



Nicholson, Dennis and McCulloch, Emma (2006) Investigating the feasibility of a distributed, mapping-based, approach to solving subject interoperability problems in a multi-scheme, cross-service, retrieval environment. In: International Conference on Digital Libraries, 5-8 Dec 2006, New Delhi, India.

<http://strathprints.strath.ac.uk/2875/>

This is an author-produced version of a paper presented at the International Conference on Digital Libraries, 5-8 Dec 2006.

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in Strathprints to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the url (<http://eprints.cdlr.strath.ac.uk>) of the Strathprints website.

Any correspondence concerning this service should be sent to The Strathprints Administrator: eprints@cis.strath.ac.uk

Investigating the Feasibility of a Distributed, Mapping-based, Approach to Solving Subject Interoperability Problems in a Multi-scheme, Cross-service, Retrieval Environment

Dennis Nicholson and Emma McCulloch

Centre for Digital Library Research (CDLR), Department of Computer and Information Sciences, University of Strathclyde, UK

Keywords: Interoperable subject descriptions; KOS; Mapping; SKOS-Core; SRW; Terminology services.

Abstract

The HILT project is researching the problems of facilitating interoperability of subject descriptions in a distributed multi-scheme environment. HILT Phase I found a UK community consensus in favour of utilising an inter-scheme mapping service to improve interoperability. HILT Phase II investigated the approach by building a pilot server, and identified a range of issues that would have to be tackled if an operational service was to be successful. HILT Phase III will implement a centralised version of an M2M pilot, but will aim to design it so that the possibility of a move to a distributed service remains open. This aim will impact on likely future research concerns in Phase III and beyond. Wide adoption of a distributed approach to the problem could lead to the creation of a framework within which regional, national, and international efforts in the area can be harmonised and co-ordinated.

1. Introduction and background: HILT in brief

Interoperability of knowledge organisation systems (KOS) is, to quote Zeng and Chan (2004), ‘an unavoidable issue’ in today’s networked environment. It is an issue likely to impact, in time, on the semantic web vision (Berners-Lee et al, 2001), but is more usually tackled at present in an information retrieval context¹. Currently, information services employ a plethora of different subject schemes to describe their resources. In some cases, they use recognised standards, in others ‘in-house’ or even uncontrolled schemes. Either way, the practice acts as a barrier to effective cross-searching by subject over distributed information services. The issue not only impacts globally, in the sense of being an issue across the networked world, it is also, in many cases, global in its scope, with users often cross-searching across national boundaries to meet their research or business information requirements.

The issue has attracted a good deal of interest in recent years. Potential solutions proposed include linking or switching between schemes, mapping, derivation/modelling (see for example Doerr, 2001; Chan and Zeng, 2002), and automatic or semi-automatic classification (see for example Koch and Vizine-Goetz, 1998; Godby et al, 1999; Ardo, 2004). CARMEN (2000), LIMBER (2000), Renardus (2002), and MACS (2005) are amongst a range of recent projects that have tackled the problem, and key international players such as OCLC (<http://www.oclc.org/> have also done relevant work (see <http://www.oclc.org/productworks/terminologiespilot.htm>).

The HILT (High-Level Thesaurus) project (<http://hilt.cdlr.strath.ac.uk/>), based at CDLR (<http://cdlr.strath.ac.uk>) is one project active in the area. It is researching the problems of facilitating interoperability of subject descriptions in a distributed multi-scheme environment, aiming, ideally, to identify a generic solution. HILT Phase I (<http://hilt.cdlr.strath.ac.uk/index1.html>) found a UK community consensus in favour of improving interoperability via an inter-scheme mapping service. This idea was followed up in HILT Phase II (<http://hilt.cdlr.strath.ac.uk/index2.html>), which built a user accessible pilot terminologies mapping service based on a Dewey Decimal Classification (DDC) spine to investigate the approach. The subsequent Machine to Machine (M2M) Feasibility Study (<http://hilt.cdlr.strath.ac.uk/hiltm2mfs/>) then investigated, proposed, and costed a project to build an M2M version of the pilot and this led to the funding of HILT Phase III

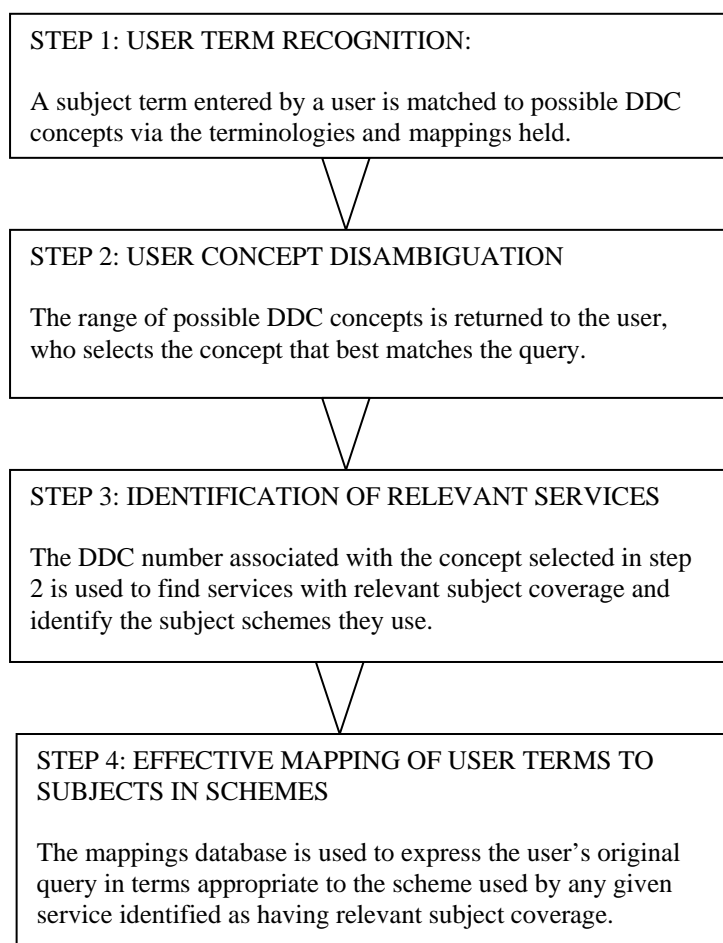
(<http://hilt.cdrl.strath.ac.uk/index3.html>). The final reports of the earlier stages of HILT are available online (HILT, 2002; 2003; 2005). Project progress prior to this paper is reported in Nicholson and Wake, 2001a, 2001b; Nicholson et al 2002, 2006; Nicholson, 2001, 2003; Nicholson and Shiri, 2003; Shiri et al, 2004; McCulloch, 2004; McCulloch et al, 2004, 2005.

HILT Phase III began in November 2005 and will run until January 2007. It aims to take a version of² the Phase II pilot service (<http://hiltpilot.cdrl.strath.ac.uk/pilot/top.php>)³, extend its functionality in various ways, and create an M2M version of it built around SRW (<http://www.loc.gov/z3950/agency/zing/srw/>) and the SWAD-Europe (<http://www.w3.org/2001/sw/Europe/>) project's SKOS-Core (Miles et al, 2005), taking the syntax and data-exchange protocol implications of eScience and semantic web developments into account in the design. The present paper utilises outcomes from HILT Phase II, the subsequent HILT M2M Feasibility Study, and early work on the Phase III centralised pilot to illuminate key facets of the subject interoperability problem and map out the context for a consideration of the likely shape and form of planned Phase III research on the issue of whether or not a distributed approach to the provision of interoperability services might be a fruitful basis for future M2M developments. One outcome of the M2M Feasibility Study was that a distributed approach, perhaps using SKOS-Core concept URIs as the basis for mapping between different schemes across services, was worth investigating. If practicable, it might ultimately lead to a matrix of servers being available internationally with mappings between schemes being based on SKOS Core concept URIs and being built up slowly over a long period of time. Such an approach is theoretically attractive in that it might implement the kind of mapping-based solution HILT had envisaged to subject interoperability issues in a way that would spread the cost and effort over many organisations and a longer period of time. Thus, whilst the primary aim of HILT Phase III is the creation of a centralized M2M pilot, a secondary aim is to investigate whether a distributed version might offer a fruitful future development path.

2. A selective issues overview: The distributed issue in context

By the end of HILT Phase II, a range of issues facing HILT and other projects dealing with the subject interoperability problem had been noted, drawn out by considering and analysing a combination of research results reported by others active in the field, results from the project itself, and insights provided through the interaction of these with the process of designing, implementing, testing, and refining a working pilot. Providing a selective summary here helps put the Phase III work in context. In particular, it provides insight into the kinds of issues that an effective service will have to face, whether centralised or distributed, and so provides a backdrop for the discussion on the feasibility of the distributed approach towards the end of the paper. Note that it is not intended to be comprehensive, only give a feel for the kinds of issues that arise.

A useful approach is to highlight key issues in a framework provided by the various design elements of the Phase II pilot, then supplement this with additional headings of relevance (user interface issues; M2M issues). The Phase II design is described in detail in Nicholson et al, 2006. However, it may be summarised in terms of the steps that occur when a user interacts with the working pilot, as follows:



2.1 User term recognition

Step 1 entails user term recognition. The user enters a subject term expressing a subject interest and the term is sought in the database of mapped schemes. In the pilot, this database consists primarily of the whole of the Dewey Decimal Classification Scheme (DDC) – captions, index, and standard subdivisions – together with a large number of LCSH mappings. Other mappings are relatively small in number and include illustrative mappings between DDC and MeSH, DDC and UNESCO, and DDC and a small selection of Scottish terms. DDC is used as a spine and all schemes are mapped to this rather than to each other. This reduces the level of mapping required, avoiding the scalability problems highlighted by Keizer (2005), although it is likely to entail as yet unidentified complications relating to possible inconsistencies in some cases between mapping types. Once found in the database, the user term is mapped to one or more possible concepts in the DDC hierarchy, thereby identifying a precise DDC number for the subject sought.

Notable issues identified here are:

- Database coverage. Clearly, the more extensive the database of schemes and mappings, the greater the chance that a user term will be recognised by the system. More important still (although the two can be linked) is the extensive inclusion of what are sometimes called entry vocabularies (Buckland et al, 1999), and the mapping of these to the ‘unfamiliar metadata vocabularies’ of standard schemes. As McCray et al (1999) note in relation to medical terminologies, the terms typically entered by users do not tend to match those in standard schemes.

- The complexity of the mapping issues encountered when mapping between schemes. Chaplan (1995), for example, identified 19 different mapping types - exact match, singular-plural match, concept match, and so on.
- Granularity issues. These are discussed briefly under 'Effective mapping of user terms to subjects in schemes' below, but are also relevant here in step 1.

2.2 User concept disambiguation

In step 2, the user is presented with one or more concepts from DDC and asked to choose which best describes his or her subject interest. The result is then used to identify the appropriate DDC number for the user's term.

Notable issues identified here are:

- A lack of knowledge of the problems faced by users making this choice and the consequent need for further research in this area. (Shiri et al, 2004)
- The strong likelihood that such a choice can be made more readily if the user is presented with representative resources likely to be covered by a given concept – an option difficult to provide in the current pilot at the disambiguation stage (ibid). It is technically feasible, but adds a level of design complexity that was beyond the resources of HILT Phase II.

2.3 Identification of services with relevant subject coverage and the subject scheme(s) they use

Once the DDC number relevant to the user's subject interest is fixed, it is used to identify information services likely to cover that subject area in a database of collections or services that has been classified for subject coverage using DDC, and to retrieve metadata on the subject schemes they employ. If, as is usually the case, there is nothing covering the subject specifically, successive truncations of the DDC number are used to search for relevant services.

Notable issues identified here are:

- Services tend to be classified at more general numbers (D510, say, for mathematics, rather than a more specific number covering a more specific subject). In theory, this number covers all of the more specific topics under mathematics. In practice, there is no guarantee that the service will actually hold resources on a particular, more specific, subject.
- More research is required to determine how good or otherwise the algorithm used in this process is in tackling the problems thrown up by a representative range of subject areas.
- In an ideal world, services would be identified in a more direct and more reliable fashion – preferably on the basis that they are known to include at least one resource covering the specific subject sought by the user.

2.4 Effective mapping of user terms to subjects in schemes

Once a service with relevant subject coverage has been identified, and metadata on the subject scheme it uses to describe the resources it holds retrieved, the next step is to determine how the user's search is best expressed in the scheme in question. The user's subject interest – as identified by the full DDC number chosen in step 2 of Figure 1 – is mapped to the appropriate term in that scheme using the mappings in the HILT database, thereby providing the user with the appropriate term to use when searching the service for the subject the user is interested in. Where possible, the service is searched automatically under that term to provide the user with the resources sought. Otherwise, the appropriate term and a link to the service is provided so that the search can be performed by the user.

Notable issues identified here are:

- The various mapping issues mentioned in relation to step 1 also apply here.
- The need to identify schemes, versions of schemes, and implementations of schemes uniquely.
- The need for the database of schemes and mappings to hold the various versions and implementations of schemes.

- Granularity issues. Mapping between schemes may be easier – and is certainly less resource intensive – at more general levels, but there is evidence to suggest (ibid), and common sense would certainly affirm, that users often need mapping at quite specific levels.

2.5 User interface issues

At each of the above steps, there are user interface design issues. Although some preliminary work on interface design was done in HILT Phase II (Shiri et al, 2004), it was recognised that at great deal more was required,

Notable issues identified here are:

- The need to deal with subjects expressed as phrases. In the current interface, only single word terms are handled. Brown (1995) found that novice searchers favour single as opposed to compound terms. However, the results of a series of interviews with potential HILT users conducted within Phase II, where users ranged from novice-level to very experienced, showed a different pattern, with only 12 of 208 responses entailing the use of a single word search term (McCulloch et al, 2004).
- The probability that a good deal of design work is needed to optimize accurate and successful disambiguation.
- The probability that some users will wish to choose more than one option at the disambiguation stage. At present, only a single choice is possible.

2.6 Machine to machine functionality

Since the HILT pilot was seen as a step towards providing terminologies support for the distributed services of the developing JISC Information Environment (<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/>) and their users, there was a known requirement to provide HILT-type services on an M2M basis at an early stage of the ongoing initiative. The requirement did not fit readily into the Phase II focus on a user-accessible pilot, but was a recognised issue throughout and was the focus of a special project report compiled by Rachel Heery of UKOLN (UKOLN, 2003). In light of this report, it was decided to focus on M2M issues before tackling some of the others mentioned above.

An M2M Feasibility Study was seen as the first step in this process, and drew out a range of issues that influenced the shape and form of the M2M demonstrator proposed in Phase III (see <http://hilt.cdlr.strath.ac.uk/hiltm2mfs/0HILTM2MFinalReportRepV3.1.doc>). A minimal functionality SOAP demonstrator (<http://nevis.ed.ac.uk:8080/asp-misc/public/hilt.asp>), built by EDINA (<http://edina.ac.uk/>), working with CDLR, was a useful offshoot of this, but its main outcomes were:

- a. A report detailing five agreed ‘use cases’ that an M2M pilot should address, and assessing the suitability of the various protocols and markup approaches considered from handling the issues they raised. This is available in full as Appendix D of the final report of the study (HILT, 2005). Its conclusions are summarised in Nicholson and McCulloch (2006).
- b. Outline proposals for an actual demonstrator that would:
 - Use the SRW protocol only, but be designed so that a possible extension offering other protocols such as Z39.50 could be introduced at a later date.
 - Use SKOS-Core as the ‘markup’ for sending out terminology and classification set responses, but be designed so that adding other formats such as MARC and Zthes would be later option.
 - Offer similar functionality to that provided by the Phase II pilot (use case 1), but extended to cover the requirements of the other four use cases.
 - Keep in mind the syntax and data-exchange protocol implications of eScience and semantic web developments.

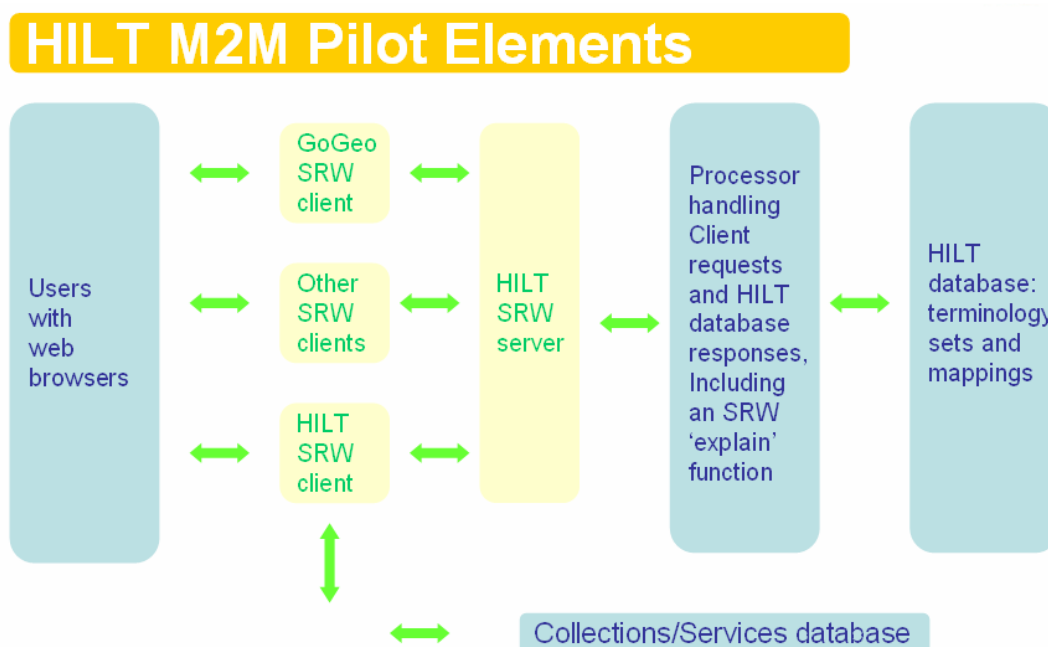
Another key finding was the recognition that the M2M environment, together with the recent availability of SKOS-Core and SKOS-Core concept URIs (W3C, 2005) as the basis for mapping between different schemes across services, was worth investigating. In theory, by opening up the

possibility of providing terminology and terminology mapping services in a distributed fashion, this approach could, amongst other things, help make high volume mappings available more quickly than might otherwise be possible, help spread the cost across a range of players and countries, and, potentially, local (national or even regional) interest in building such services in a way that might otherwise be difficult.

As a result of this finding it was agreed that, whilst the pilot created in Phase III would be based on the centralised model, it could be designed so that a future move to a distributed approach would be possible, a decision that raises its own issues, and that will, as a consequence, impact on Phase III research work in the various ways considered in section 4 below.

3. The Core Design of the Centralised Pilot: A Final Contextual Element

Early work on the design and development of the core of the centralised pilot has already begun, and will provide the final element of the context within which the feasibility of the distributed model will be considered. The detail of this aspect of the project is described elsewhere⁴. In outline, what is proposed is an M2M pilot based on the various elements shown in the diagram below.



The expectation is that the SRW clients for accessing the service (based at EDINA in Edinburgh and CDLR in Glasgow) will be embedded in web pages presented to users of participating JISC information services and will communicate with the (SOAP-based) HILT requests and responses handler (at CDLR) via an SRW server (at EDINA). This, in turn, will pass requests to the HILT database (at CDLR), and send responses from it back to the SRW server for onward communication to the SRW clients. The HILT database will include a range of subject or class schemes – DDC, UNESCO, LCSH (partial), AAT (partial), MeSH (partial), IPSV, JACS, and others – together with illustrative mappings between schemes via the DDC spine. The collections database is required in instances where it is necessary to identify collections with subject coverage relevant to a user query as described in section 2.3 above.

A move to a distributed version in future is likely to impact on all elements of the centralised model shown. The SRW clients will have to be able to deal with requests to, and responses from, a range of terminology servers, via a range of SRW servers, (some (but probably not all) of which might handle multiple terminology servers). The collections and services database may have to serve up information

to SRW clients and SRW servers alike on the existence and connection details of a range of terminology services, and the databases and requests and responses handler will be affected by the distribution of terminologies and mappings across many servers.

4. Discussion and conclusion: the possibilities of a distributed approach

4.1 Investigating the Distributed Approach

Just what the nature of the impact of a move to a distributed approach will have on these various elements of the centralised pilot depends on the results of research in a range of inter-dependant areas:

- a) Work to identify alternative designs for a distributed service and determine whether they are operationally feasible
- b) An examination of the issues raised by these designs, whether they be new and unique to the distributed approach or a particular distributed design, or of a type already encountered and covered in section 2 above.
- c) Work to determine how and where any given design or issue impacts on the M2M design outlined in section 3 above.
- d) An examination of whether the resulting design itself offers the possibility of an effective service
- e) An initial examination of whether it appears to offer a significant advantage over the centralised version in key areas.

The detail of this work has yet to be determined, but is expected to emerge from an examination of how the various functional elements of the HILT II pilot described above might best be realised via a distributed approach, and of how the results of this would then map onto the elements of the M2M pilot as currently envisaged.

Particular areas requiring detailed examination include the likely impact of the distribution of terminologies and mappings across many services on inter-scheme mapping issues, and on the efficacy of mapping, issues related to user concept disambiguation, and issues arising from the need to identify services relevant to particular subject queries should a distributed approach rule out the use of a single spine based only on DDC.

The current interest in the distributed option is also likely to impact on research (and associated development) concerns beyond Phase III, delaying a detailed consideration of many of the other issues identified in Phase II until the direction in this respect is clear. Obviously, the design and creation of a distributed M2M pilot will be an early concern, followed closely by an examination of how the distributed and centralised versions of the pilot compare on questions such as effectiveness, cost, rate of development possible, the difficulty of issues raised, and so on. More significantly, perhaps, if the research shows that the distributed approach is feasible, and has exploitable advantages worth exploring, the concerns of the project will shift significantly. The aim of seeking to involve as many players as possible in the creation of a single, ultimately multi-lingual, mapping service as envisaged in Phase II will fall away. In its place will be a focus on work that will facilitate, in time, the creation of internationally distributed network of inter-working subject interoperability services. This work could take several forms, but the development of a framework within which regional, national, and international efforts in the area can be harmonised and co-ordinated is likely to be a key element.

4.2 Summary and conclusion

The HILT project has been investigating the subject interoperability problem since 2000. HILT Phase I found a UK community consensus in favour of utilising an inter-scheme mapping service to improve interoperability. HILT Phase II investigated the approach by building a pilot server, and identified a range of issues that would have to be tackled if an operational service was to be successful. The M2M Feasibility Study focused on one of these issues – the need to build M2M functionality into an operational service. It found such an approach to be feasible and scoped out a follow up project – HILT Phase III – to build an M2M pilot. A secondary outcome of the M2M study was a recognition of the

possibilities of a distributed approach to the provision of an M2M service. As a result of this, HILT Phase III will implement a centralised version of an M2M pilot, but will aim to design it so that the possibility of a move to a distributed service remains open.

This requirement will impact on likely future research concerns in Phase III and beyond. Neither Phase III itself, nor any subsequent project aimed at building a distributed M2M pilot will solve the many other issues identified in HILT Phase II (see section 2 above). If a distributed approach is shown to be feasible and attractive, however, it may ultimately impact upon the speed with which such issues can be tackled and resolved, opening up opportunities for wider national and international collaboration, and enabling a larger, more co-ordinated approach to research and development in this important area.

Acknowledgments

The authors would like to thank JISC for their ongoing financial support of HILT and OCLC for kindly providing the DDC-LCSH mappings (as referred to in section 2.1) and for their continued support throughout the HILT project.

References

- Ardo, A. (2004), **Automatic Subject Classification and Topic Specific Search Engines - Research at KnowLib**, Presented at *DELOS Regional Awareness Event: Between Knowledge Organization and Semantic Web: Semantic Approaches in Digital Libraries*, Lund, Sweden, 2004. Available online at: http://www.delos.info/eventlist/LUB1/Anders_Ardo/DELOS.PDF (accessed 20 April 2006).
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001), **The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities**, *Scientific American*, 2001, Available online at: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> (accessed 20 April 2006).
- Brown, M. E. (1995), **By any other name: accounting for failure in the naming of subject categories**. *Library and Information Science Research*, 1995, **17**(4), 325-429.
- Buckland, M., K., Chen, A., Gebbie, M., Kim, Y. and Norgard, B. (2000), **Variation by Subdomain in Indexes to Knowledge Organization Systems**. Presented at *International Society for Knowledge Organization, ISKO6*, Toronto, 2000. Available online at: <http://metadata.sims.berkeley.edu/papers/iskopaper.html> (accessed 20th April, 2006).
- Chan, L. and Zeng, M. (2002), **Ensuring interoperability among subject vocabularies and knowledge organisation schemes: a methodological analysis**. Presented at *68th IFLA Council and General Conference*, Glasgow, Scotland, 2002. Available online at: <http://www.ifla.org/IV/ifla68/papers/008-122e.pdf> (accessed 19th April, 2006).
- Chaplan, M., A. (1995), **Mapping Laborline thesaurus terms to Library of Congress subject headings: implications for vocabulary switching**. *Library Quarterly*, 1995, **65**(1), 39-61.
- CARMEN**. Available online at: <http://www.mathematik.uni-osnabrueck.de/projects/carmen/index.en.shtml> (accessed 18 April 2006).
- Doerr, M. (2001), **Semantic problems of thesaurus mapping**. *Journal of Digital Information*, 2001, **1**(8). Available online at: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/> (accessed 20 April 2006).
- Godby, C. J., Miller, E. J. and Reighart, R. R. (1999), **Automatically Generated Topic Maps of World Wide Web Resources**, *Annual Review of OCLC Research*, 1999. Available online at <http://digitalarchive.oclc.org/da/ViewObject.jsp?fileid=0000002655:000000059193&reqid=9300> (accessed 20 April 2006).

Heery, R., Carpenter, L. and Day, M. (2001), **Renardus Project Developments and the Wider Digital Library Context**. *D-Lib Magazine*, 2001, 7(4) Available online at: <http://www.dlib.org/dlib/april01/heery/04heery.html> (accessed 20 April 2006).

HILT, HILT Phase I Final Report, 2002. Available online at: <http://hilt.cdrl.strath.ac.uk/reports/finalreport.html> (accessed 20 April 2006).

HILT, HILT Phase II Final Report, 2003. Available online at: <http://hilt.cdrl.strath.ac.uk/hilt2web/finalreport.htm> (accessed 20 April 2006).

HILT, HILT M2M Feasibility Study II Final Report, 2005. Available online at: <http://hilt.cdrl.strath.ac.uk/hiltm2mfs/0HILTM2MFinalReportRepV3.1.doc> (accessed 20 April 2006).

Keiser, J. (2005), **Coherence and Interoperability In Agricultural Information Systems: Issues, Experiences and a Project**. Presented at *DC (International Conference on Dublin Core and Metadata Applications)*, Madrid 2005. Available online at: http://dc2005.uc3m.es/program/presentations/2005-09_DC-keynote.ppt (accessed 20 April 2006).

Koch, T., Neuroth, H. and Day, M. (2001), **Renardus: Cross-browsing European subject gateways via a common classification system (DDC)**. Presented at *IFLA Satellite Conference: Subject Retrieval in a Networked Environment*, Dublin, Ohio, 2001. Available online at: <http://www.lub.lu.se/tk/renardus/preIFLA-demo.html> (accessed 20 April 2006).

Koch, T. and Vizine-Goetz, D. (1998), **Automatic Classification and Content Navigation Support for Web Services: DESIRE II Cooperates with OCLC**. *Annual Review of OCLC Research*, 1998. Available online at: <http://www.lub.lu.se/tk/demos/class-ws/automatic.htm> (accessed 20 April 2006).

LIMBER: Language Independent Metadata Browsing of European Resources. Available online at: <http://www.limber.rl.ac.uk/> (accessed 18 April 2006).

MACS, Multilingual Access to Subjects. Available online at: <https://ilmacs.uvt.nl/> (accessed 18 April 2006).

McCulloch, E. (2004), **Multiple terminologies: an obstacle to information retrieval**. *Library Review*, 2004, 53(6). pp. 297-300.

McCulloch, E., Shiri, A. and Nicholson, D. (2004), **Subject searching requirements: the HILT II experience**. *Library Review*, 2004, 53(8) 408-414.

McCulloch, E., Shiri, A. and Nicholson, D. (2005), **Challenges and issues in terminology mapping: a digital library perspective**. *Electronic Library*, 2005, 23(6) 671-677.

McCray, A. T., Loane, R. F., Browne, A. C. and Bangalore, A. K. (1999), **Terminology Issues in User Access to Web-based Medical Information, 1999**. Available online at: <http://www.amia.org/pubs/symposia/D005626.PDF> (accessed 20 April 2006).

Miles, A., J., Rogers, N. and Beckett, D. (2005), **SKOS-Core 1.0 Guide, An RDF schema for thesauri and related knowledge organisation systems, 2005**. Available online at: <http://www.w3.org/2001/sw/Europe/reports/thes/1.0/guide> (accessed 20 April 2006).

Nicholson, D. and McCulloch E. (2006), **Interoperable Subject Retrieval in a Distributed Multi-Scheme Environment: New Developments in the HILT Project**. Paper presented at *Ibersid*, Zaragoza, Spain, 2006.

Nicholson, D. (2003), **Subject-Based Interoperability: Issues from the High Level Thesaurus (HILT) Project**. *International Cataloguing and Bibliographic Control*, 2003, 32(1).

- Nicholson, D. and Shiri, A. (2003), **Interoperability in subject searching and browsing**. *OCLC Systems and Services*, 2003, **19**(2) 58-61.
- Nicholson, D. and Wake, S. (2001), **Interoperability in Subject Terminologies: The HILT Project**. *New Review of Information Networking*, 2001, 7.
- Nicholson, D. and Wake, S. (2001). **High-level Thesaurus Project: an overview**. Paper presented at *IFLA Satellite Conference: Subject Retrieval in a Networked Environment*, Dublin, Ohio, 2001.
- Nicholson, D., Dawson, A. and Shiri, A. (2006), **HILT: A Pilot Terminology Mapping Service with a DDC Spine**. *Cataloguing and Indexing Quarterly*, 2006 In press.
- Nicholson, D., Dunsire, G. and Neill, S. (2002), **HILT: moving towards interoperability in subject terminologies**. *Journal of Internet Cataloguing*, 2002, **5**(4), 97-111.
- Nicholson, D. (2001), **High-level Thesaurus Project: Investigating the Problem of Subject Cross-searching and Browsing between Communities**. In, *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*, edited by Chin-chih Chen, pp219-226, 2001 (Tsinghua University Press: Beijing).
- Renardus. Available online at: <http://www.renardus.org/> (accessed 18 April 2006).
- Saeed, H. and Chaudhury, A. S. (2002), **Using Dewey Decimal Classification scheme (DDC) for building taxonomies for knowledge organisation**. *Journal of Documentation*, 2002, **58**(5) 575-583.
- Shiri, A., Nicholson, D. and McCulloch, E. (2004), **User evaluation of a pilot terminologies server for a distributed multi-scheme environment**. *Online Information Review*, 2004, **28**(4) 273-283.
- UKOLN (2003), **HILT Final Report Appendix J, 2003**. Available online at: <http://www.ukoln.ac.uk/metadata/hilt/m2m-report/hilt-final-report.pdf> (accessed 20 April 2006).
- Vizine-Goetz, D., Hickey, C., Houghton, A. And Thompson, R. (2004), **Vocabulary Mapping for Terminology Services**. *Journal of Digital Information*, 2004, **4**(4) Available online at: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/> (accessed 20 April 2006).
- WC3 (2005), **SKOS Core Guide: W3C Working Draft, 10 May 2005**. Available online at: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/> (accessed 20 April 2006).
- Zeng, M.L. and Chan, L.M. (2004), **Trends and Issues in Establishing Interoperability Among Knowledge Organisation Systems**. *Journal of the American Society for Information Science and Technology*, 2004, **55**(5).

¹ Recent examples of work in the area are reported in Heery et al, 2001; Koch et al, 2001, Saeed and Chaudhury, 2002; and Vizine-Goetz et al 2005, but see Zeng and Chan, 2004 for a more comprehensive list.

² The initial pilot was based around the Wordmap software (<http://www.wordmap.com/>) and this software may be used again in future. However, the plan for HILT Phase III is to use a PHP and SQL Server based version of the pilot (see <http://hilt.cdrl.strath.ac.uk/hilt3/top.cfm>)

³ See also the worked examples in Appendix I of the HILT Phase II Final report (HILT, 2003) and at <http://hilt pilot.cdrl.strath.ac.uk/pilot/examples/>.

⁴ There is a detailed requirements document at on the HILT web-site at <http://hilt.cdrl.strath.ac.uk/hilt3web/reports/h3requirements.pdf> and a paper on the pilot will be presented at ECDL in September