

(Metadata and) Vocabulary Registries

Rachel Heery, UKOLN, University of Bath
r.heery@ukoln.ac.uk



UKOLN is supported by: **JISC** 
MUSEUMS LIBRARIES ARCHIVES



www.ukoln.ac.uk



a centre of expertise in digital information management

Contents

- Why are vocabulary registries needed?
- Registries as part of a digital infrastructure

- What should they contain
- How they might be used

- Overview of current initiatives

Metadata Schema Registry

- Application that provides services based on information about "metadata terms" (and related resources)
- "Metadata term" = "unit of meaning" deployed in metadata descriptions
- Functions might include
 - Disclosure/discovery of information about "terms"
 - Verification of provenance/status of "terms"
 - Discovery of relationships between "terms"
 - support for mapping, inferencing
 - Pointers to related resources
 - usage in metadata application profiles, guidelines for use, bindings
- Support for services to human readers, software agents

My perspective

- DESIRE, SCHEMAS, CORES, MEG Registry (2000-2004)
 - DC application profiles
 - Working towards m2m interface
- JISC Metadata Schema Registry (2004 ongoing)
 - DC application profiles
 - IEEE LOM application profiles
 - Schema creation tool, registry server, web interface
 - Query API (SPARQL)
 - <http://www.ukoln.ac.uk/projects/iemsr/>
- DCMI Registry (2001 ongoing)
 - DC vocabularies
 - Web site for navigation, query API
 - Multilingual translations



'Metadata vocabularies' and 'KOS vocabularies'

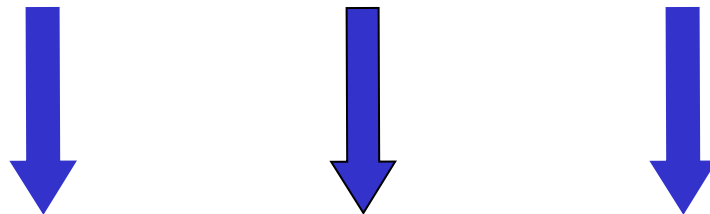
- Differences between 'metadata element sets' and 'subject vocabularies'
 - Relationship between terms, use cases, communities
- Different standards, different conventions

.... but complementary

- Contribute to same 'business processes' e.g. enterprise portal, records management, resource discovery
- Contribute to same workflows and choreographed services

Common concerns

- Proliferation of vocabularies
- Requirement for extension and localisation
- Potential to build 'programmatic interfaces', APIs, web services



Opportunities for re-use, data exchange, innovative interfacing with applications from 'other domains'

Interoperability between vocabularies

- Common approach to syntax
- Common approach to semantic modelling i.e. common structural relationships between entities
- Availability of authoritative versions of vocabularies
- m2m access to vocabularies

Layers of interoperability

Layer 3	(a) Attribute Space (e.g LOM, Dublin Core MES, indecs)	(b) Value Space (e.g. ontologies, classifications, controlled vocabularies, taxonomies)
Layer 2	Representation (e.g. XML, RDF, DAML-OIL)	
Layer 1	Transport and Exchange (e.g HTTP Get, OAI Protocol for Metadata Harvesting)	

Table 1

Baker, T. et al. *Principles of metadata registries*. A white paper of the DELOS working group on registries, 2002.

Delos-noe.iei.pi.cnr.it/activities/standardizationforum/registries.pdf



Moving from declaring terms to building registries

- Machine readable structured vocabulary



- Machine readable vocabulary with ‘added value service(s)’



- ‘Well managed, trusted Registry’
 - Policies
 - Status, persistence, identification, quality, standards compliant
 - Rich services

Registry services usage scenarios (1)

Registry services can focus on

- An individual vocabulary– providing authoritative information on a particular vocabulary
- Extensions of a vocabulary – providing information on how a standard has been extended and localized by communities of use
- Data warehouses – storing definitions of data elements and data types for the purpose of interworking between databases
- Usage within domains – providing access to schemas of interest to a particular domain such as education, cultural heritage, or commerce



Registry services usage scenarios (2)

- Metadata functions – providing access to schemas of use for particular tasks, such as resource discovery, digital rights management, or user profiling
- Corporations— providing access to knowledge frameworks or taxonomies used in enterprise portals or corporate intranets
- Application-based – providing schemas available in a particular syntax or format for use in specific software applications
- Mappings and conversions – providing services for translating metadata between different metadata systems

Types of registry ?

- ‘Shallow registry’ (thin registry)– metadata about vocabularies linking to structured vocabulary instances

In contrast to:

- Building ‘registry services’ : added value services built on structured vocabularies
- Note: location of vocabulary storage is design decision, might be centralised, federated, distributed

Characteristics of 'registry services'

- Provides 'programmable access' (APIs, web services)
- Uses open standards
- Common structure of data:
 - Vocabulary registry: Zthes, SKOS, MARC
 - Metadata registry: RDFS, OWL...

Business issues

- What is the 'ROI'? Will funders and service providers save money? Will users (service providers and end users) get improved services?
- Who is the obvious funder for a 'shared' service?
- Practicalities...
 - Who owns content?
 - Who is responsible for transforming content to m/readable structure? And maintaining...
 - Is there commercial motivation for KOS owners to 'work together'?

Technical issues

- Services need to be built on a standard data - requires consensus on data model
 - Note application profiles still have no formal model
- Can a single registry deal with more than one underlying model ?
- Large vocabularies have commercial business model – issues around managing IPR and copyright
- Scalability???



Conclusion

Most potential for reaching consensus and building services where

- Small similarly structured vocabularies
- Liberal licensing

Thin registry activity

- NKOS Registry plans Draft Set of Attributes developed by Linda L. Hill and Interconnect Technologies (last revision: 7/30/98)
- Traugott Koch, Controlled vocabularies, thesauri and classification systems available in the WWW
<http://www.lub.lu.se/metadata/subject-help.html>
- Taxonomy warehouse
<http://www.taxonomywarehouse.com/>
- Who compiles?
- Inclusion criteria (available on web, quality etc)... who decides?
- Maintenance...

Becta Terminology Studio

- Uses the English National Curriculum as a 'spine' terminology
- Supports a number of terminology management facilities, including versioning, mapping between terminologies and distributed management of candidate terms
- Built by Knowledge Integration using SchemaLogic's Schema Server with proprietary add-ons



HILT terminology service

Aim is to explore disambiguation, query expansion, mapping between terminologies – prototyping..

- Have experimented with Wordmap – proprietary mapping structures
- Future plans to structure terminologies using SKOS? ‘Use SKOS-Core as the ‘mark-up’ for sending out terminology sets and classification data responses’
- Offer web-services access via the (SOAP-based) SRW protocol starting November 2005

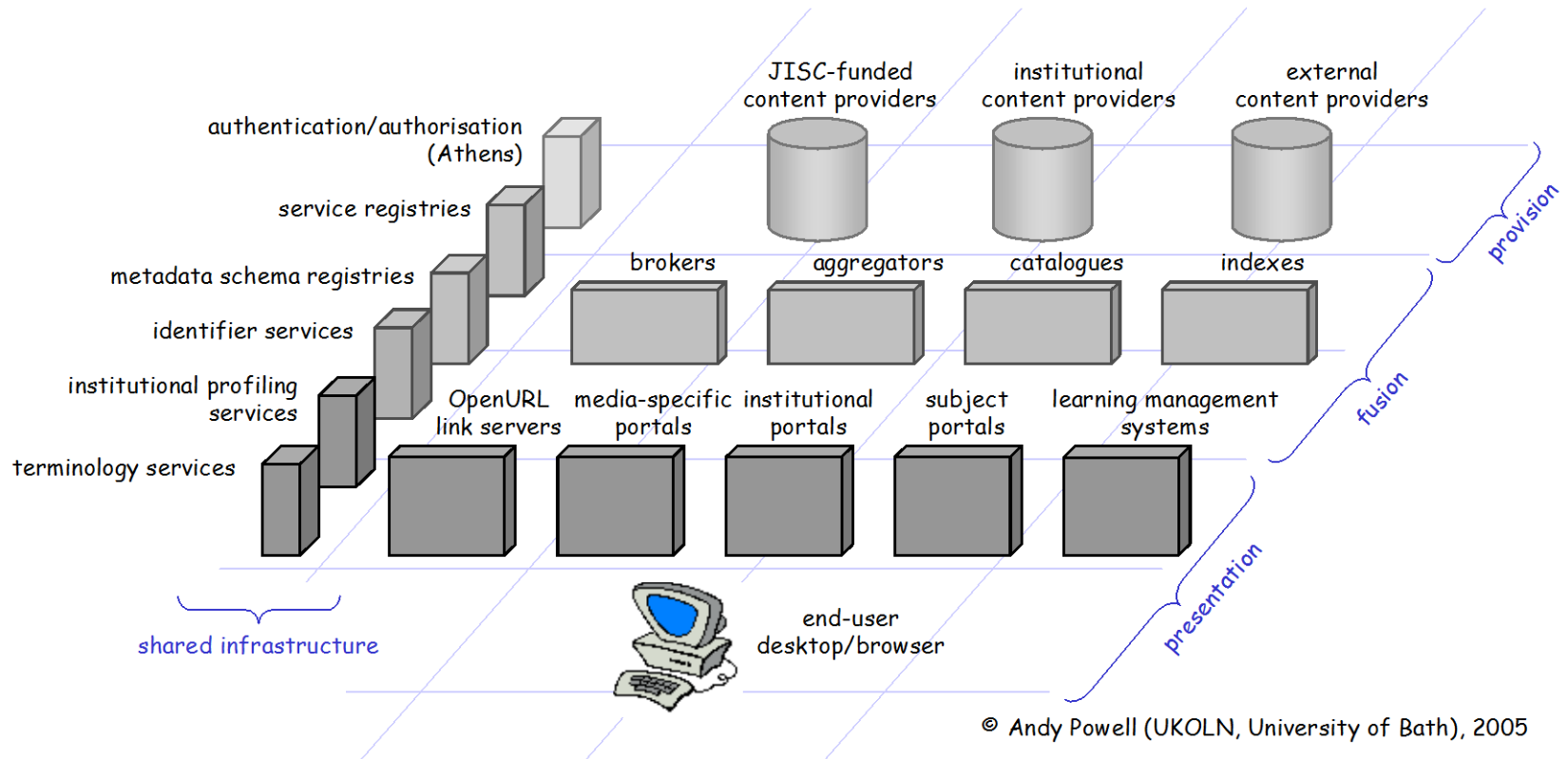


XMDR Extended Metadata Registry

- ISO/IEC 11179-3 : Registry metamodel and basic attributes
- Specifies a conceptual data model for a registry to keep information about
 - data elements and...
 - data element concepts, conceptual domains, value domains
- XMDR taking forward ISO11179 to
 - Register complex structures (concept systems, terminologies)
 - Inter-relations between concept systems and data elements
 - Inter-relations between concept systems
 - <http://xmdr.org>



JISC IE Schema Registry

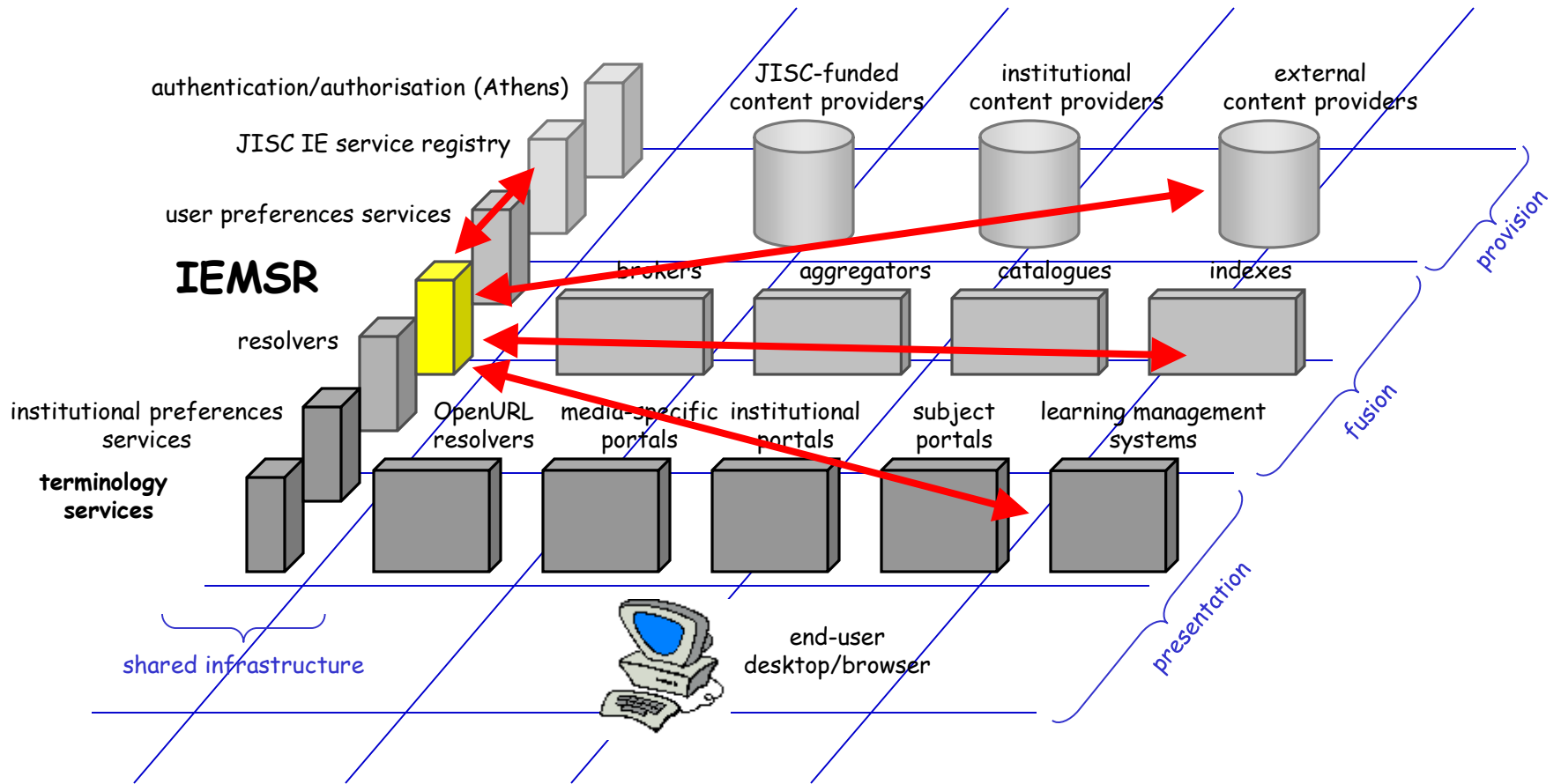


© Andy Powell (UKOLN, University of Bath), 2005

This work is licensed under a [Creative Commons License](https://creativecommons.org/licenses/by-sa/2.0/)
Attribution-ShareAlike 2.0



JISC IEMSR: machine to machine use



NSDL Metadata Registry

“to enable collection holders creating metadata for their collections and various applications that generate, consume and process metadata to

- *identify, declare* and *publish* their metadata schemas (element/property sets) and schemes (value spaces/controlled vocabularies)
- in support of *discovery, reuse, standardization* and *interoperability* within NSDL and globally.”

From: Diane Hillmann and Stuart Sutton, Collaborative Project: an NSDL Registry: supporting interoperable metadata distribution. Proposal to NSF, 2005.
<http://eg2.ischool.washington.edu/registry/documents/proposalDocs/>



NSDL proposes to register...

- GEM Exchange controlled vocabularies
- Eisenhower National Clearinghouse (ENC) list of math and science topics used on its web site as a browsing vocabulary. May evolve to NSDL Learning Resource Type Vocabulary.
- DLESE vocabularies, exposed primarily in their cataloging tools.
- KMODDL Project has developed several vocabularies, some very specific to the project but others have potential for re-use
- The Macaulay Library at the Cornell Laboratory of Ornithology vocabulary for animal behavior

DEST Metadata SchemaRegistry, Australia

- DSTC, University of Queensland rewarded Australian Research Information Infrastructure grant (12-18 months).
- Includes developing a metadata schema registry for DEST (Dept. Education Science and Training)
- Aims is to reduce wasted effort on creating metadata schemas and improve interoperability of adopted metadata schemas.
- Will 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



Questions...

