



A SKOS Core approach to implementing an M2M terminology mapping server

George Macgregor, Anu Joseph & Dennis Nicholson
Centre for Digital Library Research
University of Strathclyde
Glasgow, UK

CENTRE FOR DIGITAL
LIBRARY RESEARCH



Introduction

- Recent growth in digital libraries, repositories, LO repositories, etc.
- Increase in the need for improved interoperability between Knowledge Organization Systems (KOS), encompassing a variety of terminologies
- Need to improve user access to discrete heterogeneous digital objects, particularly in distributed retrieval contexts
- Research exploring techniques to aid effectiveness of simultaneous searching / browsing of multiple distributed repositories remains a key area of study (e.g. Binding & Tudhope, 2004; Doerr, 2001; Nicholson & McCulloch, 2006; Liang & Sini, 2006; Vizine-Goetz et al. 2004, etc.)

Introduction cont.

- Simple Knowledge Organization System Core (SKOS Core) – offers opportunities for experimenting with terminology services capable of being deployed in an M2M web services context
- This paper / presentation will describe and propose an approach to implementing a pilot M2M terminology server employing terminology mappings and using SKOS Core to ‘mark up’ terminological data (conducted under **HILT project**: <http://hilt.cdjr.strath.ac.uk/>)

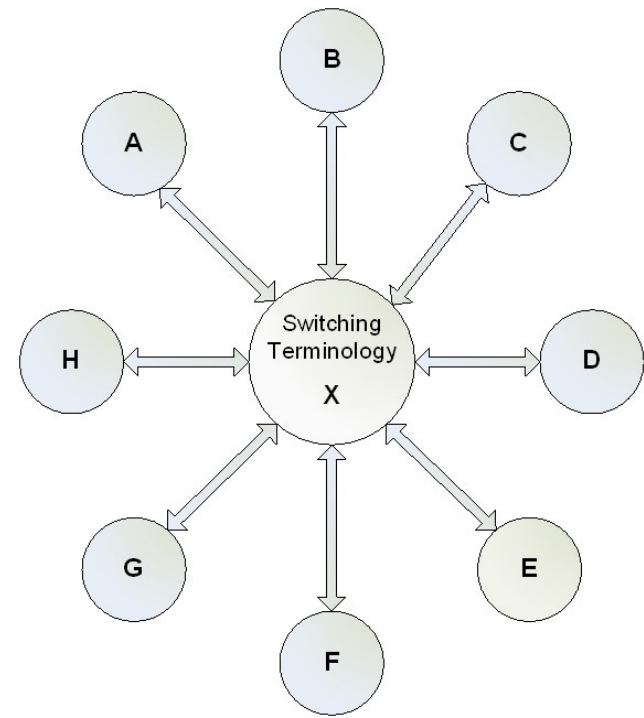
First it is necessary to provide some context: terminological mapping, SKOS Core, etc.

Terminology mapping

- Mapping: relating equivalent terms, concept or hierarchical relationships from one terminology to another (Doerr, 2001)
- Remains largely an intellectual process, even where automated support has been explored (See: Vizine-Goetz et al., 2004; McCulloch et al., 2005)

Focus on **terminology switching**:

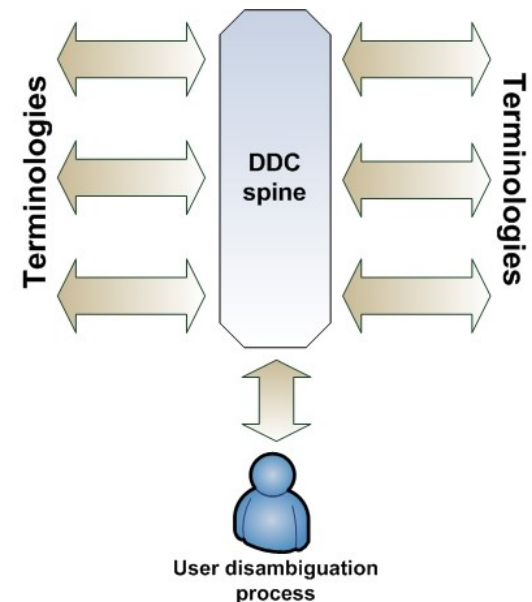
Figure: typical switching model



Terminology mapping cont.

- Terminology server documented operates similarly to switching model (using DDC spine). Our original model provides users with **disambiguation** processes (Shiri et al., 2004). This can be invoked by clients (if desired, by **no means** essential!).
- Offers other terminological functions, such as terminology-based interactive query expansion

Figure: diagram of DDC spine-based model, employing 'disambiguation'



Disambiguation example...

1. User enters search query into client user interface (or browses DDC hierarchies)
2. M2M terminology mapping server identifies best matches and asks the user to clarify which is most appropriate
3. User chooses most appropriate term
4. HILT retrieves the most appropriate terms (from disparate terminologies) to match the local indexes of the repositories to be searched in the federation
5. Depending on how the client is configured for cross-repository searching (and services), client can then use the terminological data requested from HILT to initiate searches across chosen repositories and present integrated results to user
 - Process is transparent and seamless for user...

SKOS Core

- Simple Knowledge Organization System (SKOS) Core: application of the Resource Description Framework (RDF) and proposed by the W3C SWD WG (Miles & Brickley, 2005)
- Flexible framework for expressing the structure and content of 'concept schemes', thus enabling efficient machine processing
- Consists of a series of RDF properties and RDF Schema classes

A simple example:

- Thesaurus = `skos:Concepts`
- And contains preferred terms = `skos:prefLabel`
- And contains non-preferred terms = `skos:altLabel`
- And contains broader and narrower terms (BT & NT), and perhaps related terms (RT) = `skos:broader`, `skos:narrower`, `skos:related` and so forth...

SKOS Core cont.

- The SKOS Core Mapping Vocabulary Specification (MVS) proposed (Miles & Brickley, 2004): draft
- Facilitates mapping of concepts between terminologies within SKOS framework: exactMatch, broadMatch, narrowMatch, majorMatch and minorMatch
- Series of classes (AND, OR, NOT) for combining or excluding concepts. For example, AND might be used to signify the intersection of two or more concepts.

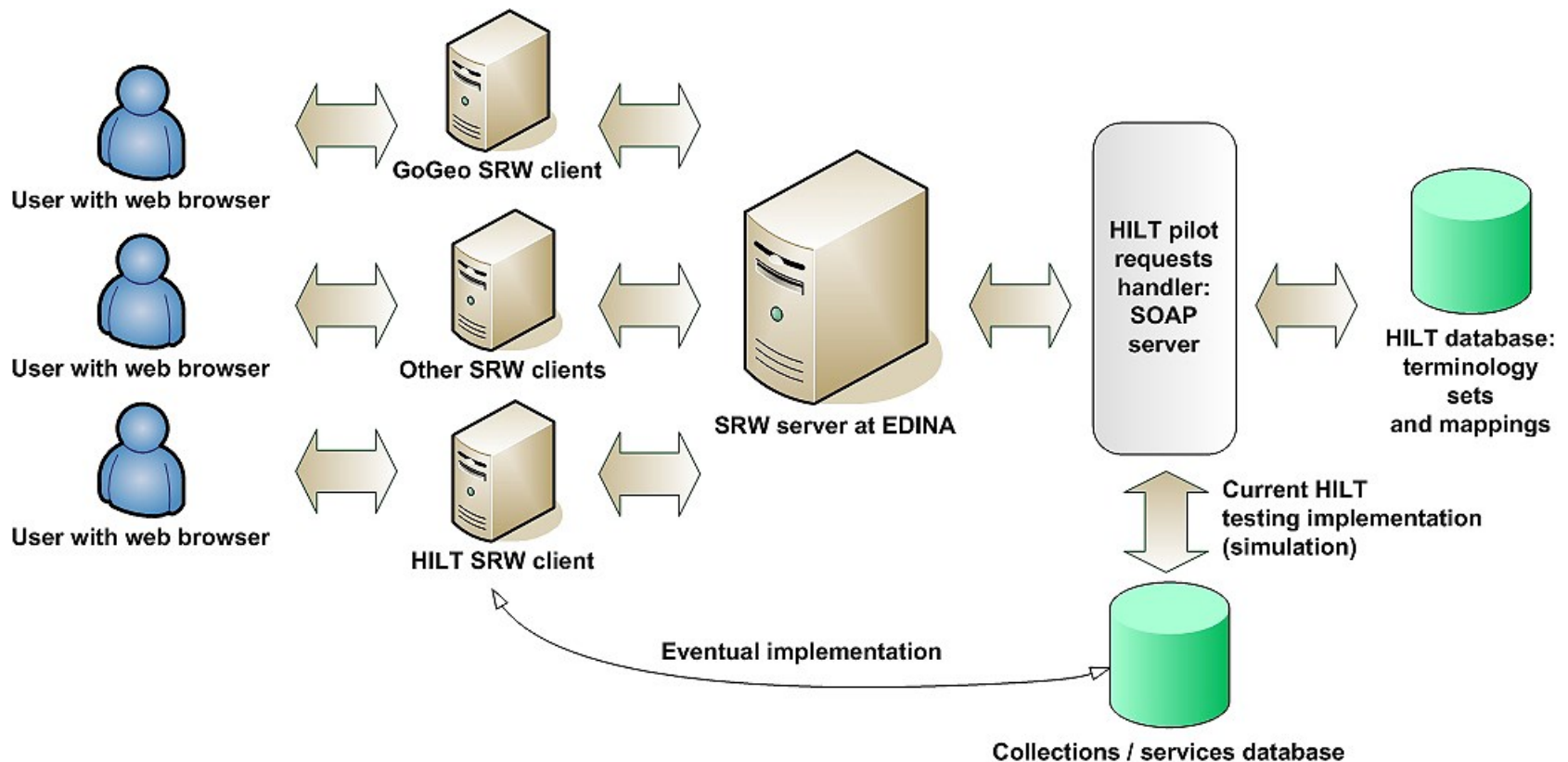
Education (India) – terminology A

Education **AND** India – terminology B

Using SKOS Core for terminology services

- Limited terminological interoperability within the federation of repositories, digital libraries, information services, etc. comprising UK JISC Information Environment
- Testing of M2M terminology server, using SOAP-based Search/Retrieve Web Service (SRW), conducted using GoGeo! (<http://www.gogeo.ac.uk/>) and a local HILT SRW client
- As we shall propose, using SKOS Core to wrap terminological data

Using SKOS Core for terminology services cont.



Wrapping terminologies using SKOS Core

- Non-standard use of SKOS Core?
- SKOS Core experimented with and proposed in order to facilitate meaningful communication with SRW clients regarding structural nature of terminological data requested and/or found in HILT database
- Results identified in database (e.g. scheme information, mapped terms, etc.) are ‘wrapped’ in SKOS Core by the SOAP server
- SRW requests (CQL - SQL) handled using XML over HTTP using SOAP, thus terminological data can easily be embedded within a SOAP XML envelope for messaging to SRW clients

Wrapping terminologies using SKOS

Core cont.

Use of SKOS Core for this application advantageous:

- Accurate modelling and maintenance of the structural and semantic properties of terminological data requested, thus enabling **flexible**, **reliable** and **simple** re-use by clients in local systems
 - **May include innovative user interfaces, browsing structures, charts, etc.**
- Can accommodate the representation of terminological mappings
- SKOS Core offers opportunities for enhanced interaction with the server, enabling added functionality such as terminology-based interactive query expansion (Efthimiadis, 1996)

Server functions in SKOS Core

- M2M implementation has 7 terminological functions (but may grow) and can (obviously) be invoked by clients as they see fit
- Some of the functions currently under experimentation:
 - `get_all_records`
 - `get_DDC_records`
 - `get_non_DDC_records`
 - `get_filtered_set`
- Too detailed to address each today (but might find time!), so focus will be on:
 - `get_non_DDC_records`
 - `get_filtered_set`

Server functions in SKOS Core cont.

- `get_filtered_set` to enable enrichment of user search vocabulary, provide user feedback, interactive query expansion, etc.
- Can provide preferred, broader and narrower terms, related terms, non-preferred, scope notes (where available), etc.
- Expected to be used by client services to enrich user search vocabulary (e.g. query expansion techniques) or build dynamic hierarchical browse structures
- Filtered query set to a specific scheme would return SKOS Core records providing necessary details to process and invoke a variety of local searching functionality



Example: filtered query set to IPSV for 'Arboriculture'

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<SOAP-ENV:Envelope SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" xmlns:SOAP-
  ENV="http://schemas.xmlsoap.org/soap/envelope/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:SOAP-
  ENC="http://schemas.xmlsoap.org/soap/encoding/">
<SOAP-ENV:Body>
<ns1:get_filtered_setResponse xmlns:ns1="http://tempuri.org">
<return xsi:type="SOAP-ENC:Array" SOAP-ENC:arrayType="xsd:string[1]">
<item xsi:type="xsd:string">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
  ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xml:base="http://hiltm2m.cdldr.strath.ac.
  uk/hiltm2m/concepts.php">
<skos:Concept rdf:about="#2715">
  <skos:prefLabel xml:lang="en">Arboriculture</skos:prefLabel>
  <skos:broader rdf:resource="#504"/>
  <skos:narrower rdf:resource="#2633"/>
  <skos:related rdf:resource="#1566"/>
  <skos:related rdf:resource="#15"/>
  <skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/hiltm2m/schemes/IPSV.xml"/>
</skos:Concept>
<skos:concept rdf:about="#504">
  <skos:prefLabel xml:lang="en">Horticulture</skos:prefLabel>
  <skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/hiltm2m/schemes/IPSV.xml"/>
</skos:concept>
<skos:concept rdf:about="#2633">
  <skos:prefLabel xml:lang="en">Tree planting</skos:prefLabel>
  <skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/hiltm2m/schemes/IPSV.xml"/>
</skos:concept><skos:concept rdf:about="#1566">
  <skos:prefLabel xml:lang="en">Woodlands</skos:prefLabel>
  <skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/hiltm2m/schemes/IPSV.xml"/>
</skos:concept>
<skos:concept rdf:about="#15">
  <skos:prefLabel xml:lang="en">Trees</skos:prefLabel>
  <skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/hiltm2m/schemes/IPSV.xml"/>
</skos:concept>
</rdf:RDF>
</item>
</return>
</ns1:get_filtered_setResponse>
</SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Server functions in SKOS Core cont.

- However – primary purpose of server is to enable accurate federated subject searching. Several mapping functions...
- `get_all_records`, `get_DDC_records`, `get_non_DDC_records`
- `get_non_DDC_records`: Can be used in conjunction with ‘disambiguation’ processes
- Provides client with details of non-DDC records that include a mapping to the DDC number sent to the server
- Enables identification of a variety of terms from disparate terminologies associated with a DDC number to be used to search relevant repositories using correct terminology to match local indexes
- (Almost) Inverse function: `get_DDC_records`
- Providing any DDC record either including the term specified, or mapped to by a record including the term specified



Example: non-DDC query for 363.34 (Disasters)

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<SOAP-ENV:Envelope SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" xmlns:SOAP-
  ENV="http://schemas.xmlsoap.org/soap/envelope/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/">
<SOAP-ENV:Body>
<ns1:get_filtered_setResponse xmlns:ns1="http://tempuri.org">
<return xsi:type="SOAP-ENC:Array" SOAP-ENC:arrayType="xsd:string[1]">
<item xsi:type="xsd:string">
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:map="http://www.w3.org/2004/02/skos/mapping#"
xml:base="http://hiltm2m.cdldr.strath.ac.uk/hiltm2m/concepts.php">
<skos:Concept rdf:about="#363.34">
  <skos:prefLabel xml:lang="zxx">363.34</skos:prefLabel>
  <skos:altLabel xml:lang="en">Disasters</skos:altLabel>
  <skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/schemes/DDC.xml"/>
  <map:narrowMatch>
    <skos:Concept rdf:about="#sh 91000441"/>
  </map:narrowMatch>
  <map:exactMatch>
    <skos:Concept rdf:about="#2256"/>
  </map:exactMatch>
  <map:exactMatch>
    <skos:Concept rdf:about="#762"/>
  </map:exactMatch>
  <map:exactMatch>
    <skos:Concept rdf:about="#143"/>
  </map:exactMatch>
</skos:Concept>
<skos:Concept rdf:about="#sh 91000441 ">
<skos:prefLabel xml:lang="en">Emergency management</skos:prefLabel>
<skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/schemes/LCSH.xml"/>
</skos:Concept>
<skos:Concept rdf:about="#2256">
<skos:prefLabel xml:lang="en">Natural disasters</skos:prefLabel>
<skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/schemes/UNESCO.xml"/>
</skos:Concept>
<skos:Concept rdf:about="#762">
<skos:prefLabel xml:lang="en">Natural hazards</skos:prefLabel>
<skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/schemes/GCMD.xml"/>
</skos:Concept>
<skos:Concept rdf:about="#143">
<skos:prefLabel xml:lang="en">Civil emergencies</skos:prefLabel>
<skos:inScheme rdf:resource="http://hiltm2m.cdldr.strath.ac.uk/schemes/IPSV.xml"/>
</skos:Concept>
</rdf:RDF>
</item>
</return>
</ns1:get_filtered_setResponse>
</SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Demonstrators

- A variety of demonstrators are publicly available, including a SOAP client demonstrator (allows viewing of SOAP requires in SKOS) and GoGeo! query expansion demonstrator

See: <http://hilt.cdlr.strath.ac.uk/hilt3web/pilots.html>

Note: Since these are demonstrators, they are under constant development and may therefore be taken down, altered at any time, or may experience sub-optimal performance!

Conclusions and further work

- The M2M implementation proposed in this paper offers terminology mapping as a principal function to enhance user access to disparately indexed heterogeneous digital objects held within multiple repositories
- But also other terminological functions...
- A demonstration that SKOS Core can function in non-standard application
- Constitute a flexible and simple means of implementing various terminological functions for third party terminology services – substitutable model?

Conclusions and further work cont.

- Wrapping of terminological data in SKOS Core provides a readable means of transport data over networks (i.e. over HTTP using SOAP) and enables such data to remain structurally correct, semantically sound, and modelled accurately – thus allowing **flexible** and **simple** re-use by clients
- We intend to continue this line of work and a proposal has been submitted for further research. Some of which will include...

Conclusions and further work cont.

- System and user evaluation of the system, as embedded within a client service (to be disseminated in a future paper)
- Investigating issues with the current MVS draft – abstract paradigm? Noted by Liang & Sini (2006) as issue also
 - MVS useful; but our analysis indicates the need for greater specificity to enable improved results ranking, to inform users during hierarchical browsing, disambiguation, etc.
 - Future research will test this hypothesis via a user study

Conclusions and further work cont.

- Investigating a more distributed model, including exploring a collaborative model to maintaining and implementing mappings to the spine (a wiki-style model for a group of cataloguers, indexers, etc. within the JISC IE)
- General work:
 - Exploring the improved modelling of terminologies: current approach is necessarily generic
 - Providing an open source SRW client, with default illustrative routines which could then be customised by clients

And other things...

Acknowledgement

- This work was funded by the UK Joint Information Systems Committee (JISC) under this Integrated Information Environment Committee funding strand (2005-2007)

JISC: <http://www.jisc.ac.uk/>

References

- Binding, C. & Tudhope, D. (2004). KOS at your Service: Programmatic Access to Knowledge Organisation Systems, *Journal of Digital Information* 4(4). Retrieved September 19, 2006 from <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/>
- Doerr, M. (2001). Semantic problems of thesaurus mapping, *Journal of Digital Information* 1(8). Retrieved September 19, 2006 from <http://jodi.tamu.edu/Articles/v01/i08/Doerr/>
- Efthimiadis, E., N. (1996). Query expansion, *Annual Review of Information Systems and Technology (ARIST)*, 31, 121-187.

References cont.

- Liang, A., C. & Sini, M. (2006). Mapping AGROVOC and the Chinese Agricultural Thesaurus: definitions, tools, procedures, *New Review of Hypermedia and Multimedia* 12(1), 51-62.
- McCulloch, E., Shiri, A. & Nicholson, D. (2005). Challenges and issues in terminology mapping: a digital library perspective, *Electronic Library* 23(6), 671-677.
- Miles, A. & Brickley, D. (Eds.). (2004). *SKOS Mapping Vocabulary Specification*. Retrieved September 19, 2006 from <http://www.w3.org/2004/02/skos/mapping/spec/>

References cont.

Miles, A. & Brickley, D. (Eds.). (2005). *SKOS Core Guide: W3C Working Draft 2 November*. Retrieved September 19, 2006 from <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>

Nicholson, D. & McCulloch, E. (2006). Investigating the feasibility of a distributed, mapping-based, approach to solving subject interoperability problems in a multi-scheme, cross-service, retrieval environment, *International Conference on Digital Libraries, 5-8 December 2006*, New Delhi, India.

Shiri, A., Nicholson, D. & McCulloch, E. (2004). User evaluation of a pilot terminologies server for a distributed multi-scheme environment, *Online Information Review* 28(4) 273-283.

References cont.

Vizine-Goetz, D., Hickey, C., Houghton, A. & Thompson, R. (2004). Vocabulary mapping for terminology services, *Journal of Digital Information* 4(4). Retrieved September 19, 2006 from <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/>

Thanks! Contact & links

George Macgregor: george.macgregor@strath.ac.uk

Anu Joseph: anu.joseph@cis.strath.ac.uk

Dennis Nicholson: d.m.nicholson@strath.ac.uk

Links:

- Demonstrators: <http://hilt.cdlr.strath.ac.uk/hilt3web/pilots.html>
- HILT project: <http://hilt.cdlr.strath.ac.uk/>
- Centre for Digital Library Research: <http://cdlr.strath.ac.uk/>
- EDINA: <http://www.edina.ac.uk/>



University of
Strathclyde
Science