Association for Library Collections and Technical Services
Subject Analysis Committee

Subcommittee on Semantic Interoperability

Subject Semantic Interoperability: Final Report

Report of the Subcommittee on Semantic Interoperability to the ALCTS Subject Analysis Committee

Submitted by Chair, Shelby E. Harken, University of North Dakota

Approved by SAC June 2006

With assistance of Subcommittee members:

Bonnie A. Dede
University of Michigan

Lois M. Chan
University of Kentucky

Anton J. Olson
Northwestern University

Ruth A. Bogan
Rutgers University

Shannon L. Hoffmann
Brigham Young University

Daniel Solomon Lovins
Yale University Libraries

With input from Subcommittee members:

Diane Dates Casey
Governors State University

Rebecca J. Dean
OCLC

Giles Stewart Martin
OCLC

Lynn M. El-Hoshy
Library of Congress

Mary C. Lasater
Vanderbilt University

and non-member Assist. Prof. Joseph Tennis
University of British Columbia

Table of Contents

Part A. Report

Introduction

An information system managing its own subject access for a single resource can relatively easily produce a successful database. However, there is an increasing need to access multiple resources in multiple languages or with multiple thesauri or controlled vocabularies. To a point, multiple controlled vocabularies and knowledge organization systems (KOS) can be made to interoperate. However, without appropriate design, the resulting search results will be 'non-semantic' and of little value to users. Given that converging information systems — with their idiosyncratic histories and social functions — are likely to produce overlaps, seams, and gaps in the composite whole, the Subject Analysis Committee formed the Subcommittee on Semantic Interoperability to investigate what techniques are currently being employed by developers to minimize loss of meaning and create true semantic interoperability.

Work of the Subcommittee

Charge

> Survey the current state of international interoperability projects which focus on subject and/or classification data. Produce a document outlining "best practices" at a level of generality that is both flexible enough to be measured against a variety of actual projects and specific enough to be made operational in current or proposed projects.
>
> Specific tasks include, but are not necessarily limited to: a) an inventory of known semantic interoperability projects, with descriptions; b) an evaluation of selected projects in terms of those projects' stated objectives; c) an investigation of the various concepts involved in the harmonization of indexing languages.

To carry out its charge, the Subcommittee undertook a number of tasks which have resulted in the documents appended to this report.

- In order to survey the current state of international interoperability projects, the Subcommittee began with an extensive literature review. This literature review along with other background information is included in Part B.
- Based on the literature review and to help guide its work, the Subcommittee developed a glossary of terms used in discussing various semantic interoperability (SI) projects (Appendix I). During its investigations and discussions the Subcommittee developed a working definition of subject semantic interoperability which is given below:

  > The ability of two or more systems or components to exchange or harmonize cognate subject vocabularies and/or knowledge organization schemes to be used for effective and efficient resource discovery without significant loss of lexical or connotative meaning and without special effort by the user.

- Using the above definition, the Subcommittee identified 37 SI projects, which were then compiled into a list that included descriptions of the projects (Part C).
- The results of investigating the various concepts and issues involved in subject semantic interoperability were used to formulate criteria for evaluating and developing SI projects (Part D).
- From the criteria in Appendix D, the Subcommittee developed a Checklist which could be used to evaluate or design a SI project (Part E).
- Using the Checklist, members of the Subcommittee evaluated seven of the SI projects that had been identified and described in its list of projects. The project evaluations are in Part F.
- The Chair of the Subcommittee served on the ALCTS Metadata Enrichment Task Force. The Subcommittee and Task Force presented a joint program at the American Library Association 2004 Annual Conference entitled: "Enriching Subject Access." For more information about the Task Force and a description of the Program, see Appendix II.
- The Subcommittee compiled an annotated bibliography (Appendix III) which included the sources cited in the Literature Review and various other appendices, as well as other background readings not

cited elsewhere in this report or its appendices. The Subcommittee believes that this bibliography could aid others in the investigation of the concepts and various projects involved in semantic interoperability.

Based on the above list of tasks and accomplishments, the Subcommittee believes that it has met its Charge with one exception, i.e., the development of a "Best Practices" document. The reasons for this are discussed below in the Subcommittee's Findings.

Subcommittee Findings

1. The 37 projects in Appendix C fall into 2 broad categories.
    - Production projects, such as the H.W. Wilson Megathesaurus and the MACS (Multilingual Access to Subjects) Project. The goal of these projects is to develop a product or system that can be used by a large number of users in a setting in which semantic interoperability is needed. Most of the 21 projects in this category are still in development, and a few have ceased or become inactive.
    - Research/demonstration projects, such as the DARPA Unfamiliar Metadata Project and the HILT (High Level Thesaurus) Project. There are 14 projects in this category, most of which are completed or have become inactive due to a lack of funding. It is possible that a few of the active research projects might evolve into working production systems. Note that the Subcommittee was unable to classify two of the projects as production or research. Furthermore, lack of documentation made it difficult to determine if some of the projects had been permanently or temporarily suspended.
2. After reviewing the literature and examining various projects, the Subcommittee decided that a best practices document for semantic interoperability was premature. There were several reasons for this.
    - The Subcommittee was unable to find any existing tool that could be used to evaluate a semantic interoperability project. Consequently, the Subcommittee would have to first develop an evaluation tool (described in 3. below and Part E).
    - Once developed, the tool could be employed to evaluate selected projects in order to identify successful methods and models. The methods and models would form the basis of a best practices document. The Subcommittee's three year term proved insufficient to accomplish three major tasks: the development of a tool; the project evaluations; and the analysis of successful projects.
    - Finally, even with a completed evaluation tool, there were still only a few SI projects in full production, and these had not been in production long enough to yield much analyzable data about how successful they were in meeting their goals and objectives.
3. For its evaluation tool the Subcommittee developed a Checklist (Part E) comprising a number of questions to evaluate projects. To test its viability, the Subcommittee used the Checklist to evaluate seven projects (Part F). Based on these evaluations, the Subcommittee found that the Checklist could serve as a useful evaluation tool.
4. Many of the questions in the Checklist are the same ones that developers of SI projects need to answer as they design their projects. Therefore, the Subcommittee also concluded that the Checklist could serve as a guide to developers of SI projects.
5. The Subcommittee has been able to identify a few semantic interoperability projects that are in full production. Some examples are listed below:
    - AGROVOC Thesaurus {Food and Agricultural Association}
    - Art and Architecture Thesaurus (AAT) {Getty Research Institute}
    - Bilingual Subject Access {Library & Archives of Canada}
    - Classification Web {Library of Congress}
    - H. W. Wilson Megathesaurus
    - Renardus {Renardus Consortium}
    - Unified Medical Language System (UMLS) {National Library of Medicine}
    - WebDewey {OCLC}

    These projects share the following attributes:
    - A well developed master plan for life-cycle management and data migration

- Reliance on international standards
- A viable business model which provides ongoing financial support for the project
- Adequate staff, computer software and hardware to support the project

Conclusion

The need for improved semantic interoperability between and among vocabularies and knowledge organization schemes is undeniable and growing in importance. There is an ever-increasing need to create an environment by which even multiple portals could be accessed via subject metadata using software that is neutral and available ubiquitously or directly to the user, that could be copied by libraries for use in their own environment. In order to develop or improve a knowledge organization system including emerging options in semantic interoperability, scholars and practitioners need to be able to evaluate a wide variety of projects and stay current with the professional literature.

Based on its findings, the Subcommittee concludes that the development of a successful subject semantic interoperability project is a long and difficult process. It requires a substantial investment of financial, human and computer resources. The Subcommittee recommends using the information and tools in this report and its appendices to assist in developing a successful project incorporating subject semantic interoperability. Finally the Subcommittee concludes that since this field of endeavor is still relatively young and immature, it is too early to generate a set of Best Practices that could be used in developing a successful project. We are past the theoretical and basic research phase and into the development phase. Even though there are some successful projects in full production, more projects need to reach maturity and much more research needs to be done.

Part B. Background Information and Literature Review

Most online library systems worldwide utilize some type of controlled and, in many cases, multiple vocabularies. From a librarian's point of view keyword searching on the Internet has its limitations. Yet online catalogs exist in the Internet environment along with other remotely accessible databases which may utilize their own controlled vocabularies. Consequently during an information seeking experience, users may be presented with a myriad of thesauri and other controlled vocabularies. These same problems were identified by Marcia Bates in her report to the Library of Congress1. Although some Internet search engines function fairly well, the Subcommittee felt it needed to limit its focus to environments using some type of structured subject-based metadata or embedded metatags, rather than random or weighted keywords.

In the ALCTS report "Subject data in the metadata record,"2 functional requirements for subject access to Internet resources include: a) to assist searchers in identifying the most efficient paths for resource discovery and retrieval; b) help users focus their searches; c) enable optimal recall; d) enable optimal precision; e) assist searchers in developing alternative search strategies; f) provide all of the above in the most efficient, effective and economical manner.

In a networked environment, interoperability among disparate systems is necessary to allow users to search among resources from multiple sources generated and organized according to different standards and approaches. Lois Chan in her paper for the Bicentennial Conference on Bibliographic Control for the New Millennium 20003 summarized the interoperability requirements as follows: a) interoperability among different systems, metadata standards, and languages; b) flexibility and adaptability to different information communities, not only different types of libraries, but also other communities such as museums, archives, corporate information system, etc; c) extensibility and scalability to accommodate the need for different degrees of depth and different subject domains; d) simplicity in application, i.e. easy to use and to comprehend; e) versatility, i.e. the ability to perform different functions; and f) amenability to computer application.

Doerr (2001)4 notes that terminological resources are increasingly important for information retrieval in the networked environment, for retrieving documents by querying databases, and for using metadata employing controlled vocabularies. There is a growing interest in developing automated intermediaries to negotiate the differences between controlled vocabulary schemes so that a user can use a familiar set of terms to search collections using other vocabulary schemes.

Hunter (2001)5 points out that networked knowledge organization systems typically contain objects of mixed media types which are described using a multitude of diverse metadata schemas. Hence machine understanding of metadata descriptions which conform to schemas from different domains is a fundamental requirement for access. Yet, problems arise from the differences in terminological semantics and hierarchical relationships within various subject schemes.

Bella Hass-Weinberg6 in *Thesaurus Design for Semantic Information Management* suggested that "semantic information management" really just means vocabulary control; that ontology usually just means classification scheme, but sometimes is used as a synonym for thesaurus, and that taxonomy is just a synonym for classification. Subject headings lists, such as LCSH, are essential tools for managing information in a print environment, while true

---

1 Marcia J. Bates. *Task Force Recommendation 2.3, Research and Design Review: Improving User Access to Library Catalog and Portal Information : Final Report.* (2003). <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>

2 American Library Association. "Subject Data in the Metadata Record: Recommendations and rationale." Division of Association for Libraries and Technical Services, Cataloging and Classification Section, Subcommittee on Metadata and Subject Analysis (1999). <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees /subjectanalysis/metadataandsubje/subjectdata.htm>

3 Lois Mai Chan. "Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources: Issues and Challenges," in *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium* (Washington, DC: Library of Congress, Cataloging Distribution Service, 2001), p. 159-178. <http://www.loc.gov/catdir/bibcontrol/chan.html>

4 M. Doerr. "Semantic Problems of Thesauri Mapping," *Journal of Digital Information*, vol. 1, no. 8 (Mar. 26, 2001) <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

5 Jane Hunter. "MetaNet - a Metadata Term Thesaurus to Enable Semantic Interoperability between Metadata Domains," *Journal of Digital Information,* v. 1, no. 8 (Feb. 2001). <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/>

6 Daniel Lovins. "Summaries and Reflections of *Thesaurus Design for Semantic Information Management,"* a day-long seminar led by Prof. Bella Hass-Weinberg in New York, April 16, 2002. [email May 6, 2002]

thesauri are often more useful in the online environment (where they can be viewed hierarchically or combined in Boolean searches). Thesauri often run into the problem of needing to distinguish homographs. The problem in the selection of thesaurus terms is largely one of determining a set of appropriate lexemes, that is, the smallest units of lexicon that can be understood on their own terms. Synonymy is a common problem, though easily managed, e.g. Cancer, see Neoplasm. Other problems: having to choose between singular and plural, parts of speech, etc.

A subject portal connects users to a site focusing on a particular subject, with access to high-quality information resources, allowing aggregated cross-searching, streamlined account management, user profiling, or additional services.7 However, the user has to know to go to the portal. The number of subject portals is growing.

Renardus is an example of a subject gateway/portal project with a goal of providing users with integrated access by searching or browsing, through a single interface, to partners' quality-controlled subject gateways. Further goals are to develop and define organizational models, business models, technical solutions and metadata standards (Renardus Application Profile, Renardus Namespaces, Renardus Collection Level Description). The following elements can be used to define a quality-controlled subject gateway: a) selection and collection development, b) collection management, c) creation, d) resource description and metadata, e) subject access, f) search and browse access, g) standards, h) value-adding features. Each participating partner is responsible for mapping its metadata format to the common Renardus metadata format, derived from Dublin Core. A generic normalization toolkit with Z39.50 configuration files and a conversion script were provided. Each participant set up a Renardus server with their content normalized to the Renardus data model. A set of screens were built for the user interface: a) homepage, b) advanced search screen, c) index scan window, d) advanced search page after index scan, e) browse by subject screen, f) (preliminary) result screen, g) sorted result screen, h) participating gateways screen, and I) help (index) screen. In order to accomplish subject browsing, the various systems are mapped to a common classification system. The Renardus service provides access to resources from all kinds of subjects, published world-wide and in many languages and it is intended to be offered to an international multi-disciplinary community of users. The *Dewey Decimal Classification and Relative Index* (DDC) was chosen because of online availability and tools, global usage, suitability of the classification system and its functionality, frequency and character of the updates, research and methodological development efforts.8

About the same time the SAC Subcommittee on Semantic Interoperability was formed, NISO decided Z39.19 *Guidelines for the Construction, Format, and Management of Monolingual Thesauri* needed changing to meet the needs of the changing information environment. Included in their rationale was that "Developers of Internet and Intranet-accessible Web pages, databases, and information systems need better metadata to support non-expert information searches, and metadata developers are recognizing the value of incorporating high-quality, interoperable controlled vocabularies and taxonomies into their schemes."9

Literature Review

Some researchers have been making close examinations of individual projects, while others focus mainly on theoretical issues. Recent noteworthy articles of both types in the library and information science domain include

7 Resource Discovery Network. "Renardus;" "Subject Portals Development Project." 2002 <http://rdn.ac.uk/projects/#Euro>

8 Heike Neuroth and Traugott Koch. *Cross-browsing and Cross-searching in a Distributed Network of Subject Gateways: Architecture, Data Model, and Classification*, 2001. <http://www.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.html>

9 National Information Standards Organization. *Developing the Next Generation of Standards for Controlled Vocabularies and Thesauri.* <http://www.niso.org/committees/MT-info.html>

those by Chan & Zeng10, Tennis11, and Zeng & Chan12 while those in the computer science and database design domain include Dhamankar, et al.13 Park & Ram14 and Parsons & Wand15.

The work of Chan and Zeng is particularly useful for breaking down the many variables that make up subject semantic interoperability. One major variable involves the selection of data types, systems, or standards, which are to be made interoperable. There are projects, for example, that harmonize different controlled vocabularies in the same language, e.g., Northwestern University's mapping of Library of Congress Subject Headings (LCSH) and Medical Subject Headings (MeSH)16, the Wilson Megathesaurus17, and CARMEN's integration of multiple German thesauri; projects that aggregate subject vocabularies from among *different* languages and classification systems, e.g., the Unified Medical Language System (UMLS)18, the High Level Thesaurus (HILT)19 &20), and the DARPA Unfamiliar Metadata Project21; projects that map a controlled vocabulary to a universal classification system such as OCLC Online Computer Library Center's (OCLC) correlation of LCSH with DDC22 and the mapping of UDC to General Finnish Subject Headings23; and projects that harmonize *heterogeneous* classification schemes such as the American Mathematical Society's mapping of Mathematics Subject Classification to Schedule 510 of the DDC24.

Some interoperability variables are more methodological in nature. Following the work of Chan and Zeng25, these may be sorted into six categories: (1) "Derivation/Modeling," where a relatively simple vocabulary is derived

---

10 Lois Mai Chan. "Ensuring Interoperability among Subject Vocabularies and Knowledge Organization Schemes: A Methodological Analysis," (by Lois Mai Chan and Marcia Lei Zeng) *IFLA Journal* 28, 5/6 (2002):323-27. Paper presented at the 68th IFLA Council and General Conference, Glasgow, Scotland, Aug. 18-24, 2002. <http://www.ifla.org/IV/ifla68/prog02.htm http://www.ifla.org/IV/ifla68/papers/008-122e.pdf>

11 Joseph T. Tennis. "Layers of Meaning: Disentangling Subject Access Interoperability." *Advances in Classification Research,* 12 (2004)

12 Marcia Lei Zeng and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems," *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.

13 R. Dhamankar, Y. Lee, A. Doan, A. Halevy, and P. Domingos. "iMAP: Discovering Complex Semantic Matches between Database Schemas," in *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data,* Paris, France, 2004, p. 383-394.

14 J. Park and S. Ram. "Information Systems Interoperability: What Lies Beneath?*" ACM Transactions on Information Systems,* 22, no. 4 (2004): 595-632.

15J. Parsons and Y. Wand. "Choosing Classes in Conceptual Modeling." *Communications of the ACM,* 40 (1997): 63-69.

16 Tony Olson. "Integrating LCSH and MeSH in Information Systems," in *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 21-24.

17 Patricia S. Kuhr. "Putting the World Back Together: Mapping Multiple Vocabularies into a Single Thesaurus" (paper delivered at the IFLA satellite meeting: Subject Retrieval in a Networked Environment, OCLC, Dublin, Ohio, USA, 14-16 August 2001).

18 National Library of Medicine. *Fact sheet: UMLS Metathesaurus,* 2005. <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>

19 HILT. *High-level Thesaurus Project Proposal*, 2005. <http://hilt.cdlr.strath.ac.uk/AboutHILT/proposal.html> Not directly accessible; see <http://hilt.cdlr.strath.ac.uk/>

20 Dennis Nicholson, Dennis, Susannah Wake and S. Currier. "HILT: High Level Thesaurus Project: Investigating the Problems of Cross-Searching Distributed Services by Subject in the UK (presented at the meeting, "New Information Technology 2001." Tsinghua University, Beijing, China. 29-31 May). <http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/hiltchina2.ppt>

21 Michael Buckland, and others. "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies." *D-Lib Magazine*, 5, no. 1 (January 1999).

22 D. Vizine-Goetz, C. Hickey, A. H. Houghton, and R. Thompson. "Vocabulary Mapping for Terminology Services." *Journal of Digital Information,* 4, no. 4 (2004)

23 J. Himanka and V. Kautto. "Translation of the Finnish Abridged Edition of UDC into General Finnish Subject Headings." *International Classification,* 19, no. 3 (1992): 131-4+.

24 H. Iyer, H. and M.D. Giguere. "Towards Designing an Expert System to Map Mathematics Classificatory Structures." *Knowledge Organization,* 22, no. 3-4 (1995), 141-147.

25 Zeng, Marcia Lei and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.

from a more complicated pre-existing source, the way Faceted Application of Subject Terminology (FAST) is extracted from LCSH, for example; (2) "Translation/Adaptation" (e.g., the Bibliothèque Nationale's Rameau system, generated through translation and adaptation of LCSH and Canadian Subject Heading (CSH)); (3) "Satellite and Leaf Node Linking," where specialized thesauri (such as *The Legislative Indexing Vocabulary* (LIV), *Thesaurus for Graphic Materials*, *Global Legal Information Network* (GLIN)) are treated as satellites of a larger entity (LCSH) or conceptualized as leaves (specialized thesauri) attached to a tree structure (the larger thesaurus or vocabulary list); (4) "Direct mapping," where equivalence between differently-sourced terms and classification numbers are established, usually requiring intensive intellectual effort; (5) linking through a "temporary union list"; and (6) linking through a "thesaurus server protocol," as with the Alexandria Digital Library project.

Other variables discussed in the literature include: How are interoperable links stored and managed? Do they rely on authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers26, or some other structure? How are data and metadata *in general* stored? This is to say, are they being gathered into a union catalog (e.g., American Memory Project, NSDL), or living in a distributed system. How is data structured? For example, do they rely on XML, MARC, Dublin Core, and/or other metadata standards?

Yet another set of variables involves difference in degree of granularity, and logical structure. In the chapter "Compatibility and Convertibility" (pp. 179-216) of his *Vocabulary Control for Information Retrieval*, W.F. Lancaster points out several difficulties with which anyone attempting semantic interoperability (or "vocabulary reconciliation," as he puts it) must contend: how to reconcile vocabularies which have different degrees of specificity, different degrees of pre-coordination, overlap in subject matter, and different arrangements of hierarchy27. Vizine-Goetz, et al.28 paraphrases Lancaster's observations, and adds to them the more recently discussed problems of common versus scientific names from Doerr29, Olson30 and "differences in meaning resulting from different classifications of terms31 &32." In an automated environment there is also the problem of different methods and standards for encoding and preserving metadata.

---

26 Joseph T. Tennis. "Layers of Meaning: Disentangling Subject Access Interoperability." *Advances in Classification Research,* 12 (2004)

27 F. Wilfrid Lancaster. *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington, W.Va.: Information Resources Press, 1986.

28 D. Vizine-Goetz, C. Hickey, A. H. Houghton, and R. Thompson. "Vocabulary Mapping for Terminology Services." *Journal of Digital Information,* 4, no. 4 (2004)

29 M. Doerr. "Semantic Problems of Thesauri Mapping." *Journal of Digital Information*, vol. 1, no. 8 (Mar. 26, 2001) <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

30 Tony Olson. "Integrating LCSH and MeSH in Information Systems," in *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 21-24.

31 M. Doerr. "Semantic Problems of Thesauri Mapping." *Journal of Digital Information*, vol. 1, no. 8 (Mar. 26, 2001) <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

32 C. Whitehead. "Mapping LCSH into Thesauri: The AAT model," in T. Peterson and P. Moholt, eds., *Beyond the book: Extending MARC for subject access.* Boston: G.H. Hall, 1990. 81

Part C. Project Inventory

Using the definition of semantic interoperability developed by the Subcommittee, 37 projects were identified. The projects, along with information about them, are listed below alphabetically by name. As can be seen from the list, the amount of information that the Subcommittee was able to find varied from extensive for some projects to very little for others. Minimally, for each project the Subcommittee attempted to provide contact information, a URL, and/or a citation, so that a reader of this report could be directed to additional sources of information about a particular project. The Subcommittee's work was performed between 2002 and 2006. This list has not been updated since Aug. 2006. Since then, information about some of these projects may have changed, and some new projects may have begun. The Subcommittee attempted to be as comprehensive as possible and include all known major SI projects in the List, but of course some projects may have been overlooked. The Subcommittee would especially like to acknowledge the work of Marcia Lei Zeng and Lois Mai Chan, whose list of 18 SI projects33 (with descriptions) was the starting point for the Subcommittee's list.

Name	**ADL Thesaurus Protocol**	**F1**
Institution or agency	University of California, Santa Barbara	
URL	project site at http://alexandria.sdc.ucsb.edu/~gjanee/thesaurus/ demonstrator page at http://www.comp.glam.ac.uk/%7Efacet /formats/skos/skos_search.htm	
Contact information	Linda Hill, Ph.D. Alexandria Digital Library Project UC Santa Barbara Santa Barbara, California 93106 lhill@alexandria.ucsb.edu	
Project type	Production	
Project dates		
Status of project	Current with demonstrator project available for public viewing	
Languages		
Knowledge organization systems (KOS)	Thesauri	
Subject Coverage	General	
Description	Protocol for exchange of thesaurus information. Thesaurus data exchange tool. The Thesaurus Protocol is based on the ANSI/NISO (1993) Z39.19 thesaurus model and supports downloading, querying, and navigating thesauri.	
Methodology	In 2001-2002, the ADL Implementation team developed a Thesaurus Service Protocol. It is a lightweight, stateless, XML- and HTTP-based protocol designed to support searching and retrieval of thesaurus data. All that is required for its use is the development of a thesaurus server that can accept the specified XML-encoded queries and return the specified standard reports. The demonstrator system loads a thesaurus of choice (from a proffered list). The thesaurus can then be searched by keyword. Displays of results take several formats--alphabetical list of retrieved terms with USE references, hierarchical display, scope notes.	
User interface	The Thesaurus Protocol is based on the ANSI/NISO (1993) Z39.19 thesaurus model and supports downloading, querying, and navigating thesauri.	
Relevant standards	Extensible Markup Language (XML), XML Schemas, Hypertext Transfer Protocol (HTTP), ANSI/NISO Z39.19-1993 (thesaurus structure), XML Path Language (XPATH), Simple Knowledge Organization Systems (SKOS)	

33 Marcia Lei Zeng and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.

Notes	
Citation	ADL Thesaurus Protocol cited in recent articles *in Cataloging and Classification Quarterly*, 37, no. 3-4 (2004) Janée, G, S. Ikeda, S. and L.L. Hill. The ADL Thesaurus Protocol. Alexandria Digital Library Project, 2002. Available: http://www.alexandria.ucsb.edu/thesaurus/protocol/specification.html Binding, Ceri and Douglas Tudhope. "KOS at Your Service: Programmatic Access to Knowledge Organization Systems." *Journal of Digital Information*, 4, no. 4, art. 265 (Feb. 5, 2004) Marcia Lei Zeng and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.

Project Name	**AGROVOC**
Institution or Agency	Food and Agricultural Organization of the United Nations
URL	http://www.fao.org/agrovoc/
Contact Information	
Project Type	Production
Project Dates	
Project Status	Operational
Languages	Multilingual: Arabic, Chinese, Czech, English, French, Portuguese, Spanish
Knowledge Organization Systems (KOS)	Thesaurus
Subject Coverage	Agriculture
Description	Multilingual agricultural thesaurus.
Methodology	
User Interface	A user selects one of the languages and submits a string in that language to the AGROVOC database. The result is a list of terms and phrases that begin with the string. On the same page is a thesaural display of the first term in the list, and a list of equivalent terms in the other languages with links to thesaural displays of the term in these languages. A user select other terms from the list.
Relevant Standards	
Notes	
Citation	

Project Name	**Art & Architecture Thesaurus (AAT)**
Institution or Agency	Getty Research Institute
URL	http://www.getty.edu/research/conducting_research/vocabularies/aat/
Contact Information	Getty Research Institute 1200 Getty Center Drive, Suite 1100 Los Angeles, CA 90049-1688 (310) 440-7335; griweb@getty.edu
Project Type	Production
Project Dates	
Project Status	Operational
Languages	Multilingual
Knowledge Organization Systems (KOS)	Thesaurus
Subject Coverage	Art, Architecture and Material Culture
Description	The AAT is one of three Getty vocabularies which provide terminology and other information about the objects, artists, concepts,

	and places important to various disciplines that specialize in art, architecture and material culture.
Methodology	The AAT is a structured vocabulary containing terms and other information about concepts. Terms for any concept may include the plural form of the term, singular form, natural order, inverted order, spelling variants, various forms of speech, equivalent terms in various languages and synonyms of different etymological roots. Among these terms one is flagged as the preferred term or descriptor for the concept.
User Interface	Online public access catalogs and/or the Getty Web Site
Relevant Standards	MARC 21, XML
Notes	The other two Getty vocabularies are: the Thesaurus of Geographic Names (TGN), which contains names and other information about places; and the Union List of Artist Names, which contains names and other information about artists.
Citation	

Project Name	**BUBL**
Institution or Agency	Centre for Digital Library Service, University of Strathclyde
URL	http://bubl.ac.uk/
Contact Information	BUBL Information Service Centre for Digital Library Service Department of Computer and Information Sciences University of Strathclyde Livingstone Tower 26 Richmond Street Glasgow G1 1XH U.K. 0141 548 4752; bubl@bubl.ac.uk
Project Type	Production
Project Dates	1990-
Project Status	Operational
Languages	English
Knowledge Organization Systems (KOS)	Subject heading list and classification system BUBL subject tree Dewey Decimal Classification (DDC)
Subject Coverage	General
Description	BUBL is an Internet-based information service for the UK higher education community. BUBL LINK is a catalog of selected Internet resources for covering all academic subject areas.
Methodology	
User Interface	A user can browse for subjects through the BUBL subject tree; browse through the DDC hierarchy; or search by author, title, subject, DDC, or resource type.
Relevant Standards	
Notes	
Citation	

Project Name	**CAMed**
Institution or Agency	Columbia University and Kent State University
URL	http://circe.slis.kent.edu/mzeng/tmshome.html
Contact Information	Marcia Lei Zeng School of Library and Information Science Kent State University Kent, OH 44242-0001 mzeng@kent.edu

Project Type	Research/prototype
Project Dates	
Project Status	Current?
Languages	Multilingual: English, French
Knowledge Organization Systems (KOS)	Thesauri AcuBase Thesaurus AMED Thesaurus JICST MiliMedicalThesaurus
Subject Coverage	Complementary and Alternative Medicine
Description	An integrated thesaurus management and cross-thesaurus search system for complementary and alternative medicine (CAM).
Methodology	Four thesauri in the areas of CAM were normalized and stored in a thesaurus repository. This system allows a database manager to manage and edit his thesaurus in his local office through a Web interface, while the thesauri are deposited and hosted on a server at Kent State University.
User Interface	The cross-thesaurus search function allows a user to enter a term and search all or any of the thesauri in this repository. Software matches the query against the thesauri and gives back all fully- or partially-matched thesaurus entries. When a term is selected from the search results, a user can see the details of a thesaurus term entry (including the broader, narrower, and related terms, as well as non-preferred terms) and continue selecting among the terms displays. The term-search eventually enables a direct search in four bibliographical databases (samples) that have been integrated in the prototype. The term search function also extends to the full-text searching of all resources in the CAMed website.
Relevant Standards	
Notes	
Citations	Zeng, Marcia Lei and Yu Chen. "Features of an Integrated Thesaurus Management and Search System for the Networked Environment." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 122-128. Zeng & Chan (2004).

Project Name	**CARMEN (Content Analysis, Retrieval and Metadata: Effective Networking)**
Institution or Agency	
URL	http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.en
Contact Information	Dr. Friedrich Geisselmann Universitätsbibliothek Regensburg 93042 Regensburg Germany friedrich.geisselmann@bibliothek.uni-regensburg,de
Project Type	Research/prototype
Project Dates	
Project Status	Current?
Languages	Multilingual: English, German
Knowledge Organization Systems (KOS)	Thesauri, classification systems and subject headings lists Informationszentrum Sozialwissenschaften (IZT) {Thesaurus} German Institute for Educational Research Thesaurus

	Schlagwortnormdatei (SWD) {Subject heading list} Dewey Decimal Classification (DDC) Regensburger Verbund Klassifikation (RVK) Mathematics Subject Classification (MSC) Physics and Astronomy Classification Scheme (PACS)
Subject Coverage	Social Sciences, Mathematics, Physics, Astronomy
Description	The goal is to provide an integrated subject search in distributed databases representing different disciplines, taking into account the conceptual differences of the applied thesauri and classifications by cross concordances.
Methodology	Starting from alphabetical lists which contain descriptors from a specific subject area, the relationships between IZT, the German Institute for Educational Research Thesaurus and SWD are determined intellectually. After the relationships have been established, they are recorded in a link management system.
User Interface	
Relevant Standards	
Notes	
Citations	Kunz, M. Sachliche Suche in verteilten Ressourcen: Ein kurzer Überblick über neuere Entwicklungen [Subject retrieval in distributed resources: a short review of recent developments]. Paper presented at the 68th IFLA Council and General Conference, Aug. 18-24, 2002, Glasgow, UK. Available: http://www.ifla.org/IV/ifla68/papers/007-122g.pdf English translation available: http: //www.ifla.org/IV/ifla68/papers/007-122g.pdf Zeng & Chan (2004).

Name	**Classification Web**
Institution or agency	Library of Congress
URL	http://classweb.loc.gov/
Contact information	Cheryl C. Cook, Product Coordinator Library of Congress Cataloging Distribution Service Washington, DC 20541-4912 ccoo@loc.gov
Project type	Production
Project dates	
Status of project	Current, in production
Language	English
Knowledge organization systems (KOS)	Classification system, Subject heading list Library of Congress Classification (LCC) Library of Congress Subject Headings (LCSH) {Subject heading list}
Subject Coverage	General
Description	This project links LCC numbers to LCSH headings and vice versa.
Methodology	LCC numbers are added to LCSH authority records; and LCSH headings are added to LCC authority records.
User interface	In Classification Web users can move across the KOS through the links that have been established.
Relevant Standards	MARC 21
Notes	
Citation	Zeng and Chan (2004).

Name	Czech National Subject Gateway Project and Uniform Information Gateway
Institution or agency	National Library of the Czech Republic
URL	Uniform Information Gateway: http://www.jib.cz (User interface)
Contact information	
Project type	Production
Project dates	2nd version released March 2003
Status of project	Current
Language	Czech
Knowledge organization systems (KOS)	
Subject Coverage	General
Description	The Czechs explored building subject portals for online resources and existing bibliographic records. They surveyed the field of national bibliographical agencies to see what sources they use for subject terminology. Like other similar projects, this is an attempt to achieve interoperability through control of descriptive cataloging.
Methodology	Mapping was being done intellectually on the main classes and principal subdivisions level: in order to reach the highest possible accuracy in mapping process, it was necessary to use common auxiliary subdivisions. Contains four files: geographic, chronological, genre/form, and topical authority files. Subject categorization of heterogeneous information resources using Conspectus method is used. The scheme consists of mapping DDC and UDC. Topics authority terms contain English equivalents.
User interface	Aleph interface allows user to search subjects authority records or conspectus records in a number of languages.
Relevant standards	
Notes	
Citations	Stoklasova, Bohdana, Marie Balikova and Ludmila Celbova. "The Relationship between Subject Gateways and National Bibliographies in International Context." Paper presented at 69th IFLA General Conference and Council, 1-9 August 2003, Berlin. http://www.ifla.org/IV/ifla69/papers/054e-Stoklasova_Balikova_Celbova.pdf

Project Name	DARPA Unfamiliar Metadata Project
Institution or Agency	University of California Berkeley
URL	http://metadata.sims.berkeley.edu/GrantSupported/unfamiliar.html
Contact Information	Michael Buckland, Professor Emeritus School of Information Management and Systems University of California, Berkeley South Hall 203A Berkeley, CA 94720-4600 (510) 642-3159; buckland@sims.berkeley.edu
Project Type	Research/prototype
Project Dates	
Project Status	Complete?
Languages	Multilingual: English, French, German, Russian, Spanish
Knowledge Organization Systems (KOS)	Thesauri and classification systems INSPEC Thesaurus Medical Subject Headings (MeSH) U.S. Patent and Trade Office Patent Classification World Intellectual Property Organization International Patent

	Classification Library of Congress Classification in the Physical Sciences Standard Industrial Classification
Subject Coverage	Biotechnology, Physical Sciences, Technology
Description	"The objective of this project is to link ordinary language queries to unfamiliar indexes and classifications."
Methodology	
User Interface	Entry Vocabulary Modules are built to respond adaptively to a searcher's query posed in ordinary language. A searcher can enter an ordinary language query to a particular database, and the searcher will be presented with a ranked list of terms from the database's vocabulary. The searcher can then use these terms to perform a search of the database.
Relevant Standards	
Notes	This project was carried out under the auspices of the Metadata Research Program of the School of Information Management & Systems, University of California, Berkeley (http://metadata.sims.berkeley.edu). Two later projects build on the work of the Unfamiliar Metadata Project: the DARPA TIDES Project, Translingual Information Management Using Domain Ontologies; and the Seamless Searching of Numeric and Textual Resources, funded by the Institute of Museum and Library Services.
Citation	Buckland, Michael, and others. "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies." *D-Lib Magazine*, 5, no. 1 (January 1999). Available: http://www.dlib.org/dlib/january99/buckland/01buckland.html Zeng & Chan (2004).

Project Name	**DESIRE**
Institution or Agency	DESIRE Consortium
URL	http://www.desire.org/
Contact Information	Tracy Hooper, DESIRE Project Manager Institute for Learning and Research Technology University of Bristol 8-10 Berkeley Square Bristol BS8 1HH UK 44 117 928 7197; t.a.hooper@bristol.ac.uk
Project Type	Research
Project Dates	1998-2000
Project Status	Complete?
Languages	
Knowledge Organization Systems(KOS)	Subject gateways
Subject Coverage	
Description	The Project's focus was on enhancing existing European information networks for research users across Europe through research and development in three main areas: caching, resource discovery and directory services. The Project proposed development and support of subject gateway services, facilitating access to high-quality internet resources and development of services that would allow cross-browsing and cross-searching across gateways.
Methodology	The Project participants proposed a representation of the conceptual relationships typical of controlled vocabularies using the Resource Description Framework (RDF). It was hoped that such an approach would enable the use of generic RDF tools as a basis for mapping

	between subject vocabularies. The Project report included a proposal for a RDF/XML thesaurus schema that attempted to demonstrate how the RDF data model could represent a web of inter-related concepts and terms from more than one thesaurus. Registries were developed for metadata application profiles (http://desire.ukoln.ac.uk/registry/ra.php3); and metadata terminology (http://desire.ukoln.ac.uk/registry/element.php3)
User Interface	
Relevant Standards	
Notes	During the second phase of the Project (DESIRE II) some background work was conducted on subject vocabularies in order to support the development of interoperable subject gateways, especially with regard to multilinguality and the mapping of different vocabularies.
Citation	

Name	**The FACET Project**
Institution or agency	Hypermedia Research Unit School of Computing University of Glamorgan Pontypridd CD37 1DL Wales, UK
URL	http://www.comp.glam.ac.uk/~FACET/default.asp
Contact information	Douglas Tudhope (dstudhope@glam.ac.uk) Daniel Cunliffe (djcunlif@glam.ac.uk)
Project type	Demonstration
Project dates	Initial funding covered three year period, 2001-2003
Status of project	Current with demonstrators available for public viewing
Languages	English
Knowledge organization systems (KOS)	Thesauri; faceted thesauri
Subject Coverage	not subject specific; uses thesaurus terms and data from AAT as demonstration
Description	The objective of the FACET Project research has been to: "Develop and evaluate retrieval tools based on a matching function incorporating thesaurus semantic closeness measures." The FACET Project attempts to find a way to present thesaurus data to a searcher, to allow the user to search for appropriate resources from displayed thesaurus terms and to provide the searcher with behind the scenes expansion of a search based on concepts of the semantic relationships among thesaurus terms. One premise of the project is the value of the facet analysis model of thesaurus building. Demonstrators for the FACET Project make use of the Art and Architecture Thesaurus, as an example of a faceted thesaurus.
Methodology	The FACET system architecture comprises client and web browser interfaces, utilities that interact with data objects, and an SQL server database that serves the thesaurus information. In a recent (2004) publication, the developers of FACET state that their intention is to "move toward and open (Web service) platform … and build on a general programmatic KOS interface … rather than the custom API employed in the Web demonstrator."
User interface	Several web based search and display interfaces are proposed in the demonstrators
Relevant standards	XML; the developers are recently acknowledging that there needs to be a standardized protocol for the presentation of representation of

	thesaurus data; they mention the ADL protocol as a step in the right direction.
Notes	In short, the project attempts to present thesaurus data in a meaningful way to searchers, to propose expanded searching options by suggesting terms in context, and to allow searchers to use the discovered terms in a query of resources. Initial funding from Engineering and Physical Sciences Research Council (EPSRC), a UK government funding agency for research and training in engineering and the physical sciences (http://www.epsrc.ac.uk/default.htm)
Citation	Tudhope, Douglas, and others. "Compound Descriptors in Context: a Matching Function for Classifications and Thesauri." Presented at International Conference on Digital Libraries, Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital libraries, Portland, Ore., 2002, 84-93. Available http://delivery.acm.org/10.1145/550000/544235/p84-tudhope.pdf?key1=544235&key2=0184227511&coll=&dl=ACM&CFID=15151515&CFTOKEN=6184618 Binding and Tudhope (2004)

Project Name	**Finnish Project**
Institution or Agency	
URL	
Contact Information	
Project Type	Research
Project Dates	
Project Status	Prototype; research
Languages	Multilingual: English, Finnish
Knowledge Organization Systems (KOS)	Subject heading list and classification system General Finnish Subject Headings (GFSH) {Subject heading list} Universal Decimal Classification (UDC)
Subject Coverage	General
Description	This project converts assigned class numbers based on the Finnish abridged edition of UDC into GFSH headings.
Methodology	A dictionary was created that maps UDC numbers to GFSH headings. The dictionary was mechanically applied to convert the bibliographic databases.
User Interface	
Relevant Standards	
Notes	
Citation	Himanka, J. and V. Kautto. "Translation of the Finnish Abridged Edition of UDC into General Finnish Subject Headings." *International Classification*, 19, no. 3 (1992): 131-134. Zeng & Chan (2004).

Project Name	**HEREIN (The European Information Network on Cultural Heritage) Thesaurus**
Institution or Agency	European Heritage Network, Council of Europe
URL	http://www.european-heritage.net/sdx/herein/
Contact Information	
Project Type	Production
Project Dates	
Project Status	In development
Languages	Multilingual: English, French, Spanish

Knowledge Organization Systems (KOS)	Thesaurus
Subject Coverage	Cultural heritage
Description	This multilingual thesaurus is attached to the HERIN Project. It intends to offer a terminological standard for national policies dealing with architectural and archaeological heritage.
Methodology	Most of the terms in the thesaurus come from reports on cultural heritage policy in Europe, supplemented with additional terms issued from specialized documentary sources. Teams from Spain, France and the UK created separate lists of terms in their own languages. The three teams then compared their lists so as to obtain a pool of words with linguistic equivalencies in the three languages.
User Interface	Through the Project Web site, a user can either search for a specific term, or browse through the hierarchical classes.
Relevant Standards	
Notes	
Citation	Therond, Daniel. "Www.European-Heritage.Net: The European Heritage Network." *Cultivate Interactive*, issue 2, no. 16 (Oct. 2000). Available: http://www.cultivate-int.org/issue2/herein/ Zeng and Chan (2004).

Name	**HILT (High Level Thesaurus Project)**	**F4**
Institution or agency	Funded by JISC (Joint Information Systems Company)	
URL	http://hilt.cdlr.strath.ac.uk	
Contact information	Dennis Nicholson, Director of Research Centre for Digital Library Research c/o Andersonian Library, University of Strathclyde 101 St. James Road Glasgow G4 0NS 44 (0) 141 548 2102; d.m.nicholson@strath.ac.uk	
Project type	Pilot Project	
Project dates	2000-	
Status of project	Current	
Language	Multilingual	
Knowledge organization systems (KOS)	Thesauri, classification systems, subject heading lists Art and Architecture Thesaurus (AAT) Dewey Decimal Classification (DDC) Library of Congress Subject Headings (LCSH) UNESCO Thesaurus RDN terminologies Wordmap taxonomies set	
Subject Coverage	General and special	
Description	The pilot project (Phase II) will develop an online terminologies route map (or TeRM) that will map subject schemes to user terminologies and to each other.	
Methodology		
User interface		
Relevant Standards		
Notes	Phase I investigated the problem of searching and browsing across a number of distributed services using different indexing vocabularies and attempted to derive a set of recommendations to help facilitate cross-searching and browsing by subject between communities, services and initiatives. The results of these investigations led to HILT Phase II, the Pilot Project described above.	
Citation	Nicholson, Dennis and Susannah Wake. "HILT: Subject Retrieval in a	

	Distributed Environment." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 61-67. Available: http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/hilt-ifla.ppt Zeng & Chan (2004).

Name	**H.W. Wilson Megathesaurus for Omnifile Project**	**F3**
Institution or agency	H.W. Wilson	
URL	www.hwwilson.com/Databases/omnifile.cfm	
Contact information		
Project type	Production	
Project dates		
Status of project	Active, in production	
Language	English	
Knowledge organization systems (KOS)	Thesauri	
Subject Coverage	General	
Description	Merges knowledge organization systems of different structural types H.W. Wilson has developed a "megathesaurus" that gathers the vocabulary for all its indexes for inclusion in its Omnifile product. The Omnifile product now includes six of the 11 Wilson periodical files, plus all of the full text from the remaining five files. Eventually Omnifile will probably include all their files, but this may take some time, since the remaining five are very specialized. Files covering non-periodical material use different indexing vocabularies and do not form part of the Omnifile product.	
Methodology	Concepts merge into single terms, while the megathesaurus retains the terminology used in the separate indexes. The individual database products use the same terms as always; in the Omnifile product, the megathesaurus equivalent appears. Wilson has changed the vocabularies for individual products where conflict between indexes used to exist. Homographs (two words that look the same though they are not necessarily pronounced the same) are clarified by means of devices such as qualifiers, and if a term was used differently in two indexes, e.g., "writing" as composition versus learning to write has been resolved. Names used as subject descriptors appear uniformly across all files; only styling rules are applied to author names.	
User interface	Web, specifically, "WilsonWeb." Megathesaurus is largely invisible to the user.	
Relevant standards	Unknown	
Notes		
Citations	Kuhr, Patricia S. "Putting the World Back Together: Mapping Multiple Vocabularies into a Single Thesaurus." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 37-42. Milstead, Jessica. "Cross File Searching: How Vendors Help--and Don't Help--Improve Compatibility." *Searche*r, 7, no. 5 (May 1999)	

Project Name	**IMesh**
Institution or Agency	UKOLN: the UK Office for Library and Information Networking
URL	http://www.imesh.org
Contact Information	UKOLN c/o The Library University of Bath Bath BA2 7AY 44 1225 38658; imesh-toolkit@imesh.org
Project Type	Production
Project Dates	Sept. 1999 - July 2003
Project Status	
Languages	
Knowledge Organization Systems (KOS)	Subject gateways
Subject Coverage	
Description	The Project will build on existing subject software to develop a configurable, reusable and extensible toolkit for subject gateway providers.
Methodology	Components evolve independently but rely on each other to accomplish larger tasks. To achieve interoperability the goal is for components to be able to call on one another efficiently and conveniently.
User Interface	
Relevant Standards	Resource Description Framework (RDF), Structured Query Language (SQL)
Notes	NSF/JISC International Libraries Initiative.
Citation	

Project Name	**LCSH/MeSH Mapping Project**
Institution or Agency	Northwestern University Libraries
URL	http://www.library.northwestern.edu/public/
Contact Information	Tony Olson, Catalog Librarian Galter Health Sciences Library Northwestern University 303 East Chicago Ave Chicago, IL 60611 (312) 503-8125; ajolson@northwestern.edu
Project Type	Production
Project Dates	1990-
Project Status	Active, in development
Languages	English
Knowledge Organization Systems (KOS)	Subject heading lists Library of Congress Subject Headings (LCSH) Medical Subject headings (MeSH) Thesaurus
Subject coverage	General and medicine
Description	The goal of this project is to integrate LCSH and MeSH in online catalogs.
Methodology	Corresponding established headings in LCSH and MeSH are mapped, and the mapping data is entered into 7XX linking fields of LCSH and MeSH MARC 21 authority records. The data in these fields can be used to generate equivalent term references in an online catalog. The mapping data is continually updated to take into account changes in the two KOS.
User Interface	In online public access catalogs *see also* references will be provided

	between equivalent LCSH and MeSH headings.
Relevant Standards	MARC 21
Notes	The project is still in development because most library management systems do not yet index 7XX fields in authority records, and consequently do not supply linking references between equivalent LCSH and MeSH headings. The mapping data is available for use in other interoperability projects. Files of enhanced LCSH and MeSH authority records with the mapping data can be downloaded from the Northwestern public http site above.
Citation	Olson, Tony and Gary Strawn. "Mapping the LCSH and MeSH Systems." *Information Technology and Libraries*, 16, no. 1 (March 1997): 5-19. Zeng & Chan (2004).

Name	**LEAF (Linking and Exploring Authority Files)**
Institution or agency	Multiple European institutions; Dept. of Manuscripts, Staatsbibliothek zu Berlin Preussischer Kulturbesitz;
URL	http://www.crxnet.com/leaf/ No longer available.
Contact information	Name: WEBER, Jutta (Dr) Tel: +49-30-2662416 Fax: +49-30-2663007 Email: jutta.weber@sbb.spk-berlin.de
Project type	Research/prototype
Project dates	2001-2004 (Fifth Framework Programme)
Status of project	Completed
Languages	Multilingual
Knowledge information systems (KOS)	Name authority files
Subject Coverage	General
Description	Utility for creating universal name authority file. [From the web site of the Fifth Framework Programme] The beneficial potential of authority information is presently only partly utilized by cultural heritage organizations: libraries, archives, museums etc. are independently working with them without jointly exploiting this valuable resource. Public users are not involved in this scenario neighboring work in the commercial sector is not integrated. LEAF proposes a model for harvesting existing authority data and person name/corporate body information in a multilingual environment. Via user queries the LEAF system will automatically and dynamically create a common name authority file with links to organizations that provide information about a person or corporate body and/or items connected to them. The LEAF model will be applicable to all projects and co-operations that are dealing with cultural heritage data in all kinds of institutions by making authority information available to everyone involved. The project results will be implemented by extending an existing, fully functional, international online Search and Retrieval service network of OPACs that provides information about modern manuscripts and letters, the MALVINE project.
Methodology	LEAF develops a model architecture for establishing links between distributed authority records and providing access to them. The system allows uploads of the distributed authorities to the central system and automatically links those authorities concerning the same entity. Information which is retrieved as a result of a query will be stored in a pan-European "Central Name Authority File." This file will grow with

	each query and at the same time will reflect what data records are relevant to the LEAF users. Libraries and archives wanting to improve authority information will thus be able to prioritize their editing work. Registered users will be able to post annotations to particular data records in the LEAF system, to search for annotations, and to download records in various formats. The local authority data that is uploaded to the central LEAF system is originally encoded in different formats. In order to be able to compare individual records and thus make them available for further operations one common exchange format data needed to be identified into which all records, independently of their native format, could be converted. LEAF has adapted EAC for this purpose. The conversion module of the central LEAF system consists of data conversion routines for each local data structure which convert the uploaded or harvested local records into EAC XML and the different character sets into Unicode (UTF-8). The converted data are then further processed in the LEAF system. In addition to the converted form records are saved in their local formats as provided by the LEAF Data Providers.
User interface	None found (12/31/2004)
Relevant standards	XML, Encoded Archival Context (EAC)
Notes	most recent newsletter is 11/03 link to MALVINE yields a blank page 2004/12/31 most scheduled documentation of last 2 years not delivered online, including a final report
Citations	Kaiser, Max, Hans-Jorg Lieder, Kurt Majcen and Heribert Vallant. "New Ways of Sharing and Using Authority Information: the LEAF Project." *D-lib magazine*, 9, no. 11 (Nov. 2003), http://www.dlib.org/dlib/november03/lieder/11lieder.html

Project Name	**Library & Archives of Canada Bilingual Cataloguing**	**F2**
Institution or Agency	Library & Archives of Canada (L&AC)	
URL	http://www.collectionscanada.ca/csh/s23-120-e.html (link to information about CSH and relation to RVM)	
Contact Information		
Project Type	Production	
Project Dates		
Project Status	Operational	
Languages	Multilingual: English, French	
Knowledge Organization Systems (KOS)	Subject heading lists Canadian Subject Headings (CSH) {Subject heading list} Répertoire de vedettes-matières (RVM) {Subject heading list} Library of Congress Subject Headings (LCSH) {Subject heading list}	
Subject Coverage	General	
Description	To support the bilingual cataloging policy of the Library & Archives of Canada (L&AC), all publications cataloged by the L&AC are assigned subject headings in both official languages, English and French. References between equivalent CSH and RVM headings are displayed in the L&AC's online public access catalog, AMICUS.	
Methodology	Equivalent RVM and LCSH headings are entered into 7XX fields of CSH MARC21 authority records. The equivalent term references displayed in the online catalog are generated from these 7XX fields.	
User Interface	Online public access catalog	
Relevant Standards	MARC 21	
Notes	URL for AMICUS: http://www.collectionscanada.ca/amicus/index-e.html	

Citation	Armstrong, Pam. "Navigating Bilingual Subject Headings in AMICUS." Presented at the program, *Getting the Most Out of Subject References in the Online Catalog: Better Than It Used to Be?* American Library Association Annual Conference, Toronto, Ontario, June 21, 2003..

Project Name	**LIMBER (Language Independent Metadata Browsing of European Organizations)**
Institution or Agency	LIMBER Consortium
URL	http://www.limber.rl.ac.uk/ No longer available.
Contact Information	Michael Wilson Project Manager m.d.wilson@rl.ac.uk
Project Type	Production, Development
Project Dates	1999-2001
Project Status	Complete
Languages	Multilingual: English, French, German, Spanish
Knowledge Organization Systems (KOS)	Thesaurus: European Language Social Science Thesaurus (ELSST)
Subject Coverage	Social Sciences
Description	The goal of the LIMBER Project was to develop tools to support multilingual access to data distributed across the world wide web by using metadata and a multilingual thesaurus of terms in a restricted vocabulary.
Methodology	LIMBER is using W3C's RDF language as the technology to define metadata and the multilingual thesaurus, and FortH's SIS multilingual thesaurus management system as the base technology for the multilingual thesaurus server. The LIMBER tools will be generic, but they will be demonstrated by enhancing the existing NESSTAR data access system with multilingual capability, for the domain of social science. Another project FASTER is enhancing the categories of data that NESSTAR can retrieve. LIMBER is using the UK Data Archive's Hasset thesaurus of terms in social science as the starting point for a multilingual thesaurus for social science in English, French, Spanish and German. LIMBER is advancing the DDI metadata format for social science data to support multilingual access as a demonstration of multilingual access in the social science domain.
User Interface	Web Interface
Relevant Standards	Resource Description Framework (RDF), Data Documentation Initiative (DDI)
Notes	LIMBER is an EU IST programme funded research and development project.
Citation	Miller, Ken and Brian Mathews. “Having the Right Connections: the LIMBER Project.” *Journal of Digital Information*, 1, no. 8 (Feb. 5, 2001) Available: http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Miller/

Project Name	**MACS (Multilingual Access to Subjects)**	**F5**
Institution or Agency	Conference of European National Librarians. Project partners are: the Swiss National Library (SNL), Bibliothèque nationale de France (BnF), the British Library (BL), and Die Deutsche Bibliothek (DDB)	
URL	https://ilmacs.uvt.nl/pub/	
Contact Information	Patrice Landry, MACS Project Leader Chef du Catalogage matières Bibliothèque nationale Suisse Hallwylstrasse 15	

	3003 Berne Suisse Tel.: +41 31 324 06 25; Fax: +41 31 322 84 63 E-mail: patrice.landry@slb.admin.ch
Project Type	Production
Project Dates	
Project Status	In development
Languages	Multilingual: English, French, German
Knowledge Organization Systems (KOS)	Subject headings lists Schlagwortnormdatei (SWD) Répertoire d'autorité-matière encylopédique et alphabétique unifié (RAMEAU) Library of Congress Subject Headings (LCSH)
Subject Coverage	General
Description	MACS aims to provide multilingual subject access to library catalogs. MACS enables users to simultaneously search the catalogs of the project's partner libraries in the language of their choice (English, French, German).
Methodology	Equivalence links are created between the three subject headings lists used in the partner libraries' catalogs. The links are stored in the MACS Links Database. There are two search interfaces for the Database. (1) The Search Interface: allows users to browse headings and retrieve bibliographic records by using the links established between the concepts. The search interface uses the Z39.50 protocol. (2) The Link Management Interface: enables the creation and management of links between headings from the subject headings lists.
User Interface	Online Public Access Catalog
Relevant Standards	NISO Z39.50
Notes	The headings from the three lists are analyzed to determine whether they are exact or partial matches, of a simple or complex nature. The end result is neither a translation nor a new thesaurus but a mapping of existing and widely used KOS.
Citation	Freyre, Elisabeth and Max Naudi. "MACS: Subject Access Across Languages and Networks." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 3-10. Zeng & Chan (2004).

Project Name	**Merimee**
Institution or Agency	Ministère de la Culture et de la Communication - direction de l'Architecture et du Patrimoine
URL	http://www.culture.gouv.fr/documentation/merimee/accueil.htm
Contact Information	
Project Type	Production
Project Dates	
Project Status	Operational
Languages	Multilingual: English, French
Knowledge Organization Systems KOS)	Thesauri Le thesaurus de l'architecture Art and Architecture Thesaurus (AAT) English Heritage Thesaurus
Subject Coverage	Cultural heritage, art, architecture
Description	For the purpose of indexing complexes, buildings and built structures,

	Le thesaurus de l'architecture was created and mapped to AAT and the English Heritage Thesaurus.
Methodology	When mapping from *Le thesaurus de l'architecture* to the other thesauri, Boolean operators "AND" and "OR" are used to indicate equivalence in addition to the exact equivalence types, exact and partial.
User Interface	
Relevant Standards	
Notes	
Citation	Doerr, M. "Semantic Problems of Thesauri Mapping." *Journal of Digital Information*, vol. 1, no. 8 (Mar. 26, 2001) Available: http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/ Zeng and Chan (2004).

Project Name	**MSC and Schedule 510 in DDC**
Institution or Agency	University at Albany, State University of New York
URL	
Contact Information	Iyer Hemalata School of Information Science and Policy University at Albany, State University of New York hi651@albany.edu
Project Type	Research/prototype
Project Dates	
Project Status	
Languages	English
Knowledge Organization Systems (KOS)	Classification systems Mathematics Subject Classification (MSC) of the American Mathematical Society Dewey Decimal System (DDC) Schedule 510
Subject Coverage	Mathematics
Description	This project maps the MSC to the DDC 20, Schedule 510.
Methodology	
User Interface	
Relevant Standards	
Notes	
Citation	Iyer, H. and M. D. Giguere. "Towards designing an expert system to map mathematics classificatory structures." *Knowledge Organization,* 22, no. 3-4 (1995): 141-147. Zeng & Chan (2004).

Project Name	**OCLC Terminology Services**
Institution or Agency	OCLC Online Computer Library Center, Inc.
URL	http://www.oclc.org/research/projects/termservices/default.htm
Contact Information	Diane Vizine-Goetz Consulting Research Scientist OCLC Online Computer Library Center, Inc. 614-764-6084
Project Type	Production
Project Dates	
Project Status	Active, in development
Language	English
Knowledge Organization Systems KOS)	Thesauri, classification systems, subject heading lists Dewey Decimal Classification (DDC) ERIC Thesaurus

	Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc. (GSAFD) genre terms Library of Congress Classification (LCC) Library of Congress Subject Headings (LCSH) Library of Congress Children's Headings (LCSHac) Medical Subject Headings (MeSH) National Library of Medicine Classification (NLMC)
Subject Coverage	General, Education, Literature, Medicine
Description	The goal of this project is to offer accessible, modular, web-based terminology services by providing mappings from a term in one KOS to one or more terms in other KOS.
Methodology	
User Interface	
Relevant Standards	MARC 21, MARC 21 XML, Dublin Core, RDF (Other Project related standards can be found at: http://www.oclc.org/research/projects/termservices/resources/standards.htm)
Notes	Selected vocabularies have been made accessible for machine interaction and for downloading. E.g., the GSAFD vocabulary with mappings is accessible using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The GSAFD records are accessible to users via a browser (http://alcme.oclc.org/gsafd/) and to machines through the OAI-PMH Web services mechanism.
Citation	Vizine-Goetz, Diane. "Terminology Services: Making Knowledge Organization Schemes More Accessible to People and Computers." *OCLC Newsletter*, 266 (October/November/December 2004). Available: http://www.oclc.org/news/publications/newsletters/oclc/2004/266/research.html

Project Name	**Polish Project**
Institution or Agency	Institute for Scientific, Technical and Economic Information (Warsaw, Poland)
URL	
Contact Information	
Project Type	Research
Project Dates	1992-
Project Status	
Languages	Multilingual: English, Polish
Knowledge Organization Systems (KOS)	Thesauri, classification systems, subject headings lists Polish Thematic Classification (PTC) Subject Heading Language (SHL) of the National Library in Warsaw Thesaurus of Common Topics (TCT) Universal Decimal Classification (UDC)
Subject Coverage	General
Description	The goal of this project was to establish concordances for the four KOS listed above.
Methodology	PTC was chosen as the master language, whereas the others served as target languages.
User Interface	
Relevant Standards	
Notes	
Citation	Scibor, E. and J. Tomasik-Beck. "On the Establishment of Concordances between Indexing Languages of Universal or Interdisciplinary Scope (Polish Experiences)." *Knowledge Organization*, 21 (1994): 203-212. Zeng and Chan (2004).

Name	RDN Subject Portals Project (SPP)	F6
Institution or agency	Funded by JISC (Joint Information Systems Company), which is supported by UK institutions of higher learning	
URL	http://www.portal.ac.uk/spp/	
Contact information		
Project type	Production	
Project dates	Phase 1: 2001-2003; Phase 2: 2003-2004	
Status of project	Active	
Languages		
Knowledge organization systems (KOS)	Thesauri, classification systems, subject portals	
Subject Coverage		
Description	Cataloging of online resources with subject categories controlled by proprietary hierarchical list. From the web site, "The Resource Discovery Network is the UK's free national gateway to Internet resources for the learning, teaching and research community." Currently eight subject portals are available: Altis (hospitality, leisure, sport and tourism), Artifact (arts and creative industries), Biome (health and life sciences), EEVL (engineering, mathematics and computing), GEsource (geography and environment), Humbul (humanities), PSIgate (physical sciences) and SOSIG (social sciences). The first phase of the project was to build a Z39.50 cross search prototype at three RDN hubs, SOSIG, EEVL, and BIOME. The second phase adds HUMBUL and PSIgate.	
Methodology	Using SOSIG, the social sciences portal as an example, the methodology is to select high quality electronic resources, e.g. web sites, and catalog them. Keyword descriptors provide subject access. In addition, each resource is assigned a subject category, chosen from hierarchically presented terms. Source of the hierarchy is not revealed. Terms probably are assigned by internal staff (as opposed to resource contributors) as cataloging record does not have a place to input subject categories. Each portal has its own cataloging guidelines and suggested thesauri.	
User interface	Web based search interfaces for each individual portal as well as a simple keyword search, from the project home page, that searches across all portals. A search on "business" retrieves over 3300 pages of brief descriptive records.	
Relevant standards	Z39.50, various subject thesauri	
Notes	The first phase of the project (2000-2001) was to build a Z39.50 cross search prototype at three RDN hubs, SOSIG, EEVL, and BIOME. The second phase ads HUMBUL and PSIgate. Sites are selected on the basis of selection criteria, cataloged following consistent practices, and analyzed by people with expertise in the relevant subject discipline. Links are checked daily in an automated process and all entries are updated regularly by subject specialists. These are classified using an appropriate controlled vocabulary. Basically, this appears to be a cataloging project, with a Z39.50 search engine that does cross-portal searching.	
Citation		

Name	RENARDUS
Institution or agency	Renardus Consortium
URL	http://www.renardus.org

Contact information	(See Web site above)
Project type	Production
Project dates	
Status of project	Active
Language	Multilingual: English, Dutch, French, Finnish, German
Knowledge organization systems (KOS)	Classification system, subject gateways Dewey Decimal Classification (DDC)
Subject Coverage	General
Description	Renardus provides integrated search and browse access to records from individual participating subject gateway services across Europe.
Methodology	The Renardus Subject Gateways map their local browsing structures and classification systems to DDC.
User interface	Users can search for Internet resources by browsing a subject hierarchy (based on DDC) or by searching for specific terms in titles, subjects, description, creator and DDC.
Relevant Standards	
Notes	
Citation	Koch, Traugott, Heike Neuroth, and Michael Day. "Renardus: Cross-browsing European Subject Gateways via a Common Classification System (DDC)." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 25-33. Available: http://www.ukoln.ac.uk/metadata/renardus/papers/ifla-satellite/ifla-satellite.html Zeng & Chan (2004).

Project Name	**ROADS (Resource Organization and Discovery in Subject-based services)**
Institution or Agency	UKOLN: the UK Office for Library and Information Networking
URL	http://roads.opensource.ac.uk/
Contact Information	
Project Type	Production
Project Dates	
Project Status	Completed, no longer active.
Languages	
Knowledge Organization Systems (KOS)	Subject gateways
Subject Coverage	
Description	**ROADS** is a set of software tools to enable the set up and maintenance of Web based subject gateways.
Methodology	
User Interface	**ROADS** is a software tool-kit allowing gateway managers to pick and choose what parts of the software they require whilst allowing the integration of other software according to requirement.
Relevant Standards	
Notes	The ROADS project has now ended and the Web pages are no longer maintained. The ROADS software is no longer available for download. It is suggested that those interested in subject gateway tools should instead look at the Scout Portal Toolkit (http://scout.wisc.edu/Projects/SPT/). The original ROADS manual is still available from the Web site.
Citation	

Project Name	**SAB and DDC**
Institution or Agency	Swedish Royal Library
URL	
Contact Information	
Project Type	Production
Project Dates	
Project Status	
Languages	Multilingual: Swedish, English
Knowledge Organization Systems (KOS)	Classification systems Klassifikationssystem för svenska bibliotek (SAB) Dewey Decimal Classification (DDC)
Subject Coverage	General
Description	A concordance between SAB, 7th ed. and DDC, 21st ed.
Methodology	
User Interface	
Relevant Standards	
Notes	
Citation	Leth, P. "Report from Sweden: Concordance Dewey - SAB." *Newsletter* (IFLA Section on Classification and Indexing), 24 (2001): 34. Zeng & Chan (2004).

Project Name	**SALT (Standards-based Access service to multilingual Lexicons and Terminologies)**
Institution or Agency	Localization Industry Standards Association (LISA)
URL	http://www.loria.fr/projets/SALT/
Contact Information	Alan K. Melby Dept. of Linguistics Brigham Young University 2129 JKHB Provo, Utah 84602 U.S. 801-422-2144; akm@byu.edu
Project Type	Production
Project Dates	2001-2002
Project Status	Inactive: Absorbed by the TBX Project
Languages	Multilingual
Knowledge Organization Systems (KOS)	Terminology databases Machine translation lexicons
Subject Coverage	General: multilingual lexicons and terminologies
Description	SALT is a consortium of academic, government, association, and commercial groups in the U.S. and Europe who are working together on the task of testing, refining, and implementing a universal putting together format for the interchange of terminology databases and machine translation lexicons.
Methodology	
User Interface	
Relevant Standards	XML, Translation Memory eXchange (TMX), Machine-Readable Terminology Interchange Format (MARTIF) (ISO 12200), Open Lexicon Interchange Format (OLIF), Unicode
Notes	An open source project
Citation	

Project Name	Simple Knowledge Organization System (SKOS) Mapping
Institution or Agency	W3C
URL	http://www.w3.org/2004/02/skos/mapping/
Contact Information	Alistair Miles E-Information, Business and Information Technology Dept. CCLRC Rutherford Appleton Laboratory Chilton, Didcot, Oxfordshire QX11 0QX UK 44-1235-445440; a.j.miles@rl.ac.uk
Project Type	Development
Project Dates	
Project Status	Inactive
Languages	Multilingual
Knowledge Organization Systems (KOS)	Thesauri Classification systems Subject heading lists Taxonomies Terminologies Glossaries
Subject Coverage	General
Description	This project is an application of the RDF that can be used to express mappings between concepts from different KOS as an RDF graph.
Methodology	
User Interface	
Relevant Standards	Resource Description Framework (RDF)
Notes	Development of SKOS Mapping was initiated by the Semantic Web Advanced Development for Europe Project.
Citation	

Project Name	TBX (TermBase eXchange)
Institution or Agency	Localization Industry Standards Association (LISA)
URL	http://www.lisa.org/standards/tbx/
Contact Information	Alan K. Melby Dept. of Linguistics Brigham Young University 2129 JKHB Provo, Utah 84602 U.S. 801-378-2144; akm@byu.edu
Project Type	Production
Project Dates	1987-
Project Status	Current, in production: Absorbed the SALT Project
Languages	Multilingual
Knowledge Organization Systems (KOS)	Terminology databases Machine translation lexicons
Subject Coverage	General: multilingual lexicons and terminologies
Description	TBX is an open XML-based standard format for terminological data. This standard provides a number of benefits so long as TBX files can be imported into and exported from most software packages that include a terminological database. This capability will greatly facilitate the flow of terminological information throughout the information cycle both inside an organization and with outside service providers. In addition, terminology that is made available to the general public will become much more accessible to humans and more easily integrated into existing terminological resources.

Methodology	An organization requires a translation, and supplies terminology for the translation. If the organization's terminology is in TBX, and the localization tool supports TBX, then there is full reuse of the data. A document production system includes multiple terminology-aware components. TBX can be used as the interchange format between components. TBX facilitates information interchange among term bases with different data models.
User Interface	
Relevant Standards	XML, TMX, MARTIF (ISO 12200), OLIF
Notes	An open source project
Citation	

Name	**UMLS (Unified Medical Language System)**	**F7**
Institution or agency	National Library of Medicine	
URL	http://www.nlm.nih.gov/pubs/factsheets/umls.html	
Contact information		
Project type	Production	
Project dates		
Status of project	Operational	
Language	Multilingual	
Knowledge organization systems (KOS)	Thesauri, classification systems, subject heading lists, coding systems	
Subject Coverage	Medicine, Health, Biological Sciences, and related areas	
Description	The UMLS consists of three Knowledge Sources: the UMLS Metathesaurus, the SPECIALIST lexicon, and the UMLS Semantic Network. The Metathesaurus is a database containing semantic information about biomedical concepts, their various names, and the relationships among them.	
Methodology	The Metathesaurus is built from over 100 biomedical source vocabularies, some in multiple languages. The 2003 edition includes 875,255 concepts and 2.14 million concept names. The UMLS Semantic Network is used for mapping index terms from different thesauri through its 134 semantic types which provides a consistent categorization of all concepts represented in the Metathesaurus.	
User interface		
Relevant Standards		
Notes		
Citation	Zeng and Chan (2004)	

Name	**VIAF (Virtual International Authority File)**
Institution or agency	OCLC Online Computer Library Center, Inc.
URL	http://www.oclc.org/research/projects/viaf/
Contact information	Edward T. O'Neill Consulting Research Scientist OCLC Online Computer Library Center, Inc 614-764-6074; oneill@oclc.org
Project type	Research
Project dates	
Status of project	Active?
Language	English
Knowledge organization systems (KOS)	Name authority files
Subject Coverage	General
Description	VIAF explores virtually combining the name authority files of the

	Library of Congress (LC) and Die Deutsche Bibliothek (DDB) into a single name authority file.
Methodology	OCLC will use its proven software to match and link authority records for personal names from DDB to corresponding authority records from LC.
User interface	
Relevant Standards	MARC 21
Notes	
Citation	Tillett, Barbara. "A Virtual International Authority File." Presentation to the Giornata di studio sul controllo di autorità nel Servizio Bibliotecario Nazionale Nov. 22, 2002. Available: http://www.iccu.sbn.it/upload/documenti/Tillett.ppt

Project Name	**VILIB (Virtual digital LIBrary on a Europe-wide level)**
Institution or Agency	University of Cologne Faculty of Management, Economics and Social Sciences
URL	http://www.islp.uni-koeln.de/aktuell/vilib/
Contact Information	Prof. Dr. Dietrich Seibt Research Group Information Systems and Learning Processes University of Cologne dietrich.seibt@uni-koeln.de
Project Type	
Project Dates	1998-1999
Project Status	
Languages	Multilingual: English, French, German, Spanish
Knowledge Organization Systems (KOS)	CANAL/LS (Catalog with natural multilingual Access / Linguistic server)
Subject Coverage	General
Description	The goal of the VILIB project was to develop a system which enables a “cross-lingual” search of catalog records and full-text documents in every library connected to the Internet.
Methodology	
User Interface	Online Public Access Catalog
Relevant Standards	Z39.50
Notes	
Citation	

Project Name	**WebDewey**
Institution or Agency	OCLC Online Computer Library Center, Inc.
URL	http://www.oclc.org/dewey/versions/webdewey/
Contact Information	
Project Type	Production
Project Dates	
Project Status	Operational
Language	English
Knowledge Organization Systems (KOS)	Classification system and subject heading list Dewey Decimal Classification (DDC) Library of Congress Subject Headings (LCSH) {Subject heading list}
Subject Coverage	General
Description	DDC numbers are linked to LCSH headings in MARC records.
Methodology	The linking is carried out intellectually or statistically where feasible.
User Interface	The linking of DDC and LCSH facilitates the subject cataloging and classification process by requiring only the identification of either the appropriate class number or subject headings for each document.

Relevant Standards	MARC 21
Notes	
Citation	Zeng & Chan (2004).

Name	**XML Topic Maps (XTM)**
Institution or agency	TopicMaps.Org Consortium
URL	http://www.topicmaps.org/xtm/1.0/
Contact information	
Project type	
Project dates	
Status of project	
Language	
Knowledge organization systems (KOS)	Topic maps
Subject Coverage	
Description	XTM 1.0 is an abstract model and XML grammar for interchanging Web-based topic maps.
Methodology	
User interface	
Relevant Standards	ISO 13250
Notes	Topic maps are a new ISO standard for describing knowledge structures and associating them with information resources.
Citation	

Part D. Criteria for Evaluating and Developing Subject Semantic Interoperability

The criteria in this appendix are based on background information and literature review. They comprise the distillations of several general discussions, during which a number of issues were raised and reviewed. From these criteria came the checklist and project reviews. For a more detailed description of the terms and concepts listed below, see the Glossary (Appendix B).

Definition of Subject Semantic Interoperability

The ability of two or more systems or components to exchange or harmonize cognate subject vocabularies and/or knowledge organization schemes to be used for the purposes of effective and efficient resource discovery without significant loss of lexical or connotative meaning and without special effort by the user.

Goal of Subject Semantic Interoperability

To enable developers to create an environment, system, or method by which multiple portals, including library online public access catalogs (OPAC), could be accessed via subject metadata, using software that is neutral and available ubiquitously or directly to the user.

Attributes of a Successful Subject Semantic Interoperability Solution

A successful subject semantic interoperability solution will address and resolve the following environmental challenges, which may occur within a single system or between multiple systems:

- multiple metadata standards
- multiple languages
- different subject domains
- different levels of indexing
- multiple subject indexing vocabularies and knowledge organization schemes

A successful subject semantic interoperability solution will address and resolve the following semantic challenges:

- synonyms
- homographs
- singulars and plurals
- parts of speech
- cultural differences affecting meaning
- narrower and broader and related terms
- syntactical differences

A successful subject semantic interoperability solution will utilize one or more of the following techniques or tools:

- harmonization of indexing languages
- mapping between different knowledge organization systems
- switching languages
- concordance tables
- front-end thesauri or front-end "cluster"
- metathesauri
- semantic networks
- multilingual thesauri
- controlled vocabularies
- authority files

A successful subject semantic interoperability solution will provide for the user:

- a single search interface

- an interface that is easy to use and comprehend
- browsing and searching capabilities
- improved retrievability of web resources
- the capability to access information in the preferred language or script
- different views of the resource description, depending on the user's particular interest, perspective or requirements.
- assistance in identifying the most efficient paths for resource discovery
- optimal recall and precision
- assistance in developing alternative search strategies
- consistent look and feel of displays
- explanations for variations and inconsistencies in terms that describe similar concepts

The subcommittee also recommends that a successful subject semantic interoperability solution should:

- be amenable to computer application
- be adaptable to gateways/portals, bibliographic databases or catalogs
- adapt to standards such as Z39.50, XML, etc.
- be supported by a business model
- develop shareable technical solutions and metadata standards (e.g. Renardus Application Profile, Renardus Namespaces, Renardus Collection Level Description)
- facilitate sharing in order to reduce cataloguing/indexing cost
- aid in the creation and maintenance of subject-related databases or authority records
- be extensible and scalable

Part E. Checklist for Evaluating and Developing Subject Semantic Interoperability Projects

Utilizing the criteria in Appendix D and building on the work of Joseph Tennis34, the Subcommittee developed the following checklist. It is intended to be used as an evaluation tool and to assist developers of semantic interoperability projects.

Checklist

1. Types of Data Being Integrated

			Does the project have:
Yes	**No**	**Unknown**	a) different controlled vocabularies in same language?
Yes	**No**	**Unknown**	b) different controlled vocabularies in different languages?
Yes	**No**	**Unknown**	c) different classification schemas (e.g. DDC, UDC, LCC)?
			If yes, which ones?
Yes	**No**	**Unknown**	d) controlled vocabularies combined with classification schemas?
Yes	**No**	**Unknown**	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
			If yes, which ones?
Yes	**No**	**Unknown**	f) different communication protocols?
Yes	**No**	**Unknown**	g) other:

2. Autonomy and Integrity of Constituent Parts

Yes	**No**	**Unknown**	a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes	**No**	**Unknown**	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
Yes	**No**	**Unknown**	b) Is full complement of metadata and indigenous subject hierarchies preserved?
			If so, how?

34 Joseph T. Tennis. "Layers of Meaning: Disentangling Subject Access Interoperability." *Advances in Classification Research,* 12 (2004)

Yes	No	Unknown	
			c) Does project rely on principle of least common denominator?
			If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core). How does the use of least common denominator affect the quality of service?
Yes	No	Unknown	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
Yes	No	Unknown	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)

3. Reconciliation of Heterogeneous Vocabularies

Yes	No	Unknown	
			a) How are correlations established when a single term in one source has no equivalent term in the other?
Yes	No	Unknown	b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
Yes	No	Unknown	c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
Yes	No	Unknown	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes	No	Unknown	
			a) Does project provide high or satisfactory levels of precision and recall?
Yes	No	Unknown	b) To what extent does project rely on pre-coordination?
			If mostly post-coordinate, then:
Yes	No	Unknown	i) by what means is recall maximized?
Yes	No	Unknown	ii) by what means is precision maximized?

Yes / No / Unknown	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
	d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes / No / Unknown	i) Locate entities in the system via surrogates (find)
Yes / No / Unknown	ii) Identify a surrogate that matches an entity (collocate)
Yes / No / Unknown	iii) Select an entity appropriate to a user's need via surrogates (choice facilitation)
Yes / No / Unknown	iv) Obtain access to the entity via the system and its surrogates (acquisition)
Yes / No / Unknown	v) Navigate the system and its surrogates (navigation)
Yes / No / Unknown	e) Has developer released beta version for general testing?
Yes / No / Unknown	f) Have user satisfaction surveys been conducted?

5. Ease of Use

	Does the project have:
Yes / No / Unknown	a) intuitive interface for data entry, searching, browsing, etc.?
Yes / No / Unknown	b) automated validation, mapping, metadata extraction, etc., as much as possible?
Yes / No / Unknown	c) availability of documentation?

6. Long-term Viability

	Does the project have a:
Yes / No / Unknown	a) master plan for life-cycle management and data migration?
Yes / No / Unknown	b) reliance on open-source international standards versus proprietary standards?
Yes / No / Unknown	c) viable business model (e.g., not based exclusively on research grant with likely expiration)?

Part F. Project Evaluations

Introduction

From its Project List (Appendix C) the Subcommittee selected seven projects to be evaluated using the Checklist. These evaluations were to test the viability of the Checklist as an evaluation tool. The criteria used to select the projects are listed below.

- The project had to be active and either in production or in development. Research/demonstration projects or production projects that had become inactive would not be evaluated.
- The Subcommittee wanted to evaluate projects exhibiting a wide range of methodologies and standards.
- The Subcommittee selected projects which attempted to integrate different types of knowledge organization systems, such as thesauri or other types of controlled vocabularies, classification systems, subject portals, etc.
- Yes/No/Unknown boxes are not checked when the response is "known" but the response is textual.

F1. ADL Thesaurus Protocol

URL: http://alexandria.sdc.ucsb.edu/~gjanee/thesaurus/

Description: Protocol for exchange of thesaurus information. Thesaurus data exchange tool.

The Thesaurus Protocol is based on the ANSI/NISO (1993) Z39.19 thesaurus model and supports downloading, querying, and navigating thesauri.

Protocol: XML- and HTTP-based protocol

Standards: ANSI/NISO Z39.19-1993: Guidelines for the Construction, Format, and Management of Monolingual Thesauri. <http://www.niso.org/standards/resources/Z39-19.html> (Sept. 1, 2006)

Hierarchy: hierarchy of terms above (broader than) or below (narrower than) a starting preferred term, including the starting term itself. The hierarchy is indicated by the nesting of XML elements.

The protocol provides five independent, stateless services.

Queries the thesaurus by term name and returns a list of the matching terms. operator is the matching operator to employ.

For more information about the project, consult the sources below:

The Alexandria Digital Earth Modeling System (ADEPT) : Towards a Distributed Digital Model of the Earth in Support of Learning. <http://www.alexandria.ucsb.edu/adept/proposal.pdf> (2003)

Buchel, Olha and Anita Coleman. "How Can Classificatory Structures be Used to Improve Science Education?" *Library Resources & Technical Services*, 47, no. 1 (2003): 4-15

Janée, Greg, Satoshi Ikeda, Linda L. Hill. *The ADL Thesaurus Protocol.* 2003. <http://alexandria.sdc.ucsb.edu/~gjanee/thesaurus/specification.html> (April 9, 2003).

Checklist

1. Types of Data Being Integrated

			Does the project have:
Yes	**No**	**Unknown**	a) different controlled vocabularies in same language?
√			
Yes	**No**	**Unknown**	b) different controlled vocabularies in different languages?
	√		
Yes	**No**	**Unknown**	c) different classification schemas (e.g. DDC, UDC, LCC)?
	√		
			If yes, which ones?

Yes	No	Unknown	
	√		d) controlled vocabularies combined with classification schemas?
Yes	**No**	**Unknown**	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
√			
			If yes, which ones?
			XML and HTTP-based protocol
Yes	**No**	**Unknown**	f) different communication protocols?
		√	
Yes	**No**	**Unknown**	g) other:
		√	

2. Autonomy and Integrity of Constituent Parts

Yes	No	Unknown	
√			a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes	**No**	**Unknown**	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
	√		
			There are 5 ways to search the data: 1. Hierarchy – Search asking for broader terms 2. Hierarchy – Search asking for narrower terms 3. String 4. Single term 5. Boolean searches
Yes	**No**	**Unknown**	b) Is full complement of metadata and indigenous subject hierarchies preserved?
√			
			If so, how?
			Hierarchies are preserved through broader and narrower terms
Yes	**No**	**Unknown**	c) Does project rely on principle of least common denominator?
√			
			If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core. How does the use of least common denominator affect the quality of service?
			Hierarchies are preserved through broader and narrower terms
Yes	**No**	**Unknown**	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
			Distributed database
Yes	**No**	**Unknown**	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
		√	

3. Reconciliation of Heterogeneous Vocabularies

Yes	No	Unknown	
			a) How are correlations established when a single term in one source has no equivalent term in the other?
			Used-for terms are listed

Yes	No	Unknown	
		√	b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
Yes	No	Unknown	c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
			Thesauri provide a basis for resolving semantic inconsistencies. By knowing which ontologies are used in different contexts, and by mapping between them, it is possible to make appropriate semantic correlations between different information sources.
Yes	No	Unknown	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?
		√	

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes	No	Unknown	
		√	a) Does project provide high or satisfactory levels of precision and recall?
Yes	No	Unknown	b) To what extent does project rely on pre-coordination?
		√	
			If mostly post-coordinate, then:
Yes	No	Unknown	i) by what means is recall maximized?
		√	
Yes	No	Unknown	ii) by what means is precision maximized?
		√	
Yes	No	Unknown	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
√			
			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes	No	Unknown	i) *Locate* entities in the system via surrogates (find)
√			
Yes	No	Unknown	ii) *Identify* a surrogate that matches an entity (collocate)
√			
Yes	No	Unknown	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
	√		
Yes	No	Unknown	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
		√	
Yes	No	Unknown	v) *Navigate* the system and its surrogates (navigation)
√			
Yes	No	Unknown	e) Has developer released beta version for general testing?
		√	

Yes	No	Unknown	f) Have user satisfaction surveys been conducted?
		√	

5. Ease of Use

			Does the project have:
Yes	**No**	**Unknown**	a) intuitive interface for data entry, searching, browsing, etc.?
	√		
Yes	**No**	**Unknown**	b) automated validation, mapping, metadata extraction, etc., as much as possible?
	√		
Yes	**No**	**Unknown**	c) availability of documentation?
√			

6. Long-term Viability

			Does the project have a:
Yes	**No**	**Unknown**	a) master plan for life-cycle management and data migration?
	√		
Yes	**No**	**Unknown**	b) reliance on open-source international standards versus proprietary standards?
	√		
Yes	**No**	**Unknown**	c) viable business model (e.g., not based exclusively on research grant with likely expiration)?
√			

F2. Library & Archives of Canada Bilingual Cataloguing

URL: http://www.collectionscanada.ca/csh/s23-120-e.html

Description: To support the bilingual cataloging policy of the Library & Archives of Canada (L&AC), all publications cataloged by the L&AC are assigned subject headings in both official languages, English and French. References between equivalent CSH and RVM headings are displayed in the L&AC's online public access catalog, AMICUS.

Canadian Subject Headings (CSH) is a list of subject access points in the English language; however, the authority records contain links to French language equivalents which allow the user to search on equivalent French language headings, from *Répertoire de vedettes-matière* (RVM), and be led to the record for the equivalent heading in CSH.

The *Répertoire de vedettes-matière* (RVM) provides access to more than 200,000 French subject headings and their English equivalents, as well as some 30 lists of subdivisions applicable to these headings.

Inaugural efforts at a separate list of subject headings for Canadian topics not adequately covered in the *Library of Congress Subject Headings* (LCSH) began in 1968.

Checklist

1. Types of Data Being Integrated

			Does the project have:
Yes	**No**	**Unknown**	a) different controlled vocabularies in same language?
√			
Yes	**No**	**Unknown**	b) different controlled vocabularies in different languages?
√			
Yes	**No**	**Unknown**	c) different classification schemas (e.g. DDC, UDC, LCC)?
√			
			If yes, which ones?
			DDC, LCC

Yes	No	Unknown	
√			d) controlled vocabularies combined with classification schemas?
Yes	No	Unknown	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
√			
			If yes, which ones?
			MARC
Yes	No	Unknown	f) different communication protocols?
		√	
Yes	No	Unknown	g) other:
		√	

2. Autonomy and Integrity of Constituent Parts

Yes	No	Unknown	
√			a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes	No	Unknown	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
√			
Yes	No	Unknown	b) Is full complement of metadata and indigenous subject hierarchies preserved?
√			
			If so, how?
			Hierarchies are preserved through broader and narrower terms
Yes	No	Unknown	c) Does project rely on principle of least common denominator?
	√		
			If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core. How does the use of least common denominator affect the quality of service?
Yes	No	Unknown	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
			Union catalog
Yes	No	Unknown	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
			Authority records

3. Reconciliation of Heterogeneous Vocabularies

Yes	No	Unknown	
			a) How are correlations established when a single term in one source has no equivalent term in the other?
			Began creating a separate list of subject headings for Canadian topics not adequately covered in the Library of Congress Subject Headings in 1968.
Yes	No	Unknown	b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
			Through authority records

Yes	No	Unknown	
			c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
			Policies are set to define how conflicts are resolved
Yes	No	Unknown	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?
			Only LCSH is used as a heading in this case

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes	No	Unknown	
√			a) Does project provide high or satisfactory levels of precision and recall?
Yes	No	Unknown	b) To what extent does project rely on pre-coordination?
			Relies on pre-coordination
			If mostly post-coordinate, then:
Yes	No	Unknown	i) by what means is recall maximized?
		√	
Yes	No	Unknown	ii) by what means is precision maximized?
		√	
Yes	No	Unknown	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
	√		
			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes	No	Unknown	i) *Locate* entities in the system via surrogates (find)
√			
Yes	No	Unknown	ii) *Identify* a surrogate that matches an entity (collocate)
√			
Yes	No	Unknown	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
√			
Yes	No	Unknown	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
√			
Yes	No	Unknown	v) *Navigate* the system and its surrogates (navigation)
√			
Yes	No	Unknown	e) Has developer released beta version for general testing?
		√	
Yes	No	Unknown	r) Have user satisfaction surveys been conducted?
√			

5. Ease of Use

Yes	No	Unknown	Does the project have:
√			a) intuitive interface for data entry, searching, browsing, etc.?
√			b) automated validation, mapping, metadata extraction, etc., as much as possible?
√			c) availability of documentation?

6. Long-term Viability

Yes	No	Unknown	Does the project have a:
√			a) master plan for life-cycle management and data migration?
√			b) reliance on open-source international standards versus proprietary standards?
√			c) viable business model (e.g., not based exclusively on research grant with likely expiration)?

F3. H. W. Wilson Megathesaurus

Project URL: www.hwwilson.com/Databases/omnifile.cfm

Project Description: Merges KOS of different structural types

H.W. Wilson has developed a "megathesaurus" that gathers the vocabulary for all its indexes for inclusion in its Omnifile product. The Omnifile product now includes six of the 11 Wilson periodical files, plus all of the full text from the remaining five files. Eventually Omnifile will probably include all their files, but this may take some time, since the remaining five are very specialized. Files covering non-periodical material use different indexing vocabularies and do not form part of the Omnifile product.

For more information about the project, consult the source below:

Kuhr, Patricia S. "Putting the World Back Together: Mapping Multiple Vocabularies into a Single Thesaurus." In Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC. ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 37-42.

Checklist

1. Types of Data Being Integrated

Yes	No	Unknown	Does the project have:
√			a) different controlled vocabularies in same language?
	√		b) different controlled vocabularies in different languages?
	√		c) different classification schemas (e.g. DDC, UDC, LCC)?
			If yes, which ones?
	√		d) controlled vocabularies combined with classification schemas?

Yes	No	Unknown	
√			e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
			If yes, which ones?
			SFX
Yes	No	Unknown	f) different communication protocols?
		√	
Yes	No	Unknown	g) other:
		√	

2. Autonomy and Integrity of Constituent Parts

Yes	No	Unknown	
√			a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes	No	Unknown	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
		√	
Yes	No	Unknown	b) Is full complement of metadata and indigenous subject hierarchies preserved?
√			
			If so, how?
			WilsonWeb displays a Thesaurus Term page, which provides information about the current term, such as a definition, related terms, previously used indexing terms, and broader and narrower terms. The information provided depends upon the database you are searching.
Yes	No	Unknown	c) Does project rely on principle of least common denominator?
	√		
			If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core). How does the use of least common denominator affect the quality of service?
Yes	No	Unknown	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
			Distributed database.
Yes	No	Unknown	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
			Includes used-for terms from thesaurus and automated switching.

3. Reconciliation of Heterogeneous Vocabularies

Yes	No	Unknown	
			a) How are correlations established when a single term in one source has no equivalent term in the other?
			Through the thesaurus. One can search keyword, browse from a dropdown or get suggestions for alternate, narrower, or broader terms from the thesaurus.

Yes	No	Unknown	
			b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
			Thesaurus is highly structured
Yes	No	Unknown	c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
			Manual editing of thesaurus records reconciles subject headings in the various specialties
Yes	No	Unknown	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?
		√	

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes	No	Unknown	
√			a) Does project provide high or satisfactory levels of precision and recall?
Yes	No	Unknown	b) To what extent does project rely on pre-coordination?
			Highly pre-coordinated
			If mostly post-coordinate, then:
Yes	No	Unknown	i) by what means is recall maximized?
		√	
Yes	No	Unknown	ii) by what means is precision maximized?
		√	
Yes	No	Unknown	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
		√	
			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes	No	Unknown	i) *Locate* entities in the system via surrogates (find)
√			
Yes	No	Unknown	ii) *Identify* a surrogate that matches an entity (collocate)
√			
Yes	No	Unknown	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
√			
Yes	No	Unknown	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
√			
Yes	No	Unknown	v) *Navigate* the system and its surrogates (navigation)
√			
Yes	No	Unknown	e) Has developer released beta version for general testing?
	√		

Yes	No	Unknown	f) Have user satisfaction surveys been conducted?
√			

5. Ease of Use

			Does the project have:
Yes	**No**	**Unknown**	a) intuitive interface for data entry, searching, browsing, etc.?
√			
Yes	**No**	**Unknown**	b) automated validation, mapping, metadata extraction, etc., as much as possible?
√			
Yes	**No**	**Unknown**	c) availability of documentation?
√			

6. Long-term Viability

			Does the project have a:
Yes	**No**	**Unknown**	a) master plan for life-cycle management and data migration?
√			
Yes	**No**	**Unknown**	b) reliance on open-source international standards versus proprietary standards?
√			
Yes	**No**	**Unknown**	c) viable business model (e.g., not based exclusively on research grant with likely expiration)?
√			

F4 Project Name: HILT (High Level Thesaurus Project)

Project URL: http://hilt.cdlr.strath.ac.uk/

Project Description: The goal of the project was to set up a pilot terminology services for the Joint Information Systems Company Information Environment. The terminologies server would be the basis of a community process that would develop, maintain and gradually improve interoperability of subject descriptions by mapping between terminology sets, and that the aim of the project was to determine specific design requirements based on this approach.

For more information about the project, consult the sources below:

HILT. (2005). *High-level Thesaurus Project Proposal.* http://hilt.cdlr.strath.ac.uk/AboutHILT/proposal.html (Jan. 7, 2005)

HILT Project Overview. <http://hilt.cdlr.strath.ac.uk/About-HILT/overview.html> (March 26, 2002).

Nicholson, Dennis, and others. *HILT: High-Level Thesaurus Project: Final Report to RSLP & JISC,* December 2001. <http://hilt.cdlr.strath.ac.uk/Reports/FinalReport.html> (Oct. 29, 2002).

Nicholson, Dennis. "HILT High Level Thesaurus Project: Interoperability and Cross-searching Distributed Services." Presented at the Thesaurus Conference organized by Waterways Trust, hosted at the Science Museum, London. 3 April 2001. <http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/HILTD%20Nicholson.ppt> (Oct. 26, 2002)

Nicholson, Dennis, Susannah Wake, and S. Currier. "HILT: High Level Thesaurus Project: Investigating the Problems of Cross-Searching Distributed Services by Subject in the UK." Presented at the meeting, "New Information Technology 2001." Tsinghua University, Beijing, China. 29-31 May. In: Global digital library development in the new millennium: Fertile ground for distributed cross-disciplinary collaboration. ed. C. C. Chen. Beijing: Tsinghua University Press, 2001. <http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/hiltchina2.ppt> (March 3, 2006)

Nicholson, Dennis and Susannah Wake. "HILT: Subject Retrieval in a Distributed Environment." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001* and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC. ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 61-67. <http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/hilt-ifla.ppt> (March 3, 2006)

Nicholson, Dennis. "Subject-based Interoperability: Issues from the High Level Thesaurus (HILT) Project." Paper presented at 68th IFLA Council and General Conference, Glasgow, Scotland, 18-24, 2002. <http://www.ifla.org/IV/ifla68/prog02.htm> (2002)

Wake, Susannah and Dennis Nicholson. "HILT - High-Level Thesaurus Project: Building Consensus for Interoperable Subject Access across Communities." *D-Lib Magazine*, 7, no. 9 (Sept. 2001). <http://www.dlib.org/dlib/september01/wake/09wake.html> (Oct. 26, 2002).

Checklist

1. Types of Data Being Integrated

	Does the project have:
Yes: √ / No: / Unknown:	a) different controlled vocabularies in same language?
Yes: / No: √ / Unknown:	b) different controlled vocabularies in different languages?
Yes: √ / No: / Unknown:	c) different classification schemas (e.g. DDC, UDC, LCC)?
	If yes, which ones?
	DDC; thesauri and classification schemes in Zthes, SKOS-Core, MARC
Yes: √ / No: / Unknown:	d) controlled vocabularies combined with classification schemas?
Yes: √ / No: / Unknown:	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
	If yes, which ones?
	DDC, LCSH, UNESCO
Yes: / No: √ / Unknown:	f) different communication protocols?
Yes: / No: / Unknown: √	g) other:

2. Autonomy and Integrity of Constituent Parts

Yes: √ / No: / Unknown:	a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes: / No: / Unknown: √	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
Yes: / No: √ / Unknown:	b) Is full complement of metadata and indigenous subject hierarchies preserved?
	If so, how?

Yes	No	Unknown	
		√	c) Does project rely on principle of least common denominator?
			If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core. How does the use of least common denominator affect the quality of service?
Yes	**No**	**Unknown**	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
			Distributed database.
Yes	**No**	**Unknown**	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
			Concordance tables.

3. Reconciliation of Heterogeneous Vocabularies

Yes	No	Unknown	
			a) How are correlations established when a single term in one source has no equivalent term in the other?
			Mapping form terms.
Yes	**No**	**Unknown**	b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
			In general, a single DDC pre-coordinated class would map to a Boolean combination of terms in a thesaurus or other post-coordinate target vocabulary.
Yes	**No**	**Unknown**	c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
			Because of the many provisions for synthesis within DDC, the number of possible classes is indeterminately large. It would be impossible to create all possible combinations and map them to other schemes. Mapping would therefore have to be done by starting from each term in each of the other vocabularies, and finding the DDC classes that contained the concept represented by that term. This would not provide the required mappings where DDC classes had to be represented by combinations of terms in the other vocabularies.
Yes	**No**	**Unknown**	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?
		√	

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes	No	Unknown	
√			a) Does project provide high or satisfactory levels of precision and recall?

Yes	No	Unknown	
			b) To what extent does project rely on pre-coordination?
			Can be pre-coordination.
			If mostly post-coordinate, then:
Yes	No	Unknown	i) by what means is recall maximized?
		√	
Yes	No	Unknown	ii) by what means is precision maximized?
		√	
Yes	No	Unknown	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
√			
			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes	No	Unknown	i) *Locate* entities in the system via surrogates (find)
√			
Yes	No	Unknown	ii) *Identify* a surrogate that matches an entity (collocate)
√			
Yes	No	Unknown	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
√			
Yes	No	Unknown	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
√			
Yes	No	Unknown	v) *Navigate* the system and its surrogates (navigation)
√			
Yes	No	Unknown	e) Has developer released beta version for general testing?
√			
Yes	No	Unknown	f) Have user satisfaction surveys been conducted?
√			

5. Ease of Use

			Does the project have:
Yes	No	Unknown	a) intuitive interface for data entry, searching, browsing, etc.?
√			
Yes	No	Unknown	b) automated validation, mapping, metadata extraction, etc., as much as possible?
√			
Yes	No	Unknown	c) availability of documentation?
√			

6. Long-term Viability

			Does the project have a:
Yes	No	Unknown	a) master plan for life-cycle management and data migration?
		√	
			Phase III study funded to begin Nov. 2005.

Yes	No	Unknown		b) reliance on open-source international standards versus proprietary standards?
√				
Yes	**No**	**Unknown**		c) viable business model (e.g., not based exclusively on research grant with likely expiration)?
√				

F5. MACS (Multilingual Access to Subjects)

URL: https://ilmacs.uvt.nl/pub/

Description: MACS aims to provide multilingual subject access to library catalogues.

MACS enables users to simultaneously search the catalogs of the project's partner libraries in the language of their choice (English, French, German). The partners are: the Swiss National Library (SNL), project leader, the Bibliothèque nationale de France (BnF), The British Library (BL) and Die Deutsche Bibliothek (DDB). The project is running under the auspices of the Conference of European National Librarians (CENL).

This multilingual search is made possible thanks to the equivalence links created between the three indexing languages used in these libraries: SWD (for German), RAMEAU (for French) and LCSH (for English). Topics (headings) from the three lists are analyzed to determine whether they are exact or partial matches, of a simple or complex nature. The end result is neither a translation nor a new thesaurus but a mapping of existing and widely used indexing languages.

On the basis of this approach, a prototype has been developed by Index Data (Denmark) and Tilburg University Library (Netherlands) which contains a small subset of data from the indexing languages and the libraries' databases so that link creation and management and subsequent searching can be explored and tested." (MACS)

Clavel-Merrin, Genevieve. "Multilingual Access to Subjects: the MACS Prototype." Paper presented at TEL Milestone Conference, April 29-30, 2002, Frankfurt am Main, Germany. <http://www.europeanlibrary.org /doc/tel_milconf_presentation_clavel.doc> (Oct. 8, 2002)

Clavel-Merrin, Genevieve. "The Need for Co-operation in Creating and Maintaining Multilingual Subject Authority Files." Paper presented at the 65th IFLA Council and General Conference, Bangkok, Thailand, Aug. 20-28, 1999. <http://www.ifla.org/IV/ifla65/papers/080-155e.htm> (Aug. 7, 2002). (MACS)

Freyre, Elisabeth and Naudi, Max. "MACS: Subject Access across Languages and Networks." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*. ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 3-10.

Hudon, Michele. "Multilingual Thesaurus Construction: Integrating the View of Different Cultures in One Gateway to Knowledge and Concepts." *Knowledge Organization,* v. 24, no. 2 (1997): 84-91.

Kunz, Martin. "Subject Retrieval in Distributed Resources: a Short Review of Recent Developments." Paper presented at the 68th IFLA Council and General Conference, Aug. 18-24, 2002. <http://www.ifla.org/IV/ifla68/papers/007-122e.pdf> (Oct. 27, 2002)

Landry, Patrice (2000). "The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD)." Paper presented at the 66th IFLA Council and General Conference, Jerusalem, Aug. 13-18, 2000. <http://www.ifla.org/IV/ifla66/papers/165-181e.pdf> (Aug. 7, 2002).

MACS (Multilingual Access to Subjects) Project, report for 2000-2001. <https://ilmacs.uvt.nl/pub/node/7?PHPSESSID=ff54ff63320ff2635357304df902dca9> (Aug. 7, 2002)

Checklist

1. Types of Data Being Integrated

				Does the project have:
Yes	**No**	**Unknown**		a) different controlled vocabularies in same language?
	√			
Yes	**No**	**Unknown**		b) different controlled vocabularies in different languages?
√				
				Subject headings in three languages: English, French, German.

Yes	No	Unknown	
√			c) different classification schemas (e.g. DDC, UDC, LCC)?
			If yes, which ones?
			Recommendation: carry out a mapping of LCSH, the UNESCO thesaurus, AAT, UDC to a DDC backbone, as the reference language
Yes	No	Unknown	d) controlled vocabularies combined with classification schemas?
√			
Yes	No	Unknown	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
√			
			If yes, which ones?
			Indicates in each authority file the equivalent preferred descriptors of the other authority files for a few chosen subject areas via the thesauri. Staff access via a Link Management Interface.
Yes	No	Unknown	f) different communication protocols?
	√		
Yes	No	Unknown	g) other:
		√	

2. Autonomy and Integrity of Constituent Parts

Yes	No	Unknown	
√			a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes	No	Unknown	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
		√	
Yes	No	Unknown	b) Is full complement of metadata and indigenous subject hierarchies preserved?
√			
			If so, how?
			Hierarchical navigation is only possible within each Subject Heading Language (SHL), so it is envisaged that searches are refined by the user in his own language until the required concept is identified and then expanded for linguistics equivalences and documents in other libraries.
Yes	No	Unknown	c) Does project rely on principle of least common denominator?
	√		
			If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core). How does the use of least common denominator affect the quality of service?
Yes	No	Unknown	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
			Union catalog.

Yes	No	Unknown	
			e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
			In three subject heading lists.

3. Reconciliation of Heterogeneous Vocabularies

Yes	No	Unknown	
			a) How are correlations established when a single term in one source has no equivalent term in the other?
			Through linking.
			b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
			Not all differences are reconciled, but linking is used where possible.
			c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
			Through linking.
			d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?
			The cross linking here is language, not subject as identical headings are not likely.

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes	No	Unknown	
√			a) Does project provide high or satisfactory levels of precision and recall?
		√	b) To what extent does project rely on pre-coordination?
			If mostly post-coordinate, then:
		√	i) by what means is recall maximized?
		√	ii) by what means is precision maximized?
		√	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
√			i) *Locate* entities in the system via surrogates (find)

Yes	No	Unknown	
√			ii) *Identify* a surrogate that matches an entity (collocate)
Yes	**No**	**Unknown**	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
		√	
Yes	**No**	**Unknown**	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
√			
Yes	**No**	**Unknown**	v) *Navigate* the system and its surrogates (navigation)
√			
Yes	**No**	**Unknown**	e) Has developer released beta version for general testing?
√			
Yes	**No**	**Unknown**	f) Have user satisfaction surveys been conducted?
√			

5. Ease of Use

			Does the project have:
Yes	**No**	**Unknown**	a) intuitive interface for data entry, searching, browsing, etc.?
√			
Yes	**No**	**Unknown**	b) automated validation, mapping, metadata extraction, etc., as much as possible?
	√		
Yes	**No**	**Unknown**	c) availability of documentation?
√			

6. Long-term Viability

			Does the project have a:
Yes	**No**	**Unknown**	a) master plan for life-cycle management and data migration?
	√		
Yes	**No**	**Unknown**	b) reliance on open-source international standards versus proprietary standards?
√			
Yes	**No**	**Unknown**	c) viable business model (e.g., not based exclusively on research grant with likely expiration)?
√			
			Recommendation for future: have a strong user focus, determine reliable costs, includes cost benefits, involve international players, look at how best to integrate semantic web and artificial intelligence developments, involved a broad range of target services.

F6. Project Name: RDN Subject Portals Project

Project URL: http://www.portal.ac.uk/spp/

Project Description: Cataloging of online resources with subject categories controlled by proprietary hierarchical list. From the web site, "The Resource Discovery Network is the UK's free national gateway to Internet resources for the learning, teaching and research community." Currently eight subject portals are available: Altis (hospitality, leisure, sport and tourism), Artifact (arts and creative industries), Biome (health and life sciences), EEVL (engineering, mathematics and computing), GEsource (geography and environment), Humbul (humanities), PSIgate (physical sciences) and SOSIG (social sciences). The first phase of the project was to build a Z39.50 cross search prototype at three RDN hubs, SOSIG, EEVL, and BIOME. The second phase adds HUMBUL and PSIgate.

For more information about the project, consult the sources below:

Clark, Judith. "Subject Portals." *Ariadne,* 29 (Oct. 2, 2001). <http://www.ariadne.ac.uk/issue29/clark/> (Jan. 21, 2003)

Day, Michael. "Metadata in Support of Subject Gateway Services and Digital Preservation." Draft version of paper presented at Electronic Resources: Definition, Selection and Cataloguing, Rome, Italy, Nov. 2001. <http://www.ukoln.ac.uk/metadata/presentations/rome-2001/paper.html> (Aug. 8, 2002)

Checklist

1. Types of Data Being Integrated

	Does the project have:
Yes: √ / No: / Unknown:	a) different controlled vocabularies in same language?
Yes: / No: √ / Unknown:	b) different controlled vocabularies in different languages?
Yes: / No: √ / Unknown:	c) different classification schemas (e.g. DDC, UDC, LCC)?
	If yes, which ones?
Yes: / No: √ / Unknown:	d) controlled vocabularies combined with classification schemas?
Yes: √ / No: / Unknown:	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
	If yes, which ones?
	SOAP, XML, OAI, XHTML, Dublin Core
Yes: √ / No: / Unknown:	f) different communication protocols?
Yes: / No: / Unknown: √	g) other:

2. Autonomy and Integrity of Constituent Parts

Yes: / No: / Unknown: √	a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes: / No: / Unknown: √	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
Yes: / No: √ / Unknown:	b) Is full complement of metadata and indigenous subject hierarchies preserved?
	If so, how?
Yes: √ / No: / Unknown:	c) Does project rely on principle of least common denominator?
	If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core). How does the use of least common denominator affect the quality of service?

	Portlets provide the main route for surfacing data to the user and for their subsequent interactions. They can take advantage of the portal framework services but at a minimum must conform to an interface that allows them to be plugged into the framework
Yes / No / Unknown:	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
	Distributed database.
Yes / No / Unknown:	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
	Full text.

3. Reconciliation of Heterogeneous Vocabularies

Yes / No / Unknown: √	a) How are correlations established when a single term in one source has no equivalent term in the other?
Yes / No / Unknown: √	b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
Yes / No / Unknown: √	c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
Yes / No / Unknown: √	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes: √ / No / Unknown	a) Does project provide high or satisfactory levels of precision and recall?
Yes / No / Unknown: √	b) To what extent does project rely on pre-coordination?
	Pre-coordinated at each hub.
	If mostly post-coordinate, then:
Yes / No / Unknown: √	i) by what means is recall maximized?
Yes / No / Unknown	ii) by what means is precision maximized?
	Sites are selected on the basis of selection criteria, cataloged following consistent practices, and analyzed by people with expertise with the relevant subject discipline. Links are checked daily in an automated process and all entries are updated regularly by subject specialists. These are classified using an appropriate controlled vocabulary.

Yes	No	Unknown	
√			c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?
			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes	No	Unknown	i) *Locate* entities in the system via surrogates (find)
		√	
Yes	No	Unknown	ii) *Identify* a surrogate that matches an entity (collocate)
		√	
Yes	No	Unknown	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
		√	
Yes	No	Unknown	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
		√	
Yes	No	Unknown	v) *Navigate* the system and its surrogates (navigation)
		√	
Yes	No	Unknown	e) Has developer released beta version for general testing?
		√	
Yes	No	Unknown	f) Have user satisfaction surveys been conducted?
		√	

5. Ease of Use

			Does the project have:
Yes	No	Unknown	a) intuitive interface for data entry, searching, browsing, etc.?
	√		
Yes	No	Unknown	b) automated validation, mapping, metadata extraction, etc., as much as possible?
	√		
Yes	No	Unknown	c) availability of documentation?
√			

6. Long-term Viability

			Does the project have a:
Yes	No	Unknown	a) master plan for life-cycle management and data migration?
	√		
Yes	No	Unknown	b) reliance on open-source international standards versus proprietary standards?
√			
Yes	No	Unknown	c) viable business model (e.g., not based exclusively on research grant with likely expiration)?
√			

F7. UMLS Metathesaurus

URL: http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html

Description: The UMLS consists of three Knowledge Sources: the UMLS Metathesaurus, the SPECIALIST lexicon, and the UMLS Semantic Network. The Metathesaurus is a database containing semantic information about biomedical concepts, their various names, and the relationships among them.

The Metathesaurus is built from over 100 biomedical source vocabularies, some in multiple languages. The 2003 edition includes 875,255 concepts and 2.14 million concept names. The UMLS Semantic Network is used for mapping index terms from different thesauri through its 134 semantic types which provides a consistent categorization of all concepts represented in the Metathesaurus.

National Library of Medicine. *Fact sheet: UMLS Metathesaurus,* 2005.
<http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html> (Jan.7, 2005)

Unified Medical Language System (UMLS). <http://www.nlm.nih.gov/research/umls/> (March 26, 2002).

Checklist

1. Types of Data Being Integrated

	Does the project have:
Yes: √ / No: / Unknown:	a) different controlled vocabularies in same language?
Yes: √ / No: / Unknown:	b) different controlled vocabularies in different languages?
Yes: √ / No: / Unknown:	c) different classification schemas (e.g. DDC, UDC, LCC)?
	If yes, which ones?
	Includes more than 100 biomedical and health-related vocabularies, classifications, and coding systems (some in multiple languages)
Yes: √ / No: / Unknown:	d) controlled vocabularies combined with classification schemas?
Yes: √ / No: / Unknown:	e) different metadata framework schemas (e.g. XML, MARC, Dublin Core)?
	If yes, which ones?
	Specialized coding schemes, e.g. SNOMED CT and LOINC
Yes: √ / No: / Unknown:	f) different communication protocols?
Yes: / No: / Unknown: √	g) other:

2. Autonomy and Integrity of Constituent Parts

Yes: √ / No: / Unknown:	a) Is standardization, reconciliation, or conversion of semantic data reversible?
Yes: √ / No: / Unknown:	a.1) can pre-coordinated strings, once filtered or deconstructed for semantic matching, later be put back together again?
Yes: √ / No: / Unknown:	b) Is full complement of metadata and indigenous subject hierarchies preserved?
	If so, how?
	Through linking.
Yes: / No: √ / Unknown:	c) Does project rely on principle of least common denominator?
	If so, many data sets may be able to coexist in the database, but given resulting stripped-down or 'dumbed-down' resource descriptions, the database may no longer serve the interests of readers. (cf. recently cited problems with Dublin Core). How does the use of least common denominator affect the quality of service?

Yes / No / Unknown:	d) How is data stored: gathered into a union catalog (e.g. American Memory Project, NSDL), vs. distributed database?
	Distributed database.
Yes / No / Unknown:	e) How are metadata (including semantically interoperable links) stored? (e.g., via authority records, concordance tables, a central switching language, semantic networks, lexical databases, semantic layers, etc.)
	Linking records.

3. Reconciliation of heterogeneous vocabularies

Yes / No / Unknown:	a) How are correlations established when a single term in one source has no equivalent term in the other?
	Linking
Yes / No / Unknown:	b) Certain vocabularies are highly structured and hierarchical, while others contain terms lacking any structure at all aside from serial numbers or other unique identifiers. How are these differences reconciled?
	By adding certain basic information to each concept and establishing new relationships between terms from different source vocabularies.
Yes / No / Unknown:	c) How are conflicts resolved when an established heading in one vocabulary matches a cross reference in other vocabularies? (E.g., Tumors is an established LCSH heading, but in MeSH it is a cross reference to Neoplasms; and vice versa)
	By linking concepts that are similar along some dimension.
Yes / No / Unknown:	d) If multiple vocabularies are used in a single bibliographic record, and the headings from such vocabularies are identical (after normalization), how are duplicate retrievals handled?
	Through linking.

4. Effective and Efficient Resource Discovery (Precision and Recall), Satisfying User Needs

Yes: √ / No / Unknown:	a) Does project provide high or satisfactory levels of precision and recall?
Yes / No / Unknown:	b) To what extent does project rely on pre-coordination?
	Highly pre-coordinated.
	If mostly post-coordinate, then:
Yes / No / Unknown: √	i) by what means is recall maximized?
Yes / No / Unknown: √	ii) by what means is precision maximized?
Yes: √ / No / Unknown:	c) Does project provide faceted approach (facilitating polysemy) while retaining option for browsable hierarchy (facilitating navigation)?

			d) Are the following objectives and functions supported in the subject semantically interoperable environment?
Yes	**No**	**Unknown**	i) *Locate* entities in the system via surrogates (find)
√			
Yes	**No**	**Unknown**	ii) *Identify* a surrogate that matches an entity (collocate)
√			
Yes	**No**	**Unknown**	iii) *Select* an entity appropriate to a user's need via surrogates (choice facilitation)
√			
Yes	**No**	**Unknown**	iv) *Obtain* access to the entity via the system and its surrogates (acquisition)
√			
Yes	**No**	**Unknown**	v) *Navigate* the system and its surrogates (navigation)
		√	
Yes	**No**	**Unknown**	e) Has developer released beta version for general testing?
		√	
Yes	**No**	**Unknown**	f) Have user satisfaction surveys been conducted?
		√	

5. Ease of Use

			Does the project have:
Yes	**No**	**Unknown**	a) intuitive interface for data entry, searching, browsing, etc.?
√			
Yes	**No**	**Unknown**	b) automated validation, mapping, metadata extraction, etc., as much as possible?
√			
Yes	**No**	**Unknown**	c) availability of documentation?
√			

6. Long-term Viability

			Does the project have a:
Yes	**No**	**Unknown**	a) master plan for life-cycle management and data migration?
√			
Yes	**No**	**Unknown**	b) reliance on open-source international standards versus proprietary standards?
√			
Yes	**No**	**Unknown**	c) viable business model (e.g., not based exclusively on research grant with likely expiration)?
		√	

Appendix I. Glossary

classification scheme. The terms classification scheme, taxonomy, categorization scheme are often used interchangeably. Though there may be subtle differences from example to example, in general these types of KOSs provide ways to separate entities into buckets or relatively broad topic levels. Some examples provide a hierarchical arrangement of numeric or alphabetic notation to represent broad topics. These types of knowledge organization systems may not follow the strict rules for hierarchy required in the ANSI NISO Thesaurus Standard (Z39.19) (NISO), and often lack the explicit relationships presented in a thesaurus.35

concept map. A diagram showing the relationships between concepts. Concepts are connected with labeled arrows, in a downward-branching hierarchical structure. The relationship between concepts is articulated in linking phrases, e.g., "gives rise to," "results in," "is required by," or "contributes to."36

concordance table. Also called a correspondence table. Methodologically, a concordance table describes the way in which terms in multiple vocabularies are related.37

controlled vocabulary. A subset of a language, consisting of pre-selected words and phrases designated as index terms. In a controlled vocabulary, each subject is represented by one valid term only; and, conversely, each term represents only one subject. References are made from equivalent or synonymous terms not selected as valid index terms. Homographs are disambiguated. In addition, a controlled vocabulary contains links among hierarchically or otherwise related terms. Examples of controlled vocabularies include Library of Congress Subject Headings, Thesaurus of ERIC Descriptors, and Medical Subject Headings. The term "controlled vocabulary" is often used in a broad sense to include scheme-based classification data, which also manifest rigorous structures and embody relationships among concepts.38

cross-domain search. A search of multiple resources from different domains through a single interface, using a single query.

crosswalk. A program or algorithm to map elements in different metadata schemes. An example is the Dublin Core/MARC/GILS Crosswalk designed by the Library of Congress.39

descriptors. Terms used in indexes, abstracts, or other databases/periodical indexes to describe the subjects of an article.

dictionary. Alphabetical lists of terms and their definitions that provide variant senses for each term, where applicable. They are more general in scope than a glossary. While a dictionary may also provide synonyms and through the definitions, related terms, there is no explicit hierarchical structure or attempt to group terms by concept.40

facet analysis. Ranganathan developed the theory of faceting to allow for the expression of compound or multi-concept subjects. A faceted system recognizes that any given subject has many aspects. Concepts are decomposed into elemental classes. A system employing faceting synthesizes these aspects in a way that best describes the subject.

gazetteer. A dictionary of place names. Traditional gazetteers have been published as books or they appear as indexes to atlases.41

---

35 Gail Hodge. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. CLIR Pub91. April 2000. <www.clir.org/pubs/abstract/pub91abst.html> viewed from <http://nkos.slis.kent.edu/KOS_taxonomy.htm>

36 Wikipedia: the Free Encyclopedia. (http://en.wikipedia.org/wiki/Main_Page)

37 UN Glossary of Classification Terms. <http://unstats.un.org/unsd/class/family/glossary_short.htm>

38 American Library Association. "Subject Data in the Metadata Record: Recommendations and Rationale." Division of Association for Libraries and Technical Services, Cataloging and Classification Section, Subcommittee on Metadata and Subject Analysis (1999). <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees /subjectanalysis/metadataandsubje/subjectdata.htm>

39 Ibid.

40 Gail Hodge. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* CLIR Pub91. April 2000. <www.clir.org/pubs/abstract/pub91abst.html> viewed from <http://nkos.slis.kent.edu/KOS_taxonomy.htm>

41 Ibid.

glossary. A list of terms, usually with definitions. The terms may be from a specific subject field or those used in a particular work. The terms are defined within that specific environment and rarely have variant meanings provided. Examples include the EPA Terms of the Environment.42

harmonization. The process of making disparate entities or systems work together. Its purpose is to resolve conflicts and to remove obstacles by overcoming idiosyncrasies of individual systems. Within the context of subject access, harmonization implies efforts to make terms from different controlled vocabularies work together for the benefit of improving retrieval results. Differences may occur in semantics and/or syntax, and among multiple languages. Harmonization provides the ability to accommodate two or more different systems, schemes, or standards to facilitate searching across databases. Methods of harmonization include linking and mapping.43

interoperability. The ability of two or more systems or components to exchange information and use the exchanged information without special effort on the part of either system.44

knowledge organization system. A general term referring to the tools that present the organized interpretation of knowledge structures; includes authority files, classification systems, concept spaces, dictionaries, gazetteers, glossaries, ontologies, subject heading sets, thesauri; often called KOS, sometimes, knowledge organization scheme.45

KOS. See *knowledge organization system.*

least common denominator. A common metadata scheme that other metadata standards can easily be mapped to and can be used to index across heterogeneous databases and support interoperability. An example is Dublin Core.

lexical database. A lexical database is organized around lexemes, which include all the morphemes of a language, even if these morphemes never occur alone. A lexeme is the minimal unit of language which has a semantic interpretation and embodies a distinct cultural concept. A lexeme is conventionally listed in a dictionary as a separate entry.

link. A mechanism for associating equivalent or associated terms.

mapping. A special form of linking, with efforts to identify equivalence or establish one-to-one and, in some instances, one-to-many relationships. Mapping facilitates automatic switching between systems or languages. Recent developments include efforts to match elements in the MARC record with those in other metadata records and efforts to identify equivalent terms among different controlled vocabularies or different languages. Examples of mapping of subject entries include the Omni File (based on the indexes to individual WILSONLINE databases) and MACS (Multi-lingual Access to Subject headings), a European project on multilingual access to subject authority files and data to develop a prototype for the mapping of subject entries based on three controlled vocabularies: Library of Congress Subject Headings (LCSH), RAMEAU, and Schlagwortnormdatei (SWD)).46

metathesaurus. A "thesaurus of thesauri," serving as a framework within which diverse controlled vocabularies are harmonized for the purpose of facilitating cross-file searching. An example is the UMLS (Unified Medical Language System) Metathesaurus developed and maintained by the National Library of Medicine, in which "alternate names [from different source vocabularies] for the same concept (synonyms, lexical variants, and translations) are linked together. Each Metathesaurus concept has attributes that help to define its meaning,

---

42 Ibid.

43 American Library Association. "Subject Data in the Metadata Record: Recommendations and Rationale." Division of Association for Libraries and Technical Services, Cataloging and Classification Section, Subcommittee on Metadata and Subject Analysis (1999). <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees /subjectanalysis/metadataandsubje/subjectdata.htm>

44 Report of the CC:DA Task Force on Metadata. <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>

45 Gail Hodge. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* CLIR Pub91. April 2000. <www.clir.org/pubs/abstract/pub91abst.html> viewed from <http://nkos.slis.kent.edu/KOS_taxonomy.htm>

46 American Library Association. "Subject Data in the Metadata Record: Recommendations and Rationale." Division of Association for Libraries and Technical Services, Cataloging and Classification Section, Subcommittee on Metadata and Subject Analysis (1999). <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees /subjectanalysis/metadataandsubje/subjectdata.htm>

e.g., the semantic type(s) or categories to which it belongs, its position in the hierarchical contexts from various source vocabularies, and, for many concepts, a definition." (National Library of Medicine 1999).47

networked knowledge organization system. An interactive information device aimed at supporting the description and retrieval of heterogeneous information resources on the internet; sometimes NKOS.48

NKOS. See *networked knowledge organization system*

ontology. A knowledge representation format. That is, an ontology is a shared understanding of the structure of a domain of interest. Ontologies make it easy both for humans to compile and maintain a body of knowledge, and for computer programs to use this knowledge to intelligently manipulate data. An ontology organizes all data using the concepts of class, object, and relationship. Classes are organized into a hierarchy, ordered by subclass, called a taxonomy. A well-known taxonomy is the biological taxonomy of all living things, in which living things are sub-classed into their kingdom: plant or animal. Plants and animals are further classified into phylum, etc. An ontology extends a taxonomy by including relationships among objects and classes, which can represent properties and values. To continue the biological example, there is a relationship "number of limbs" between certain classes of animals and integers. Many taxonomies have been developed to organize knowledge in particular areas.49

ontology mapping. The process of ontology mapping concerns how classes from one ontology can be mapped to classes of another taxonomy in an automated way.50

polysemy. Polysemy refers to a word or phrase that has two or more meanings which are related. An individual word or phrase can be used in different contexts to express different meanings.

query term. The word or term with which a user begins a search.

semantic interoperability. The ability of two or more systems or components to exchange or harmonize cognate subject vocabularies and/or knowledge organization schemes to be used for the purpose of effective and efficient resource discovery without significant loss of lexical or connotative meaning and without special effort by the user

semantic layer. A semantic layer is an interface in a program or application that allows the user to create and use words that make sense to them, and then takes that set of names and superimposes it on top of the true physical structure of the data. This abstract view or semantic layer working with semantic relationships allows the users to view, work with and otherwise maintain the data in a vernacular that is efficient and quick to use.

semantic network. A type of KOS that structures concepts and terms not as hierarchies but as a network or a web; concepts are thought of as nodes with various relationships branching out from them; the relationships generally go beyond the standard BT, NT and RT and may include specific whole-part relationships, cause-effect, parent-child, etc. Examples of semantic networks include Princeton's WordNet, which is now used in a variety of search engines, and the Unified Medical Language System (UMLS) Semantic Network.51

subject authority file. An internal tool for catalog or database management. It contains authority records and provides documentation of a body or list of authorized and authoritative indexing terms in the context and framework of its vocabulary.52

subject authority record. A record of a subject heading that shows its established form, cites the authorities consulted in determining the choice and form of the heading, and indicates the cross-references made to and from the heading.53

subject headings. A set of controlled terms to represent the subjects of items in a collection. Subject heading lists can be extensive, covering a broad range of subjects. In use, subject headings tend to be pre-coordinated,

---

47 Ibid.

48 Marcia Lei Zeng and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.

49 *XSB, Inc. Glossary*. <http://www.xsb.com/glossary.html>

50 Ibid.

51 Gail Hodge. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* CLIR Pub91. April 2000. <www.clir.org/pubs/abstract/pub91abst.html> viewed from <http://nkos.slis.kent.edu/KOS_taxonomy.htm>

52 Hope A. Olson and John J. Boll. *Subject Analysis in Online Catalogs*. Englewood, CO: Libraries Unlimited, 2001.

53 Lois Mai Chan. *Cataloging and Classification: an Introduction*. 2nd ed. New York: McGraw-Hill, c1994.

with rules for how subject headings can be joined to provide more specific concepts. Examples include MeSH and LCSH.54

switching language. Intermediary terms that serve as a mechanism for moving between vocabularies; unlike links, which are internal, switching language is external to records for the terms being associated

taxonomy. A hierarchical data structure or a type of classification schema made up of classes, where a child of a taxonomy node represents a more restricted, smaller, subclass than its parent.55

term list. A list of words or phrases, often with definitions; examples include authority files, glossaries, gazetteers, and dictionaries.56

thesaurus. A type of KOS, which is based on concepts that show relationships between terms. Relationships commonly expressed in a thesaurus include hierarchy, equivalence, and associative (or related). These relationships are generally represented by the notation BT (broader term), NT (narrower term), SY (synonym), and RT (associative or related).57

54 Gail Hodge. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* CLIR Pub91. April 2000. <www.clir.org/pubs/abstract/pub91abst.html> viewed from <http://nkos.slis.kent.edu/KOS_taxonomy.htm>

55 *XSB, Inc. Glossary*. <http://www.xsb.com/glossary.html>

56 Marcia Lei Zeng and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems," *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.

57 Gail Hodge. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.* CLIR Pub91. April 2000. <www.clir.org/pubs/abstract/pub91abst.html> viewed from <http://nkos.slis.kent.edu/KOS_taxonomy.htm>

Appendix II. Program Summary: Enriching Subject Access

The Chair served on the ALCTS Metadata Enrichment Task Force formed following the conference, *Bibliographic Control of Web Resources: a Library of Congress Action Plan*, in particular section 2.3. This Task Force was focused on subject/topical access to Web subject portals and partly focused on methods of achieving semantic clarity in terms presented to users.

The Subcommittee presented a program jointly with the Task Force at the 2004 Annual Conference entitled "Enriching Subject Access." A report of the program was published in *Technical Services Quarterly.*

> Harken, Shelby E. (2005). "Enriching subject access: a report on the Joint Program of the ALCTS Metadata Enrichment Task Force and the ALCTS Subject Analysis Committee's Subcommittee on Semantic Interoperability. American Library Association Annual Conference, Orlando, June 2004." *Technical Services Quarterly*, 22(3) 2005, p. 75-87.

Enriching Subject Access

A Joint Program of the ALCTS Metadata Enrichment Task Force and the ALCTS Subject Analysis Committee's Subcommittee on Semantic Interoperability

Report by Shelby E. Harken, University of North Dakota

The program was presented in two parts. Part 1, Metadata Enrichment for Subject Access; and Part 2, Bringing subject access together through interoperability

The program, Enriching Subject Access grew out of work by two different committees. Part 1 of the program was developed by the ALCTS Metadata Enrichment Task Force (METF). It was established to cooperate with the Library of Congress on projects resulting from the "Bibliographic Control of Web Resources: A Library of Congress Action Plan" which stemmed from the Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenge of Networked Resources and the Web, held on November 15-17, 2000.58 In particular it is involved with Action item 2, Enhance the access to and display of records for selected Web resources across multiple systems. Specifically it was to address action item 2.3. Part 2 of the program was developed by the ALCTS Subject Analysis Committee's Subcommittee on Semantic Interoperability. Part of its charge was to survey the current state of international semantic interoperability projects which focus on subject and/or classification data. The Chair of the Subcommittee served on the METF. Although each has a quite specific goal, they converge in efforts to improve subject access to information resources accessible via the Internet by enhancing subject metadata and investigating methods of manipulating and presenting the data to users.

Part 1. Metadata Enrichment for Subject Access

Action item 2.3 states "Explore ways to enrich metadata by focusing on providing additional subject and other access mechanisms (e.g., front-end user thesauri) and increasing granularity of access and display (e.g., by enabling progression through hierarchy and versions and by additional description information including summaries)."

The Charge for Metadata Enrichment Task Force (Action Item 2.3): An ALCTS Task Force59

Action item 2.3 is intended to further objective 2 of the LC Action Plan: "enhanced record display and access across multiple systems." This action item is related to Action 3 and 6, but its focus is on access mechanisms and the ways in which existing metadata might be enriched to improve and extend them. As such, it promotes development in three related functions associated with a user's ability to identify relevant resources across systems: 1) making use of thesauri; 2) representing relationships among resource versions and formats and; 3) the relating of summaries to records. METF was to identify what kinds of front-end mechanisms will best support user access in these areas, what metadata is currently lacking or machine-inaccessible in current metadata records that would be needed for

58 Bibliographic Control of Web Resources: A Library of Congress Action Plan. <http://lcweb.loc.gov/catdir/bibcontrol/actionplan.html>

59 Bibliographic Control of Web Resources: a Library of Congress Action Plan's Charge for Metadata Enrichment Task Force (Action Item 2.3), and ALCTS Task Force. Email July 2, 2002.

such mechanisms to operate and how that missing metadata might be located and associated with existing records. To accomplish this, METF sorted this very large work item into three functionalities: front-end user thesauri; access mechanisms that represent relationships among resource versions and formats; and additional descriptive information, including relating summaries to records for digital objects.

In related work at the Library of Congress, the Library of Congress Portals Applications Issues Group (LCPAIG) was formed to work on item 2.1 that calls for development and enhancement of portal functionality for the benefit of the library community in general. In part it states: Support development of common user interfaces for searching, sorting, and retrieving relevant search results across a range of discovery tools.60

The Principal Investigator chosen to serve as consultant to the Metadata Enrichment Task Force was Marcia Bates. Her report, "Research and Design Review: Improving User Access to Library Catalog and Portal Information: The Final Report" (version 3) was completed June 1, 2003 incorporating comments from METF and ALCTS TF members.

Summary of Bates' recommendations[61]

It is recommended that with regard to access vocabulary:

a) Cluster vocabulary be created, based on the searcher vocabulary developed by Sara Knapp, if she and her publisher agree.

b) For the price of a share of the maintenance of the database, libraries and commercial firms may subscribe to the searcher vocabulary database, and install it in their catalogs, portals, and websites.

c) Access to catalogs and portal information should be available both directly through and around the vocabulary database.

d) Institutional users may link the searcher vocabulary with their own controlled vocabulary. As a result, users of these sites may input their search term(s), be shown a cluster of terms, including "legitimate" controlled terms, and use the clusters as a basis for selecting terms for either controlled vocabulary or keyword searching

e) With this vocabulary as a core, one or two lexicographers are hired cooperatively to maintain the searcher vocabulary

It is recommended that with regard to bibliographic families:

a) Preliminary agreement be gained on what shall constitute bibliographic families, probably based on the work of Tillett, Smiraglia, and others.

b) Based on pilot studies, criteria are finalized for the creation of bibliographic families. The families are numbered, and libraries may acquire the cataloging information for the families in a manner similar to the currently existing cooperative cataloging arrangements.

c) Further experience will also provide enlightenment regarding just how far down the chain of family size the cooperative effort should go

d) It [could] become possible that whenever a searcher happens on a record that is part of a bibliographic family, the searcher may click on a "related records" link and see displayed on the screen the progenitor record plus links to all the different types of bibliographically related records arrayed around the core record.

It is recommended that with regard to staging of access to records:

a) Libraries and other information institutions take as an objective the approach of providing staged access to information that drops down into the information in a 1:30 ratio.

b) Cooperation with publishers can be explored.

60 Library of Congress Portals Applications Issues Group http://www.loc.gov/catdir/lcpaig/ (July 1, 2004)

61 Marcia Bates. "Improving User Access to Library Catalog and Portal Information." 2003.http://lcweb.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf

Summary of Bates' report and vendors responses

Judith Ahronheim, University of Michigan Graduate Library

Judith Ahronheim provided an introduction by summarizing Bates' report. She began by defining the purpose of the report: to explore metadata enrichment, e.g. via a front-end user thesauri of clustered terms and then providing progression through a hierarchy and version of terminology to allow the user to select the most appropriate search term(s). The most problematic aspect of OPAC searching is subject searching. People have a lot of "no match" or "poor match" searches, yet they still like to use subject searches. The process focuses on subject enrichment by inserting clusters of terms in the search transaction. Research shows that users will use poor quality information if it is easy to find, rather than good quality information if it is hard to find. The likelihood that the same term will be used by a user for a topic is between 10% and 20%. Searching by skilled searchers for same queries show only 1% to be identical and less that 60% had a realistic match. Users who fail with a search will practically never try another search. The system rarely responds with a suggestion if the search fails. The user now has a hard time thinking up a new thought. Almost all efforts to include thesauri don't use user terms. Those who have developed interfaces with thesauri did not test with users or compare with OPACs.

In the clustered approach, the user sees both controlled and uncontrolled words in the cluster, the cluster would bring up a group of related words, allowing the user to select the most precise term. This would be enhanced by computer operations or algorithms. It would also include misspellings. Sarah Knapp has developed a cluster approach. A base line database would need to be developed and maintained and continually augmented.

Three vendors gave their responses.

Michael Kaplan, Ex Libris

Research is often slow to reach the catalog and portal decision-makers. We as librarians need to move faster. Libraries are a very small part of a much larger information business. He cited Gresham's law: payment will always as far as possible, be made in that medium of which the cost is least and the more valuable will tend to disappear from circulation. The same is true for keyword vs. subject and skimpy vs. full bibliographic records. His version of the law is that the lowest level bibliographic records will drive out fuller cataloging following the more-better-cheaper-fast approach.

What should we do with this cluster idea? Should we do FRBR? Is it a higher priority? What should we do with the subjects, keyword use, etc. we already have? In a study of Aleph searches, 87% of all index searches were keyword or keyword author, keyword title, or keyword subject. Librarians and users search very differently. Do we want to control free-form searching or let it go on? How do you additionally include resources outside of the OPAC? Metalib allows a federated search beyond on the OPAC; the words in records need to be available to search. Online metadata (bibliographic records) is no longer enough. Users want more full-text access. He questioned how this new clustered vocabularies database would be maintained. He wants to see research on users' search patterns. He wants to see analysis. Could Aleph help users? Aleph uses synonyms which approaches the cluster to some extent. Are we too late to catch the current future users' behavior?

Deb Bendig, OCLC (Product Manager for FirstSearch)

She views this from the FirstSearch WorldCat product. Is there a cost-effective way to improve subject access? What is the value to helping users? There is always a value, but … a) the cost has to be less than the benefit; b) can it get in the user's way? -- too many words may appear confusing; c) is it useful? is it integrated? d) she wants to see interoperability with established thesauri, e) she wants to see how the clusters will be updated in relation to other terms and other clusters. She also asked if there are other alternatives, such as: a) "find more like this," b) need to expose people to controlled vocabulary, c) what about a browsable interface vs. searching? d) categorized results are helpful but not if there is too much to read d) users are probably expecting full-text searching but the databases they are accessing don't usually have it.

She wants to see a prototype and usability testing. Users are getting used to accepting something close as okay. They can't see what Google or Amazon have hidden behind the scenes to do their searching. She asks: a) how would this be configured? It needs interoperability with thesauri and classification schemes; b) would there be enough subject access in a general overall subject database? c) there needs to be a business model to maintain it, probably a subscription; d) users are getting more sophisticated with their searching - this seems suited to perpetually novice users; e) how does one include other interfaces? Z39.50, Yahoo, Google? OCLC's work with Google is an effort to get out to users who normally don't come in the library and bring them in. She believes usability testing is necessary.

The testing needs to assess different implementations, different interfaces, existing vocabularies. Will users drill down? Will they quit before the get to the 'right' word? Proof of concept is needed to obtain funding or adoption.

Steve Neilsen, Dynix

One of the number one user questions is: why did that record come up? I don't see it. There are a number of issues with this proposal including expense, maintenance, and the need to multiply by language. Dynix actually has a feature that will cluster but he sees problems from their experience. They tried to put options up for users to pick from or chose add/modify/or not use. They already ask, why did I get this hit? Will a cluster of many words make it harder? Users start with top of drop-down search boxes so what is put in a list in what order is important. Even though Dynix has something of a cluster search, he thinks it is early for a release of clustered techniques. Users don't understand relevancy, adjacency or proximity, at least certainly not those terms. Funding would be enormous. He is concerned about the speed of search response time. If it takes seven seconds to get 15 million hits it is too slow. Who would build these clusters and at what cost?

Questions

Clustering sounds like extensive cross-referencing. Why can't we just add to existing subject heading authority records? Bendig agreed but that has its own expense. Judy explained clustering gives a core concept with words surrounding it without hierarchy. Nielson said full cataloging should approach adequate terms but Bates included more keywords, e.g. title words. He agreed more of the work should be done in authority records, particularly enhancing the cross-referencing structure so less needs to be in the bibliographic record. Bendig suggested we could do a better job of explaining what our related terms mean.

Can an ILS do an internal on-the-fly re-search of related keywords in subjects [and/or] authority records to create a cluster? Kaplan pointed out that all projects cited by Bates have died. Bendig explained that OCLC doesn't show some of the pairing that takes place behind the scenes.

There is a value to meeting the needs of undergraduate, low-level searching, but that we also have sophisticated searchers we need to help. The patron needs to be able to meaningfully limit and refine searches. Kaplan is concerned about maintaining and including other languages and even the evolution of language. Bendig likes the idea of clustering but not the cost of maintaining it.

Work is being done by the Library of Congress to increase terminology in reference structures. Couldn't keyword searching of authority records be used instead of clustering? Kaplan said authority keyword searching is not generally done and that too would need analysis.

Thesaurus descriptors are an intermediary between the authority term and pre-coordinated strings. The problem is precision. Couldn't we make it easier to do keyword searching of 1XX 4XX 5XX and scope note fields of authority records, then allow users to narrow their search? Kaplan stated that bibliographic instruction also has a role in teaching better searching.

Part 2. Bringing Subject Access Together Through Interoperability

The ALCTS Subject Analysis Committee (SAC) established the Subcommittee on Semantic Interoperability to address subject access issues across thesauri and/or multiple languages among systems. This can be applied to both online library systems and web interfaces. Whereas the report by Marcia Bates, among other things, proposes a clustered vocabulary approach that has as its foundation a managed vocabulary, the Subcommittee on Semantic Interoperability investigated methods where separately existing thesauri might be managed to assist users seeking information.

Subcommittee Charge

Specific tasks include, but are not necessarily limited to:

a) An inventory of known semantic interoperability projects, with descriptions;

b) An evaluation of selected projects, in terms of those projects' stated objectives;

c) An investigation of the various concepts involved in the harmonization of indexing languages, such as switching languages, concordance tables, front-end thesauri, meta-thesauri, and mapping.

The Subcommittee had several goals. Among them, one was to provide recommendations to serve as guidelines in structuring a system that supports semantic interoperability among vocabularies by employing one or several methods such as: 1) harmonization of indexing languages; 2) switching languages; 3) concordance tables; 4) front-

end thesauri or front-end "cluster"; 5) metathesaurus; 6) mapping (methodologies). Examples of mapping methodologies include: a) among controlled vocabularies in the same language: thesauri, controlled lists of keywords, ontologies, clustering approaches, taxonomies, lexical databases, concept maps/spaces, semantic road maps, etc.; b) among multiple vocabularies in different languages and classification systems; c) between a controlled vocabulary and a universal classification system; d) between classification systems; e) to a new system/metathesaurus; f) within (e.g. LCSH/Mesh); g) to another thesaurus or classification not used by the participants. Another goal is to guide development of database management structures to allow automated artificial intelligence and manual methods to create the appropriate relational links addressing: 1) multilinguality; 2) synonyms; 3) homographs; 4) singulars and plurals; 5) parts of speech; 6) cultural differences affecting meaning; 7) narrower and broader and related terms.

To address these issues, four speakers already working on research and/or implementation of projects employing semantic interoperability were invited to speak.

Lois Mai Chan, Professor, School of Library and Information Science, University of Kentucky

Interoperability Among Knowledge Organization Systems: Projects and Methodology

Knowledge organization systems may include term lists, thesauri, or classification schemes. The purpose of semantic interoperability is to enable cross-domain and cross-vocabulary searching. In information retrieval, users should not need to know terms used or even be restricted to searching one database at a time. What is needed is an interoperable search approach application. There are two types: a) either verbal (vocabulary foundation) or classification scheme (numerical foundation).

Chan discussed a number of projects:

- Renardus uses Dewey Decimal Classification (DDC) as a switching language and additionally has a graphic navigation structure also based on DDC.
- HEREIN is used by eleven countries. They created a thesaurus, Interlingua, based on the different projects' definitions. Terms were then matched semantically as to whether they were equivalent, hierarchical, or associate. Selecting one term from many meant there were certain issues in mapping that needed consensus.
- HILT has a goal to have a terminology server that will identify a resource likely to match the term entered by the users, which results in a list of terms (mapped to DDC) that can be limited, and then can go to a specific online resource already identified with DDC numbering.
- DARPA has a query vocabulary, Unfamiliar Metadata Vocabularies, which tries to map a user's input term to entry vocabulary (one based on several databases). The terms are mapped to the class number appropriate to the database, e.g. a patent number if searching a patent database or LCSH if searching bibliographic records.

She discussed a number of methods of improving interoperability:

- Derivation/modeling - new vocabulary created known vocabularies
- Translation/adaptation - translation to a new language with or without modification
- Satellite or leaf notes - specialized thesauri are treated as satellites of a super thesaurus which maintains a uniform structure
- Direct mapping - equivalence between words or word classification
- Switching language
- Co-occurrence mapping - works on an application level - mapped from bibliographic records based on co-occurrence of the thesauri
- Link - closely linked linguistically; linking is done through a temporary union list generated on-the-fly. An example is CAMed - a user enters "acupuncture" then can select a database or all databases
- Linking through a thesauri server protocol. An example is Alexandria Digital Library.

She continued with the following points:

- Once mapping is established, there is a need to manage authority records, e.g. Northwestern University's LCSH and Mesh database or Wilson's megathesaurus.
- Concordances utilize one master scheme others are matched to.

- In semantic networks, such as UMLS, each unit in the network has clustered around it equivalent terms based on a hierarchical structure.
- In lexical databases, such as WordNet, words are organized into synonym sets.

Cultures have an impact. One has to be careful to not stretch the meaning of words to the point that a term is no longer understandable in the local language or meaningless in the target language. The challenge is one of defining equivalence within cultural linguistics.

Jean-Frédéric Jauslin, CENL Chairman, Director Swiss National Library

Cross-language subject access to information: challenges and solutions, the example of MACS < https://ilmacs.uvt.nl/pub/ >

Jauslin pointed out that trying to be multilingual also has political issues. There are 25 countries in the European Union. There is also a Council of Europe with 45 countries. He explained the principles behind MACS: a) maximize investment; b) use existing headings; c) follow standards (XML, Z39.50) - you just can't create something new. They began mapping LCSH at the British Library with Rameau at the Bibliothèque Nationale and will add SWD from the Deutsche Bibliothek, followed by more languages in other countries.

What are they doing? They use a link management interface and a user interface.

The link management interface features:

- Management system using federative management - each library is responsible for linking in their own language; no central management
- Annotations are possible to indicate problems
- All languages are of equal value
- Have mapping clusters
- Use a link to hierarchical numbering
- Only headings are linked - hierarchies can't be mapped
- If there is no match, create new authority records for future mapping

User interface features:

- User can enter a term, pick a language, and pick a library
- System shows related terms
- User picks what he wants and terms are searched in selected catalogs
- There are 30,000 LCSH-Rameau links
- There is a new project underway in Europe, The European Library, to allow access to all libraries in Europe.

Pat Kuhr, Wilson Company

Putting the World Back Together: Mapping Multiple Vocabularies into a Single Thesaurus

<http://www.hwwilson.com/Databases/omnifile.cfm>

We librarians used to be very generalized in our provision of information services. We have become very specialized trying to match services to special users. With the advent of computers, users want it all back together - one search among multiple thesauri. Wilson settled on a megathesaurus of terms used across all their databases. On a first pass they created a thesaurus with 580,000 terms but were able to merge like terms to create a file of 340,000. Now they have a file of 900,000 terms.

Kuhr discussed the issues they had to address to create a functional megathesaurus. Singular vs. plural - are they the same or different? Should word phrases be inverted or direct? Headings and subdivisions don't mean the same thing both ways. They developed a "dotted" term list in which they used the computer to create an unpunctuated word string for review. They tried synonym merging, but free-floating subdivisions got in the way, so decided not to use them. Separating the subdivisions resulted in 20,000 new headings. This allowed them to have synonyms with only one variable that could be mapped.

Kuhr said cross-references are very important. Homonyms needed to be qualified to clarify one from another. They had to build hierarchical structures to show relationships. Once they had a structure, they could add broader and/or narrower terms.

They developed a product, OMNI. It maps thesauri to bibliographic records. The thesaurus is visible and it can be worked up and down. Terms are ranked by being a heading, then a cross-reference, some of the terms, and finally words in the title.

Diane Vizine-Goetz, OCLC Research Scientist

Advancing Semantic Interoperability through Terminology Services

Web services are modular, web-based, machine-to-machine applications that can be combined. Mappings are done from one term in one vocabulary across languages or across domains. There are many vocabularies in a variety of formats. Some are free, some are proprietary; some are easily accessible by PCs, some not. Vizine-Goetz addressed schema transformations, record enhancement, and access.

To address schema transformations, they began by encoding MARC21 in XML following principles of Simple Knowledge Organization Systems (SKOS) used for developing the semantic web. They also employed Zthes, Z39.50 profile for thesaurus navigation (Z39.19). They have been studying where there is overlap in schemes, for example they might find matching non-preferred terms and broader terms, but find a related term doesn't always appear. First they map to MARC21, then maybe add some information. Then they add additional vocabulary mappings, e.g. Zthes. SKOS has adopted definitions used in the Renardus project, e.g. "is exact match" or "is format of." They enhance the record with provenance information using the MARC21 Organization Code and persistent identifiers of the type "info:uri" and then perform inter-vocabulary mapping.

The Info:URI is a registration of namespace. In the current project, they must identify LCSH, DDC, or GSAFD in defining a concept. MARC21 code lists are used where appropriate for codes, language and organization identifiers. Descriptions are at the expression level. They have done direct equivalence between terms. For co-occurrence they have used statistical mappings via computer.

How are these mapped vocabularies being made available? OAI protocol is being used. GSAFD has an OAI harvestable form. GSAFD has a user interface and an OAI viewer. They have developed a searchable database generated via computer. It is not really a user interface as such, but a way a system could do it. So far they have GSAFD available publicly. They are working on a Dewey version wherein the top three levels are being translated into French, German, and Spanish. See OCLC's Terminology Services for more information.62

Questions followed.

Chan is still doing research on the usability of different projects, but how they are set up depends somewhat on user needs.

Vocabulary lists are dynamic. Has Wilson built in a method of dealing with change? Kuhr said they work with 500 new headings a week. They keep old ones as cross-references. They are also working backwards to add indexes retrospectively and label old terms as historical. When linking to LCSH, they automatically update bibliographic records. Jauslin said that since all headings are to be linked, if something is missing, it is visible to staff. Vizine-Goetz said OCLC has information about the history and links. The encoding schemes all allow you to know dates of changes and by whom they were done. A guest (Tony Olson) noted that Northwestern's LCSH-MeSH database is checked and updated weekly.

Is there more information about the European Library? It has a web site63 and the European Union has set up an office in Holland which will be open at the end of the year.

62 Terminology Services. http://www.oclc.org/research/projects/termservices/default.htm

63 The European Library (TEL) - The Gate to Europe's knowledge. http://www.europeanlibrary.org/ (July 1, 2004)

Appendix III. Annotated Bibliography

Note: Text is often recorded as it appears in the article.

The Alexandria Digital Earth Modeling System (ADEPT): Towards a Distributed Digital Model of the Earth in Support of Learning. <http://www.alexandria.ucsb.edu/adept/proposal.pdf> (2003)

> ADEPT is being developed as an integrated learning environment based on ADL geospatial digital library technology. It is currently used to teach Physical Geography to undergraduate students at University of California, Santa Barbara.
>
> ADEPT will provide gazetteer, thesaurus, and geo-ontology services. The gazetteer will be built from the ADL project gazetteer and serve as an index supporting transformations between named places and geographic coordinates. Thesauri provide a basis for resolving semantic inconsistencies, for example between alternative names for geographic feature types. They will build a set of core thesauri covering geographic representations of regions in space and space relations with objects. Geo-ontology services: the vocabularies used to describe geographic features and phenomena vary by discipline. By knowing which ontologies are used in different contexts, and by mapping between them, it is possible to make appropriate semantic correlations between different information sources. They will build 1) a set of domain-specific ontologies for geospatial information; and 2) a set of domain-independent ontologies supporting system, syntactic, and structural interoperability.

American Library Association. "Subject Data in the Metadata Record: Recommendations and Rational." Division of Association for Libraries and Technical Services, Cataloging and Classification Section, Subcommittee on Metadata and Subject Analysis (1999). <http://www.ala.org/ala/alctscontent /catalogingsection/catcommittees/subjectanalysis/metadataandsubje/subjectdata.htm> (Sept. 19, 2005)

Ardö, Anders, Godby, Jean, Houghton, Andrew, Koch, Traugott, Reighart, Ray, Thompson, Roger and Vizine-Goetz, Diane. "Browsing Engineering Resources on the Web : *a General Knowledge Organization Scheme (Dewey) vs. a Special Scheme (EI)039.*" In *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth ISKO Conference*, 10-13 July, 2000: 385-390

> The goal of the DESIRE II project is to explore automated methods for gathering and organizing Web resources to improve resource discovery on the Internet. Researchers at NetLab and OCLC provided searching and browsing of a test collection of engineering documents on the Web. The goal of the project is to explore simple methods of automatic classification to provide subject browsing of a robot-generated engineering index. At NetLab the documents were automatically classified and organized using an engineering-specific scheme, the Engineering Index (Ei) Thesaurus and Classification; at OCLC the Dewey Decimal Classification (DDC), a general knowledge organization scheme was used. The enhanced DDC database includes several mechanisms for incorporating new terminology. Scorpion is used to do automatic class number assignment. WordSmith software was used to create a small set of high-quality topical vocabulary suitable as an index or browse display and that can supplement the subject indexes provided by the Ei Thesaurus or the DDC.

Ardo, Anders, Marten Berggren, Traugott Koch Reidun Kringstad. *Nordic Interconnected Subject-based Information Gateways (NISBIG): Final report* (2002). <http://www.lub.lu.se/nisbig/slutrapport.html> (Oct. 8, 2002). No longer available.

> Project final report addresses all types of metadata including subject access for use in a quality-controlled subject gateway. It discussed problems and limitations, and recommended pursuing Renardus, IMesh Toolkit, etc. Subject gateways were developed in order to support discovery and retrieval of Internet resources as well as to integrate Internet resources with "traditional" library resources. Apart from gaining experience with content and metadata profiling and classification mapping for cross-browsing, the main technical goal of the project was to explore the applicability of the

LDAP-based Isaac Network software developed by the US Internet Scout Project to provide cross-searching between the involved three Nordic subject gateways and other gateways joining the Isaac Network.

Baker, Thomas and Makx Dekkers. "Identifying Metadata Elements with URIs : the CORES Resolution." *D-Lib Magazine,* 9, no. 7/8 (July/August 2003). <http://www.dlib.org/dlib/july03/baker/07baker.html> (2003).

At a meeting organized by the CORES Project (Information Society Technologies Programme, European Union), several organizations regarded as maintenance authorities for metadata elements achieved consensus on a resolution to assign Uniform Resource Identifiers (URIs) to metadata elements as a useful first step towards the development of mapping infrastructures and interoperability services. The maintainers of GILS, ONIX, MARC 21, CERIF, DOI, IEEE/LOM, and Dublin Core reported on their implementations of the resolution and highlighted issues of relevance to establishing good-practice conventions for declaring, identifying, and maintaining metadata elements more generally. In November 2002, they committed to implementing the agreement to define URI assignment mechanisms, assign URIs to elements, and formulate policies for the persistence of those URIs.

Baker, Thomas. "What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives." *Journal of Digital Information,* 2, no. 2 (Nov. 6, 2001). <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/> (Aug. 6, 2002).

Rachel Heery and Manjula Patel have defined application profiles as 'schemas which consist of data elements drawn from one or more namespaces, combined together by implementers, and optimized for a particular application.' By definition, such profiles depend for their elements on namespaces. Namespaces, in this context, are element sets maintained as stable points of reference. They serve to 'identify the management authority for an element, support definition of unique identifiers for elements, [and] uniquely define particular data element sets or vocabularies." The registry prototyped in the DESIRE Project focused on the disclosure of information about the authoritative use of metadata -- element definitions, usage notes, allowed schemes, and mappings to other namespaces -- and explored typical user queries. The SCHEMAS registry builds on the DESIRE experience.

Bates, Marcia. *After the Dot-bomb: Getting Web Information Retrieval Right this Time*. 2002 <http://www.firstmonday.org/issues/issue7_7/bates/> (Sept. 28, 2002).

The author proposes using systems already design for information retrieval, e.g. faceted classification and information resources thesauri, which have an internal structure, concept clusters, etc. The long-term solution to index the Web is probably overlapping methods of classifying and indexing knowledge. She disapproves of the use of the word "ontology" since it refers to the philosophical issues surrounding the nature of being.

Bates, Marcia J. “The Cascade of Interactions in the Digital Library Interface.” *Information Processing & Management*, 38 (2002): 381-400.

---. “Indexing and Access for Digital Libraries and the Internet: Human,

Database, and Domain Factors.” *Journal of the American Society for Information Science*, 4913 (1998): 1185-1205.

---. “Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution,” *Emerging Frameworks and Methods : Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)*, edited by Harry Bruce, Raya Fidel, Peter Ingwersen, and Pertti Vakkari. Greenwood Village, CO: Libraries Unlimited, 2002, p. 137-150.

Bates, Marcia. *Task Force Recommendation 2.3, Research and Design Review: Improving User Access to Library Catalog and Portal Information : Final Report.* (2003). <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf> at: http://www.loc.gov/catdir/bibcontrol/actionplan.html (Feb. 7, 2006)

Selections:

It is recommended that with regard to access vocabulary:

- A cluster vocabulary be created, based on the searcher vocabulary developed by Sara Knapp (1993, 2000), if she and her publisher agree.
- For the price of a share of the maintenance of the database, libraries and commercial firms may subscribe to the searcher vocabulary database, and install it in their catalogs, portals, and websites.
- With experience, other types of clusters are added--for names, works, geographical locations, etc.
- Access to catalogs and portal information should be available both directly through and around the vocabulary database. In this way, searchers may choose to use the database or not, and, if they do choose it, they do not have to enter and exit a separate database (a violation of the ever-present Principle of Least Effort).
- Institutional users may link the searcher vocabulary with their own controlled vocabulary. As a result, users of these sites may input their search term(s), be shown a cluster of terms, including "legitimate" controlled terms, and use the clusters as a basis for selecting terms for either controlled vocabulary or keyword searching.
- With this vocabulary as a core, one or two lexicographers are hired cooperatively to maintain the searcher vocabulary, adding popular new terms as they come along, and adding terms found by cooperating organizations in "zero hit" searches. As changes are made in the vocabulary, rather than in millions of individual cataloging records, cultural and research changes can be accommodated much more rapidly and cheaply.
- These vocabularies become part of a "Vocabulary Headquarters" (VHQ) website, supported by the library community or organizations therein.

It is recommended that with regard to bibliographic families:

- Preliminary agreement be gained on what shall constitute bibliographic families at the work level, probably based on the work of Tillett, Smiraglia, Hickey, and others. It may be found that work-sets, as described by Hickey et al. should also be considered.
- As these bibliographic families probably follow the Bradford Distribution, there will be some few that are very large, and many that are very small or singletons. As the larger families are much more likely to cause difficulties for searchers, and as they are also often around canonical works that attract a great deal of research and cultural interest, the larger families should be grouped first.
- At first on an experimental basis, individual libraries or other institutions offer each to do the work to collect just one large family (from records already created at the individual level). The results of these experiences are shared at conferences and other meetings.
- Based on these experiences, criteria are finalized for the creation of bibliographic families. Libraries may acquire the cataloging information for the families in a manner similar to the currently existing cooperative cataloging arrangements.
- Further experience will also provide enlightenment regarding just how far down the chain of family size the cooperative effort should go.

- Eventually, with further technological advances, it becomes possible that whenever a searcher happens on a record that is part of a bibliographic family, the searcher may click on a "related records" link and see displayed on the screen the progenitor record plus links to all the different types of bibliographically related records arrayed around the core record.

It is recommended that with regard to staging of access to records:

- Libraries and other information institutions take as an objective the approach of providing staged access to information that drops down into the information in a 1:30 ratio. For example, in a catalog a book has a title of a few words and an abstract of about 30 times the number of words in the title. With this ratio specifically in mind, the effectiveness of catalogs so designed can be tested.
- Current cooperation with publishers can be extended, including use of book flap and contents information that is already in electronic form for catalog records.
- The online bookstore, amazon.com, contains within it many of the design features that have been recommended by catalog and database user studies over the years. Amazon.com can be seen as a source of ideas and prior testing of design features.

Bates, Marcia J. "Toward an Integrated Model of Information Seeking and Searching." *The New Review of Information Behavior Research*, 3 (2002): 1-15.

Becker, Hans J. "Cultural Heritage Projects: Renardus." Paper presented at TEL Milestone Conference, Frankfurt am Main, Germany, April 29-30, 2002 <http://www.europeanlibrary.org/ppt/tel_milconf_presentation_becker.ppt> (Aug. 7, 2002).

Goals: a) to improve access to existing academic subject gateway services in Europe; b) to develop a 'broker' service that will allow integrated searching and browsing of distributed resource collections; c) to develop models for sharing metadata, agreement on technical solutions and other standards. Subject gateway definition: quality controlled subject gateways and resource discovery broker systems with a target audience which is predominantly higher education and academic research communities across Europe: a) selection and collection development (human intellectual effort, certain policy with regard to collection development, documented selection criteria); b) collection management (maintaining or improving the level of quality of the collection, certain policy with regard to maintenance); c) resource description (all selected resources are described according to a fixed and documented metadata set, metadata are structured in well-defined semantic fields to enable structured searching); d) subject classification (all resources are indexed according to a subject classification scheme in order to enable subject browsing). Various aspects of the project will be addressed by different groups in areas called 'work packages'.

Beghtol, Clare. "The Iter Bibliography: International Standard Subject Access to Medieval and Renaissance Materials (400-1700)." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 74-80.

The Iter Bibliography contains unique provisions for subject analysis and access. It uses a combination of multiple LCSH headings and multiple DDC notations for subject specification in order to incorporate the strengths of each system, and it also provides uncontrolled keywords to cater to terms that would likely to be used by Medieval and Renaissance scholars.

Bird, Steven and Gary Simons. "The OLAC Metadata Set and Controlled Vocabularies." *ArXiv*, May 21, 2001. <http://arXiv.org/abs/cs/0105030> (Feb. 16, 2005)

This paper describes a new digital infrastructure for language resource discovery, based on the Open Archives Initiative, and called OLAC - Open Language Archives Community. The OLAC Metadata Set and the associated controlled vocabularies facilitate consistent description and focused searching.

Brickley, Dan and Libby Miller. *Imesh Tk: Subject Gateway Review Plan*, 2000. <http://www.ilrt.bris.ac.uk/discovery/2000/07/itk-sgr/> (Aug. 7, 2002)

The objective of the Subject Gateway Review is to ensure that the IMesh Tk architectural and technical strategies are well-grounded in the documented needs and practical requirements of the Internet cataloging community as they stand now, with a view to the next 2-3 years. The Review will be responsible for producing scope and prioritization guidelines and a literature review. The relationship between XML-based metadata systems, notably RDF and other traditions such as LDAP and X39.50 is not yet clear. XML's popularity stems in large part from its cross-domain generality: XML representations of white pages data, bibliographic metadata, structured documents etc. can (to some extent) exploit common tools and software components. One issue that the Subject Gateway Review will need to address is the distinction between data-format based interfaces and API/protocol interfaces. The latter addresses the possibility of tools such as on-the-fly adaptors that translate (say) Z39.50 queries into LDAP queries or vice versa, while the former addresses the need for common data formats/information models for data exchange. Need to address: Do gateway managers prefer query-time protocol mapping to scenarios in which they 'batch convert' (given some standard data format, e.g. some flavor of qualified Dublin Core) records to make them available in multiple search protocols?

Buchel, Olha and Anita Coleman. "How Can Classificatory Structures be Used to Improve Science Education?" *Library Resources & Technical Services*, 47, no. 1 (2003): 4-15

The Alexandria Digital Earth Prototype (ADEPT) project provides the test bed for instructional materials and user analyses. ADEPT is supported by the National Science Foundation Digital Libraries Initiative, Phase 2 and is a successor to the Alexandria Digital Library (ADL) project. <http://alexandria.sdc.ucsb.edu/~gjanee/thesaurus/>

Buckland, Michael, and others. "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies." *D-Lib Magazine*, 5, no. 1 (January 1999). <http://www.dlib.org/dlib/january99/buckland/01buckland.html> (Feb. 16, 2005)

Buckland proposes an entry module to help the user get started. Mapping entry vocabulary modules use classification clustering, exploit the combination of linguistic analysis with statistical methods, and is based on searching fragments within the metadata and databases, performing statistical and linguistic analysis, presenting the user with a familiar term.

There is always one additional vocabulary in play - the User's.

The network environment is leading to an increasing number of heterogeneous repositories, using diverse metadata vocabularies (categorization codes, classification numbers, index and thesaurus terms). This is creating more and more unfamiliar sets of terms users must employ to access Internet resources. It has been argued that the most cost-effective single investment for improving effectiveness in the searching of repositories would be technology to assist the searcher in coping with unfamiliar metadata vocabularies.

A DDC number is a word/meaning. The Relative index provides the English to DDC number translation. What is now needed is a natural language index ('ordinary English') to the Relative Index and/or DDC numbers. The Entry Vocabulary Module helps the searcher be more effective and, thereby, provides a value-added enhancement.

Research has focused on: development of tools to support the creation of Entry Vocabulary Modules; creation of a set of prototype Entry Vocabulary Modules for a challenging range of examples, including subdomains; deployment; use of natural language processing techniques in addition to statistical term co-occurrence; recommendations for the improvement of metadata documentation for numeric databases. Prototype available at: http://www.sims.berkeley.edu/research/metadata/oasis.html

Chan, Lois Mai. "Ensuring Interoperability among Subject Vocabularies and Knowledge Organization Schemes: A Methodological Analysis." (by Lois Mai Chan and Marcia Lei Zeng) *IFLA Journal* 28, 5/6 (2002):323-27. Paper presented at the 68th IFLA Council and General Conference, Glasgow, Scotland, Aug. 18-24, 2002. <http://www.ifla.org/IV/ifla68/prog02.htm http://www.ifla.org/IV/ifla68/papers/008-122e.pdf> (2002)

The ideal approach would be to provide "one-stop" seamless searching instead of requiring the user to search individual databases or collections separately. To enable such an approach, it is important to render the different knowledge organization systems, such as controlled vocabularies and classification schemes, interoperable within a single search apparatus. A number of projects are trying to achieve interoperability between and among different subject vocabularies (including both controlled and uncontrolled vocabularies) and knowledge organization systems. They include efforts at establishing interoperability among vocabularies in the same language or in different languages, among different classification schemes, and between controlled vocabularies and classification schemes.

---. "Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources: Issues and Challenges." In *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium* (Washington, DC: Library of Congress, Cataloging Distribution Service, 2001), p. 159-178. <http://www.loc.gov/catdir/bibcontrol/chan.html>

(Sept. 25, 2005) Vocabulary control for improved precision and recall and structured organization for efficient shelf location and browsing have contributed to effective subject access to library materials. The question is whether existing tools can continue to function satisfactorily in dealing with web resources. To meet the challenges of web resources, certain operational requirements must be taken into consideration, the most important being the ability to handle a large volume of resources efficiently and interoperability across different information environments and among a variety of retrieval models. Schemes that are scalable in semantics and flexible in syntax, structure, and application are more likely to be capable of meeting the requirements of a diversity of information retrieval environments and the needs of different user communities.

Chan, Lois Mai, Eric Childress, Rebecca Dean, Ed T. O'Neill, and Diane Vizine-Goetz. "A Faceted Approach to Subject Data in the Dublin Core Metadata Record." *Journal of Internet Cataloging,* 4, no. 1 / 2 (2001): 35-47.

For the Dublin Core metadata record, a new approach to subject vocabulary was investigated. Faceted Application of Subject Terminology (FAST) is based on the existing vocabulary in *Library of Congress Subject Headings*. It is applied in a simpler syntax. In FAST, non-topical (geographic, chronological, and form) data are separate from topical data and placed in different elements provided in the Dublin Core metadata record.

Chan, Lois Mai, Xia Lin, Marcia Zeng. "Structural and Multilingual Approaches to Subject Access on the Web." Paper presented at the 65th IFLA Council and General Conference, Bangkok, Thailand, Aug. 20-28, 1999. <http://www.ifla.org/IV/ifla65/papers/012-117e.htm> (Aug. 7, 2002)

A report in three parts.

Part I. Structural approaches to organizing web resources.

Using hierarchical or classification-based formats to organize web resources should have important advantages, among which are improved subject browsing facilities,

potential multi-lingual access and improved interoperability with other services. In the web environment, subject data often are separate from or reside outside the resources themselves. It can be stored in interfaces that link subject data to the resources but do not affect them otherwise. The advantage of "linking-to" rather than "storing-with" is flexibility. Desirable characteristics: a) intuitive, logical and easy to use … with expressive captions; b) flexible, adjustable, and expandable; c) useful in a wide range of settings; d) relatively easy to maintain and revise.

Part II. Knowledge Class.

The purpose of this research project is to create and test a device called "Knowledge Class," designed for customizing knowledge organization and access, to supplement and complement existing devices for Web users. Knowledge Class contains two basic components: a) an organizing framework, and b) interface for access to and retrieval of web resources. The organizing framework is a classified mini-thesaurus, consisting of a hierarchically structured collection of terms on a specific topic or discipline of interest or concern to an individual user. The user can initiate searches by selecting the display terms or by using pre-stored search strategies, which often contain synonyms and can also connect to sites previously discovered by clicking on links with pre-stored URLs.

Part III. Multilingual approach to subject access.

Multilingual processing has emerged as a key issue in the evolution of search engine technologies. Major search engines have developed new services functional as regional search guides in these areas: a) domain filtering, b) domain direction, c) mirror sites, d) language specific search, e) multilingual search, f) regional interfaces, g) localized subject directories.

The road towards a fully functional cross-lingual subject access is both optimistic and sophisticated. Many other technical issues as well as social and cultural issues also need to be addressed. These include character encoding support, user interface linguistic translation, support of culture-specific data formats (date, currency, etc.), user interface graphical modification (color, images), foreign products support (e.g. databases), and operating system compatibility. In summary, there has been an increasing need for effective mechanisms to organize web resources for exploration, discovery, and retrieval.

Cherry, Steven M. "Weaving a Web of Ideas." *IEEE spectrum*, Sept. 2002.

Software agents, robots, were not successful in dealing with semantics, with multiple meanings of words. The Semantic Web idea instead suggests that Web pages should contain their own semantics. Successful search engines have developed sophisticated methods of delivering documents. The Semantic Web aims to get to the information in the documents by using an ontology - a collection of related RDF statements, which together specify a variety of relationships among data elements and ways of making logical inferences among them. It addresses syntax, which is the set of rules or patterns according to which words are combined into sentences. Semantics is the meaningfulness of the terms - how the terms related to real things. Search engines have room for improvement. One method of improving Web searches proposed by user interface researchers at the Palo Alto Research Center is scatter/gather, which takes a random collection of documents and gathers them into clusters, each denoted by a single topic word. The user then picks several of the clusters and the software re-scatters and re-clusters them until a user gets a desirable set. Another method (Autonomy) is using a Bayesian network which is a pattern-matching engine that distinguishes different meanings of the same term and so "understands" them as different concepts.

Clark, Judith. "Subject Portals." *Ariadne,* 29 (Oct. 2, 2001). <http://www.ariadne.ac.uk/issue29/clark/> (Jan. 21, 2003).

The author describes a 3-year project to develop a set of subject portals or hubs, part of the Development Programme of the Distributed National Electronic Resource (DNER), funded by the JISC. The project aims to enhance resource discovery by developing a series of portals focused on the requirements of end-users located in a variety of learning environments within higher education sectors. The first phase of the project (2000-2001) was to build a Z39.50 cross search prototype at three RDN hubs, SOSIG, EEVL, and BIOME. The second phase adds HUMBUL and PSIgate. Sites are selected on the basis of selection criteria, cataloged following consistent practices, and analyzed by people with expertise with the relevant subject discipline. Links are checked daily in an automated process and all entries are updated regularly by subject specialists. These are classified using an appropriate controlled vocabulary.

RDN portals (http://www.rdn.ac.uk/projects/) are primarily concerned with technologies that broker subject-oriented access to resources. Effective cross-searching depends on consistent metadata standards. Z39.50 is the standard that has been adopted for preliminary cross-search functionality. Further functionality is being developed using RSS (Rich Site Summary) and OAI (Open Archives Initiative). Other standards applications that underpin the portals are notably Dublin Core and a variety of subject-specific thesauri such as the CAB Thesaurus and MeSH.

Clavel-Merrin, Genevieve. "Multilingual Access to Subjects: the MACS Prototype." Paper presented at TEL Milestone Conference, Frankfurt am Main, Germany, April 29-30, 2002. <http://www.europeanlibrary.org/doc/tel_milconf_presentation_clavel.doc> (Oct. 8, 2002)

National and other libraries have invested heavily in encyclopedic subject heading languages that offer a complementary access to their collections. The tasks of creation, management and maintenance of these subject heading languages require significant resources, and rely generally on co-operation so that this approach is naturally considered as a way to extend access to users from other linguistic areas. Therefore, the CoBRA+ Working Group on Multilingual Subject Access conducted a feasibility study between Autumn 1997 and February 1999 on linking headings between the three *Subject Heading Languages* (SHL's) used in the Bibliothèque Nationale, Die Deutsche Bibliothek, the Swiss National Library and the British Library. The SHLs used were RAMEAU, SWD/RSWK and LCSH. As a result the MACS (Multilingual Access to Subjects) project was set up to develop a prototype system testing the recommendations and findings of the feasibility study.

---. "The Need for Co-operation in Creating and Maintaining Multilingual Subject Authority Files." Paper presented at the 65th IFLA Council and General Conference, Bangkok, Thailand, Aug. 20-28, 1999. <http://www.ifla.org/IV/ifla65/papers/080-155e.htm> (Aug. 7, 2002).

In 1997, the Conference of European National Librarians (CENL) asked Computerized Bibliographic Record Actions (CoBRA+) to consider the problem of multilingual subject access to bibliographic databases and conduct a pilot study in French, German and English. The aim of the study was to establish equivalents between RAMEAU, SWD/RSWK and LCSH: 1) establish a methodology for the selection and linking of headings, 2) link headings and analyze the results in the selected subject areas, 3) see the practical applications of these linked headings by indexing a test group of titles, 4) compare the indexing of titles in other subject fields. The study did confirm the following: 1) the number of headings and subdivision which may be combined and the complexity of the strings which may result varies from language to language, 2) the number of strings that may be applied to a document also varies according to the different rules applied.

CORES - A Forum on Shared Metadata Vocabularies. <http://www.ukoln.ac.uk/metadata/cores/> (Aug. 6, 2002)

The CORES project is funded within the Information Societies Technology (IST) Programme, managed by the Information Society Directorate-General of the European Commission. The central objective of the CORES project is to encourage

the sharing of metadata semantics. CORES will address the need to reach consensus on a data model for declaring semantics of metadata terms in a machine-readable way. A consensus of the ground-rules for declaring standard definitions of terms, as well as local usage and adaptations, will enable the diversity of existing standards to "play together" in an integrated, machine-understandable Semantic Web environment. In order to achieve this level of interoperability, CORES will support applications re-using and adapting terms maintained by key organizations and standardization initiatives.

Day, Michael. "Metadata in Support of Subject Gateway Services and Digital Preservation." Draft version of paper presented at Electronic Resources: Definition, Selection and Cataloguing, Rome, Italy, Nov. 2001.<http://www.ukoln.ac.uk/metadata/presentations/rome-2001/paper.html> (Aug. 8, 2002).

This paper provides an introduction to two of the metadata-related projects in which UKOLN has been a partner. It first describes the development of services known as quality controlled subject gateways and looks in more detail at the Resource Discovery Network and the EU Renardus project. It then provides an outline of recent preservation metadata initiatives and describes the way the OAIS model has been used in the Cedars project.

DESIRE Information Gateways Handbook (2000). <http://www.desire.org/handbook> (Aug. 7, 2002).

This is a thorough guide to creating a high quality portal or gateway on the Internet. Section 2 of the handbook covers important decisions to be made when setting up a new gateway (such as choosing a metadata format, designing a user interface, writing a selection policy) but also covers issues such as cataloging and resource discovery. Subject gateways should aim to guarantee high quality resources and facilitate subject-based access to the collection. Information gateways are characterized by their creation of third-party metadata records - individual descriptions of Internet resources held in a database that have separate fields for different attributes of the resources, such as title, author, URL, etc. The role of cataloging rules or guidelines is to specify how the content of a metadata format is entered in accordance with certain rules and will often include additional features such as classification, subject analysis and authority control. Once a metadata format is selected, a metadata content standard needs to be selected or developed to address dates, language codes, name authority files, and subject information. The use of classification schemes, keywords and thesauri are central features of the formal resources descriptions provided by a gateway service. Browsing (through a directory-like structure) is usually based on subject classification schemes or thesauri. Classification schemes differ from other subject indexing systems, such as subject headings and thesauri, by trying to create collections or related resources in a hierarchical structure. Cross-browsing two or more gateways is useful, but difficult. Mapping methods can be used, e.g. DESIRE II and has been tested by ROADS. "As with cross-browsing using classification schemes, cross-searching only becomes possible if either of the different catalogs use the same vocabulary or if a mapping has been done between two or more different schemes." Gateways need to address the language needs of their audiences. Users may want to search a multilingual collection by using queries in one language or to retrieve documents in a number of specific languages, preferably also via an interface in the language of their choice. There are two issues: the storing, processing, and presentation of information in many languages; and multilingual search and retrieval. Each chapter includes a bibliography.

Dhamankar, R., Y. Lee, A. Doan, A. Halevy, and P. Domingos. “iMAP: Discovering Complex Semantic Matches between Database Schemas.” In *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on management of data,* Paris, France, 2004, p. 383-394.

Doerr, M. "Semantic Problems of Thesauri Mapping." *Journal of Digital Information*, vol. 1, no. 8 (Mar. 26, 2001) <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/> (Sept. 25, 2005)

With networked information access to heterogeneous data sources, the problem of terminology provision and interoperability of controlled vocabulary schemes such as thesauri becomes increasingly urgent. Solutions are needed to improve the performance of full-text retrieval systems and to guide the design of controlled terminology schemes for use in structured data, including metadata. Thesauri are created in different languages, with different scope and points of view and at different levels of abstraction and detail, to accommodate access to a specific group of collections. In any wider search accessing distributed collections, the user would like to start with familiar terminology and let the system find out the correspondences to other terminologies in order to retrieve equivalent results from all addressed collections. This paper investigates possible semantic differences that may hinder the unambiguous mapping and transition from one thesaurus to another.

Dunsire, Gordon. "Joined up Indexes: Interoperability Issues in Z30.50 Networks." Paper presented at the 68th IFLA Council and General Conference, Glasgow, Scotland, Aug. 18-24, 2002. <http://www.ifla.org/IV/ifla68/prog02.htm> <http://www.ifla.org/IV/ifla68/papers/022-144e.pdf> (Jan. 18, 2003).

The paper discusses issues in the interoperability of indexes to metadata records in distributed information retrieval networks, based on the findings of Cooperative Academic Information Retrieval Network for Scotland (CAIRNS) and Scottish Collections Network Extension (SCONE) projects. The two have evolved services which together provide user-driven collection identification and selection mechanisms and the ability to cross-search related metadata for item discovery and access. The CAIRNS Cataloguing Issues Working Group identified a number of factors affecting cross-searching of metadata indexes for authors, titles, subjects and control numbers, including local cataloging policies, content standards, and index structures. The SCONE project has identified issues in subject indexing at the collection level, in particular the relationship between collections with specific subject content and general collections for which Conspectus-type subject strength mappings are appropriate.

Duval, Erik, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. "Metadata Principles and Practicalities." *D-Lib Magazine*, 8, no. 4 (April 2002). <http://www.dlib.org/dlib/april02/weibel/04weibel.html> (May 1, 2002).

The focus of the article is metadata in general, but some information is apropos to subject analysis. The use of controlled vocabularies is another important approach to refinement that improves the precision for descriptions and leverages the substantial intellectual investment made by many domains to improve subject access to resources. The Dewey Decimal Classification System, for example, affords a multilingual classification system long used in traditional library environments that can be applied to electronic resources as well. There are hundreds of domain-specific thesauri and classification systems, as well, that can be imported into the Web metadata architecture to support subject descriptions. Specifying the use of a particular vocabulary in a given collection of metadata will allow applications to provide more coherent search and browsing facilities. It is essential to adopt metadata architectures that respect linguistic and cultural diversity. However, unless such resources can be made available to users in their native languages, in appropriate character sets, and with metadata appropriate to management of the resources, the Web will fail to achieve its potential as a global information system.

By elucidating shared principles and practicalities of metadata, the authors hope to raise the level of understanding among our respective (and shared) constituents. The ideas in this paper are divided into two categories: a) *Principles,* and b) Pr*acticalities.*

Eden, Brad. "Metadata and its Application." *Library technology reports*, 38, no. 5 (Sept./Oct. 2002): p. 1-77.

This report is a guide to current metadata standards and their application. Major standards are included. The report examines: which metadata is suitable for certain

libraries, linking initiatives and how they relate to metadata, how to use metadata to build an enriched library catalog, how metadata assists in natural language recognition technology.

Creating metadata is important because metadata facilitates the discovery of relevant information and resources. Metadata help identify resources, distinguish among dissimilar resources, bring similar resources together, allow resources to be found by relevant criteria and give location information. Metadata promotes interoperability if accompanied by careful mapping of data elements and crosswalking of standards. Interoperability allows multiple systems to exchange data with minimal loss of content and functionality, regardless of different hardware and software platforms, data structures, and interfaces. The use of metadata allows resources to be searched seamlessly across networks through crosswalks and shared transfer protocols. Metadata ensures resources will be accessible into the future, can provide persistent and unique digital identification, can track rights and reproduction information, and organize information. Problems with polysemy (words with multiple meanings), ambiguity of meaning, and synonymy can all be alleviated by the proper application of metadata, either manually or through selected harvesting. Interoperability has become the key shared focus if multiple metadata standards are to survive.

Fitch, Kent. *Taking RDF and Topic Maps Seriously.* <http://ausweb.scu.edu.au/aw02/papers/refereed/fitch2/> (July 18, 2002)

One of the core ideas behind the Semantic Web is the creation of machine-processable relationships between resource identifiers (URI's). Two often discussed ways of representing those relationships are RDF and Topic Maps. A topic is simply a representation of any subject or concept of interest; it is the 'proxy' of that subject in the topic map. Topics have characteristics: names of different types, roles played by the topic in associations with other topics, occurrences, which are resources pertinent to the topic, also of different types. Topic characteristics can be asserted as being valid with in a "scope" which acts as a context for assertions. Topics in a Topic Map each play an identified "role." Topic Maps tend to start with the 'abstract' and optionally extend to include concrete resources, whereas RDF tends to start with defining relationships between concrete resources and optionally building abstract conceptual links between those relationships.

Fra^ncu, Victoria. "The Impact of Specificity on the Retrieval Power of a UDC-based Multilingual Thesaurus." *Cataloging & Classification Quarterly,* v. 37, no. 1 / 2 (2003): p. 49-64.

Summary: The article describes the research done over a bibliographic database in order to show what impact the specificity of the knowledge organizing tools may have on information retrieval. For this purpose two multilingual Universal Decimal Classification (UDC) based thesauri having different degrees of specificity are considered. Issues of harmonizing a classificatory structure with a thesaurus structure are introduced, and significant aspects of information retrieval in a multilingual environment are examined.

Franklin, Rosemary Aud. "Re-inventing Subject Access for the Semantic Web." *Online Information Review,* 27, no. 2 (2003): 94-101.

Second generation web research is beginning to model subject access with library science principles of bibliographic control and cataloging. Harnessing the Web and organizing the intellectual content with standards and controlled vocabulary provides precise search and retrieval capability, increasing relevance and efficient use of technology. Current research points to a type of structure based on a system of faceted classification. This system allows the semantic and syntactic relationships to be defined. Controlled vocabulary can be assigned, not in a hierarchical structure, but rather as descriptive facets of relating concepts.

Freyre, Elisabeth and Max Naudi. "MACS: Subject Access Across Languages and Networks." In Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 3-10.

This paper explains how MACS meets the challenge of multilingualism created by the new network environment. Based on the equality of languages and making use of work already carried out by the partners, the MACS project sets up equivalences between subject heading languages. It enables in this way, with a monolingual subject search, to retrieve all the pertinent documents held in catalogs in different languages.

Garrison, William A. "Retrieval Issues for the Colorado Digitization Project's Heritage Database." *D-Lib Magazine*, 7, no. 10 (Oct. 2001). <http://www.dlib.org/dlib/october01/garrison/10garrison.html> (Oct. 26, 2002).

The Colorado Digitization Project (CDP) is a collaborative initiative involving Colorado's archives, historical societies, libraries and museums. The project is creating a union catalog of metadata records and has developed tools for the creators of metadata records, the assignment of subject headings, and the use of name headings. The CDP is also investigating the use of Dewey Decimal Classification number through WebDewey to allow linkage of general subject terms and highly specialized subject terms within a subject browse feature of the union catalog.

Geisselmann, Friedrich. *CARMEN. WP12: Cross Concordances of Classifications and Thesauri*, 2004. <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.en> (Jan. 2005)

The goal is to allow an integrated search for subject aspects in distributed data holdings with different intentional emphases taking into account the conceptual differences of the applied thesauri and classifications by cross concordances.

Godby, Carol Jean and Ray Reighart. "Terminology Identification in a Collection of Web Resources." *Journal of Internet Cataloging*, 4, no. 1 /2 (2001): 49-65.

The primary goal of OCLC's WordSmith project was to obtain subject terminology directly from raw text. The hypothesis was that reliable subject terms can be automatically collected, re-used, and organized into thesaurus-like objects that enhance access to Internet material that is too time consuming to catalog by hand.

Godby, C. Jean. *The WordSmith Indexing System*. <http://www.oclc.org/research/publications/arr/1998/godby_reighart/wordsmith.htm> (Dec. 27, 1999).

The OCLC WordSmith indexing system uses the results of research in computational linguistics to implement a series of largely statistical filters to identify descriptive vocabulary in collections of English-language text of arbitrary subjects.

Godby, Carol Jean and Jay Stuler. "The Library of Congress Classification as a Knowledge Base for Automatic Subject Categorization." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 163-169. < http://staff.oclc.org/~godby/auto_class/godby-ifla.html> (accessed Oct. 26, 2002)

This paper describes a set of experiments in adapting a subset of the Library of Congress Classification for use as a database for automatic classification. A high degree of concept integrity was obtained when subject headings were mapped from OCLC's WorldCat database and filtered using the log-likelihood statistic. The project had three goals: 1) to adapt the LCC for use as a knowledge base for automatically classifying full text, 2) to exploit the LCC's structure for online subject-oriented browsing, and 3) to make the results of the work freely available to the library community.

Harding, Chris. "3 questions: Semantic Interoperability Defined." *ITBusinessEdge*, (June 16, 2005). <http://www.itbusinessedge.com/item/?ci=1172> (July 18, 2005).

An example from the business arena for the need for semantic interoperability among records.

Harken, Shelby E. "Enriching Subject Access: a Report on the Joint Program of the ALCTS Metadata Enrichment Task Force and the ALCTS Subject Analysis Committee's Subcommittee on Semantic Interoperability.

American Library Association Annual Conference, Orlando, June 2004."
Technical Services Quarterly, 22(3) 2005, p. 75-87.

Harken, Shelby E. *SAC subcommittee on semantic interoperability: Introduction/criteria [draft]*, 2005. <http://www.und.nodak.edu/dept/library/Departments/abc/SACSEM-Criteria.htm> (Sept. 25, 2005)

Heery, Rachel, Leona Carpenter, and Michael Day. "Renardus Project Developments and the Wider Digital Library Context." *D-Lib Magazine,* 7, no. 4 (April 2001). <http://www.dlib.org/dlib/april01/heery/04heery.html> (Aug. 8, 2002)

> A subject gateway provides a search service to high quality web resources selected from a particular subject area. This work was informed by earlier modeling work carried out in the context of Moving to Distributed Environments for Library Services (MODELS). It is hoped that results of the Renardus work will feed back to the ongoing development of the MODELS application framework, and also to the Imesh Toolkit project. The IMEsh Toolkit project is providing subject gateway developers with a systems framework for an extendable set of interoperable tools and components.
>
> Enhanced subject access is considered a key difference offered by subject gateways, and an important part of the Renardus service will be its attempt to provide some kind of subject directory browsing service across the participating gateways. In order to achieve this, a classification scheme has been chose to act as an 'interlingua' within the Renardus pilot. The scheme chosen is the Dewey Decimal Classification (DDC). Gateways participating in the Renardus system will be invited to map DDC terms to the subject terms used in their own browse hierarchies. In order to facilitate this process, the project established a small working group to prepare guidelines for this work. In addition, the software tool developed as part of the German CARMEN project has been adapted to facilitate the relevant workflow. The Renardus browse system will link directly into the subject hierarchies of individual gateways. If a part of an individual gateway's browse structure has been mapped to this DDC term, the gateway's name is visible and this becomes a hyperlink to the relevant part of the local browse structure. It relates to work currently taking place within the UK HILT project which is studying the problem of cross-searching and browsing by subject across a range of communities, services, and service or resource types. HILT will assist with consensus building on best practice in the short to medium term perspective as regards working with existing or new subjects schemes and thesauri. Renardus will feed back experience to Network Knowledge Organization Systems/Services (NKOS), a loose coalition of people and organizations concerned with the use of knowledge organization systems such as classification systems, thesauri, gazetteers, and ontologies, to support description and retrieval of resources via the Web.
>
> A draft Renardus application profile has been agreed upon to form the basic metadata schema. Definitions of the semantics of these elements are based, where possible, on the Dublin Core Metadata Element Set. There is the possibility of expanding the scope of the Renardus search service to the end-user. One proposal suggests that it would be possible to combine a brokered gateway service with Web indexes based on harvesting techniques. Within Renardus they intend to explore the possible benefits of collaborative cataloging for creating metadata about web resources. There may no longer be a need to duplicate metadata describing the same resource in so many locations, rather original metadata will be created and further enhancements to that metadata will be linked to an original authoritative metadata instance. One possible methodology to achieve this is to use XML/RDF annotations. Within Renardus they may explore linking local metadata enhancements to metadata residing in a central 'union catalog'.

Heery, Rachel and Harry Wagner. "A Metadata Registry for the Semantic Web." *D-Lib Magazine*, v. 8, no. 5 (May 2002). <http://www.dlib.org/dlib/may02/wagner/05wagner.html> (May 17, 2002).

The article primarily deals with schema registries. Registries essentially provide an index of terms. RDF provides the basis for declaring the schema in use. Work is underway to add richness and fullness to the schema language, to incorporate the features of the DARPA Agent Markup Language (DAML) and the Ontology Interface Layer (OIL) <http://www.ontoknowledge.org/oil> ontology language, and to bring this work to recommendation status. The Dublin Core Metadata Initiative (DCMI) has defined a relatively small set of data elements (referred to within the DCMI as the DCMI vocabulary or DCMI terms) for use in describing Internet resources as well as to provide a base-line element set for interoperability between richer vocabularies. The aim was to enable registration, discovery, and navigation of semantics as defined by DCMI. Two of several goals were: 1) automating identification of relationships between terms in vocabularies, 2) be multilingual. DCMI tried several prototypes including using the Extensible Open RDF Toolkit (EOR) for database management and Extensible Stylesheet Language Transformation (XSLT) for the user interface. A multi-lingual schema language must always be identified when registering a schema; it helps enable discovery and navigation; a multi-lingual user interface is accomplished using XSLT 'translate' stylesheet. Prototype 3 used BerkeleyDB which performed better than relational databases.

HILT. *High-level Thesaurus Project Proposal,* 2005. http://hilt.cdlr.strath.ac.uk/AboutHILT/proposal.html (Jan. 7, 2005) Not directly accessible; see <http://hilt.cdlr.strath.ac.uk/>

HILT Project Overview. <http://hilt.cdlr.strath.ac.uk/About-HILT/overview.html> (March 26, 2002).

The project is jointly funded by the RSLP and the JISC. The purpose of the first-year of the project was to study and report on the problem of cross-searching and browsing by subject across a range of communities, services, and service or resource types. Phase II aims to move the findings of Phase I into a "Pilot Project" stage. The project encompasses partners and stakeholders from a wide range of communities including archives, museum and libraries, amongst others.

Himanka, J. and V. Kautto. “Translation of the Finnish Abridged Edition of UDC into General Finnish Subject Headings.” *International Classification,* 19, no. 3 (1992): 131-134.

Hudon, Michele. "Multilingual Thesaurus Construction: Integrating the View of Different Cultures in One Gateway to Knowledge and Concepts.” In *Knowledge Organization,* v. 24, no. 2 (1997): 84-91.

The article focuses on the social/political aspects of treating multiple languages in egalitarian fashion, along with the technical implications.

Hunter, Jane. "MetaNet - a Metadata Term Thesaurus to Enable Semantic Interoperability between Metadata Domains." *Journal of Digital Information,* v. 1, no. 8 (Feb. 2001). <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/> (Feb. 17, 2005)

Abstract: Metadata interoperability is a fundamental requirement for access to information within networked knowledge organization systems. The Harmony international digital library project has developed a common underlying data model (the ABC model) to enable the scalable mapping of metadata descriptions across domains and media types. The ABC model provides a set of basic building blocks for metadata modeling and recognizes the importance of 'events' to describe unambiguously metadata for objects with a complex history. To test and evaluate the interoperability capabilities of this model, we applied it to some real multimedia examples and analyzed the results of mapping from the ABC model to various different metadata domains using XSLT. This work revealed serious limitations in the ability of XSLT to support flexible dynamic semantic mapping. To overcome this, we developed MetaNet, a metadata term thesaurus which provides the additional semantic knowledge that is non-existent within declarative XML-encoded metadata descriptions. This paper describes MetaNet, its RDF Schema representation and a hybrid mapping approach which combines the structural and syntactic mapping

capabilities of XSLT with the semantic knowledge of MetaNet, to enable flexible and dynamic mapping among metadata standards.

Huxley, Lesly, Leona Carpenter, Marianne Peereboom. *Collaborative Systems and Tools: Renardus Case Study*. (2002) Abstract, <http://www.internet-librarian.com/presentations/huxley.pdf> No longer available.

Renardus builds on existing trends towards greater collaboration, standardization, and interoperability between information services. The ability to cross-search and particularly to cross-browse participating gateways' records led to development of tools to support the integration and 'sensible' presentation of records from a wide range of services, each using unrelated classification systems and data models, providing interfaces and data in different languages, based on different technical solutions.

IFLA. Classification and Indexing Section, Division of Bibliographic Control. *Newsletter,* 27 (May 2003).

Sect. 2.2 Changing Roles of Subject Access Tools describes several projects: a) FAST, faceted Library of Congress Subject Headings; b) UDC implementation (UK) - role of classification in information retrieval systems to serve as an underlying knowledge structure to provide systematic subject organizations and thus complement the search using natural language terms; c) SWD/RSWK (SZ) after 5 years. Dewey Decimal Classification is being translated into German and is being used for the ePrint UK project. A subject indexing and classification project at the National Library of the Czech Republic involves Subject categorization of heterogeneous information using the Conspectus method based on intellectual mapping of DDC and UDC notations. The authority file contains four types of files: a) geographic, b) chronological, c) genre/form, d) topical

IFLA. Classification and Indexing Section. Working Group on Multilingual Thesauri. *Guidelines for Multilingual Thesauri.* http://www.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf (Apr. 20, 2005)

The IFLA Working Group on Guidelines for Multilingual Thesauri started to prepare this document in 2002. The objective of the document is to add to the existing Guidelines for Multilingual Thesauri as worded in the ISO-standard for multi-lingual thesauri (ISO-5964-1985) or in handbooks on thesaurus building. The general principles for the building of monolingual thesauri are assumed.

There are three approaches in the development of multilingual thesauri: 1) building a new thesaurus from the bottom up: a) starting with one language and adding another language or languages, or b) starting with more than one language simultaneously; 2) combining existing thesauri: a) merging two or more existing thesauri into one new (multilingual) information retrieval language to be used in indexing and retrieval, or b) linking existing thesauri and subject heading languages to each other, using the existing thesauri and/or subject heading languages both in indexing and retrieval; 3) translating a thesaurus into one or more other languages.

IFLA. Section on Classification and Indexing, Division of Bibliographic Control. *Newsletter,* 24 (Dec. 2001).

Czechia

More detailed subject access to documents to get a piece of information has become the vital need in the online environment where the best solution seems to be combination of keywords with a controlled vocabulary. Merging many external documents into the database of Union Catalog gives rise to discrepancies between index terms (lexical units), application syntax and hierarchical structure of original indexing systems.

Subject authority file: a) an integrated indexing and retrieval tool, in which the verbal terms of a thesaurus (controlled vocabulary) are combined with equivalent notations of a classification scheme (e.g. UDC); it enables subject access to documents either via verbal terms (searching) or through the classification notation (browsing; b)

application of this integrated tool in online (Web) environment may support automatic indexing and classification of web resources; in this case would be very useful to apply such verbal expressions and UDC notations that are reflecting real situations. Since subject access depends on national languages … it was difficult to find and apply any international recipe. After much debate LCSH system has been finally chosen. However it was considered useful at that time to meet local needs and requirements as well, so some modifications of LCSH were formulated such as: direct form of geographical subdivisions, form subdivisions were made separate headings, used generic headings for classes of persons or types of corporate bodies more often, etc.

France

RAMEAU is not the subject authority file of the Bibliothèque nationale de France, but the common French indexing language. We are classifying our RAMEAU subject headings in about sixty broad subject fields, named RAMEAU Domains, which are more or less arranged on the basis of DDC numbers. This work is partly done thanks to an automatic mapping between call numbers and subject indexing. It will allow to propose thematic views of RAMEAU and to provide consistent files of headings for our multilingual subject access project MACS.

Royal Library in Sweden is mapping is Swedish subject to LCSH.

Imesh Toolkit. <http://www.imesh.org/toolkit> (Aug. 6, 2002).

The Imesh Toolkit project evolved out of discussions within the Imesh community which was set up to encourage international collaboration amongst subject gateways. The project will build on existing subject gateway software to develop a configurable, reusable, and extensible toolkit for subject gateway providers.

The project plan: a) manual selection, description and classification; b) a structured record format; c) some search and retrieve protocol; d) mechanism for routing queries between gateways. For this reason, in the subject gateways review interviews are restricted to the needs of the Renardus definition of quality controlled subject gateways. It is a subject-based resource discovery guide which provides links to information resources (documents, collections, sites or services), predominantly accessible via the Internet, and applies a documented set of quality measures to support systematic resource discovery. It is also managed, collected by humans according to documented selection criteria, with maintenance criteria, with a fixed metadata set and controlled subject classification. It will eventually be a broker system for simultaneous access to quality-controlled subject gateways and other Internet-based, distributed services.

Current and possible future technologies and standards: a) Z39.50 is the protocol of choice for the majority of the services; b) Whois++ is a very simple search and retrieval protocol which provides a profile and a protocol at once; c) LDAP is light-weight directory access protocol. XML offers the possibility of combining QSBIG records with other non-QSBIG sources. XML is not sufficient on its own; analogously to Z39.50 requiring a profile for interoperability, XML requires a syntax upon which to be agreed. Some form of DC/RDF/XML protocol was strongly supported in the Renardus survey. SOAP is a remote procedure call proposal which uses XML and http as the carrying mechanism. Queries are couched in XML and results are received in XML.

IMesh Toolkit project evolved out of discussions aimed at encouraging international collaboration amongst subject gateways and subject-based resource discovery services. To include: Resource collection, cataloging, management and discovery (e.g. academic guides, virtual libraries and subject gateways); sharing technical, marketing, standards and cataloging effort, investigating cross-searching, cross-browsing and developing standards for related software and information issues.

IMesh Toolkit: an architecture and toolkit for distributed subject gateways. <http://www.imesh.org/toolkit> (Aug. 6, 2002).

The project will build on existing subject gateway software to develop a configurable, reusable and extensible toolkit for subject gateway providers.

IMesh Toolkit: subject gateway requirements. <http://www.imesh.org/toolkit/work/requirements> (Aug. 6, 2002).

The objective of this work package is to ensure that the IMesh toolkit architectural and technical strategies are well-grounded in documented needs and practical requirements of subject gateways.

IMesh toolkit: General architectural overview of the IMesh Toolkit. <http://www.imesh.org/toolkit/work/architecture/notes.php3> (Aug. 9, 2002).

Focuses on discussion of how to achieve interoperability for the IMesh toolkit, particularly in regards to architecture and functionality of query languages, etc.

IMesh toolkit: architecture. <http://www.imesh.org/toolkit.work/architecture> (Aug. 6, 2002).

Architectural diagram.

Information and Documentation - a Reference Ontology for the Interchange of Cultural Heritage Information. ISO/CD 21127. <http://www.niso.org/international/SC4/n491.pdf> (Oct. 27, 2002).

The primary purpose of ISO 21127 is to offer a conceptual basis for the mediation of information between cultural heritage organizations such as museums, libraries, and archives. The standard aims to provide a common reference point against which divergent and incompatible sources of information can be compared and, ultimately, harmonized. It is designed to be explanatory and extensible rather than prescriptive and restrictive. Consequently, the model has been formulated as an object-oriented semantic model, which can easily be converted into other object-oriented models. All cross-references and inheritance of properties are explicitly resolved. The exchange of information relevant to museum collections with libraries and archives falls within the scope of the standard.

ISO 2788-1986. *Documentation - Guidelines for the Establishment and Development of Monolingual Thesauri.* <http://www.nlc-bnc.ca/iso/tc46sc9/standard/2788e.htm> (July 1, 2002).

Iyer, H. and M. D. Giguere. "Towards Designing an Expert System to Map Mathematics Classificatory Structures." *Knowledge Organization,* 22, no. 3-4 (1995): 141-147.

Janée, Greg, Satoshi Ikeda, Linda L. Hill. *The ADL Thesaurus Protocol.* 2003. <http://alexandria.sdc.ucsb.edu/~gjanee/thesaurus/specification.html> (April 9, 2003).

The document describes an XML- and HTTP-based protocol for accessing thesauri: structured, controlled vocabularies of words and phrases that represent conceptual categories. The protocol is intended to allow programmatic clients to easily access and utilized existing thesauri, and thus the services offered by the protocol are oriented around querying thesauri and navigating within thesauri. The protocol does not support creation, maintenance, or sharing of thesauri, or mapping between thesauri. It does address the *term* that represents a conceptual category which may have a scope note. Terms may be preferred or non-preferred. It includes the reciprocal term relations of narrower, broader, related, use (use instead) and used-for. Eight XML formats are used. The hierarchy feature describes the hierarchy of terms above (broader) or below (narrower) including the starting term itself. Operators include "equals", "contains-all-words," "contains-any-word," "matches-regexp" (a perl-like regular expression). The protocol provides five independent, stateless services which are invoked over the HTTP protocol.

Koch, Traugott and Neuroth, Heike. *Classification Mapping for Cross-browsing in the European Subject Gateway Broker Renardus.* Presentation at the NKOS workshop at JDL, June 28, 2001. <http://www.lub.lu.se/tk/renardus/NKOS01-pres.htm> (Nov. 7, 2002).

Koch, Traugott. *Controlled Vocabularies, Thesauri and Classification Systems Available in the WWW. DC Subject,* 2001. <http://www.lub.lu.se/metadata/subject-help.html> (July 29, 2002).

Lists a large number available on the web.

---. Traugott. *Quality-controlled Subject Gateways on the Internet*, 2000. <http://www.lub.lu.se/tk/demos/Sgin.html> (Aug. 8, 2002).

> This paper summarizes DESIRE approach, software solutions, cooperative subject gateway projects, broker architectures, metadata mapping and cross-searching, browsing structure in a subject gateway, and classification mapping and cross-browsing problems and issues. "Quality-controlled subject gateways" are Internet services which apply a rich set of quality measures to support systematic resource discovery. Considerable manual effort is used to secure a selection of resources which meet quality criteria and to display a rich description of these resources with standards-based metadata. Regular checking and updating ensure good collection management. A main goal is to provide a high quality of subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing.

---. "The Renardus Broker: a 'Meta Subject Gateway.'" Presentation at ELAG 2001. <http://www.lub.lu.se/tk/renardus/tokyoren.html> (Nov. 7, 2002).

Koch, Traugott, Heike Neuroth, and Michael Day. "Renardus: Cross-browsing European Subject Gateways via a Common Classification System (DDC)." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*, ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 25-33. <http://www.ukoln.ac.uk/metadata/renardus/papers/ifla-satellite/ifla-satellite.html> (June 11, 2002).

> The paper presents the approach and first results of the classification mapping process in the EU project Renardus. The outcome is a cross-browsing feature based on the Dewey Decimal Classification (DDC) and improved subject searching across distributed and heterogeneous European subject gateways. The project aims to develop a Web-based service to enable searching and browsing across a range of distributed European-based information services designed for the academic and research communities - and in particular those services known as subject gateways. Predecessor projects like the EU project DESIRE have already developed solutions for the description of individual resources and for automatic classification at the level of an individual subject gateway using established classification systems. Renardus intends to develop a service that can cross-search and cross-browse a number of distributed subject gateways through the use of a common metadata profile and by mapping all locally-used classification schemes to a common scheme.

Kriewel, Sascha, and others. "DAFFODIL - Strategic Support for User-oriented Access to Heterogeneous Digital Libraries." *D-Lib Magazine*, 10, no. 6 (June 2004) <http://www.dlib.org/dlib/june04/kriewel/06kriewel.html> (June 2004).

> DAFFODIL (Distributed Agents for User-Friendly Access to Digital Libraries) is a search system for digital libraries aiming at strategic support during the information search process. It is a system for integrated search with the heterogeneous digital libraries of a scientific community with merging of results. It combines browsing and searching strategies in a natural way. It uses a classification tool which provides users with access to a hierarchical, topic oriented representation of the search domain. It allows the browsing of classification schemes like the ACM Computing Classification system. The thesaurus tool can be used to get more general or more specific terms (hypernyms or hyponyms), or semantic definitions for a search term. Subject specific and web-based thesauri are used for finding related terms. The resulting terms can then be used in other tools for further queries.

Kuhr, Patricia S. "Putting the World Back Together: Mapping Multiple Vocabularies into a Single Thesaurus." In Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC, ed. I.C. McIlwaine. München: K.G. Saur, 2003. p. 37-42.

This paper describes an ongoing project by the H.W. Wilson Company in which the subject headings contained in twelve controlled vocabularies covering multiple disciplines from the humanities to the sciences and including law and education among others are being collapsed into a single vocabulary and reference structure. Wilson decided on a megathesaurus format and automatic switching.

Kunz, Martin. "Subject Retrieval in Distributed Resources: a Short Review of Recent Developments." Paper presented at the 68th IFLA Council and General Conference, Aug. 18-24, 2002. <http://www.ifla.org/IV/ifla68/papers/007-122e.pdf> (Oct. 27, 2002).

Subject searching across distributed resources is a current challenge when carrying out online searches for bibliographic data. The construction of portals for comparable sources is only the first step; the subsequent navigation of disparate search interfaces still presents problems. Both broad and specialist vocabularies exist. If retrieval is to be improved, there must be some adaptation of these differing resources. There are techniques for relating various subject terminologies, but they have their problems and limitations. Whether you call it a cross-concordance or a crosswalk, it is about creating links between equivalent terms describing similar concepts in two (or more) thesauri AND it is about affiliation of documentary languages. New developments in MACS, CARMEN, and Economics cross-concordance are discussed.

One part of the CARMEN Project concerns itself with the association of the thesaurus of the Informationszentrum Sozialwissenschaften (IZT) with the SWD. Starting from alphabetical lists which contain the keyword material from a specific subject area, the relationships between the two thesauri are determined intellectually and recorded in a link management system.

The aim of MACS is to study the links between the three extensive subject heading authority files - LCSH, RAMEAU, and SWD. The immediate objective is to indicate in each authority file the equivalent preferred descriptors of the other authority files for a few chosen subject areas. The process being developed for MACS will not affect the structure of the individual national authority files. It uses intellectually-determined equivalencies to link the content of the bibliographic databases which use a controlled vocabulary to describe their content and present in an ordered, structured way. MACS is based on the assumption that the users accesses the results of intellectually assigned subject descriptions via a thesaurus. Thesauri can distinguish to a better and more comprehensive degree between material to be indexed than a method based on syntactical indexing.

Kwasnik, Barbara H. and Victoria L. Rubin. "Stretching Conceptual Structures in Classifications across Languages and Cultures." *Cataloging & Classification Quarterly,* v. 37, no. 1 / 2 (2003): 33-47.

Summary: The authors describe the difficulties of translating classifications from a source language and culture to another language and culture. To demonstrate these problems, kinship terms and concepts from native speakers of fourteen languages were collected and analyzed to find differences between their terms and structure and those used in English. At issue are vocabulary, syntax, and semantics. In harmonizing classification schemes across languages and culture, one must address the way these terms are bound up in knowledge representations.

Lancaster, F. Wilfrid. *Vocabulary control for information retrieval*. 2nd ed. Arlington, W.Va.: Information Resources Press, 1986.

Landry, Patrice. "The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD)." Paper presented at the 66th IFLA Council and General Conference, Jerusalem, Aug. 13-18, 2000.

<http://www.ifla.org/IV/ifla66/papers/165-181e.pdf> (Aug. 7, 2002). <MACS : Multilingual access to subjects. https://ilmacs.uvt.nl/pub/>

> This is a report on the progress of the project during the previous year. Based on the final report of the CoBRA+ working group on multilingual subject access, the importance of co-operation in the quest for multilingual subject access was stressed. The goal is to allow the user to conduct a subject search in catalogs in their preferred language. The link management software should have a file management and a maintenance structure that allows data to be easily added and amended. The prototype should provide for any user the possibility to choose a source language and or more target catalogs. The Link Management Interface should only be accessed by the partner libraries to add and to manage the links between the different subject heading lists. The Search Results screen shows which links have been made to a particular subject heading in the focus subject heading list. The View Link function is primarily an editorial function. From this screen, a term (authority) or the link can be modified. The Search Interface was designed to give the library users the possibility of using their preferred subject heading list and doing their search in the catalogs of one or many libraries. The Browse button will show all the headings where a particular heading term is used and the links to these headings. The library user can access the full bibliographic record by clicking on the title. The interface will retrieve the bibliographic record in the selected library and will display the record in the bibliographic format used by that library.

Lauser, Boris, and others. "A Comprehensive Framework for Building Multilingual Domain Ontologies: Creating a Prototype Biosecurity Ontology." Paper presented at Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002: 113-123 <http://www.bncf.net/dc2002/program/ft/paper13.pdf> or <http://www.bncf.net/dc2002/program/papers.html> (2002)

> This paper presents ongoing work in establishing a multilingual domain ontology for a biosecurity portal. The project is embedded into the bigger context of the Food and Agriculture Organization's Agricultural Ontology Service (AOS) project of the FAO. The paper focuses on introducing a comprehensive, reusable framework for the process of semi-automatically supported ontology evolvement. An extendable layered ontology modeling approach will address multilinguality issues. In the context of the AOS, an ontology is a system of terms, the definition of these terms and the specification of relationships between the terms. It extends the approach of classical thesauri by providing the opportunity of creating an infinite number of different semantic relationships. Semantic robustness towards representational changes, as well as multilingualism, is crucial for the development of the domain ontology. Therefore, they distinguish between terms, and the concepts these terms represent. These are called Lexical Entries with two attributes, the concept it refers to and its language. RDFS (http://www.w3.org/TR/rdf-schema/#intro) is used to define vocabularies of resources and relationships amongst them. Using several tools a list of terms is developed. This is combined with terms in AGROVOC, a multilingual agricultural thesaurus, in which all terms have also been converted to concepts. Hence automated and manual processes have been used to create a single ontology which is reviewed by specialists.

Lee, Jonghoon, David S. Dubin, Michael J. Kurtz. "Co-occurrence Evidence for Subject Vocabulary Reconciliation in ADS Databases." In *ASP Conference Series, vol. 172, Astronomical Data Analysis Software and Systems, VIII, 1999.* <http://monet.astro.uiuc.edu /adass98/Proceedings/leej/> (Sept. 27, 2002)

> The article reports on a project to reconcile heterogeneous indexing vocabularies in the NASA Astrophysics Data System (ADS) which mixes controlled vocabularies and keywords. The mixture of different descriptor vocabularies in ADS defeats the standardization goal, and the merging of the abstract and key word indexes limits the search precision function of the subject indexing. Descriptors representing identical concepts can stand in several different relationships to each other. A project at the

University of Illinois investigates sources of evidence to support the automatic and/or computer-assisted reconciliation of the heterogeneous indexing in ADS. Two sources of evidence have been investigated: 1) lexical resemblance between descriptors, and 2) consistent assignment of descriptors from different vocabularies. The consistent assignment of two or more terms from different vocabularies to the same documents suggests some kind of semantic relationship among the terms. The activation of terms at input is spread through the network to the connected documents and from there to the output terms. By assessing multiple occurrence, using weighting, and setting a cut-off resulted in fairly good matches.

Lee, Maria, Stewart Baillie, Jon Dell'Oro. "TML: a Thesaural Markup Language." Paper presented at Proceedings of the 4th Australasian Document Computing Symposium, Coffs Harbour, Australia, Dec. 3, 1999. <http://www.ted.cmis.csiro.au/omt/tml.pdf> (April 7, 2003). No longer available.

Thesauri are used to provide controlled vocabularies for resource classification. Their use can greatly assist document discovery because thesauri mandate a consistent shared terminology for describing documents. A particular thesaurus classifies documents according to an information community's needs. As a result, there are many different thesaural schemas. This has led to a proliferation of schema-specific thesaural systems. In the authors' research, they exploit schematic regularities to design a generic thesaural ontology and specify it as a markup language. The language provides a common representational framework in which to encode the idiosyncrasies of specific thesauri. This approach has several advantages: it offers consistent syntax and semantics in which to express thesauri; it allows general purpose thesaural applications to leverage many thesauri; and it supports a single thesaural user interface by which information communities can consistently organize, store and retrieve electronic documents.

An ontology, in computer science, has come to denote an explicitly specified conceptualization of part of the world. In software, an ontology is implemented as a data structure. What distinguishes the ontology from the data structure is semantics: that it talks about something in the world. An ontology provides users with a representation which is essential to effective communication and coordination.

The general thesaural ontology gives us a conceptual representation of thesauri. A thesaural markup language (TML) manifests this as a grammar in which to express the content and structure of specific thesauri. TML is specified as an XML schema which defines the permitted markup element types and embedding structure. The TML syntax consists of the element names and structure.

TML provides a way to represent task-domain specific thesauri and make them available to a document management system. In order to demonstrate this generality, the authors developed a Thesaural Explorer application. The Explorer reads a thesaurus from its TML file, presents it graphically, and supports browser style term navigation. The user selects a thesaurus to explore and then can navigate the structure along inter-term relations by clicking on terms or using various look up tables such as ordered lists by class, term alphabetic, and browsing history.

Library of Congress Portals Applications Issues Group. *List of Portal Application Functionalities for the Library of Congress*, 2003. <http://www.loc.gov/catdir/lcpaig/> ; <http://www.loc.gov/catdir/lcpaig/portalfunctionalitieslist4publiccomment1st7-22-03revcomp.pdf> (Feb. 7, 2006).

The list represents the results of market analysis to study portal functionality of particular products. Functionalities include: a) general requirements, b) client requirements, c) searching and search results, d) knowledge database, e) patron authentication, and f) portal administration and vendor support. One aspect of a portal is its database and the subject metadata used within it and its maintenance. LCPAIG focused its explorations and testing on portals as tools for organized knowledge

discovery rather than as enterprise interfaces. Portals may be characterized by their ability to: a) assist users in identifying and selecting appropriate target resources, b) help users determining the target resources most useful to their research by providing effective search interfaces and an architecture that supports groupings and rich descriptions of resources, c) provide federated searching and information retrieval of descriptive metadata from multiple, diverse target resources, including but not limited to commercial or licensed electronic resources, databases, Web pages, and library catalogs, d) integrate and manage search results, e) save and export search results, f) link search results to full-text or other content delivery options, g) manage access to target resources and portal functionalities for authenticated users.

Several relevant requirements that should be present: a) the vendor must maintain descriptive metadata and configuration information for core target databases, including target title or name, subject terms, etc.; b) the ability to locally define and configure composite search qualifier groupings, e.g. name/author; c) the ability for user to search descriptive metadata in multiple metadata forms; d) the ability for user to search by specific fields in advanced searches; e) the ability to integrate metadata for target resources from more than one source; f) it should support keyword and browse searches, including: 1) the ability to browse a list of targets, 2) the ability to search target descriptions by keyword, 3) the ability to present different views of targets (e.g. by subject, user group, etc.), 4) the ability to browse target resources in hierarchical displays, 5) the ability to browse a composite list of target resources (aggregated databases), 6) the ability to present different views of the target resources.

Lin, Dekang and Patrick Pantel. "Induction of Semantic Classes from Natural Language Text." In *KDD-2001, Proceedings of the seventh ACM SIGKIDD International Conference on Knowledge Discovery and Data Mining*, Aug. 26-29, 2001, San Francisco, Calif., p. 317-322. <http://www.acm.org/sigkdd/kdd2001> (Oct. 26, 2002).

Lovins, Daniel. *Thesaurus Design for Semantic Information Management*. A day-long seminar led by Prof. Bella Hass-Weinberg in New York, April 16, 2002, email (May 6, 2002). Published: *Cataloging and Classification Quarterly*, 34, no. 4 (2003) <http://catalogingandclassificationquarterly.com/ccq36nr1news.html>

Bella suggested that "semantic information management" really just means vocabulary control; that ontology usually just means classification scheme, but sometimes gets used as a synonym for thesaurus, and the taxonomy is just a synonym for classification. Subject headings lists such as LCSH are essential tools for managing information in a print environment, while true thesauri are often more useful in the online environment (where they can be viewed hierarchically or combined in Boolean searches). Thesauri often run into the problem of needing to distinguish homographs. The problem in the selection of thesaurus terms is largely one of determining a set of appropriate lexemes, that is, the smallest units of a lexicon that can be understood on their own terms. Synonymy is a common problem, though easily managed, e.g. Cancer, see Neoplasm. Other problems: having to choose between singular and plural, parts of speech, etc.

MACS: Multilingual Access to Subjects, 2002. <https://ilmacs.uvt.nl/pub/> (Mar. 26, 2002).

MACS aims to provide multilingual subject access to library catalogs. It enables users to simultaneously search the catalogs of the project's partner libraries in the language of their choice (English, French, German). Partners are: Swiss National Library (SNL), Bibliothèque nationale de France (BnF), British Library, Die Deutsche Bibliothek (DDB), and it is running under the auspices of Conference of European National Librarians (CENL). This multilingual search is made possible thanks to the equivalence links created between the three indexing languages used in these libraries: SWD, RAMEAU, LCSH. Topics (headings) from the three lists are analyzed to determine whether they are exact or partial matches, of a simple or complex nature. The end result is neither a translation nor a new thesaurus but a mapping of existing and widely used indexing languages.

MACS (Multilingual Access to Subjects) Project, report for 2000-2001. <https://ilmacs.uvt.nl/pub/node/7?PHPSESSID=ff54ff63320ff2635357304df902dca9> (Mar. 1, 2006)

MACS is a cooperative Conference of European National Libraries (CENL) project to develop a prototype system for providing multilingual subject access searching between the catalogs of the partner libraries to: 1) research the technical and organizational issues involved in managing a working system for creating and maintaining links between the three subject headings lists (SHL), and 2) demonstrate the effectiveness of the linked SHLs for retrieving results for the end-user. The CoBRA study group defined a specific approach to mapping headings based on a number of core principles including: 1) all SHLs are equal, 2) headings are only mapped to equivalent headings judged to be synonymous in meaning, 3) hierarchical structures and thesaural relationships are not mapped or reproduced as part of the process of linking individual headings, 4) only headings at the authority level are linked, 5) where an equivalence cannot be found a proposed heading should stand alone in the system to represent the concept (for future possible mapping). Items are cataloged in the local library's language and SHL. Hierarchical navigation is only possible within each SHL, so it is envisaged that searches are refined by the user in his own language until the required concept is identified and then expanded for linguistics equivalences and documents in other libraries. Two interfaces proposed: 1) A Link Management Interface to support management of the links, their creations, and maintenance; 2) User Search Interface to support end user searching and links to the partners' catalogs. Partners share equal responsibility for authorization of links and validation of links proposed to their own SHL. MACS is to be an external link database, with each SHL remaining independent and linked to other SHLs only through MACS.

Mai, Jens-Erik. "The Future of General Classification." *Cataloging & Classification Quarterly,* v. 37, no. 1 / 2 (2003): 3-31.

Summary: Discusses problems related to accessing multiple collections using a single retrieval language. The article surveys the concepts of interoperability and switching language. It finds that mapping between more indexing languages will always be an approximation.

The paper treats the issues related to subject representation and focuses on the use of general classification schemes for accessing documents across domains and collections. The goal of interoperability is to build coherent services for users, from components that are technically different and managed by different organizations. This requires agreements on three levels: technical, content, and organizational. The problem with using switching languages is in mapping meaning of words in context of the language. Mapping will always be an approximation due to pre-coordination, hierarchical structure, and the absence of concepts to match.

Maniez, Jacques. "Database Merging and the Compatibility of Indexing Languages." In *Knowledge Organization,* 24, no.4 (1997): 213-224.

This article contains succinct and critical descriptions of concordance tables, switching languages, and reference languages, and their usability in the harmonization of information languages.

McKiernan, Gerry. *Beyond Bookmarks: Schemes for Organizing the Web,* 2001. <http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm> (Aug. 6, 2002).

Schemes for Organizing the Web is a clearinghouse of World Wide Web sites that have applied or adopted standard classification schemes or controlled vocabularies to organize or provide enhanced access to Internet resources. Topics cover Classifications systems: Alphabetic, Numeric, Alphanumeric; and Controlled vocabularies.

Medical Subject Authority in OCLC: Background and Resources. Informal discussion during ALA Midwinter 2002, January 18, 2002. <http://corc.oclc.org/WebZ/XpathfinderQuery?sessionid=0:term=3049:xid=LTM> (March 26, 2002). No longer available.

An OCLC pathfinder listing resources dealing with inclusion of medical subject heading authority records in OCLC services.

MetaSearch Initiative. <http://www.niso.org/committees/MS_initiative.html> (May 10, 2003).

Metasearch, parallel search, federated search, broadcast search, cross-database search, search portal are terms which have become commonplace in the information community's vocabulary. They speak to a common theme of allowing search and retrieval to span multiple databases, sources, platforms, protocols, and vendors at once.

One-search access to multiple resources holds the promise of enabling libraries to offer portal environments so their users can enjoy the same easy searching found in web-based services like Google.

Michel, Dee and Pat Kuhr. *Taxonomy of Subject Relationships.* Appendix B, 1996. <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/msrscu2.pdf> (March 26, 2002).

Shows associative, equivalence, and hierarchical relationships.

Miles, Alistair and Dan Brickley. *SKOS Core Guide.* < http://www.w3.org/TR/swbp-skos-core-guide/> (Aug. 25, 2005).

SKOS stands for Simple Knowledge Organization System. The name SKOS was chosen to emphasize the goal of providing a simple yet powerful framework for expressing knowledge organization systems in a machine-understandable way.

A 'concept scheme' is defined here as: a set of concepts, optionally including statements about semantic relationships between those concepts. Thesauri, classification schemes, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabulary are all examples of concept schemes.

SKOS Core provides a model for expressing the basic structure and content of concept schemes (thesauri, classification schemes, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabulary).

The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), which can be used to express a concept scheme as an RDF graph. Using RDF allows data to be linked to and/or merged with other RDF data by semantic web applications.

This document is a guide using the SKOS Core Vocabulary, for readers who already have a basic understanding of RDF concepts.

See also Quick Guide to Publishing a Thesaurus on the Semantic Web http://www.w3.org/TR/swbp-thesaurus-pubguide/. See also the SKOS Core Vocabulary Specification http://www.w3.org/TR/swbp-skos-core-spec

Miller, Libby, Dan Brickley and Martin Hamilton. *Imesh Tk: Subject Gateway Review Literature Review,* 2002. <http://www.ilrt.bris.ac.uk/discovery/2000/09/imesh/> (Aug. 6, 2002)

The goal of the literature review is: a) to try to define the scope of the IMesh Toolkit, b) its purpose - improve speed of searching, c) enable cross-searching more easily between gateways, or enable portalization of gateways, d) draw together existing research, e) summarize current and possible future technologies, f) form preliminary conclusions about possible architectures which could be used in IMesh Toolkit.

Miller, Ken and Brian Matthews. "Having the Right Connections: the LIMBER Project." *Journal of Digital Information*, 1, no. 8 (Aug. 2001). <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Miller/> (Aug. 2, 2002).

Cross-discipline interoperability will be provided via a uniform metadata description. In addition, the provision of multilingual user interfaces and the controlled vocabulary of a multi-lingual thesaurus will make these datasets globally accessible in a range of end-user natural languages. LIMBER will use the multi-lingual European Language Social Science Thesaurus (ELSST) derived and translated from HASSET. Tools developed in LIMBER will work with any thesaurus marked up in the LIMBER RDF format, and the semi-automatic indexing tool will apply keywords from these thesauri to any metadata record marked up in either XML or RDF. LIMBER will still be able to provide multi-lingual interfaces to thesaurus-aided searching across domains, using thesauri conforming to the LIMBER RDF schema and retrieving metadata mapped to the Dublin Core with assigned keywords translated back to the user's native language, the underlying metadata having been semi-automatically indexed by terms from the conforming thesauri. The project plans to develop a high-level object-oriented conceptual model that could be translated in whichever format becomes internationally accepted. All screens and drop-down menus will be available in German, French, Spanish and English to begin with, but defined in a standard format that can easily be translated to other languages in the future. LIMBER is designed as three stand-alone products: 1) multi-lingual thesaurus management tool, 2) user browsing interface, 3) semi-automatic indexing tool.

Miller, Joseph. "An Overview of Subject Cataloging and the Absence of a Code." Presented at ARLIS/NA Annual Conference, Pittsburgh, March 2000. <http://artcataloging.net/arlisna/miller.html> (March 26, 2002).

Subject cataloging deals with what a book or other library item is about, and the purpose of subject cataloging is to list under one uniform word or phrase all the materials on a given topic that a library has in its collection. A subject heading is that uniform word or phrase used in the library catalog to express a topic. The use of authorized words or phrases only, with cross-reference from unauthorized synonyms, is the essence of bibliographic control in subject cataloging.

Miller, Paul. "I Say What I Mean, but Do I Mean What I Say?" *Ariadne*, 23 (2000). <http://www.ariadne.ac.uk/issue23/metadata/> (Aug. 7, 2002).

Addresses: 1) issues surrounding the use of controlled vocabulary, 2) recent MODELS 11 workshop, 3) some recommendations for future work. First, there is a need for some mechanism for querying multiple resources simultaneously. Second, there is a need for some commonality of content or description across information resources being made available for searching. To ensure common meanings across applications and between users and between applications, the normal solution is to impose a degree of control upon the terms used by both parties. At its most basic, this control will involve no more than defining a list of words, from which application and user have to select. In more complex instances, fully formed thesauri may be employed, rich with hierarchy, synonyms, and relationships. In an uncontrolled environment, users will consistently either use the wrong terms or use right terms in wrong contexts. In the same uncontrolled environment, creators will potentially use terms inconsistently. Terminology tools are: controlled vocabularies (created manually or generated automatically by harvest keywords), alphanumeric classification schema, and thesauri. Thesauri follow the structural guidelines in ISO 2788 or ISO 5964, including synonyms, complex hierarchies, scope notes, and inter-relationships (equivalence, hierarchy, association). MODELS 11's aim was to explore the value practicality of creating a single high-level thesaurus. There is a need to study user behavior with respect to terminology.

Milstead, Jessica. *Report on the Workshop on Electronic Thesauri, November 4-5, 1999*. Presented at NISO/APA/ASI/ALCTS. <http://www.niwo.org/news/events_workshops/thes99rprt.html> (March 26, 2002). No longer available.

The definition of "thesaurus" for purposes of this meeting was broader than that of the present standard for thesauri ANSI/NISO Z39.19-1993 (R1998). The meeting

considered vocabularies that meet two basic criteria: 1) use to facilitate analysis of texts and their subsequent retrieval (or retrieval of the information which they contain); 2) and inclusion of a rich set of semantic relationships among their constituent terms. The scope included: standard thesauri, subject headings lists, semantic networks, and taxonomies (Internet directories). It excluded: simple term lists without equivalence relationships and dictionaries.

They identified 4 key issues:

1) the need for (and feasibility of developing) a standard that speaks to criteria and/or methods for generating thesauri by machine-aided or automatic means

2) the need for (and feasibility of developing) a standard set of tools which show semantic relationships among terms, as aids to text and information analysis and retrieval

3) the need for (and feasibility of developing) a standard structure that supports a variety of electronic thesaurus displays

4) the need for (and feasibility of developing) a standard that supports interoperability protocols, structures, and/or semantics applicable to thesauri.

Mongin, Larry, Yueyu Fu and Javed Mostafa. "Open Archives Data Service Prototype and Automated Subject Indexing Using D-lib Archive Content as a Testbed." *D-Lib Magazine*, 9, no. 12 (Dec. 2003). <http://www.dlib.org/dlib/december03/mongin/12mongin.html> (Dec. 18, 2003).

The Indiana University School of Library and Information Science's laboratory has as its purpose to work in areas of information retrieval and information visualization. They decided to use OAI-PMH as a resource discovery tool. Since the D-Lib metadata file does not contain a subject term, they decided to use IR algorithms to generate them. After running the Java program that computed subject terms, they read each article to make a judgment on whether the computed subject terms were relevant to that article. The criterion was not whether the program selected the best subject terms for that text, but rather whether the term generally reflected the semantic meaning of the article. The resulting scores varied from 70-95%.

Murata, Masaki, and others. "Meaning Sort - Three Examples: Dictionary Construction, Tagged Corpus Construction, and Information Presentation System." *ArXiv,* 12 March 2001 <http://arxiv.org/abs/cs/0103012> (Feb. 17, 2005)

It is often useful to sort words into an order that reflects relations among their meanings as obtained by using a thesaurus. In this paper, the authors introduce a method of arranging words semantically by using several types of "is-a" thesauri and a multi-dimensional thesaurus.

Murray-Rust, Peter and Lesley West. *Terminology in a Global Context: VHG and XML. Part II,* 2002. <http://www.vhg.org/uk.pub/vhgnews2.html> (March 26, 2002). No longer available.

The aim of this article is to set out the technical aspects of the Virtual HyperGlossary (VHG). XML is ideally suited to delivering terminology over the web. Thus, in the spirit of XML, a simple subset of ISO FDIS 12620 data categories is chosen to represent the communality of the semantics of a majority of web-based glossaries. VHG is a platform- and convention-independent specification. We put a high value on interoperability and achieve this by reliance on several current W3C initiatives in XML. Semantics are added through a mechanism which would link any tags starting with <VHG: to the semantics in the Unique Resource Locator (URL). This distinguishes the VHG approach, so that when someone encounters a VHG glossary it is self-identifying and can be processed with VHG-compliant software. In a related manner, a document can link to a number of glossaries simultaneously. It might use absolute URLS or it might use a namespace mechanism. An element in a document linked to any number of glossaries may provide complementary or even conflicting

views. In the spirit of the WWW, the reader of the document resolves the appropriate ontology.

National Information Standards Organization. *Developing the Next Generation of Standards for Controlled Vocabularies and Thesauri,* 2005. <http://www.niso.org/committees/MT-info.html> (Feb. 15, 2005)

---. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies: ANSI/NISO Z39.19-2005,* 2005. <http://www.niso.org/standards/standard_detail.cfm?std_id=814> (Feb. 20, 2006).

A thesaurus is a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associate relationships among terms are displayed clearly and identified by standardized relationship indicators that are employed reciprocally. The primary purposes of a thesaurus are a) to facilitate retrieval of documents, and b) to achieve consistency in the indexing of written or otherwise recorded documents and other items, mainly for post-coordinate information storage and retrieval systems. This standard provides guidelines for constructing monolingual thesauri: formulating the descriptors, establishing relationships among terms, and effectively presenting the information in print and on a screen. It presents guidelines and conventions for the contents, display, construction, testing, maintenance, and management of monolingual controlled vocabularies. It focuses on controlled vocabularies that are used for the representation of content objects in knowledge organization systems including lists, synonym rings, taxonomies, and thesauri.

National Library of Medicine. *Fact sheet: UMLS Metathesaurus,* 2005. <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html> (Jan.7, 2005)

Neuroth, Heike and Traugptt Koch. *Cross-browsing and Cross-searching in a Distributed Network of Subject Gateways: Architecture, Data Model, and Classification*, 2001. <http://www.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.html> (Aug. 8, 2002).

The aim of the Renardus project is to provide users with integrated access by searching or browsing, through a single interface, to partners' quality-controlled subject gateways. Further goals are to develop and define organizational models, business models, technical solutions and metadata standards (Renardus Application Profile, Renardus Namespaces, Renardus Collection Level Description). The following elements can be used to define a quality-controlled subject gateway: Selection and collection development, Collection management, Creation, Resource description and metadata, Subject access, Search and browse access, Standards, Value-adding features. Each participating partner is responsible for mapping his metadata format to the common Renardus metadata format, derived from Dublin Core. A generic normalization toolkit with Z39.50 configuration files and a conversion script were provided. Each participant set up a Renardus server with their content normalized to the Renardus data model. A set of screens were built for the user interface: Homepage, Advanced Search screen, Index scan window, Advanced search page after index scan, Browse by subject screen, (Preliminary) Result screen, Sorted result screen, Participating gateways screen and Help (index) screen. In order to accomplish subject browsing, the various systems will be mapped to a common classification system. The Renardus service will give access to resources from all kinds of subjects, published world-wide and in many languages and it is intended to be offered to an international multi-disciplinary community of users. The Dewey Decimal Classification was chosen because of online availability and tools, global usage, suitability of the classification system and its functionality, frequency and character of the updates, research and methodological development efforts.

Neuroth, Heike. *Metadata issues: Renardus*. Presented at Cultural Heritage Projects Concertation Event, Bundesamtsgebaude Wien, June 30, 2000. <http://www.cscaustria.at/events/documents/renardus.ppt> (Aug. 6, 2002).

The data model is mostly Dublin Core compatible with some Renardus specific extension. The definition of a Renardus Schema is in progress. Still need to address how to handle mutli-linguality.

Neuroth, Heike and Traugott Koch. Metadata Mapping and Application Profiles: Approaches to Providing the Cross-searching of Heterogeneous Resources in the EU project Renardus, 2001. <http://www.lub.lu.se/~traugott/drafts/DC2001-neuroth.pdf> (Nov. 7, 2002).

The paper presents the approach and results of a mapping process to define a common metadata format for cross-searching distributed and heterogeneous subject gateways in the heterogeneous subject gateways in the EU project Renardus. The outcome is a well defined data model with semantic and syntactical definitions of each metadata element. It results in richer and semantically controlled cross-searching. The metadata elements are mainly based on Dublin Core. The aim of Renardus is to provide the user with integrated access, through a single interface, to high-quality Internet resources. It is also to provide high quality subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing. All gateways participating in Renardus apply resource descriptions and subject classification to all their records. Participants have agreed to use a core set of metadata elements and qualifiers: Title, Creator, Description, Subject, Identifier, Language, and Type; plus Country. Further, they focused on the following characteristics for each metadata element: semantic definition, syntactic definition, associated qualifiers, cataloging rules, namespace definition, repeatability of elements, form of obligation, language qualifiers. For Subject, Renardus has four different namespaces plus they will develop a cross-browsing structure based on the Dewey Decimal Classification with added European specific captions.

Nicholson, Dennis. "HILT High Level Thesaurus Project: Interoperability and Cross-searching Distributed Services." Presented at the Thesaurus Conference organized by Waterways Trust, hosted at the Science Museum, London. 3 April 2001. < http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/HILTD%20Nicholson.ppt> (Oct. 26, 2002)

---. "Subject-based Interoperability: Issues from the High Level Thesaurus (HILT) Project." Paper presented at 68th IFLA Council and General Conference, Glasgow, Scotland, 18-24, 2002. <http://www.ifla.org/IV/ifla68/prog02.htm> (2002)

HILT Phase 2 will create a pilot terminologies mapping service or route map with a specific focus on current concerns in the developing Distributed National Electronic Resource (DNER), covering primarily higher education information resources. HILT Phase I discovered that the various service providers use a range of subject schemes (LCSH, UNESCO, DDC, AAT, MeSH). If cross-searching and browsing is to function coherently for users of the Information Environment (IE), these (multiple, varied) subject schemes must be mapped to one another, perhaps using a common 'spine' such as DDC with international and multi-lingual application and the potential to facilitate machine to machine interworking. The terminologies must be disambiguated, then translated into the service-assigned terms the users need to cross-search browse the group of services of relevance to their query. The aim of HILT Phase II is to build and evaluate a pilot service that will mediate as a DNER shared service in the IE. The pilot TeRM would be built using commercially available Wordmap software (http://www.wordmap.com); examples at: http://www.oingo.com or http://vivisimo.com) The initial illustrative TeRM would be based on the RDN (http://www.rdn.ac.uk/cgi-bin/browse) terminologies available as part of the Wordmap taxonomies set, which include, in particular, a set of terms used by general Internet users, and on selective subsets of LCSH, DDC, UNESCO, and AAT. At issue is the question of whether a spine such as DDC should be used to map everything else to, and also is it better to adopt (adapt) an existing scheme or create a new one. The aim is to utilize 'native subject schemes' for the collections in the environment users use them, and to use the pilot TeRM to “disambiguate" user terms and resolve

differences between schemes. TeRM supports creation, editing, display, and User [user interface], staff, and system interaction with terminologies map showing terms in use and inter-relationships. It interacts with users and systems to establish term and service context of search (e.g. archives only), provides synonyms, broader, narrower, related terms, other contexts and service-set navigational aids for cross searching browsing as required. See: http://hilt.cdlr.strath.ac.uk/Reports/FinalReport.html

Nicholson, Dennis, and others. *HILT: High-Level Thesaurus Project: Final Report to RSLP & JISC,* December 2001. <http://hilt.cdlr.strath.ac.uk/Reports/FinalReport.html> (Oct. 29, 2002).

There is evidence of growing agreement that interoperability in respect of subject schemes in a distributed environment is recognized as an issue and that a standards-based approach is the answer, but no evidence to suggest that one particular scheme or single approach will provide the answer. There is very little information available on the needs and behavior of users as regards subject searching in a distributed environment. It is suggested that a mix of controlled vocabularies and free text in searching gives the best results and is preferred by users. HILT's recommended option is to map LCSH, AAT, UNESCO Thesaurus, UDC, to DDC. Set up a mapping service, ideally with international participation and support, and gradually build towards a complete mapping of LCSH, UNESCO, UDC, and ATT to a DDC backbone. The conclusion was that the best way forward for HILT was a pilot mapping services as described in option 5.2 of the report. The pilot should have a strong user focus, determine reliable costs, include costs and benefits, involve international players, look at how best to integrate semantic web and artificial intelligence developments, involve a broad range of target services, use existing machine-readable mappings wherever possible, be closely linked to a cross-sectoral and cross-domain task force, use contexts, relationships, clustering, etc. look at user terminology as against DDC as the central spine to which other schemes were to be mapped. The use of DDC by itself is not a solution, but being mapped to more specific subject schemes was worth being a pilot project.

Nicholson, Dennis, Susannah Wake, S. Currier. "HILT: High Level Thesaurus Project: Investigating the Problems of Cross-Searching Distributed Services by Subject in the UK." Presented at the meeting, "New Information Technology 2001," Tsinghua University, Beijing, China. 29-31 May. In *Global digital library development in the new millennium: Fertile ground for distributed cross-disciplinary collaboration*. ed. C. C. Chen. Beijing: Tsinghua University Press, 2001. < http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/hiltchina2.ppt> (March 3, 2006)

Presentation describes background of HILT and the HILT Stakeholder Survey.

Nicholson, Dennis and Susannah Wake. "HILT: Subject Retrieval in a Distributed Environment." In Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 61-67. <http://hilt.cdlr.strath.ac.uk/Dissemination/Talks/hilt-ifla.ppt> (March 3, 2006)

Normore, Lorraine and Mark Bendig. "Using a Classification-based Information Space." Presented at "Subject Retrieval in a Networked Environment," an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, Aug. 14-16, 2001. <http://staff.oclc.org/~normorel/ppt/ifla_preconf_2001.htm> (Oct. 26, 2002).

The goals of the project were to: 1) use information visualization to help searchers understand and explore information spaces, and 2) use the metadata in library records to accomplish this end, specifically to explore the use of a classification system. One approach is a cluster-based space which uses clustering to coalesce documents/topics, multidimensional scaling techniques to create space and spatial metaphors to show relationships. Users infer the semantics of the space from the characteristics of the clusters.

OCLC. *Metadata Switch*. http://www.oclc.org/research/projects/mswitch/default.htm

The Metadata Switch was an umbrella activity for a set of projects which designed to construct experimental modular services that add value to metadata. A partial listing of the project services included: harvesting metadata, 'fusion' of metadata from different sources, schema transformation, enrichment or augmentation of records with various types of data, terminology and name authority services.

Of particular relevance are the terminology and name authority services. As a subproject of the Metadata Switch Project, this project explored techniques for mapping knowledge organization resources, encoding vocabularies using existing and emerging standards, and searching and distributing vocabulary resources via web-based protocols. The project was completed in December 2003.

This has been succeeded by Terminology Services. Web services are modular, web-based, machine-to-machine applications that can be combined in various ways. Web services can be accessed at various points in the metadata lifecycle, for example, when a work is authored or created, at the time an object is indexed or cataloged, or during search and retrieval. Terminology services are web services involving various types of knowledge organization resources, including authority files, subject heading systems, thesauri, web taxonomies, and classification schemes. A web service that provides mappings from a term in one vocabulary to one or more terms in another vocabulary is an example of a terminology service. This project extends the capabilities for enhancing and mapping vocabularies developed under the terminology services sub-project of the Metadata Switch.

Olson, Tony. “Integrating LCSH and MeSH in Information Systems.” In Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC, ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 21-24.

---. "The Integration of Information Languages and Interoperability." Presented at "Real World Steps to Interoperability in Libraries,” ALCTS/LITA Authority Control in the Online Environment Interest Group, ALA Annual Conference, June 16, 2002. <http://www.ala.org/ala/lita/litamembership/litaigs/authorityalcts/2002authcontrol.pdf> (Nov. 8, 2002).

There are two types of indexing languages: 1) information languages and 2) natural languages. Information languages include classification systems (e.g. DDC), controlled vocabularies (e.g. thesauri like AAT), and subject headings lists (e.g. LCSH). Issues regarding controlled vocabularies are discussed. By their very nature different controlled vocabularies are incompatible. While controlled vocabularies promote consistency within the systems for which they are designed, they tend to reduce intersystem and database compatibility. Major problems are: 1) conflicts between cross references in one vocabulary and established headings in the other vocabularies; 2) no references or links between corresponding headings from different vocabularies; 3) differences in syntax in the construction of subject heading strings; 4) correspondences between some terms in different vocabularies may be one-to-one but there is a significant number of correspondences that are not; 5) differences in semantic relationships between vocabularies, which in turn also lead to one-to-many correspondences; 6) identical headings in different vocabularies can cause the retrieval of duplicate entries.

Some of the methods used in an effort to integrate various information languages include: 1) mapping to a larger megathesaurus, 2) linking between equivalent subject headings, 3) using a reference language, 4) mapping multiple subject thesauri or lists to a backbone reference language, 5) mapping classification schemes to a backbone, 6) integrating controlled vocabularies.

Examples of these methods are discussed. Two examples of megathesauri are the Unified Medical Language System (UMLS) and the H.W. Wilson megathesaurus. The UMLS integrates over 60 biomedical vocabularies and classifications and links

many different names for the same concepts. H.W. Wilson maps twelve different Wilson vocabularies. The Multilingual Access to Subjects (MACS) Project is an example of integrating multiple subject languages by providing links between equivalent subject headings. Another method of integration is to use a reference language. In this case terms from various information languages are mapped to a term (or classification number) in a single particular information language (called a reference language). The High Level Thesaurus Project (HILT) project was to study the problems of incompatibility among various information languages utilized by various libraries and information centers. One of the recommendations was to set up a mapping service that would eventually carry-out a mapping of LCSH, the UNESCO thesaurus, AAT, and UDC to a DDC backbone, as the reference language. In the Renardus Project local classification schemes that are used in subject gateways are mapped to DDC. The LCSH/MeSH mapping project at Northwestern University is another approach to the integration of controlled vocabularies.

In the LCSH/MESH mapping project, instead of creating a separate database that contains the linking data, the data is entered into the authority records of the vocabularies being mapped. The LCSH/Mesh project developed a combination of computer-assisted techniques and human editorial review. The 750 and 788 linking entry fields are used to record “equivalent" headings. A difficult problem is mapping one-to-multiple correspondence between headings in different controlled vocabularies. Another aspect is differing semantic relationships in different vocabularies. It was decided to map at similar levels and use each vocabulary's structure to trace relationships. Two issues had to be addressed. First, broader/narrow term relationships are not explicit in MeSH but are implicit in category (tree) numbers. A program was written to generate 550 fields in MeSH authority records from category data in 072 fields. Second, the syndetic structure of LCSH is not complete (especially as distributed) containing only narrow term references and not explicit broader term references. Another problem is the syntactical differences between subject heading strings in the two vocabularies.

Olson, Tony and Gary Strawn. "Mapping the LCSH and MeSH Systems." *Information Technology and Libraries*, 16, no. 1 (March 1997): 5-19.

In an effort to resolve problems of two subject systems in one online catalog, this project maps the LCSH and MESH vocabularies. The two systems are integrated by a) mapping terms and headings from one system to corresponding headings in the other system; b) adding the mapping data to authority records, c) enhancing the library management system software so that mapping data in authority records can be used to develop syndetic structures that relate the systems smoothly and consistently, while enhancing subject retrieval.

Open Metadata Registry. <http://avalon.ulis.ac.jp/~sugimoto/RPs/dc2001.pdf> (Feb. 17, 2005)

The Open Metadata Registry has much in common with SCHEMAS [SCHEMAS has provided a forum for metadata schema designers involved in projects under the IST Programme and national initiatives in Europe]. It will be used to promote the discovery and reuse of semantics within existing vocabularies and the creation of new vocabularies. It will register vocabularies relating to the Dublin Core Metadata Initiative.

Park, J., and S. Ram. “Information Systems Interoperability: What Lies Beneath?*” ACM Transactions on Information Systems,* 22, no. 4 (2004): 595-632.

Parsons, J., and Y. Wand. “Choosing Classes in Conceptual Modeling.” *Communications of the ACM,* 40 (1997): 63-69.

Patton, Glenn. "International Efforts to Improve Interoperability.” Pesented at "Real World Steps to Interoperability in Libraries,” ALCTS/LITA Authority Control in the Online Environment Interest Group, ALA Annual

Conference, June 16, 2002. < www.ala.org/ala/lita/litamembership/litaigs/authorityalcts /pattoninteroperability.ppt>(Nov. 8, 2002).

RDF Topicmaps : Theory. OCLC Research. <http://topicmap.oclc.org:5000/theory.html> (Oct. 31, 2002).

The goal of the Topicmaps is to bootstrap the efforts to meld natural-language-processing technologies with Semantic Web development. It is comprised of: 1) the noun phrase extractor, 2) noun phrase filter, and 3) relationship generator, the goal of which was to identify simple, thesaurus-like relations such as "broader-than" using only a list of words as input.

Reisthuis, Gerhard J. A. "Information Languages and Multilingual Subject Access." In Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 11-17.

In this paper the possibilities for a multilingual thesaurus, in which not all descriptors in a given language have equivalent descriptors in all other languages and in which the hierarchical structure can have variations in the different languages, are explored. A small model of such a thesaurus is given. It is argued that the searching possibilities, which more recent programs for bibliographic databases offer, make such non-identical thesauri possible.

Renardus. <http://www.renardus.org> (March 26, 2002).

Renardus is a collaborative project that aims to improve academic users' access to a range of existing Internet-based information services across Europe. The aim is to provide users with integrated access, through a single interface, to access selected, quality resources and other Internet-based distributed services. Renardus exploits the success of subject gateways, where subject experts select quality resources for their users, usually within the academic and research communities. Renardus is based on a distributed model where major subject gateway services across Europe can be searched and browsed together through a single interface provided by the Renardus broker. A special feature of Renardus is the option to "Browse by Subject" through hierarchical trees of topics and subsequently to jump to one or several related sub-collections of the contributing Subject gateways. The Renardus service allows the user to search several Subject Gateways simultaneously. This means that you are searching the "catalogue records" (metadata), not the actual resources, of quality controlled Web resources. The user can also browse through a hierarchy of subject categories in order to explore parts of the participating Subject Gateways which contain Internet resources relevant to the user's area of interest.

Renardus Project Deliverables (2000?)

This project deliverable intends to ensure that any chosen broker architecture for Renardus is based on existing models and/or emerging developments. It provides an extensive and comprehensive review of 18 existing brokers models that have been developed for a variety of existing services, projects, or initiatives.

Renardus Project deliverable: specification of functional requirements for the broker system. <http://www.renardus.org/about_us/deliverables/d1_3/titlePage.html> (Aug. 7, 2002).

Evaluation report of existing broker models. <http://www.renardus.org/about_us/deliverables/d_1/D1_1summ.html> (Aug. 7, 2002).

Specification of functional requirements for the broker system. <http://www.renardus.org/about_us/deliverables/d1_3/D1_3bsumm.html> (Aug. 7, 2002).

Data model: requirements and specification. <http://www.renardus.org.about_us/deliverables/d6_4/D6_4summ.html> (Aug. 7, 2002).

Resnik, Philip. "Disambiguating Noun Groupings with Respect to WordNet Senses." *ArXiv*, (Nov. 29, 1995). <http://xxx.lanl.gov/abs/cmp-lg/9511006> (Feb. 17, 2005).

In word groupings within online thesauri, one is interested in the relationships among word senses, not just words. The paper presents a method for automatic sense disambiguation of nouns appearing within sets of related nouns - the kind of data one finds in online thesauri or as the output of distributional clustering algorithms.

Report of the SAC Subcommittee on Subject Reference Structures in Automated Systems: Recommendations for Providing Access to, Display of, Navigation within and among, and Modifications of Existing Practice Regarding Subject Reference Structures in Automated Systems. 2003. <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectreference/subjectreference.htm> (Feb. 17, 2005).

The subcommittee concentrated on maximizing the use of existing subject reference structures in automated systems. The recommendations are divided in four sections: access to reference structures, display of reference structures, navigation among and within reference structures, and changes to the policies and practices that govern creation of the authority records that underlie these reference structures in automated systems.

Resource Discovery Network. "Renardus Project" ; "Subject Portals Development Project". (2002) <http://rdn.ac.uk/projects/> (Completed projects) (Sept. 25, 2005)

Resource Organisation and Discovery in Subject-based Services: ROADS, 2000. <http://www.ukoln.ac.uk/metadata/roads/> (Aug. 6, 2002)

The overall object of the ROADS projects was to design and implement a user-oriented resource discovery system. It investigated the creation, collection, and distribution of resource descriptions, to provide a transparent means of searching for and using resources. The object was not to create an individual and idiosyncratic system but to draw on, and help create, standards of good practice which can be widely adopted by subject communities to aid and automate the process of resource organization and discovery. See http://www.ukoln.ac.uk/roads/

ROADS: Interoperability and Metadata, 1998. <http://www.ukoln.ac.uk/metadata/roads/interoperability/inter-meta.html> (Aug. 7, 2002).

ROADS began work in a context where interoperability is becoming increasingly important as a means to integrate the wide range of information services. Users require distributed information services to inter-work in terms of search, location and delivery. ROADS supports semantic interoperability. Users will be searching a variety of indexes constructed from a number of different underlying database structures. Effective searching across services requires that semantically equivalent fields in these indexes are mapped to each other. In addition semantics in the search (client) must be managed so that they match the semantics in the indexes (targets). Z39.50 allows indexed to be mapped to standard sets of attributes, hiding the underlying structure of the target database. A common indexing protocol enables routing of queries to the most appropriate database via a mesh of centroids or index summaries. Resource Description Framework (RDF) aims to provide a framework for expression machine-readable metadata about resources. It is designed to enable different applications to interoperate by using a common data model. RDF uses Extensible Markup Language (XML) as the encoding syntax.

Russell, Rosemary and Michael Day. *Automated and Manual Approaches to the Provision of Thesauri and Subject Vocabularies,* 2001. <http://www.ukoln.ac.uk/metadata/hilt/interfaces/> Accessed June 11, 2002. Final report <http://hilt.cdlr.strath.ac.uk/Reports/FinalReport.html> (Feb. 25, 2003).

The term thesaurus is used in different contexts to describe tools that fulfill different functions. From an information science point of view, thesauri were originally developed as tools to allow terminology control of detailed subject indexing of printed documents. What distinguishes thesauri from some other subject vocabulary types is that they show relationships between concepts. Relationships commonly

expressed in thesauri include hierarchy, equivalence (synonymy), and association or relatedness. These relationships are generally represented by the notation BT (broader term), NT (narrower term), SY (synonymy), and RT (associated or related term)

In addition to thesauri, there is a range of other types of controlled subject terminologies (or vocabularies). One can either browse alphabetical lists or the hierarchy of subject terms that may be hyperlinked, or one can search terms and if a non-preferred term is used, the user will be taken to the preferred term.

SCHEMAS Project <http://www.schemas-forum.org/>

Currently, it is in development by the United Kingdom Office for Library and Information Networking (UKOLN). Its goal is the development of a comprehensive database of RDF schemas, application profiles, and related semantics that have been used by programs under the IST Program and other related European initiatives. The SCHEMAS database will be used to promote the reuse and interoperability of semantics for existing and new projects. It will register RDF schemas and namespaces used by projects within the European Union.

SCHEMAS Registry, 2002. <http://www.schemas-forum.org/registry/> (Aug. 6, 2002).

One important focus of the SCHEMAS Project (to provide standards for metadata schema designers) is provision of a registry of metadata schemas. The registry itself will serve as a good-practice example of registry use and benefits. Workpackage6 aims to promote the deployment of metadata registries defined with the Resource Description Framework (RDF), promote standards and methods for creating and processing schemas in multiple languages and writing systems, encourage re-use and adaptation of global metadata elements in local schemas, formulate and disseminate good-practice guidelines and investigate the process for managing the evolution of multilingual registries.

Sheikholeslami, Gholamhosein, Wendy Chang, and Aidong Zhang. "SemQuery: Semantic Clustering and Querying on Heterogeneous Features for Visual Data." IEEE transactions on knowledge and data engineering, v. 14, no. 5 (2002): 988-1002.

The effectiveness of content-based image retrieval can be enhanced using heterogeneous features embedded in the images. However, since the features in text, color, and shape are generated using different computation methods, and thus may require different similarity measurements, and integration of the retrievals on heterogeneous features is a nontrivial task. In this paper the authors present a semantics-based clustering and indexing approach, termed SemQuery, to support visual queries on heterogeneous features of images.

Slater, Jenny. *References - Taxonomies and thesauri. CETIS, Metadata Special Interest Group,* 2002. <http://cetis-metadata.lboro.ac.uk/vocab-ref.htm> (July 29, 2002). No longer available.

Lists a large number available on the web.

Stoklasova, Bohdana, Marie Balikova and Ludmila Celbova. "The Relationship between Subject Gateways and National Bibliographies in International Context." Paper delivered at World Library and Information Congress, 69th IFLA General Conference and Council, Berlin, 2003. <http://www.ifla.org/IV/ifla69/papers/054e-Stoklasova_Balikova_Celbova.pdf> (Sept. 19, 2003).

The paper examines the relationship between subject gateways and national bibliographies together with general principles of universal bibliographic control in the broader context of the need for integration of heterogeneous information sources. The paper gives examples from the Czech Republic's experience and illustrates problems with integrating heterogeneous resources from different countries covering different subjects. The Czech National Subject Gateway Project is connected with the Uniform Information Gateway Project (http://www.jib.cz) which integrates heterogeneous information resources including full texts and digital objects from

different countries. The paper concludes with recommendations for improvement of bibliographic control. The subject authority system is multilingual and uses Aleph software: http://sigma.nkp.cz/F/?func=file%file_name=find-b&local_base=auv&con

Subcommittee on Subject Relationships/Reference Structures. *Report to the ALCTS/CCS Subject Analysis Committee. Appendix A, 1996*. <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/appendix.htm> (March 26, 2002).

The charge was to investigate: 1) the kinds of relationships that exist between subjects, the display of which are likely to be useful to catalog users; 2) how these relationships are or could be recorded in authorities and classification formats; 3) options for how these relationships should be presented to users of online and print catalogs, indexes, etc.

One conclusion was there is a need for BT and NT and related browsing or exploding. Because Library of Congress only distributes only the broader code, OPACs can display only broader-to-narrower references. However, Gary Strawn has demonstrated that systems can be programmed to generate narrower-to-broader references without anyone having to add "narrower" 5XX fields to the authority records. Non-specific "see also" relationships can be generated by coding the byte used for reference relationships coding "n." Indexing databases often use an alphabetical browsing list which then displays broader, narrower, and related terms for a chosen subject. In addition an "explode" function employs these term relationships along with several others (synonym, abbreviation and language equivalent) to automatically retrieve all records bearing on the chosen term or related terms.

Subject Gateways, 1999. <http://www.desire.org/html/subjectgateways/subjectgateways.html> (Aug. 7, 2002).

What is a subject gateway? "Subject gateways are online services and sites that provide searchable and browsable catalogs of internet based resources. Subject gateways will typically focus on a related set of academic subject areas." Many of the activities and research project within DESIRE are focused on developing the ideas behind this definition of a subject gateway, as well as developing methodology and tools that provide the functionality needed for a subject gateway to function.

Sugimoto, Shiegeo, and others. "Developing Community-oriented Metadata Vocabularies: Some Case Studies." Paper presented at International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society (DLKC'04), 2004. <http://www.kc.tsukuba.ac.jp/dlkc/>; <http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp128.pdf> (Sept. 13, 2004).

This paper presents two case studies which include the development of domain-specific subject vocabularies - a core subject vocabulary for a subject gateway for library and library-and-information science (LIS) resources, and subject vocabularies of a portal service for a regional community. These case studies show that small subject vocabularies are useful for these community-oriented services, and that maintenance is a crucial issue for the development and use of the vocabularies. In order to build a community-oriented information environment in the Internet, the authors have to solve two contradictory requirements for metadata schemas - specialization (or localization) in a community and interoperability among communities.

Metadata, which has been widely recognized as a key component for the Web and digital libraries in local or domain-specific communities would need to define metadata schemas and controlled vocabularies in accordance with their requirements in the case that their requirements are difficult to be satisfied only by those defined for the global communities. On the other hand, community-oriented specialization of schemas and vocabularies would raise a bar for interoperability issues for cross-community use of metadata and information resources. In addition, long-term maintenance of the schemas and vocabularies is a crucial aspect for the communities.

Svenonius, E. *The Intellectual Foundation of Information Organization.* Cambridge, Mass.: MIT Press, 2001.

Intellectual access to information organization is discussed in the first part of the book. The second part addresses three bibliographic languages: work languages, document languages, and subject languages. It looks at these languages in terms of their vocabulary, semantics, and syntax.

Taylor, Mike. *Zthes: a Z39.50 Profile for Thesaurus Navigation*. Ver. 4.0, 2000. <http://www.lcweb.loc.gov/z3950/agency/profiles/zthes-04.html> (March 26, 2002).

This document describes an abstract model for representing and searching thesauri - semantic hierarchies of terms as described in ISO 2788 - and specifies how this model may be implemented using the Z39.50 protocol. It also suggests how the model may be implemented using other protocols and formats.

This profile is laid out in two main sections. The first is concerned solely with the abstract representation of thesaurus terms and how they may be searched; and the second with the implementation of these abstract concepts in Z39.50: how thesaurus terms are encoded in the GRS-1 record structure, how searches are encoded in the type-1 query, etc. It is intended that the abstract model described here is sufficiently general that it can also be implemented by protocols and data formats other than Z39.50. This profile does not mandate any relationship between a thesaurus and any other database. The model is that terms from any thesaurus database may be used to search any other database (called a *target database*). This profile represents a thesaurus as a database of inter-linked terms. If multiple thesauri are to be supported by a single server, then they must be presented as separate databases.

Tennant, R. "Metadata's Bitter Harvest." *Library Journal,* 129, no. 12 (2004): 32.

Tennis, Joseph T. "Layers of Meaning: Disentangling Subject Access Interoperability." *Advances in Classification Research,* 12 (2004)

Therond, Daniel. "Www.European-Heritage.Net: The European Heritage Network." *Cultivate Interactive*, issue 2, no. 16 (Oct. 2000). <http://www.cultivate-int.org/issue2/herein/> (Aug. 7, 2002).

The European Information network on cultural heritage policies (HEREIN Project) recommended setting up a permanent information system for authorities, professionals, researchers and training specialists. The aim of the project was to convert the Council of Europe's paper databank on architectural and archaeological heritage into a system a) with fast, easy access via the Internet, and b) which correspondents in member countries would be able to update easily by email.

Tillett, Barbara. "A Virtual International Authority File." Presentation to the Giornata di studio sul controllo di autorità nel Servizio Bibliotecario Nazionale Nov. 22, 2002. <http://www.iccu.sbn.it/upload/documenti/Tillett.ppt> (April 1, 2003).

Objectives: a) facilitate sharing to reduce cataloguing costs to libraries, museums, archives, rights management agencies, etc. b) simplify creation and maintenance of authority records internationally, c) enable users to access information in the language, script, form they prefer.

Authority control virtues: a) "Precision" in searching, b) syndetic structure of references to help navigate (the variant forms of name/title/subject/etc.), c) displays to collocate works, d) links to forms used in particular resources, e) bring library catalogs into the mix of tools available on the Web.

There are a number of projects to facilitate or that incorporate aspects of authority control on a international scale: EU: AUTHOR Project, LEAF, <indecs>, INTERPARTY, HKCAN, IFLA: MLAR, GARR, FRANAR, Dublin Core "Agents," DELOS/NSF Working Group "Actors/Roles," EAC (Encoded Archival Context), CORC/Connexion, Unicode/Multiple Scripts, NACO/SACO for AACR2 and LCSH. There is increased need for interoperability exemplified by efforts to map different

communication formats with Z39.50 protocols, create crosswalks to the "MARCs," XML, ONIX. The Virtual International Authority File (VIAF) supports IFLA UBC authority principles. Each country is responsible for authority headings for its own personal and corporate authors. National authority records are available for everyone to use. The same form and structure would be used worldwide.

VIAF proposes using programs to facilitate authority work, that would do an automatic check of headings against existing local authority files, and if not found, would automatically check against a "virtual" international authority file. It would display found matches for editing or reference and insert authorized forms into local authority record for future linking. The author would like to test the use of unique, persistent record control numbers such as the International Standard Authority Number or the International Standard Authority Data Number and see if that works or possibly use the number assigned to an information package for an entity under OAI (Open Archive Initiative) protocols. There are many models that can be envisioned for a virtual international authority file to help with cataloging. Some of which are: a) a distributed system with the independent National Bibliographic Agencies (NBA's) being searchable using the next generation of Z39.50 protocols; b) a linked model that would use a search protocol, such as Z39.50 going to any one of the linked authority files (LEAF is testing this model); c) a centralized model that uses Open Archive Initiative protocols to harvest the metadata from authority files of the National Bibliographic Agencies on one or more servers; or d) providing a centralized link, where one authority file is viewed as the central point to which all others are linked.

Tudhope, Douglas, Harith Alani and Christopher Jones. "Augmenting Thesaurus Relationships: Possibilities for Retrieval." *Journal of Digital Information*, 1, no. 8 (Feb. 5, 2001). <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/> (June 27, 2002).

The paper discusses the augmentation of thesaurus relationships. First the authors discussed a case study that explored the retrieval potential of an augmented set of thesaurus relationships by specializing standard relationships into richer subtypes, in particular hierarchical geographical containment and the associative relationship. Various attempts to build taxonomies of thesaurus relationships are discussed. They concluded by discussing the feasibility of hierarchically augmenting the core set of thesaurus relationships, particularly the associate relationship. They discussed the possibility of enriching the specification and semantics of Related Term (RT relationships), while maintaining compatibility with traditional thesauri via a limited hierarchical extension of the associative relationships. They first illustrated how hierarchical spatial relationships can be used to provide more flexible retrieval for queries incorporating place names in applications employing online gazetteers and geographical thesauri. The work described was part of a larger project, Ontologically Augmented Spatial Information System (OASIS). Another aim was to explore the potential of reasoning over the semantic relationships in thesauri to assist retrieval. The three main types: a) equivalence (equivalent terms), b) hierarchical (broader/narrower terms: BT/NT's), c) Associative (related terms: RT's)

UK Interoperability Focus, 2000. <http://www.ukoln.ac.uk/interop-focus/about/> (Aug. 7, 2002).

UK Interoperability Focus is hosted by UKOLN. It is a team of people responsible for exploring, publicizing and mobilizing the benefits and practice of effective interoperability across diverse information sectors. Interoperability is a broad term, encompassing many of the issues impinging upon the effectiveness with which diverse information resources might fruitfully coexist. The issues are many and may be may be defined as:

1) Technical Interoperability: consideration of technical issues includes ensuring an involvement in the continued development of communication, transport, storage and representation standards such as Z39.50, ISO-ILL, XML, etc.

2) Semantic Interoperability: … individual resources - each internally constructed in their own semantically consistent fashion - are made available through gateways or catalogs. Almost inevitably these discrete resources use different terms to describe similar concepts, or even identical terms to mean very different things. The development and distributed use of thesauri such as those from Getty is worthy of further consideration.

3) Political / Human Interoperability: there are implications for the organizations concerned who may see it as a loss of control or ownership. Staff may need extensive training or retraining to ensure effective long-term use of any service

4) Inter-community Interoperability: researchers require information from a wide variety of institutions, disciplines, and information resources

5) International Interoperability: existing issues magnified with varied languages, differences in technical approach, working practices, etc.

Unified Medical Language System (UMLS). <http://www.nlm.nih.gov/research/umls/> (March 26, 2002).

NLM's Unified Medical Language System (UMLS) project develops and distributes multi-purpose, electronic "knowledge sources" and associated lexical programs. The Metathesaurus provides a uniform, integrated distribution format for more than 100 biomedical and health-related vocabularies, classifications, and coding systems (some in multiple languages) and links many different names for the same concepts. System developers can use the UMLS products to enhance their applications. There are three UMLS Knowledge Sources: the Metathesaurus®, the Semantic Network, and the SPECIALIST lexicon. They are distributed with flexible lexical tools and the MetamorphoSys install and customization program.

Van de Sompel, Herbert Van, Jeffrey A. Young and Thomas B. Hickey. "Using the OAI-PMH … Differently." in *D-Lib magazine,* 9, no. 7/8 (July 3, 2003). <http://www.dlib.org/dlib/july03/young/07young.html> (July 23, 2003).

The Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) was created to facilitate discovery of distributed resources. The OAI-PMH achieves this by providing a simple, yet powerful framework for metadata harvesting. The OAI-PMH has been widely accepted, and until recently, it has mainly been applied to make Dublin Core metadata about scholarly objects contained in distributed repositories searchable through a single user interface. Initially, the descriptive metadata provided by OAI-PMH repositories was to a large extent limited to the mandatory unqualified Dublin Core, but an evolution towards the provision of more extensive descriptive metadata, such as MARC21, is becoming apparent. Metadata records in the OAI-PMH are any data that can be validated against a W3C XML Schema. Therefore, the OAI-PMH can be a medium for incremental, data-sensitive exchange of any form of semi-structured data. The metadata contained in OAI-PMH repositories is typically gathered by harvesters that process it and make it searchable through a user interface. In these uses of the OAI-PMH, repositories are never directly accessed by end-users; the "customers" of the repositories are robots. A section of the article describes an approach to overlay OAI-PMH repositories with an interface allowing users to directly navigate the repository content. The authors also show how this approach has been used to make the GSAFD Thesaurus, the OpenURL Registry and the XTCat Thesis Catalog user-accessible.

Veen, Theo van and Robina Clayphan. "Metadata in the Context of the European Library Project." Presented at Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002: 19-26 <http://www.bncf.net/dc2002/program/papers.html>

The European Library sponsored by the European Commission, brings together 10 major European national libraries and library organizations to investigate the technical and policy issues involved in sharing digital resources.

Vizine-Goetz, Diane. "Dewey in CORC: Classification in Metadata and Pathfinders." *Journal of Internet Cataloging*, 4, no. 1 / 2 (2001): 67-80.

The Cooperative Online Resource Catalog (CORC) project provided an opportunity for OCLC research and Dewey editors to explore the potential of the Dewey Decimal Classification (DDC) system for organizing electronic resources. The mapped vocabulary was used in the following ways: 1) to improve access to Dewey by expanding the indexing vocabulary; 2) to assist in the assignment of subject elements during metadata creation; 3) to provide supplemental terminology for automated classification; 4) to provide alternative access mechanisms for views to resources in the CORC database.

Vizine-Goetz, D., C. Hickey, A. H. Houghton and R. Thompson. "Vocabulary Mapping for Terminology Services." *Journal of Digital Information,* 4, no. 4 (2004)

Vizine-Goetz, Diane. "Terminology Services." Presentation. 2003. <http://www.oclc.org/research/projects/mswitch/> ; <http://www.oclc.org/research/projects/mswitch/4_termservs.shtm> (Feb. 18, 2005),

Presentation discussed research at OCLC to add value to metadata. Metadata Switch is a project involving a set of projects: harvesting metadata, merging metadata from different sources, schema transformation, terminology and name authority services, enrichment or augmentation of records with various types of data. DDC, Thesaurus of ERIC descriptors, GSAFD genre terms, MeSH, LSCH, and LCSHAC were converted to a common content model and linked using intellectual and automated mapping techniques.

Wagner, Harry R. "The EOR Toolkit: an Open Source Solution for RDF Metadata." *Information Technology and Libraries*, 21, no. 1 (March 2002): 27-31.

RDF provides solutions that will enable a significantly higher degree of reliability, relevance, and accuracy for applications and services focused on resource discovery and management of Web sites and other Internet resources. Through its use of machine-understandable semantics, RDF enables the automated discovery, management, and exchange of metadata. It significantly improves resource discovery by enabling a finer degree of granularity and improved precision. In addition to facilitating the creation of new resource descriptions, RDF builds on the established work of various resource communities by enabling the interoperability of existing metadata vocabularies within those communities.

EOR is one of a large and growing number of open resource applications that are being used to develop applications and services focused on the discovery, management, integration, and navigation of electronic resources. http://eor.dublincore.org No longer available.

Wake, Susannah and Dennis Nicholson. "HILT - High-Level Thesaurus Project: Building Consensus for Interoperable Subject Access across Communities." *D-Lib Magazine*, 7, no. 9 (Sept. 2001). <http://www.dlib.org/dlib/september01/wake/09wake.html> (Oct. 26, 2002).

The article provides an overview of the work carried out by the HILT Project http://hilt.cdlr.strath.ac.uk in making recommendations towards interoperable subject access, or cross-searching and browsing distributed services amongst the archives, libraries, museums and electronic services sectors. The article discusses the consensus achieved at the June 19, 2001 HILT Workshop. The best way forward for HILT was the pilot mapping service combined to an extent with a terminologies task force. The service envisaged would map key schemes like LCSH, UNESCO, DDC, Universal Decimal Classification, Art and Architecture Thesaurus, and possibly user and regional terminologies, and local adaptations of standard schemes. Users would be able to: a) input the term or terms that describe their problem using the terminology that is most meaningful to them; b) specify their query more closely if necessary by

specifying a context; and c) obtain a list of equivalent or near-equivalent terms with which they could then cross-search or cross browse the various services.

Whitehead, C. “Mapping LCSH into thesauri: The AAT model.” In T. Peterson & P. Moholt, eds. *Beyond the book: Extending MARC for subject access.* Boston: G.H. Hall, 1990, p. 81.

Willpower Information. *Publications on Thesaurus Construction and Use*. <http://www.willpower.demon.co.uk/thesbibl.htm> (July 1, 2002).

This is a list of printed and electronic publications about the principles of constructing and using information retrieval thesauri. It is not a list of existing thesauri, although some thesauri have been included when they are good examples or illustrate the results of different approaches to thesaurus construction. References to lists of thesauri and systems that provide for thesaurus use by combining terms from multiple facets in search interfaces are given at the end.

WordNet: a Lexical Database of the English Language, 2001. <http://www.cogsci.princeton.edu/~wn/> (Aug. 6, 2002).

WordNet is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets. It was developed by the Cognitive Science Laboratory at Princeton University under the direction of Prof. George A. Miller.

Xiaoming Liu, [OAI-Implementers] "Dublin Core XML and OAI." March 29, 2002, personal email to listserv. <http://arc.cs.edu/edu>

Xiaoming Lui's work on ARC, building on Open Archives Initiative work includes a subject file from various schemas.

Young, Iain. "Da Chanan / Two Languages: Creating Bi-lingual Name Authorities." Paper presented at the 68th IFLA Council and General Conference, Glasgow, Scotland, Aug. 18-24, 2002. <http://www.ifla.org/IV/ifla68/prog02.htm>; <http://www.ifla.org/IV/ifla68/papers/025-144e.pdf> (Jan. 18, 2003).

The issue is how to create standard name authorities in a bilingual environment. Using as a specific example the project undertaken by the Scottish Poetry Library to create name authorities for Gaelic poets, some with Gaelic and English forms of their names, issues raised are examined.

Zeng, Marcia Lei and Yu Chen. "Features of an Integrated Thesaurus Management and Search System for the Networked Environment." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*. ed. I.C. McIlwaine. München: K.G. Saur, 2003, p. 122-128.

A report on an integrated system, CAMed, that employs an open structure for managing distributed resources (thesauri and databases) and integrates a thesaurus management system with a cross-thesaurus search system. This paper describes the functions of the system that highlight the unique design for the networked environment.

Zeng, Marcia Lei and Lois Mai Chan. “Trends and Issues in Establishing Interoperability among Knowledge Organization Systems.” *Journal of the American Society for Information Science and Technology*, 55, no. 5 (2004), 377-395.