TRSS

Terminology Registry Scoping Study

Koraljka Golub, Douglas Tudhope

ECDL NKOS Workshop, 19 Sep 2008

Project context

- UK JISC

- Partners
  - University of Glamorgan
  - UKOLN

  - Non-funded: OCLC Office of Research, USA

- Period: Feb 2008 -- Aug 2008

Aims

- To inform the development of JISC shared infrastructure for resource discovery

- To describe the scope and potential use of a terminology registry

- To analyse requirements for services based on a terminology registry

- To help stakeholders understand the need for this component of a shared infrastructure

Overall approach

- information from prior efforts and project documentation

- consultation with key services, projects and executives across digital library, research and learning domains
  - 28 responses collected

Definitions

- Terminologies
  - Controlled vocabularies often referred to as terminologies with regard to registries and web services

- Terminology services
  - Web services: return/apply vocabularies and their content
- Terminology registry
  - lists, describes, and points to sets of vocabularies
  - can hold vocabulary information: member terms, concepts and relationships, provide terminology services, for both human inspection and m2m access

Architecture

- **Option 1**: Registry provides metadata for each vocabulary and links to vocabulary owner/provider

- **Option 2:** Registry provides metadata on (and links to) any available terminology services

- **Option 3**: Registry provides access to vocabulary content (by downloading or providing access to vocabulary's concepts, terms and relationships)

- orthogonal (independent) facets which can be combined
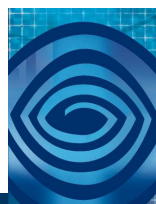
Collected use cases

- Discovery of vocabularies (Option 1)
  - Discovering suitable vocabularies for a collection

- Access and discovery of terminology services and concepts (Options 2, 3)
  - Discovery of terminology services
  - Leveraging terminology for search optimization, browsing, query expansion

…Collected use cases…

- Creation and modification of vocabularies (Option 3)
  - Managing local terminologies
  - Contributing to a published terminology
  - Capturing locally contributed end-user terminology
  - Sharing local terminologies

…Collected use cases…

- Publication of vocabularies (Option 3)
  - Services for topical crawlers

- Integration of vocabularies (Options 2, 3)
  - Searching different collections through 'one' vocabulary
  - Combining local, shared or published terminologies

…Collected use cases

- Cataloguing: Indexing/classification (Options 2,3)
  - Metadata validation
  - Browsing, searching and retrieving terms
  - Automated controlled terms suggestion

- Archiving and preservation of vocabularies (Option 3)

Metadata

1 General information
- Vocabulary name, author or editor, type etc.

2 Scope and usage
- Subjects covered, purpose, rating etc.

3 Characteristics
- Type of terms, relationships etc.

4 Terms and conditions
- Availability etc.

5 Provider
- Contact name etc.

Existing TRs…

- Taxonomy Warehouse
  - Option 1, interactive access
  - claims to host more than 670 taxonomies (73 subject domains) from 288 publishers in 39 languages

…Existing TRs…

- BECTA Vocabulary Bank
  - Options 1, 2, 3, interactive and m2m access
  - supports creating, editing and maintenance of educational vocabularies supporting UK National Curriculum

- BioPortal and OBO Foundry
  - Options 1, 2, 3, interactive and m2m access
  - US OBO – over 60 life-science ontologies
  - UK BioPortal – search and browsing access to its ontologies and experimental data

…Existing TRs…

- Cendi Terminology Locator
  - Option 1, interactive access
  - Points to terminology resources of CENDI federal science research agencies, spanning agriculture to medicine to the environment

- EHKOS
  - Options 1, 3, interactive access
  - English Heritage for their vocabularies

…Existing TRs…

- FAO KOS Registry
  - Options 1, 2, 3, interactive access (and m2m access to Agrovoc)
  - Holds over 90 KOS, in areas related to agriculture and administration

- NERC Data Grid's Vocabulary Server
  - Options 1, 2, 3, m2m access
  - The British Oceanographic Data Centre (BODC) has a TR which supports interoperability of scientific datasets in 43 international data centres
  - with more than 100 vocabularies

…Existing TRs

- NSDL registry
  - Options 1 and 3, interactive access
  - a prototype TR, with an integrated metadata registry
  - 29 vocabularies, mainly educational

- OCLC's Terminology Services Pilot
  - Options 1, 2, 3, interactive and m2m access
  - Current vocabularies held include FAST, GSAFD, LC AC SH, LCSH, MeSH, TGM

Immediate recommendation

- JISC consider possibility of funding a TR providing Option 1 for human inspection and m2m access.

- This development would be an extension of IESR, with a TR as a first class entity within IESR. Thus it would support IESR services to data collections but the TR would also give information about vocabularies in their own right.

- A focused design project for IESR and relevant stakeholders to consider the implications, propose the design and tender for the work packages.

Medium term recommendation

- In the medium term, a pilot Option 2 (for both human and m2m access) should be considered after a collaborative study on an initial set of appropriate metadata elements for terminology services.

- With this comes the need for further research into the metadata for types of services. This should be a collaborative effort. This should include further development of TR related use cases within IESR, particularly from the point of view of developers for future m2m access.

Other recommendations

- Investigate arrangement with OCLC services to access content and terminology services

- A JISC TR Support Project
  - Advice, promotion
  - Continuing to develop the metadata set, particularly for mappings, terminology services and typology of vocabularies
  - Etc.

Questions

- Metadata set core/optional for TR?

- Metadata for terminology services?

- Relationship with ontology registries?

- When is Option 3 feasible?

TRSS: More info

- Final draft submitted to JISC > early October 2008

- A paper

- http://www.ukoln.ac.uk/projects/trss