UNIVERSITY OF
BATH

Document Version
Publisher's PDF, also known as Version of record

Link to publication

University of Bath

Knowledge Discovery in an Agents Environment

Manjula Patel and Monica Duke, UKOLN

European Semantic Web Symposium 2004

10-12th May 2004, Heraklion, Crete

Manjula Patel
Research and Development
UKOLN, University of Bath, UK

UKOLN is supported by:

MUSEUMS LIBRARIES ARCHIVES

JISC

UNIVERSITY OF BATH

UKOLN

Overview

- Motivation
- Background work
- Ontologies and metadata vocabularies
- A meta-model for metadata vocabularies
- Knowledge repository
- Internal data model
- Architectural components
- Interactive web interface
- Agents interface
- Issues and further work

Motivation

- **Proliferation** of metadata schemas (developed in closed environments, tower of Babel situation for applications)

- **Disclosure** –adaptations by practitioners and communities of practice not readily and openly available

- **Investigation** of individual terms as well as whole vocabularies for adaptations, local usages and relationships with other vocabularies

- **Interoperability**
  - **Harmonisation** of semantics and usage
  - **Convergence** within specific domains e.g. education, cultural heritage, publishing, rights management etc.

- **Reasoning and inference** -automated querying of metadata vocabularies by software agents to acquire the semantics associated with specific terms

Background work at UKOLN

UKOLN has been involved in several projects in the area of metadata vocabulary repositories or knowledge bases (mainly from the perspective of digital libraries):

DESIRE II (1998-2000) –interactive browsing by users
http://www.ukoln.ac.uk/metadata/desire/

SCHEMAS (2000-2002) –machine processible format
http://www.ukoln.ac.uk/metadata/schemas/

CORES (2002-2003) –annotation service
http://www.ukoln.ac.uk/metadata/cores/

Agentcities.NET (2002-2003)-deployment in an agents environment
http://www.ukoln.ac.uk/metadata/agentcities/

DCMI Registry (on-going) –management of DC vocabulary
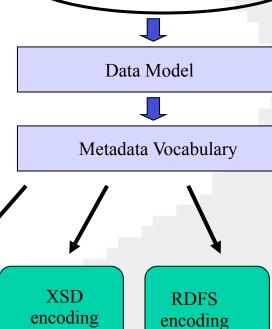http://dublincore.org/dcregistry/index.html

MEG Registry (on-going) –UK Education domain
http://www.ukoln.ac.uk/metadata/education/regproj/

Ontologies and metadata vocabularies

- Parallel developments in the computer science and digital library worlds

- Conceptual model identifies domain of discourse (knowledge level)

- Data model is a concrete formalisation of a view onto the abstract model

- A metadata vocabulary or schema:
  - declares a set of concepts or terms and their associated definitions and relationships
  - terms are often known as elements, attributes and qualifiers
  - definitions provide the semantics
  - in effect a light-weight form of ontology

- Implementation is an encoding –ideally human-readable and machine-processable

Conceptual or Abstract Model

Data Model

Metadata Vocabulary

Natural Language

XSD encoding

RDFS encoding

HTML encoding

UKOLN

Types of vocabularies

Vocabularies range from international standards to implementation specific schemas

- Single element sets
- Combinations of vocabularies
- Cross-domain (Dublin Core)
- Specific domains (IEEE LOM/IMS, OAIS, CIDOC CRM, MPEG, INDECS)
- Particular applications or implementations (Open Archives (OAi-PMH), Government Information Locator Service (GILS), MathNet)

A meta-model for metadata vocabularies

- Becoming widely adopted in the DL world
- Encourages **modular** organisation of knowledge
- Distinguishes *where* and *how* terms are defined as opposed to how they are *used* and *adapted* in practice

Element sets: **declare a unique set of elements and definitions**
- Provides semantic knowledge for reuse
- Ideally, addressed on the Web with a URI
- May be expressed in XSD, RDFS etc.

Application profiles: **declare which terms are used by a particular application or project**
- may mix-and-match terms from multiple element sets (not other application profiles)
- 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 term is mandatory, optional or repeatable

UKOLN

A knowledge base of vocabularies and ontology server

- Enables individual terms as well as whole vocabularies to be explored
- Data mining –analysis of patterns of metadata usage
- Potential creation and inference of new information
- Essential for support of ontological engineering process
- An ontology server
  - Agentcities.NET project
  - Deployment grant: Sept 2002-Feb 2003
  - Technical Report: An Ontology Server for Agentcities.NET
    http://www.agentcities.org/note/00008

Repository contents

- Metadata vocabularies or ontologies
- Contextual information relating to vocabularies
- Notion of *Element Sets* and *Application Profiles* as basis for vocabulary encodings
- Specification language currently used is RDF Schemas
- Classes or entities recorded (based on findings in SCHEMAS):
    - Agency
    - Element Set
    - Application Profile
    - Element
    - Element Usage
    - Encoding Scheme
    - Values in controlled vocabularies

Internal data model

- A model for describing the structure of metadata vocabularies
- A normalisation model which lies above the RDFS layer
- A means for integrating multiple vocabularies into a common knowledge base:
  - **Element Sets** are owned and maintained by **Agencies**
  - **Element Sets** are made up of **Elements**
  - An **Element** is a term defined in order to describe a characteristic or attribute of a resource
  - An **Element Usage** may:
    - introduce constraints on the value of an Element by associating it with one or more **Encoding Schemes**
    - introduce constraints on the obligation to use an Element (e.g. make its use mandatory) or the occurrence of an Element (e.g. whether it is repeatable)
    - refine the semantic definition of an Element to make it narrower or more specific to the local application domain
  - **Encoding Schemes** constrain the value space of **Elements**
  - An **Application Profile** defines a set of **Element Usages** of **Elements** drawn from one or more **Element Sets**

UKOLN

Information recorded

The repository holds information on each of the following entities and their relationships:

- – **Element Sets**: intended scope; area of use; relationship to other Element Sets
- – Elements: semantic definitions; recommended usage; relationship to other Elements
- – **Application Profiles**: intended scope; area of use; relationship to other Element Sets
- – **Usages of Elements**: the Element used; any prescription of Encoding Schemes; other constraints on element use
- – **Encoding Schemes**: intended scope; area of use; where an Encoding Scheme takes the form of an enumerated list, the **values** prescribed by that Encoding Scheme may be recorded
- – **Agencies**: who owns, creates, maintains Element Sets, Application Profiles and Encoding Schemes
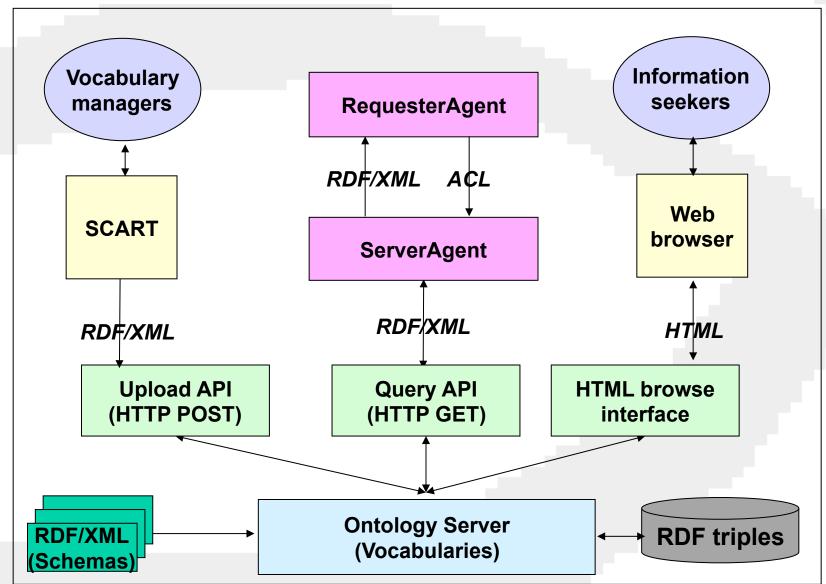
Architecture

- Centralised

  e.g. ISO/IEC 11179 based registries (Environmental Protection Agency, Australian Health Information Knowledgebase), DESIRE Registry,

  Dublin Core Metadata Initiative(DCMI)

  - management of single vocabularies
  - multiple language translations
  - support standardisation processes

- Decentralised

  e.g. SCHEMAS, MEG and CORES

  - multiple vocabularies
  - content and maintenance of vocabularies is decentralised
  - evolution of vocabularies is devolved to those committed to their development
  - based on a harvesting model
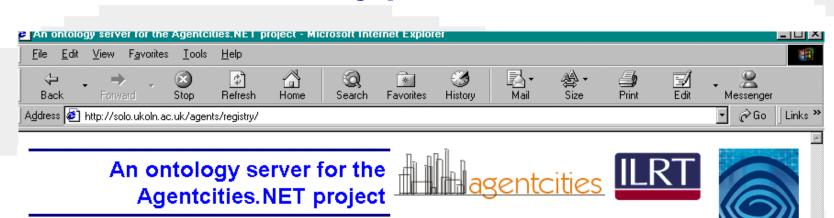
UKOLN

Architectural components

Interactive interface

- Support for ontological engineering (discovery, disclosure, reuse, harmonisation)
- Disclosure or publication environment for vocabularies
- Enable queries across a range of vocabularies
- 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
- Web-interface:

  http://agentcities.ukoln.ac.uk/server/

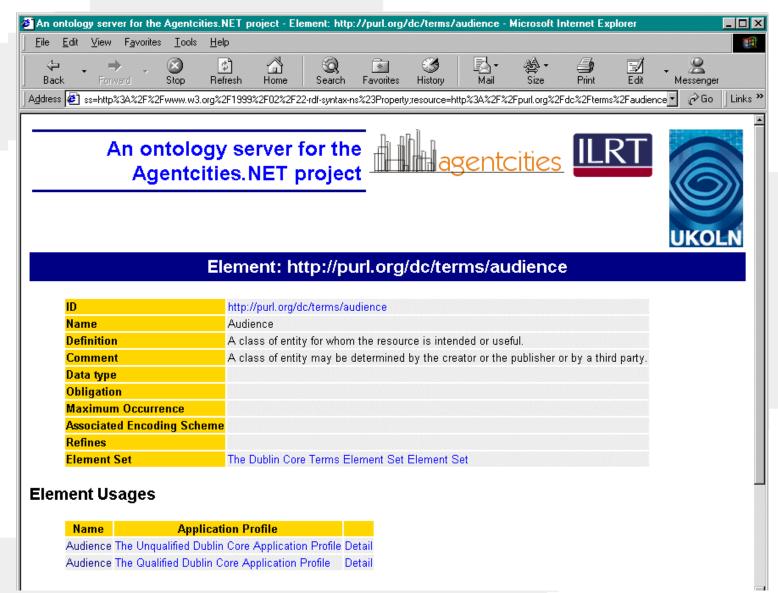# *Web interface: Entry point*

Web interface: Element sets

Web interface: DC audience term

Agents interface

- Agents interface allows software agents to query, search and navigate metadata vocabularies
  - Enables retrieval of semantics
  - Provides potential for inference and reasoning tasks
- Essential infrastructure for the composition and coordination of automated services over the Web
- Java Agent Development Environment (JADE)
  - Based on FIPA standards for intelligent software agents
  - Communication via an ontology
  - Requests in Agent Communication Language (ACL)
  - Results returned in INFORM slot of return message
- Two examples of *RequesterAgents*
  - GUI
  - Command line

Agents interface: GUI

Example: agent query and result

Example search for term "audience":

```
(action
 (agent-identifier :name UKOLNServer@agentcities.ukoln.ac.uk:1099/JADE)
 (ReturnSearchResults (Search :Scope element :SearchTerm audience))
 )
```

Example result in RDF:

```
<rdf:Description rdf:about="http://purl.org/dc/terms/audience">
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
ns#Property"/>
    <rdfs:label>Audience</rdfs:label>
    <rdfs:comment>A class of entity for whom the resource is intended or
                  useful.
    </rdfs:comment>
    <reg:useComment>A class of entity may be determined by the creator
                    or the publisher or by a third party.
    </reg:useComment>
    <reg:isElementOf rdf:resource="http://www.ukoln.ac.uk/metadata/
education/regproj/reg/elementSet/dcterms"/>
</rdf:Description>
```

Issues and further work

- Version tracking: metadata vocabularies evolve over time:
  - Semantics change
  - New terms need to be added or deleted
  - Agency information binds versions together
- Vocabulary Data Models
  - data model implicit in a schema encoding
  - differing data models are a challenge to reconcile!
- Annotation is important for trust, authority, provenance
- Rights management is important for sharing and reuse
- Investigation of interaction with external FIPA agents
- Semantic search and inference of new information

Conclusions

- Contents of repository are simple forms of ontology as well as contextual information
- Vocabularies need to adhere to meta-model described and use RDFS for encoding
- Need for a service which enables discovery and disclosure of semantics used in applications
- Provides information regarding how terms are defined, how standards are used in practice and how terms and vocabularies are related to each other
- Essential part of infrastructure required to enable exchange and reuse of semantic information
- Support for ontology engineering –semantic interoperability requires domain-level consensus on the structure, concepts and terminology to be used in knowledge representation

UKOLN

Acknowledgements

DESIRE Project

SCHEMAS Project

MEG Registry Project

Agentcities.NET Project

UKOLN

Thank you!

Manjula Patel
Research and Development
UKOLN, University of Bath, UK
m.patel@ukoln.ac.uk

http://www.ukoln.ac.uk/

UKOLN