
<td>Date</td>
<td>2000-08-22</td>
</tr>
<tr>
<td>Creator</td>
<td><a href="http://www.schemas-forum.org/registry/Person/ManjulaPatel">http://www.schemas-forum.org/registry/Person/ManjulaPatel</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>The SCHEMAS Project</td>
</tr>
<tr>
<td>Type</td>
<td>Application Profile</td>
</tr>
<tr>
<td>Language</td>
<td>en-uk</td>
</tr>
<tr>
<td>Relation</td>
<td><a href="http://hosted.ukoln.ac.uk/biblink/">http://hosted.ukoln.ac.uk/biblink/</a></td>
</tr>
<tr>
<td>Format</td>
<td>RDF/XML</td>
</tr>
</tbody>
</table>

### isProfileOf
- http://hosted.ukoln.ac.uk/biblink/wp8/fs/be- semantics.html
- http://purl.org/dc/elements/1.1/description
- http://purl.org/dc/elements/1.1/identifier
- http://purl.org/dc/elements/1.1/subject
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#placePublication
- http://purl.org/dc/elements/1.1/language
- http://purl.org/dc/elements/1.1/title
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#systemRequirements
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#edition
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#creatorOrganisation
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#contributorOrganisation
- http://purl.org/dc/elements/1.1/creator
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#price
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be#checksum
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be-ap#genid12
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be-ap#genid13
- http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/be-ap#genid14
- http://purl.org/dc/elements/1.1/rights
BIBLINK application profile

declare reuse of terms from multiple vocabs ...

From Dublin Core:


From BIBLINK:

- `<smes:uses rdf:resource="http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc#price"/>
- `<smes:uses rdf:resource="http://www.schemas-forum.org/registry/schemas/BIBLINK/1.0/bc#extent"/>
BIBLINK application profile

<smes:uses>
  <rdf:description>
    rdf:about = "http://purl.org/dc/elements/1.1/title">
  <smes:comment>
    The title of the publication
  </smes:comment>
  </rdf:description>
</smes:uses>

...adapt the definition of dc:title

Dublin Core:  Title: The title of the resource
BIBLINK :  Title: The title of the publication
BIBLINK application profile

mandate a scheme with a specific term ...

BIBLINK vocabulary:
    identifier scheme: URI, URL, DOI, ISBN, ISSN, SICI

<smes:uses>
  <rdf:description
      rdf:about = "http://purl.org/dc/elements/1.1/identifier">
    <rdfs:range
      rdf:resource="bc:IdentifierScheme"/>
    <rdfs:domain
      rdf:resource="http://purl.org/dc/elements/1.1/identifier"/>
  </rdf:description>
</smes:uses>
Interoperability

Broad-brush interoperability entails:

- **Partial** understanding (Semantic Web)
- Shared grammar for **simple statements** (RDF)
- **Core** vocabularies, **pidgin** metadata (e.g., Dublin Core)
- Broadly understood conventions for publishing metadata vocabularies
Issues: XML & RDF

XML (lacks underlying data model)

RDF (lacks explicit data typing, structuring and constraint modeling)

DAML-OIL
WebOnt WG
RDFcore
Summary

• Forum for discussion of metadata schemas
• Described landscape of metadata activities and standards
• Developed notion of application profiles
• Construction of a knowledge base of vocabularies
• Clarified the issues around interoperability
• Use of RDF to enhance cross-domain interoperability as well as convergence within domains
Beyond SCHEMAS

- Register core vocabularies and profiles
- Identify and annotate activities
- Maintain and enhance registry
- Help develop core profiles
- Bring together metadata activities
- Best-practice and guidance material
- Awareness and dissemination
Selected references

Tim Berners-Lee, James Hendler and Ora Lassila, 
*The Semantic Web*, Scientific American, May 2001

Rachel Heery & Manjula Patel, *Application Profiles ...*
http://www.ariadne.ac.uk/issue25/app-profiles/


Jane Hunter, *Combining RDF and XML Schemas....*

Thomas Baker, *A Grammar for Dublin Core*
Dlib Magazine, 6(1)) October 2000

*The SCHEMAS Registry*
http://www.schemas-forum.org/registry/

*Resource Description Framework Schema Specification*
http://www.w3.org/TR/PR-rdf-schema/
...a national focus of expertise in digital information management...

Dr. Manjula Patel
Technical Research and Development
m.patel@ukoln.ac.uk
http://www.ukoln.ac.uk/