The European Library

TEL-ME-MOR

Modular Extensions For Mediating Online Resources

Report on Cross-Language Subject Access Options

Deliverable 3.4

May 19, 2006

Report Authors: **Patrice Landry – Swiss National Library**
Maja Žumer - National and University Library, Slovenia
Genevieve Clavel-Merrin – Swiss National Library

Information Society Technologies

The TEL-ME-MOR project is funded under
the European Commission, IST Programme.

Table of contents

Summary 3
Context 3
Subject access in national collections 5
Subject access issues in a networked environment 7
Overview of the NMS subject access solutions 9
Methodology 10
Approaches in establishing interoperability 10
Overview of current practices projects and research on interoperability between subject access tools 13
- Interoperability between different subject heading languages 15
  - MACS (Multilingual Access to Subjects) 15
- Interoperability between different thesauri 20
  - Merimee 20
  - UMLS Metathesaurus : 22
  - TermSciences: portail terminologique multidisciplinaire 24
- Interoperability between subject headings and classification (UDC) 27
  - MSAC (Multilingual Subject Access to Catalogues of National Libraries) 27
- Interoperability between subject headings and classification (DDC) 32
  - WebDewey 32
  - CrissCross 35
- Interoperability between various subject indexing tools 36
  - HILT (High Level Thesaurus Project) 36
  - Renardus: 38

Conclusion and recommendations 40
Appendix A: 44
- Projects investigated but not reviewed: 44
  - CARMEN (Content Analysis, Retrieval, Metadata : Effective Networking) 44
  - HEREIN 44

Appendix B: References 46

Summary

This report presents the results of desk-top based study of projects and initiatives in the area of linking and mapping subject tools. While its goal is to provide areas of further study for cross-language subject access in the European Library, and specifically the national libraries of the Ten New Member States, it is not restricted to cross-language mappings since some of the tools used to create links across thesauri or subject headings in the same language may also be appropriate for cross-language mapping. Tools reviewed have been selected to represent a variety of approaches (e.g. subject heading to subject heading, thesaurus to thesaurus, classification to subject heading) reflecting the variety of subject access tools in use in the European Library. The results show that there is no single solution that would be appropriate for all libraries but that parts of several initiatives may be applicable on a technical, organisational or content level.

Context

One of the goals of the TEL-ME-MOR project is to ensure better access to essential European information resources by developing a networked environment that will extend access to collections and resources of the New Member State (NMS) national libraries. The goal is to facilitate access to metadata, library services and digital documents of the NMS that use at least 9 languages and a variety of character sets, including Greek

Work package 3 (WP3) aims at developing the multilingual capacity of the NMS network. The tasks are organised in a way that the various access problems are identified and investigated and appropriate solutions or future approaches are proposed. Work package 3 advises on future multilingual support requirements for improving access to and from NMS databases.

The different tasks of WP3 were designed to analyze different areas of the access problem: The two first tasks are primarily concerned with identifying capabilities in displaying data correctly in all of the languages while the remaining three tasks explore the issue of cross-language searching.

In Task 1, the UNICODE capabilities of NMS partners were analyzed to identify issues brought up by portalling databases with different scripts, character sets or individual character processing. Replies to a questionnaire on UNICODE led to a set of recommendations to The European Library office for the integration of new resources, applicable retroactively to existing ones as well (see D.3.1 Report on TEL requirements / Pierre Clavel, (www.telmemor.net/results.php)

Task 2 covers the translation of metadata labels and the user interface into languages used by the NMS.

Task 3 investigated which subject indexing tools are used by the NMS national libraries. An overview of subject access tools and practices was prepared from the responses to a questionnaire filled in by all of the NMS members from which an analysis was prepared.(see D.3.3 Report on subject access tools / Maja Žumer, Genevieve Clavel-Merrin and Patrice Landry (www.telmemor.net/results.php)). This report was an essential step in the future developments of WP3, in particular for work in Tasks 4 and 5.

The TEL-ME-MOR Report on subject access tools (Deliverable 3.3) has documented all the subject access tools used by the NMS. This includes a description of the different kinds of classifications schemes, subject headings and other means of access (pick lists, key words, etc.) used by the ten NMS. The report concluded that, even though there are a variety of subject access tools and access points used by the NMS, there were nevertheless some commons tools used in nine different languages. The report identified areas that should be further studied in the context of multilingual and cross-language access.

The main purpose of this report is to document approaches used in establishing interoperability between subject access tools and to present an overview of practices, projects and research in that field. The report will focus on work carried out in recent years and highlight solutions that have been applied successfully. Recommendations will focus on the solutions or approaches that could be further studied for establishing interoperability between subject access tools of the 10 NMS. In particular, the report

will focus on current access tools used for ensuring access to printed collections as well as digital collections. It will not address multilingual full text retrieval solutions.

Subject access in national collections

The traditional mandate of national libraries to collect, preserve and give access to printed resources has in recent years been challenged by the impressive development and increase in the amount and diversity of electronic documents. National libraries, like most public and research libraries, have had to integrate these new publications in their mandate, while at the same time working hard in trying to maintain their commitment to their core paper-based collections. In the context of bibliographic control, national libraries have made great strides in the last 30 years in developing and applying cataloguing and subject access standards and practices. Today, they are extending these standards and practices to electronic documents.

This challenge in expanding the library's mandate to ensure access to digital collections has been undertaken in a climate of rapid changes in technology as well as in information seeking behaviour and research. In a recent report commissioned by the Library of Congress, Karen Calhoun addresses this new reality facing all kind of libraries. She states:

> "Today, a large and growing number of students and scholars routinely bypass library catalogs in favour of other discovery tools, and the catalog represents a shrinking proportion of the universe of scholarly information. The catalog is in decline, its processes and structures are unsustainable, and change needs to be swift. At the same time, books and serials are not dead, and they are not yet digital. Notwithstanding widespread expansion of digitization projects, the role of catalog records in discovery and retrieval of the world's library collections seems likely to continue for at least a couple of decades and probably longer"[1]

While the prospect of being able access large digital collections and conduct full text searches is attractive to researchers, it will certainly take many years and decades to achieve the goal of converting parts of national libraries' collections to digital format. Even with the increase of digital publication, it can be assumed that published

[1] Karen Calhoun. « The changing nature of the catalog and its integration with other discovery tools » Final report, March 17, 2006. p. 5

resources would continue to make up the core of national libraries' collections in the medium to long term.

National libraries must therefore continue to ensure that access to their printed resources meets users' expectations and needs. The new generation of users is confronted with an enormous choice of information and scholarly resources, both in print and electronic formats. In this new information environment, library online catalogues must continue to offer support for browsing collections. As stated by Calhoun,

> "The predictable and consistent structure of catalog records can facilitate serendipitous discovery and offer a wide variety of ways to browse. The catalog contains good metadata, in the sense that it uses authority control, classification, and content standards to describe and collocate related materials"[2]

Subject control through subject indexing is essential for researchers seeking an extensive overview of relevant sources, regardless of the size of the collections, the languages and types of documents held. Classification schemes used in bibliographic data allow for hierarchical searches by browsing subject categories.

The traditional indexing approach ensures that searchers can access sources in large collections in a structured. The use of subject headings provides a coherent overview of the literature of a topic. Browsable alphabetic indexes enable searchers to have a global view of how a topic is subdivided and expanded. Subject indexing also provides searchers with other conceptual options relating to the topic (related terms). A search on "Europe" and "History" will yield millions of hits in Web search engines where sources contain these two words anywhere in the text – but only English language texts. On the other hand, a similar search in a library online catalogue will retrieve pertinent items in several languages. A further advantage of subject headings is that searchers are provided with a controlled vocabulary environment in which the problems of synonyms, variant phrases and different languages used in text are solved. Scholarly research using only keywords would miss not only documents published in other languages but also retrieve text in the wrong conceptual context.

[2] Ibid, p. 31.

Thomas Mann (2005)[3] provides a detailed explanation of these points underlining the need to retain subject indexing in a "Google world".

Ensuring subject access in national collections should continue to be part of national libraries mandate. However, these commitments will need to take into account the changes that are occurring in the way information is produced and accessed. This report takes the view that the heritage collections of national libraries will continue to be searched by scholars needing in-depth access to documents in all subjects. Traditional subject access tools have played a key role in ensuring that access. At the same time, the new means of ensuring access through computer assisted indexing (full text indexing) will continue to offer supplementary and alternate sources for providing better services to the users.

The 10 New Member States (NMS) all provide subject access to their collections. The current subject headings used by most are fairly recent, dating from the late 1990s / early 2000s and were established using more recent subject authorities and subject heading guidelines and standards. The catalogues are already accessible online but with their integration into The European Library, there are now opportunities to broaden the scope of access to scholars and researchers from other parts of Europe. These libraries (and others) have already invested considerable time and effort in the creation and maintenance of their indexing tools (subject heading languages, thesauri and classification schemes) which offer subject access to millions of documents. What needs to be done now is to add value to these subject access data. Establishing some type of interoperability between these subject languages and others may be the answer to improve access to their collections and ensure successful integration in The European Library portal.

Subject access issues in a networked environment

Access to library catalogues and databases has improved significantly in the last decade through advances in network technology. The Web has made it easier for library users to access online databases and online library catalogues. Users conducting searches in an online catalogue or in a virtual union catalogue can easily

[3] Will Google's keyword searching eliminate the need for LC cataloging and classification? Available from: http://www.guild2910.org/searching.htm

find and locate documents as long as they are searching for a known item for which author, title or any other factual information is known. For example, someone looking for documents by Victor Hugo in the Karsruher Virtuelle Katalog KVK (http://www.ubka.uni-karlsruhe.de/kvk/kvk/kvk_fr.html) should in a few seconds locate many documents by that author in many hundreds of libraries. However, if that same person is conducting a subject search in the KVK or in other networks, where documents and resources are indexed using different subject vocabularies and organised according to different classification schemes, the heterogeneous environment will confront him/her. In this situation, the user can only rely on his basic knowledge of some foreign languages, or rely on the use of dictionaries. In addition, he would need to query each online database separately since the vocabularies used are not compatible with each other.

As we had noted in the "Report on subject access tools"[4], subject searches have increased in importance in this new information environment. The Web environment has seen a steady development of subject search tools to access Internet information sources. These subject guides, subject directories or subject trees have been refined in the last few years, often relying on traditional library subject tools to improve access to the ever increasing amount of information. As for online catalogues, research indicates that subject access, i.e. searching on a topic is still the predominant type of end-user searching in bibliographic data.[5] Even though "keyword search" accounts for most of the searches in online catalogues, most are topical in nature and the quality of search results is often dependant on the existence of subject headings in the metadata searched[6].

Effective information retrieval is however limited by the lack of semantic interoperability in the network environment. It appears that the lack of compatibility between many subject-indexing tools is one of the major constraints facing the research and library communities today. Subject interoperability is now a key research topic for the library, computer science and web systems specialists. These information professionals have been looking for solutions to facilitate cross-domain searching in

[4] Maja Žumer, Genevieve Clavel-Merrin and Patrice Landry. Report on subject access tools, deliverable 3.3, January 30, 2006.
[5] Ibid. p.6-7.
[6] Gross, Tina and Arlene G. Taylor. 2005. What have we got to lose? The effect of controlled vocabulary on keyword searching results. College & Research Libraries Vol. 66, no. 3, 212-230.

multiple repositories (databases, online catalogues). In the last 15 years, many projects and initiatives have been set up to allow users to conduct subject searches in a “one-stop” seamless search process across different diverse resources.

Overview of the NMS subject access solutions

The Report on subject access tools (deliverable 3.3) has given an overview of subject access tools and practices in NMS national libraries and shown the wide variety of tools and practices applied. Each library has developed and adapted the tools that best reflected its cultural and linguistic specificities. The situation is probably similar in the other European national libraries.

LCSH, UDC and DDC are subject tools that are used by many of the NMS national libraries and many local languages are used in connections with these subject tools. This report has taken in account this situation when evaluating the various projects dealing with interoperability between subject access tools.

The situation may be summarised as follows (for details of the catalogues and subject access coverage, see the Report on subject access tools, deliverable 3.3, January30, 2006):

Subject headings:

- 4 different types of subject heading languages
- LCSH (or adaptations of) in 6 languages, including English
- Sears Subject Headings in English
- 2 national subject heading languages (Poland and Slovenia)
- 10 languages used

Thesauri

- 4 different thesauri used
- 4 languages used, including English

Classification:

- 2 classification schemes used (UDC –DDC)
- Captions in 5 languages

Pick lists and keywords

- Vocabularies in 6 languages

Methodology

This report on interoperability between subject access tools or cross-language searching by subject was principally compiled from published sources (articles, reports) on both the theory and methodology of cross language by subject and from studying relevant projects that have been undertaken in the last 10 years. There have been numerous projects and research in the field of interoperability. This report has limited its scope to the ones dealing with subject access. The projects retained have successfully achieved some of the goals set out and all have potential pertinence to the challenge in establishing some form of interoperability between the tools used in the NMS, and those used in other partner libraries in The European Library. All of the projects aiming to establish interoperability between subject languages were further searched on the web in the first 3 months of 2006 to ascertain their status. Most of the literature and documents reviewed are in English, French and German.

Approaches in establishing interoperability

The current interest in establishing interoperability between indexing languages as manifested by the numerous projects in this field, can be traced back to the development of the online catalogue and databases in the 1970s and the subsequent development of online networks and shared cataloguing programmes in the 1980s. In general terms, interoperability can be defined as the capacity of systems to communicate and exchange information without ambiguity and without difficulty. In the context of bibliographic data, the task is to establish some form of connexion between subject tools in order to offer a seamless access to collections.

As reported by Zeng and Chan (2004), many online services and libraries saw the need to establish some compatibility between the many subject-indexing tools that had been created in the previous 20 years. During that period, efforts had been invested in developing thesauri for the indexing and retrieval of specialised and subject oriented citation databases (indexes to periodical literature and reports). Most

bibliographic agencies created and maintained their unique controlled vocabularies. Chan and Pollard (1988) reported on the existence of over 150 thesauri in the early 1980s in Europe and in North America.

Many projects in the 1980s focused on merging and mapping various thesauri. The ISO 5964 standard supported these efforts with the publication in 1985 of the "Guidelines for the establishment and development of multilingual thesauri". It recognised the trend towards the international exchange of information and particularly tried to address the need to establish multilingual thesauri as a way to achieve improve access to specialised database and library catalogues. Some initiatives led to successful integration of thesauri some of which are still currently in use and being developed, as for example, the UNESCO thesaurus developed from the Integrated Multilingual Thesaurus project (http://www2.ulcc.ac.uk/unesco/). In the United States, many specialised libraries developed a subject access based on the linking of LCSH (Library of Congress Subject Headings) and MeSH (Medical Subject Headings) while in Canada, bilingual subject access was made possible by co-occurrence of LCSH / CSH (Canadian Subject Headings) and RVM (Répertoire de vedettes-matière de l'Université Laval) in bibliographic records. In Europe, the translation / adaptation method was used for the establishment of the French indexing standard RAMEAU. It developed as an adaptation of the Répertoire de vedettes matières (RVM) of Laval University, which in turn is a translation / adaptation of the Library of Congress Subject Headings (LCSH). Current RAMEAU authorities contain LCSH equivalences.

Practices developed since the early 1990s have been directly linked to the Internet and the creation of networked environments. These new networked environments are varied and include collections of data that often fall outside the traditional libraries' mandate. While most libraries' online catalogues have web access capabilities and many of them are part of virtual union catalogue gateways, there are a growing number of networks that bridge several cultural communities, such as museums, archives and libraries. Many of these networks have developed new standards and subject tools to describe their resources. There are now more than ever a variety of classification schemes and thesauri that give access to a wide array of collections.

The task of establishing interoperability between more varied sources of languages is probably more challenging today. Information specialists are not only dealing with metadata relating to printed resources, but working toward establishing efficient retrieval in multiple collections and repositories, in cross-domain resources and across languages. The availability of metadata and digital collections have somewhat simplified the task by increasing the possibilities of linking, mapping and merging various subject tools. Zeng and Chan (2004), Doerr (2001) and the HILT Project (2001) have in recent years presented an extensive list of projects and methods used to establish some form of interoperability. Some of these projects, such as WebDewey, Carmen, UMLS, and HILT have tried to resolve interoperability between different kind of subject indexing tools, such as classification schemes, thesauri, subject heading lists. Other projects have aimed at establishing interoperability between similar kinds of indexing languages, for example between subject headings (MACS), Merimee (Thesauri), Renardus (Classification).

In the literature reviewed, the mapping approach was the one often prescribed to link different controlled vocabularies (subject headings, thesauri, classification schemes).[7] The British Standard BS 8723, Part 4 draft document that is still under review states "interoperability is often achieved by mapping from one vocabulary to another". Zeng and Chan (2004) identified two types of mapping that are used successfully in many projects:

- Direct mapping: Establishing equivalence between terms in different controlled vocabularies or between verbal terms and classification numbers. This method can be achieved intellectually (human mediation) and by automatic means (computer assisted).
- Co-occurrence mappings: Establishing mappings from the co-occurrence of terms from different schemes in the same metadata or catalogue record. Compared to direct mapping, this method works at the application levels, in the metadata. For example, the WebDewey, part of the mapping of DDC notations and LCSH is achieved by to the co-occurrence of DDC and LCSH data in bibliographic records.

[7] McCulloch, Shiri and Nicholson (2004) states: "It seems there is strong support for mapping as a potential solution to the terminology problem and it is widely recognized that mapping does improve retrieval"

To achieve this goal of improving users' retrieval effectiveness, mapping efforts will have to overcome different factors affecting vocabulary compatibility (Doerr 2001):

- Difference in word use (in different natural languages, controlled vocabularies)
- Difference in coverage (scope)
- Difference in semantics (meaning)
- Different semantic relations (preferred terms / levels of terms)

These factors are the elements that make mapping such a labour intensive activity. To achieve accurate results, most of the mappings projects have taken the intellectual manual approach. This approach follows the principle that since most subject indexing tools are created and applied by subject cataloguers, the mapping should be done by professionals that understand the problems of subject retrieval and that are fluent in subject indexing languages.

Some projects have adopted a software-assisted approach to achieve results when the variety and quantity of resources and metadata are far too important to consider a manual approach. Some projects such as HILT and Renardus have successfully achieved some linking between resources through a classification system. These search results were achieved automatically without relying on manual assistance from indexers.

Overview of current practices projects and research on interoperability between subject access tools

The project and initiatives that are reviewed are the ones more closely related to the types of subject indexing tools used by the 10 NMS but also have relevance to the other European Library partners. Of the more than 20 projects and practices were reviewed, the following have some common characteristics with the 10 NMS subject-heading languages (LCSH), classification schemes (Universal Decimal Classification (UDC) and Dewey Decimal Classification (DDC) and offer potential solutions for linking thesauri. This overview will examine how interoperability is established using a subject tool or a combination of these subject tools. Each will be described as follows: goal, method used, work management, status, documentation and their

potential interest for the European Library. The list of some projects that were considered but not reviewed in this report can be found in Appendix A. The References are in Appendix B.

Interoperability between different subject heading languages

MACS (Multilingual Access to Subjects)

Project description:

MACS is a project of the Conference of European National Librarians (CENL) and financed by four national libraries: the Swiss National Library (SNL), the Bibliothèque nationale de France (BnF), the Deutsche Bibliothek (DDB) and the British Library (BL). Its goal is to provide multilingual subject access to library catalogues by establishing equivalence links among three lists of subject headings: SWD/RSWK (Schlagwortnormdatei / Regeln für den Schlagwortkatlog) for German, RAMEAU for French and Library of Congress Subject Headings (LCSH) for English. With these links, MACS will enable users to simultaneously search with a single query in the language of their choice (initially in English, French or German) for all relevant documents indexed by the project partner libraries in their own language.

MACS Prototype

From 1999 to 2002, work has been carried out on a MACS prototype. This prototype has currently been superseded by the MACS production system, but for archival reasons, the full prototype is kept online. Additionally, the production system currently is not linked to a Search Interface, while the prototype is.

The MACS prototype contains:

- around 3,000 headings covering theatre, athletics and those topics from all fields of knowledge which have been most often used in indexing documents
- 1,200 links created between the above concepts
- 30,000 bibliographic records extracted from the catalogues of the 4 partner libraries indexed with the appropriate subjects.
- Additional links are being prepared for headings in the fields of mathematics, psychology and information science.

Two interfaces demonstrate how a multilingual subject access system should work:

- 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.

- The Link Management Interface: enables the creation and management of links between headings from the indexing languages. Read-only access to these links is available to users outside the project.

The prototype:

- does not yet offer direct access to the partner libraries' catalogues
- does not yet contain all the concepts held in the indexing languages
- only contains authorised headings (i.e. those used to index documents) and not yet any see references (lead-in terms)

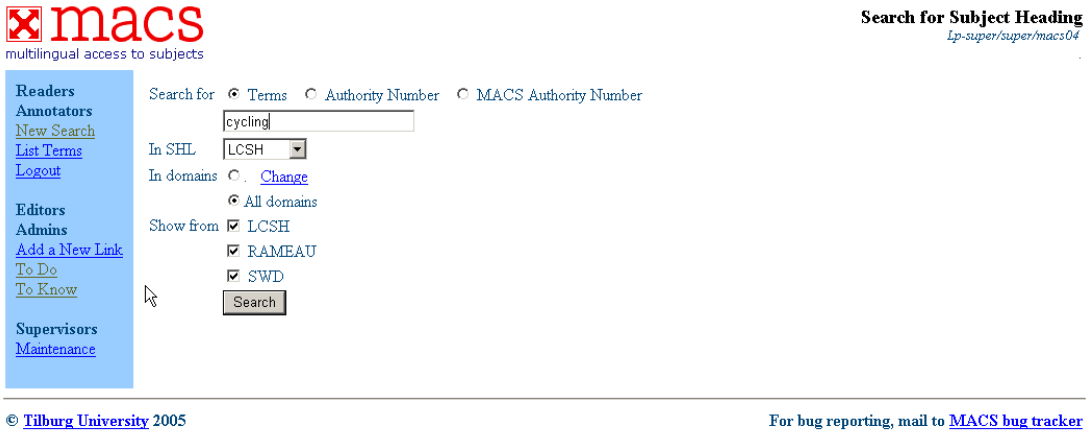
Method used:

Mapping (or equivalence linking) is the method used by MACS. Work is carried out according to subject field (using broad DDC domains) where headings from the three lists are analysed on three levels: terminological level (subject heading), semantic level (authority record) and syntactic level (application). The work follows the following principles:

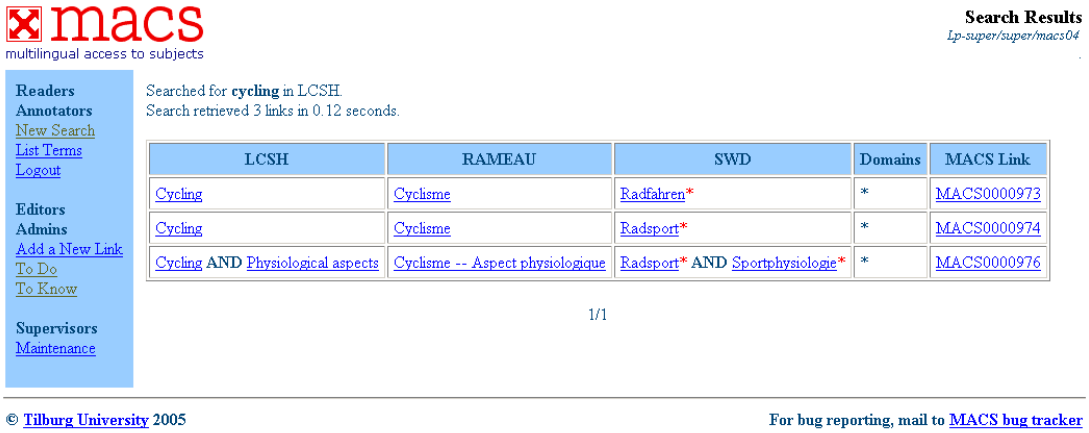
- Equality of languages and subject heading languages (SHLs): No language or SHL is used as a pivot. A language can be a source language or a target language.
- Only the establishment of equivalences between the SHLs is allowed: there is no translation of headings
- Equivalence links conceived as concept clusters: the MACS system only contains "mappings considered as conceptual clusters"
- Consistency of results: the results of subject search validate the established links

A link management database has been developed and is operational since December 2004. The Link Management Interface is a web-based database that allows each partner to add and manage links in a decentralised work environment. Partners share equal responsibility for link authorisation and each partner is responsible for adding and validating links proposed to their own SHL. Using "Cycling" as a example (see below the results in the Search Interface), the equivalent SWD heading (German heading) can be added as follows:

The term is found using the search function in the LMI (Link Management Interface)



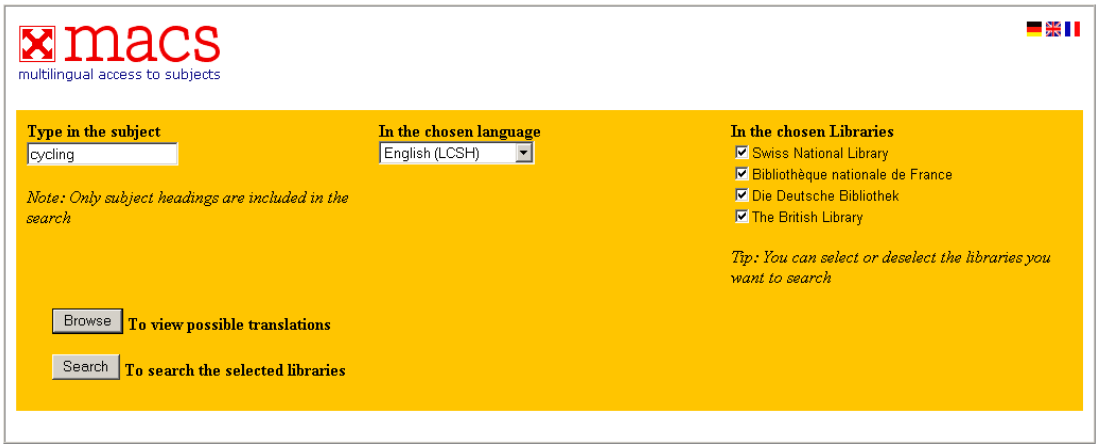
The SWD term can be added to an existing link containing (in this case RAMEAU (French heading) and LCSH (English heading) and the results achieved is:



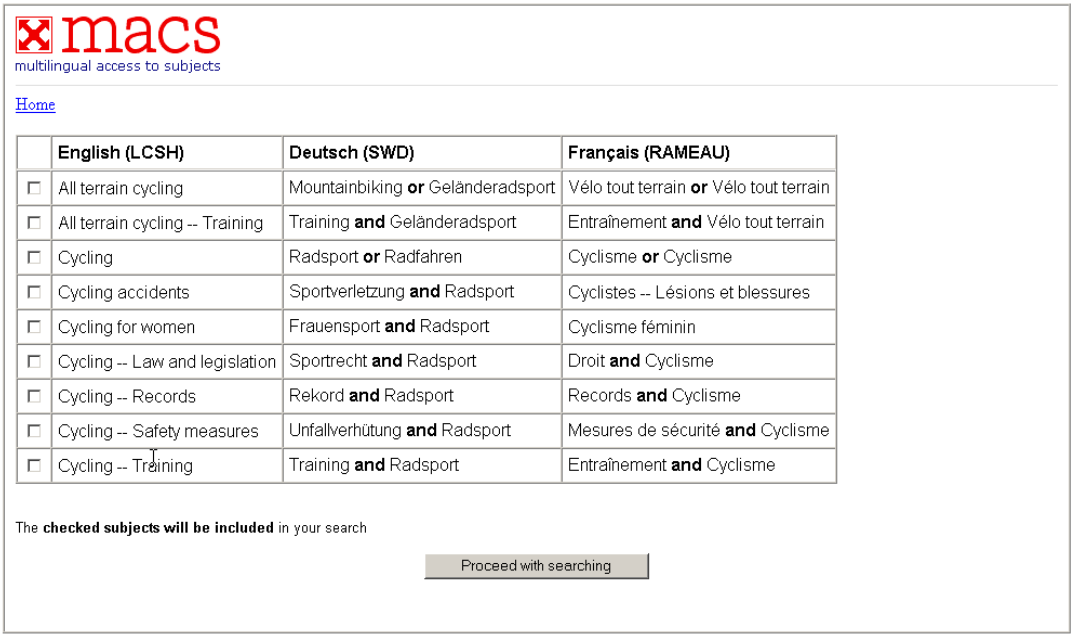
In this case, “Cycling” and “Cyclisme” are linked to two German headings: “Radfahren” and “Radsports”. These two headings are equivalent to the French and English headings. Other links using “Cycling” can be found using the “List terms” function:

Search Interface:

A prototype "Search Interface" was developed using the Zthes Z39.50 profile for Thesaurus Navigation. While the search interface was tested successfully in two partner's library catalogues, it is not presently used. The prototype should be further developed before it can be implemented in each library's OPAC.



The interface has browse and search functionalities and the users can select which library catalogues will be searched. Using the example "cycling", the library user will use the "Browse" and view and select the different links with the heading "cycling".



Work management:

The Swiss National Library manages the project. Since MACS is a CENL project, a MACS Steering Committee oversees the global management of the project. Each partner library is represented in this steering committee. Each library contributes to the project according to its own resources and plan. In recent years, the BnF has contributed RAMEAU-LCSH links while the SNL and DDB have started to add SWD headings to existing links.

The LMI is hosted at the Tilburg University (The Netherlands) and that institution will probably do further development. The prototype, including the Search Interface, is hosted by Index Data (Copenhagen)

Status:

There are approximately 40'000 RAMEAU-LCSH links (extracted from the BnF authority file) and about 5'000 with SWD equivalents in the LMI. Work started in the Fall 2005 to add SWD headings. In order to proceed in a coherent manner, the SNL and DDB are presently refining the linking methodology. Issues of source and target languages are discussed as well as work procedures. Initial discussions have started with the BNI regarding Italian subject headings. There are plans to load the remaining 30'000 RAMEAU-LCSH links in the LMI in 2006 and to test the incorporation of MACS in the European Library.

Documentation:

Information about the projects and documentation at: macs.cenl.org (https://ilmacs.uvt.nl/pub/)

Interest for The European Library

This project offers a linking approach at the subject headings level and is one of the few projects (with MSAC) that try to resolve multilingual subject access to library catalogues. The project is now at the operational level and should have solved some methodological aspects in 2006. The LMI is operational and offers a good potential for adding new subject heading languages. Since it is planned to test the incorporation

of the linked headings into The European Library portal, if the test access to bibliographic data is successful then this approach could be extended to other collections and languages.

Interoperability between different thesauri

Merimee

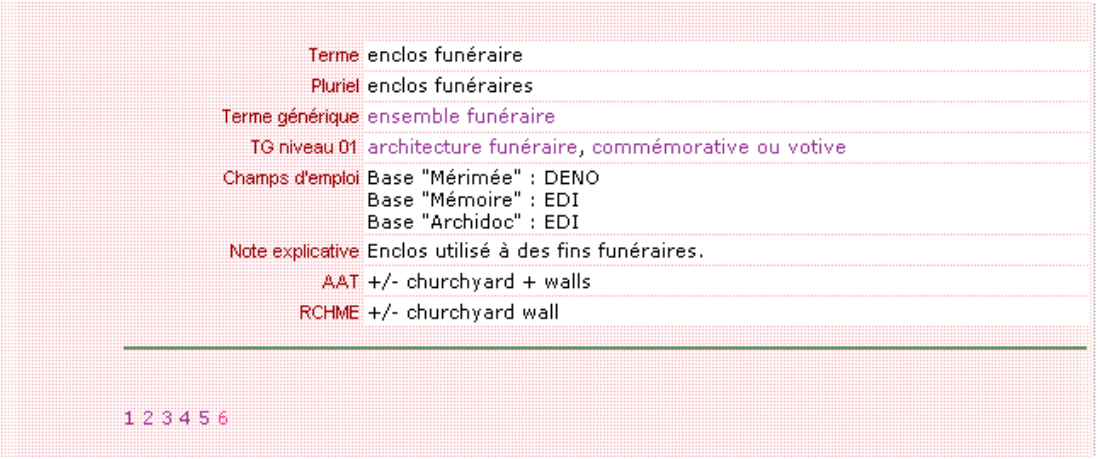
Project description:

Merimee (160'000 records) is the database developed by the Ministère de la culture et de la Communication, Direction de l'Architecture et du patrimoine and for the purpose of indexing buildings, monuments and complexes, Le thesaurus de l'architecture (in English the Thesaurus of Architecture) was created. To improve access in English, its 1135 terms have been mapped to the Art and Architecture Thesaurus (AAT) and to the English Heritage Thesaurus published by the Royal Commission on the Historical Monuments of England. It has also been recently mapped to Italian terms by the Istituto centrale per il Catalogo e la Documentazione.(the Italian terms are not yet available for searching on the Thesaurus site).

Method used:

Mapping from French to English is done at the intellectual level. The three thesaurus are linked at some points in their own hierarchies. In linking thesauri, it seems that

many of the terms in thesauri do not vary very much in the middle levels of their hierarchies and this allow for exact matches. When there is no one-to-one equivalence, mapping may be made to logical expressions made up of several terms using Boolean operators “AND and “OR” to indicate the equivalence as seen in the example below: “enclos funéraire”(Merimee) = “church” AND (+) “wall” (AAT).



Terme	enclos funéraire
Pluriel	enclos funéraires
Terme générique	ensemble funéraire
TG niveau 01	architecture funéraire, commémorative ou votive
Champs d'emploi	Base "Mérimée" : DENO Base "Mémoire" : EDI Base "Archidoc" : EDI
Note explicative	Enclos utilisé à des fins funéraires.
AAT	+/- churchyard + walls
RCHME	+/- churchyard wall

1 2 3 4 5 6

Search interface:

The Merimee database can be searched using French and English terms.

Work management:

No information

Status:

Mapping work is complete (French – English, no indication on Italian mapping work)

Documentation:

Brief description at:
http://www.culture.gouv.fr/documentation/thesarch/pres.htm, for methodological description, Doerr (2001)

Interest for The European Library

While the mapping project is a rather small project, it does have the advantage of having dealt with issues in vocabulary mappings. The mapping approach has been extensively described (Doerr (2001)) and further projects could benefit from the experiences and results of this project.

UMLS Metathesaurus :

Project description:

The UMLS Metathesaurus is developed and distributed by the National Library of Medicine as part of the United Medical Language System (UMLS) project. Zeng and Chan (2004) consider this project as “probably one of the most ambitious in harmonizing different vocabularies”. The Metathesaurus is a very large, multi-purpose, and multi-lingual vocabulary database that contains information about biomedical and health-related concepts, their various names, and the relationships among them. It is built from the electronic versions of many different thesauri, classifications, code sets, and lists of controlled terms used in patient care, health services billing, public health statistics, indexing and cataloguing biomedical literature, and/or basic, clinical, and health services research. These are referred to as the "source vocabularies" of the Metathesaurus. In the Metathesaurus, all the source vocabularies are available in a single, fully specified database format.

The Metathesaurus is organized by concept or meaning and aims at linking alternative names and views of the same concept together and to identify useful relationships between different concepts. All concepts in the Metathesaurus are assigned to at least one semantic type from the Semantic Network (linking device). This provides consistent categorization of all concepts in the Metathesaurus at the relatively general level represented in the Semantic Network. Many of the words and multi-word terms that appear in concept names or strings in the Metathesaurus also appear in the SPECIALIST Lexicon.

Method used

The mapping of terms from the different thesauri is done using a software programme called Semantic Network. This programmes uses a series of semantic types that provides a consistent categorisation of all of the concepts in the ULMS Metathesaurus.

The scope of the UMLS Semantic Network is broad, allowing for the semantic categorization of a wide range of terminology in multiple domains. Major groupings of semantic types include organisms, anatomical structures, biologic function,

chemicals, events, physical objects, and concepts or ideas. The links between the semantic types provide the structure for the network and represent important relationships in the biomedical domain. The primary link between the semantic types is the 'isa' link. The 'isa' link establishes the hierarchy of types within the Network and is used for deciding on the most specific semantic type available for assignment to a Metathesaurus concept. There is also a set of non-hierarchical relationships, which are grouped into five major categories: `physically related to,' `spatially related to,' `temporally related to,' `functionally related to,' and `conceptually related to.

Example of an 'isa' link record:

> UI: T052
> STY: Activity
> ABR: acty
> STN: B1
> DEF: An operation or series of operations that an organism or machine carries out or participates in.
> EX: Expeditions; Information Distribution; Social Planning; Return Migration
> UN: Few concepts will be assigned to this broad type. Wherever possible, one of the more specific types from this hierarchy will be chosen. For concepts assigned to this type, the focus of interest is on the activity. When the focus of interest is the individual or group that is carrying out the activity, then a type from the 'Behavior' hierarchy will be chosen. In general, concepts will not receive a type from both the 'Activity' and the 'Behavior' hierarchies.
> HL: {isa} Event;
> {inverse_isa} Behavior;
> {inverse_isa} Daily or Recreational Activity;
> {inverse_isa} Occupational Activity;
> {inverse_isa} Machine Activity

Access to the metathesausus data requires a license agreement between each institution and the National Library of Medecine (NLM). This license is at no cost to the institutions wishing to use the UMLS Metathesaurus

Search interface

UMLS Metathesaurus can be accessed at:
http://umlsks.nlm.nih.gov/kss/servlet/Turbine/template/admin,user,KSS_login.vm.
(Restricted to registered users)...

Work management:

Work is managed by the NLM. Each institution using the UMLS Metathesaurus and its applications can expand the content of the thesaurus.

Status:

On-going, new concepts are added on a regular basis. There are now close to 900'000 concepts mapped in the UMLS.

Documentation:

These is a rich documentation on this project which can be accessed at:
www.nlm.nih.gov/pubs/factsheets/umlsmeta.html
http://www.nlm.nih.gov/research/umls/umlsdoc.html
http://www.nlm.nih.gov/research/umls/meta2.html

Interest for the European Library

This is example of a successful large scale-mapping project. This project has been going for a long period of time and has developed a federated approach to increasing the number of links. It would be useful to study the organisational approach to such large-scale cooperation in order to assess its application to the European Library environment. Additionally, the linking software 'Semantic network' should be analysed in more detail in order to see if it could be used in other projects.

TermSciences: portail terminologique multidisciplinaire

Project description:

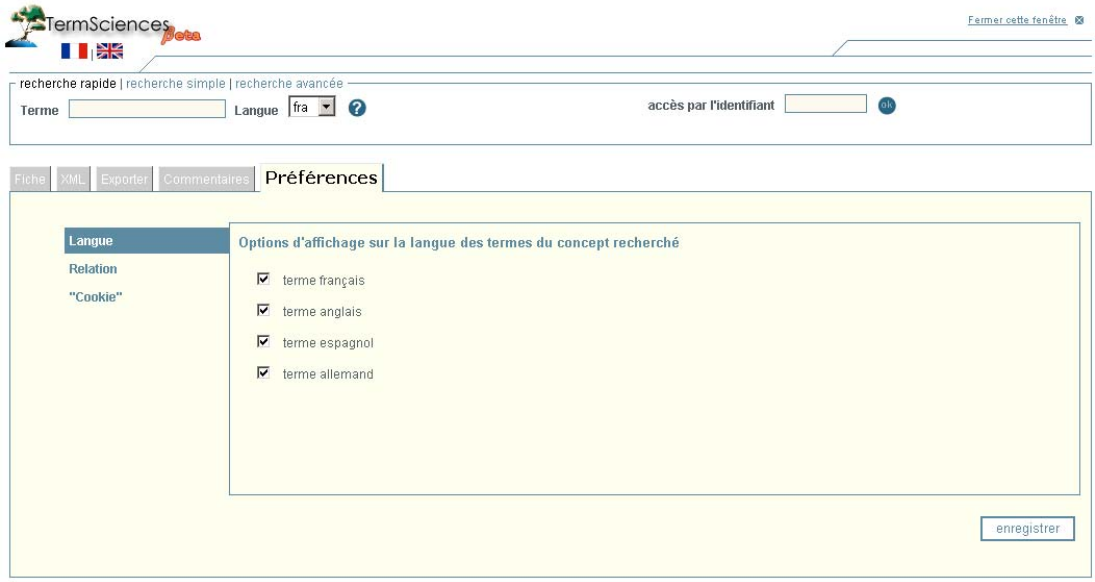
TermSciences is a multidisciplinary terminology database created by the French institution Direction de l'Information Scientifique (CNRS) and developed by l'INIST in collaboration with the LORIA and the ATILF. A BETA version of the database was made available in November 2005 and a newer version was issued in March 2006.

Method used:

Using a merging software (information not available), the database contains equivalences between the following languages sources (thesaurus, dictionaries) developed or used by the following institutions: Banque de données de la santé publique (BDSD), Institut de l'information scientique et technique (INIST), Institut national de la recherche agronome (INRA) and National Library of Medecine (NLM). The terms are mapped to "concept" terms.

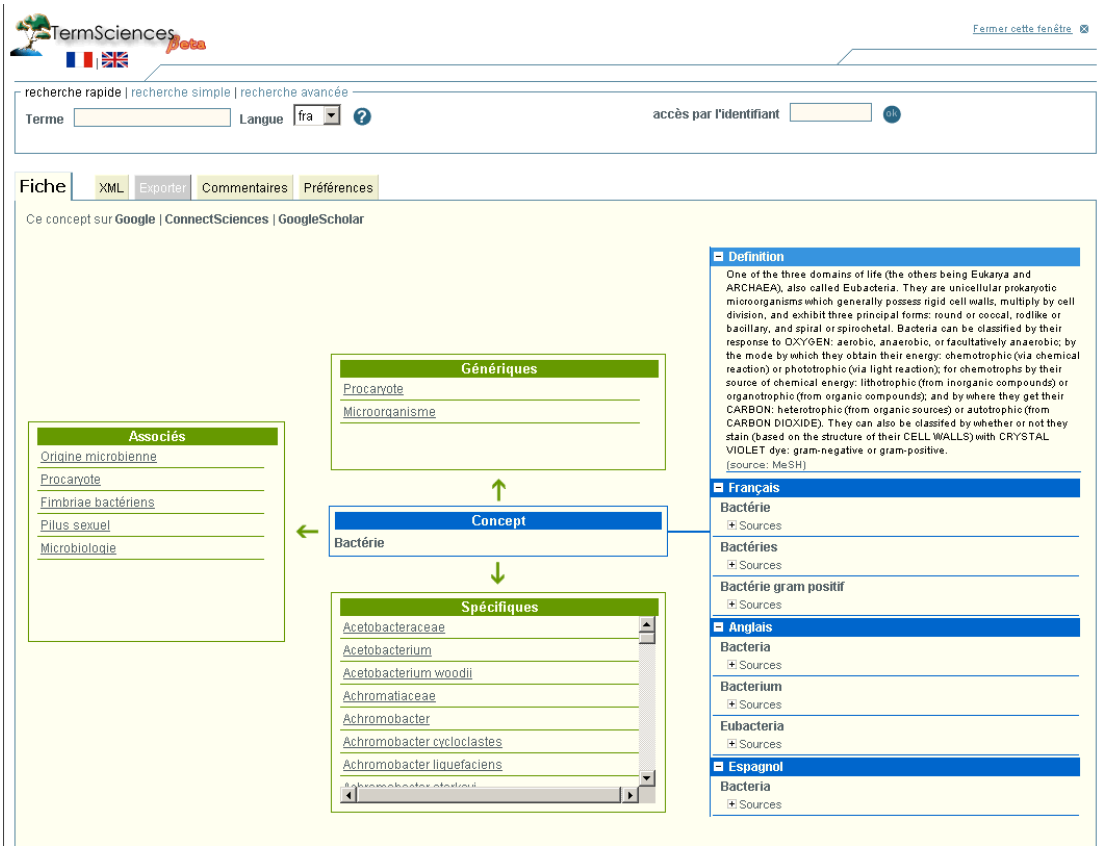
Search Interface:

1. The search interface is available in French and English and searches are possible in four languages (French, English, Spanish and German) and the terms can also be displayed in these languages.



2. The search results give the following information (when available in the sources): definition, related terms in each languages, concept term, broader and related terms. For each term displayed, the source languages are also

indicated.



This database is not yet linked to bibliographic databases.

Work management:

INIST coordinates the project. It oversees the contributions by participating members and is hosting the database. Work is organised in three phases (phase 2 completed) and the project should be completed in 2008.

Status:

On-going

Documentation:

The following web site gives access to the project documentation as well as to the search interface: http://termsciences.inist.fr/

Interest for The European Library

This project could be of some interest because of its use of a linking software. This project has succeeded in mapping successfully terms from a few thesauri in different

languages. Further investigation should be conducted to find more about the methodology used and how the mapping is maintained and managed. It would also be interesting to see how this could be connected to a bibliographic database.

Interoperability between subject headings and classification (UDC)

MSAC (Multilingual Subject Access to Catalogues of National Libraries)

Project description:

MSAC is a cooperative venture of three libraries in the Czech republic: National Library of the Czech Republic, Morovian Library and Olomouc Research Library (Czech National Subject Authority File – CZENAS) and the other participating national libraries (Croatia, Latvia, Lithuania, Macedonia, Slovakia and Slovenia). The goal of this project is to provide multilingual subject access through the Czech National Subject Authority File (CZENAS). This authority file of topical terms serves as a base for establishing equivalences between the different subject headings languages used by the partners libraries through UDC notations used as a switching language.

Method used:

The Czech National Library team is responsible for the mapping of Czech verbal expressions, UDC notations and English equivalents (mostly LCSH). The other participating libraries are responsible for mapping their own subject headings to the UDC and English headings. The links are established between preferred terms represented by isolated lexical units only. Subject headings strings are excluded form the mapping process.

Czech authority file of topical terms is intended to form a base for multilingual controlled vocabulary. The aim of the proposal is to provide users of online library catalogues and internet services of cooperating institutions with an indexing and retrieval tool which enables multilingual and cross-domain searching (“one-stop” seamless searching)

Format – MARC 21

Special indexes (language based)
Subjects-Czech
Subjects-English

Subjects-Croatian
Subjects-Latvian
Subjects-Lithuanian
Subjects-Macedonian
Subjects-Slovenian
Subjects-Slovak

Mapping process

The mapping process between Czech verbal expressions, UDC numbers and English equivalents is being done intellectually. Czech National Library is responsible for UDC notations and English equivalents (comments to both of them are invited).

Participating libraries are responsible for subject terms in their native language. Mapping process between English verbal expressions, UDC numbers and native tongue equivalents is being done intellectually as well. Cooperating libraries receive special tables of preferred terms in English accompanied by UDC numbers; they add data in their language, then verbal terms will be incorporated in authority records.

Principles of mapping

The project aims to establish equivalents between subject controlled terms used in the indexing systems of participating libraries through a switching language.

- Switching language: UDC notations based on UDC MRF (+ English equivalents)
- Mapping links are defined between controlled vocabulary terms represented by isolated lexical units only
    - subject headings strings are not mapped
    - authority records are not mapped, the links are established only between topical main headings.

MSAC is being developed in two phases
The first phase includes

- the development of Czech topical authority file
- integration of language variants of participating libraries in Czech subject authority file

The project is still at an early stage of development: there are verbal expressions of controlled terms linked to the UDC class number 52 (in Croatian, Latvian, Lithuanian, Macedonian, Slovenian, Slovak), UDC class number 314 Demography, 316 Sociology, 32 Politics, 34 Law (Slovak, Macedonian, Croatian)

In the second phase of the project the combination of UDC-natural language verbal expressions or UDC-English expressions, (or both of them) are to be inserted into the special fields of respective bibliographic records in the databases of cooperating libraries. The process could be done (semi)automatically, but intellectual checking of data will be required.

Cooperating libraries should provide access via the Z39.50 protocol. All the libraries, having the working Z39.50 server will be integrated directly, while for those not supporting this protocol a small testing database can be created at the CNL. CNL can offer access to the databases of cooperating libraries via one single interface in the Uniform Information Gateway (UIG). After joining the system, UIG users can (after accomplishing the procedure of authorisation and authentication) work in their own environment. They can work in both Czech and English environment and at the moment, can use both Czech and English languages for searching.

Example of record

LDR	-----nz--a22-----n--4500
001	ph120088
003	CZ-PrNK
005	20040809155912.0
008	000811|n|anznnbbbn----------n-a|a------
040	|a ABA001
089	|a 314.15-026.49 |c demografie
089	|a 343.264 |c právní hledisko
150	|a exil
550	|a diaspora (sociologie)
550	|a emigrace
550	|a emigrace a imigrace
550	|a exilová politická činnost
550	|a exilová vláda
550	|a exilový tisk
550	|a imigrace
550	|a politická emigrace
550	|a vyhoštění
75007	|a exile |2 eczenas
75017	|a egzil |b 314.15-026.49 - demografija |2 scr
75017	|a egzil |b 343.264 - law |2 scr
75047	|a Egzil |b 314.15-026.49 - demografija |2 mac
75047	|a Egzil |b 343.264 - pravo |2 mac

75057	|a exil |b 314.15-026.49 - demografia |2 slo
75057	|a exil |b 343.264 - právo |2 slo

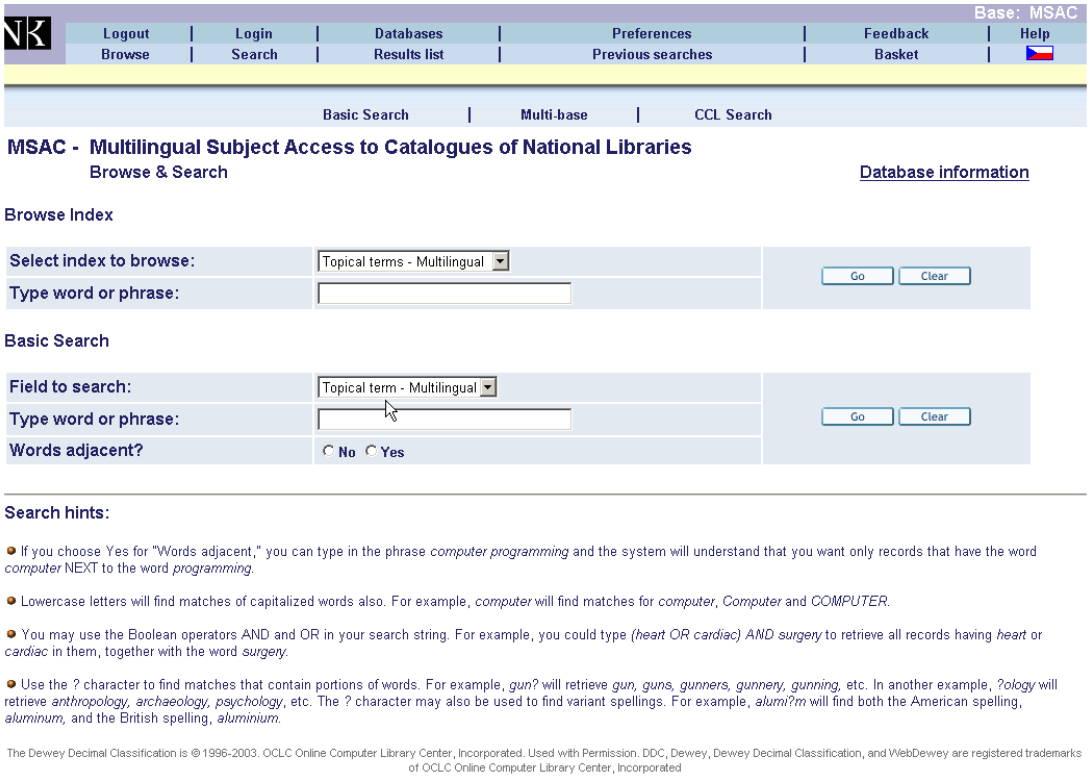
Search interface:

A search interface has been developed and is operational. The MSAC authority file is accessible in an English web gateway “Catalogues and Databases of the National Library of the Czech Republic”

See:

http://sigma.nkp.cz/F/7553YXR7ERLDRUUE2HDTG4F64I65PB6D4EUST6FM7FSRD69G71-01479?func=file&file_name=base-list

The search interface contains browse and search commands and allows Boolean search and right truncation.



The search interface can be seen at: http://sigma.nkp.cz/F/RJF43AKX11DQFGVNNAHAG2XS57H92A9QYBUG7JI1D4P58P5F7Y-02881?func=file&file_name=find-b

Work management:

Work is done on a voluntary basis and has not received financial support. There is no indication of the amount of work done so far. No strategic plan was located.

Status:

Linking work is proceeding. No additional information could be found.

Documentation:

Presentation by Marie Balikova (**http://www.ifla.org/IV/ifla71/papers/044e-Balikova.pdf**)

Interest for The European Library :

MSAC has the advantage in working with some of the subject headings languages and classification schemes used by the NMS Similarly to the MACS project, MSAC is dealing with multilingual subject access. These are the only two European projects

that are working at ensuring this access at the subject heading level. There have been some discussions between MACS and MSAC which have not yet led to concrete work initiatives.

More information about MSAC is needed. There are questions about the scalability of the linking at the authority record level. It may be difficult to manage the link data as more and more languages are linked in the authority record.

Interoperability between subject headings and classification (DDC)

WebDewey

Project description:

LCSH headings are linked to DDC notations in order to assist subject cataloguers in locating DDC notations according to LCSH headings and LCSH headings according to DDC notations**.**

"The enhanced DDC database contains data from the DDC 22 database and additional terms that have been associated with DDC numbers in the database. These terms are Library of Congress Subject Headings (LCSH) extracted from a variety of sources and associated either intellectually or statistically with the DDC".

WebDewey Abbreviation	**Source of Term**	**Association Method**	**Revised for DDC 22?**
EM	Editorially mapped LCSH	Intellectual	Yes
PPT	People, Places & Things	Intellectual/Statistical	Selected areas (see paragraph below)
SM	OCLC WorldCat	Statistical	No

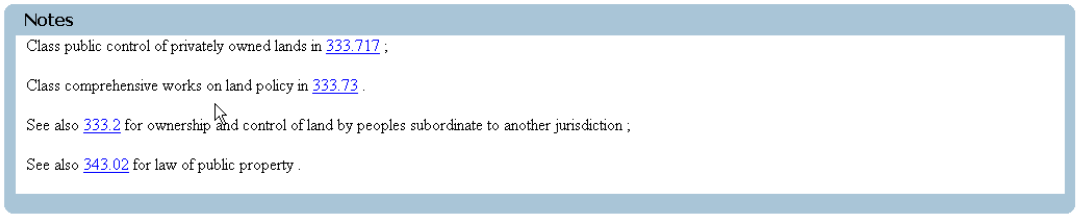
Dewey editorial staff members have intellectually linked terms with the EM designation to the DDC, and have revised these mappings to fit DDC 22. In 610 Medicine and health, a few headings designated EM will be Medical Subject Headings (MeSH) rather than LCSH; these headings can be identified because no part of the heading is linked to an authority record.

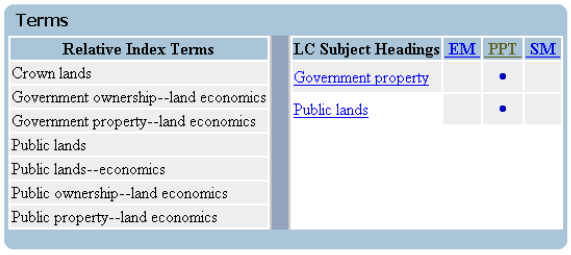
Terms with the designation PPT have been either intellectually associated with the DDC, or statistically associated and then reviewed by Dewey editorial staff. Mappings in the areas of computer science, religion, and law have been updated to match DDC 22 numbers; the rest of the PPT mappings reflect DDC 21, and have not yet been revised to fit DDC 22.

The remaining terms, identified with the abbreviation SM, have been automatically linked to the DDC using statistical techniques. These terms have been associated with the DDC using a term co-occurrence measure. This measure was applied to WorldCat records that contain MARC 082 (Dewey Decimal Classification number) and a subject heading field, MARC tags 600-651. The SM mappings have not been revised to fit DDC 22." (source: WebDewey (paid subscription)

Example of DDC-LCSH link:

For class number 333.1 Public ownership of land, the following LCSH headings were mapped:
Government property
Public lands





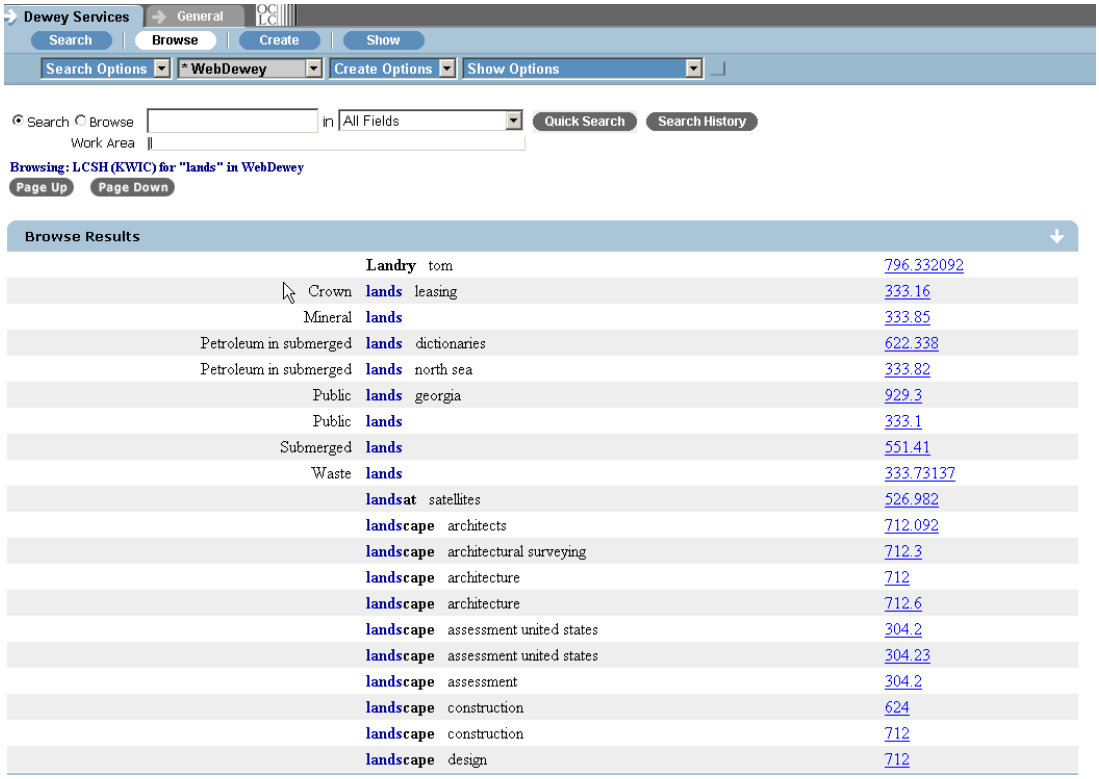
Method used:

Manual and statistical based linking

Search interface:

The DDC numbers linked to LCSH headings can be search in WebDewey:

By Browse command: from DDC numbers and from LCSH headings



By search command:
DDC numbers search will give access to complete record including LCSH

Status:

On-going

Work management:

Under the responsibility of OCLC staff

Documentation:

www.oclc.org/dewey/products/webdewey and the WebDewey website (by subscription)

Interest for The European Library

The mapping work undertaken by OCLC in collaboration with the Library of Congress should be explored in more detail. The methodology used could be of some interest for MSAC even though another classification scheme is used in that project. The CrissCross project will also investigate more closely the mapping work done in WebDewey.

CrissCross

Project description:

The CrissCross project started in early 2006 at the Deutsche National Bibliothek aims at mapping SWD headings to DDC notations. The project is funded by the DFG (Deutsche Forschungsgemeinschaft)

Method used:

The method is presently being developed at the Deutsche National Bibliothek and Köln University. The preliminary approach is to use the linking results of WebDewey (DDC-LCSH) in order to match the SWD to the LCSH. CrissCross will also use the results of the MACS project (SWD links to LCSH and RAMEAU).

Search interface:

The results will be accessed through MELVIL, the German electronic version of DDC22.

Work management:

The linking work will be done at Köln University and at the Deutsche-National Bibliothek over a period of two years (funding period). The project participants will work closely with the Swiss National Library as the SWD partner in the MACS project. The linking work will be managed in the ILTIS system, the Deutsche National Bibliothek authority and bibliographic database.

Status:

On-going

Documentation:

Not yet available.

Interest for The European Library

This project is dealing with the DDC, SWD and LCSH vocabularies. The methodology that will be developed and used could be used as another model for cross-language applications. The date produced in this project could also be integrated in the European Library.

Interoperability between various subject indexing tools

HILT (High Level Thesaurus Project)

Project description:

The HILT started as a one-year project to study and report on the problem of cross searching and browsing by subject the collections of a wide range of communities, services in the UK (libraries, museum, archives and Internet services). Phase I concluded that there was a consensus between the different members of the networks involved in the project to pursue the feasibility of developing a "user and machine responsive interactive route map to the terminologies used by these communities and the relationships between these terminologies" (Nicholson (2002).

One of the important tasks developed in Phase II was the creation of the "terminologies pilot" based on an adaptation of the Wordmap software (see search interface). The project is now at Phase III and aims at developing the current Pilot into a distributed server version of the project, which would entail building a web-services version of the current HILT pilot. The target date is 2007.

Method Used:

The HILT pilot "terminologies route map" was built using the commercially available Wordmap software. This company specialises in the development of systems based on knowledge maps, where various vocabularies are linked together.

Search Interface:

The prototype (pilot) search interface is operational and can be consulted at http://hilt.cdlr.strath.ac.uk/. The site offers search and browse capabilities. Browse is organised by DDC captions and searches are done in the BULBL Link Catalogue of Internet resources.

Work management:

The HILT project is managed by Centre for Digital Library Research (CDLR),University of Strathclyde .

Status:

On-going, in phase III

Documentation:

The project is fully documented at the HILT web site: http://hilt.cdlr.strath.ac.uk/

Interest for The European Library

The use of software to achieve interoperability between many kinds of vocabularies should be further investigated. The project management approach could also be use as a basis for other projects.

Renardus:

Project description:

Renardus was developed and funded in the context of the EU Information Society Technologies 5th framework programme. The project that lasted from January 2000 to June 2002 developed a broker information service, which allowed users to cross-search and cross-browse via a single interface a distributed range of existing European subject gateways. Each subject gateways involved in the project has created metadata sets that include thesaurus terms and classification that are linked to the DDC. The DDC is in fact the switching language used for browsing the different collections.
By 2003, the project had covered about 65'000 digital web based collections.

Method Used:

As noted above, the Dewey Decimal Classification (DDC) was chosen to act as an 'interlingua' within the Renardus pilot. Gateways participating in the Renardus system were 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, a software tool developed as part of the German CARMEN[8] project was adapted to facilitate the relevant workflow.

Search interface:

A search interface to the collections has been developed and is functional. Browsing is possible by Dewey Decimal Classification captions as well as a "graphic navigation" that gives an overview of the subject domains. Access is given to the web resources classed under the domains. Access is available at: www.renardus.org

Work management:

The project is managed by the Renardus Consortium. Since 2002, the project has been maintained with minimal funding. There is no recent information on the Renardus website.

[8] CARMEN (Content Analysis, Retrieval, Metadata : Effective Networking : www.mathematik.uni-osnabrueck.de/projects/carmen

Status:

On-going (need to be confirmed)

Documentation:

The project is fully documented at the Renardus website: www.renardus.org

Interest for The European Library

The software tool should be studied in order to assess its applicability to linking within the European Library framework.

Conclusion and recommendations

There have been a variety of projects, studies and initiatives undertaken in the last 30 years in the area of interoperability between subject heading languages. There have been successful results in clearly defined environments. For example, in Canada, bilingual subject access was achieved by linking two similar subject-heading languages (LCSH /CSH and RVM) in libraries that needed to offer bilingual access to collections. In some libraries in Belgium, multilingual access was made possible through the translation of LCSH headings in French and Flemish and in the Swiss Federal Institute of Technology libraries in Switzerland, by translating German headings (linked to UDC numbers) in English and French. Note, however that these cases are all examples of translating headings not mapping them and therefore these approaches cannot be transferred to the European Library partners as they have existing indexing languages in use covering thousands of documents.

This report, which surveyed widely the current literature on subject indexing, gathered the most current initiatives in mapping subject data. The nine projects reviewed tried to establish interoperability between different types of subject tools from different sources (libraries, museums, information centres). They represent projects that have set up cooperative initiatives between libraries and information from different countries. As such, they are very representative of typical European projects.

The survey has also found that there have been very few projects that have attempted to establish interoperability between subject indexing tools in different languages. Zeng and Chan (2004) listed 10 projects that have achieved some form of interoperability between different subject indexing tools. The MACS project was the only one mentioned linking subject headings from different languages. Many projects and initiatives have successfully created bilingual and multilingual thesauri but most are used in specialised databases and in specialised subject areas and thereby avoid the typical problems of mapping in generalised vocabularies: synonyms, partial overlap in meaning, ambiguity (especially in title material where there is not enough context to disambiguate). While national libraries have some specialised collections,

most have encyclopaedic collections indexed by general vocabularies Therefore, the relevance of specialised thesauri to national libraries was deemed to be too low for them to be included in the study.

The nine projects reviewed show the limited choice of multilingual cross language options for the European Library. Many projects are still at the “proof of concept” stage (MSAC, CrissCross) while others have not migrated from the prototype stage (CARMEN) or it is not clear that they are updated (Renardus). HILT and MACS are still building their links. Three of the projects reviewed have succeeded in establishing some interoperability in very specialised domains. The UMLS Metathesaurus and TermSciences are mostly thesauri based and oriented in the scientific fields, while Merimée deals with architecture terminology. Libraries and information providers can use the UMLS Metathesaurus and the TermSciences for indexing purposes but not for accessing bibliographic records. No search interface using these tools that allow for searching and browsing across many collections using different headings was found.

Four of the analysed projects are considered relevant to the collections of the European Library in content and methodology and we recommend they should be considered in more detail: MACS, MSAC, CrissCross and WebDewey. One is interesting from a methodological point of view: HILT. Recommendations are given below.
The four linking projects are relevant for the following reasons:
MACS and MSAC are linking subject headings in several languages. Both involve European national libraries with the goal to offer multilingual subject access to general and encyclopaedic-based collections. There are methodological differences between these projects: MACS links subject headings while MSAC uses the UDC classification as a pivot to link subject headings, but synergies can be found.

The CrissCross and WebDewey projects also link subject headings and classification numbers. The CrissCross project managed by the Deutsche National-Bibliothek will create links between DDC notations and at least LCSH, RAMEAU and SWD headings, working in collaboration with MACS. The WebDewey project is a non-European project principally managed by OCLC. It is embedded in the work process

at OCLC and at the Library of Congress and is ongoing. The CrissCross project will interact informally with this project.
All four projects are potential sources of mapped data that could be re-used within the European Library.

The following recommendations draw on selected elements from the projects studied that could contribute to successful cross-language searching in the European Library. Detailed requirements, cost and time estimates will be provided in D3.6 Implementation proposal and scenario for multilingual access. The goal is to first establish feasibility of large-scale searching using existing data before investing heavily in mapping. The first stage will determine the subsequent extension or not of work in this area.
There are currently three new or revised standards available or in preparation covering monolingual or cross-language vocabularies[9] which should be taken into account when mapping.

Recommendation 1: Proof of concept – searching

Test cross-language searching in the European Library using the 70'000 RAMEAU/LCSH links against data from British Library and Bibliothèque nationale de France.

Recommendation 2: Test large-scale linking and searching

In order to demonstrate the feasibility of large-scale linking, we recommend using links already available in the following subject areas: Sports and Theatre. Links in these areas have already been created in MACS and MSAC giving potentially 8 –10 languages to be tested.

Recommendation 3: Mapping (manual)

If the tests carried out in recommendations 1 and 2 are successful, mapping projects may be started. It is recommended that the MACS Link Management System be used

[9] ANSI/NISOZ39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.* Bethesda, Maryland: NISO Press, 2005. 172 p.
British Standard BS 8723 Part 4 (DRAFT) Interoperability between vocabularies.(2006) (restricted access). 42 p.
IFLA Working Group on Guidelines for Multilingual Thesauri, Classification and Indexing Section 2005. *Guidelines for multilingual thesauri* (in consultation) http://www.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf 21p.

as a central source (clearinghouse) of mapping results between subject headings used in European national libraries. Data can be loaded in one or many connected Link Management databases. It can be numeric (DDC or UDC) or text. The bulk loader developed to load data from the Bibliothèque nationale de France is generic and can be parametered to take different formats, such as MSAC for example. Output in xml can be loaded in an index for the European Library. The MACS Link Management System is based on a decentralised management organisation and offers each Subject Heading Language full responsibility in creation and validation of links. The choice of subject areas should be coordinated among partners according to collection strengths and, if possible, users' search areas as indicated in the European Library web statistics. Partners with subject data already linked to LCSH (Spain, Sweden, some NMS) may use this as a bridge to link to the existing mapped data.

Recommendation 4: Test automated mapping

Manual or intellectual linking is time consuming and costly. In parallel to work on manual mapping, investigate HILT, TermSciences and WebDewey as tools to speed up link creation. Note that automatic mapping requires a large corpus of data – if as planned more partners make their data available in OAI for the European Library central index this could be a source for comparison, providing enough data elements are present. Note that in the feasibility study for MACS, a small sample of data from the partner libraries was analysed to see if link creation could be carried out by matching the indexing used in each language for the same document. The study reported a concordance of between 29 % and 55 % of headings and concluded that this was not sufficient to develop automated tools for those collections at that time.

The issue of interoperability between controlled vocabularies is an important issue for libraries and information services. In the context of the European Library, multilingual subject access is one of keys to ensure that users are able to retrieve information from the collections of European national libraries. Efforts in bridging subject indexing languages used in European libraries should lead to a better understanding of the value of these indexing tools and the role they play in connecting library users to the information contained in the rich collections of documents held in libraries.

Appendix A:
Projects investigated but not reviewed:

The following projects were investigated but not retained because of:

- Lack of current information
- Status of project uncertain
- Not relevant to the scope of the WP.3.4

CARMEN (Content Analysis, Retrieval, Metadata : Effective Networking)
www.mathematik.uni-osnabrueck.de/projects/carmen

This German project started in October 1999 and finished in February 2002. Information from its WEB site indicated that the aim of this project was to improve access to decentralised data repositories. Organised in 12 work packages, the project looked at the issues of Metadata, retrieval and heterogeneity of data. WP12 (Cross concordances of classifications and thesauri) under the responsibility of F. Geisselmann looked at finding solution for establishing cross concordances of classifications and thesauri in the areas of mathematics and physics. One part of the WP12 proposed to establish a concordance between the DDC, the RVK (Regensburger Verbund Kalsiifikation in the areas of mathematics and physics and the special classifications, Mathematics Subject Classification (MSC) and the Physics and Astronomy Classification Scheme (PACS). Another part of the project looked at linking the IZT (Informationszentrum Sozialwissenschaften) thesaurus with the SWD (Schlagwortnordatei). According to Zeng and Chan (2004), several thousand descriptors were linked up to 2002.

Not retained for review: No information on results was found and no answers were received from an inquiry to the project coordinator. The results could not be evaluated and proposed to the NMS members. The tool developed in the project was subsequently used in Renardus and should be studied in that context.

HEREIN
www.european-heritage.net/sdx/herein/thesaurus/introduction.xsp
The HEREIN Project focused on building a new multilingual thesaurus in the area of archaeology and archaeological heritage. The thesaurus was constructed using terms from reports on cultural heritage policy in Europe (Spain, France, Hungary, Ireland,

Norway and UK). The three teams involved in the project (Spain, France and UK) created a separate list of terms in their own language. The three teams compared their list so as to obtain a pool of words with linguistic equivalence in the three languages. The thesaurus has been created and can be searched at the following web site: http://www.european-heritage.net/sdx/herein/thesaurus/consult.xsp

Not retained for review: Even though this seems to be a successful project, the NMS does not require the creation of a new thesaurus.

Appendix B: References

ANSI/NISOZ39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.* Bethesda, Maryland: NISO Press, 2005. 172 p.

British Standard BS 8723 Part 4 (DRAFT) Interoperability between vocabularies.(2006) (document en consultation restreinte). 42 p.

Balikova, Marie. 2005. Multilingual Subject Access to Catalogues of National Libraries (MSAC): Czech Republic's collaboration with Slovakia, Sovenia, Croatia, Macedonia, Lithuania and Latvia. 71th IFLA General Conference and Council. Available from: http://www.ifla.org/IV/ifla71/papers/044e-Balikova.pdf

Bates, Marcia J. 2003. Task force recommendation 2.3 research and design review : Improving user access to library catalog and portal information. Final report (version 3), June 1, 2003. Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium. 57 p.

Calhoun; Karen. 2006. The Changing Nature of the Catalog and its Integration with other Discovery Tools. Cornell University Library, March 17, 2006. 52 pages. Available from : http://www.loc.gov/catdir/calhoun-report-final.pdf

Chan, Lois Mai.2002. Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources: Issues and Challenges. Available from: http://lcweb.loc.gov/catdir/bibcontrol/chan_paper.html

Chan, L. and Zeng, M., 2002, Ensuring interoperability among subject vocabularies and knowledge organisation schemes: a methodological analysis. 68th IFLA Council and General Conference, August 18-24, Glasgow, Scotland, UK. Available from: http://www.ifla.org/IV/ifla68/papers/008-122e.pdf

Clavel-Merrin, Genevieve. 1991. Multilingual Subject Access in Switzerland. in *Library Automation and Networking: New Tools for a New Identity*, Herman Liebaers and Marc Walckiers ed. München : K.G. Saur, 215-228

Clavel-Merrin, Genevieve. 1999. The Need for Co-Operation in Creating and Maintaining Multilingual Subject Authority Files. *65th IFLA Council and General Conference, Meeting 155.* <http://WWW.ifla.org/IV/ifla65/papers/080-155e.htm>

Cordeiro, Maria Inês. 2001. From Library Authority Control to Network Authoritative Metadata Sources. in *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting Sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC, .

Crane, Gregory. 2006. What do you do with a million books? *D-Lib Magazine* Vol. 12, no. 3. Available from: http://www.dlib.org./dlib/march06/crane/03crane.html

Doerr, Martin. 2001. Semantic problems of thesaurus mapping. *Journal of Digital Information* 1 (8) Available from: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

Freyre, Elisabeth and Max Naudi. 2001. MACS: Subject Access Across Languages and Networks, in *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting Sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC, p. 3-10.

Garrett, Jeffrey. 2006. KWIC and dirty: human cognition and the claims of full text searching. *JEP: The Journal of Electronic Publishing* Vol. 9, no 1. 16 p. Available from: http://www.hti.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0009.106

Gross, Tina and Arlene G. Taylor. 2005. What have we got to lose? The effect of controlled vocabulary on keyword searching results. College & Research Libraries Vol. 66, no. 3, 212-230.

Hudon, Michèle. 1997. Multilingual Thesaurus Construction : Integration of Views of Different Cultures in one Gateway to Knowledge Concepts. *Knowledge Organization* Vol. 24, no 2, 84-91.

IFLA Working Group on Guidelines for Multilingual Thesauri, Classification and Indexing Section 2005. *Guidelines for multilingual thesauri* (in consultation) http://www.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf 21p.

Koch, Traugott and Heike Neuroth and Michael Day. 2003. Renardus: cross-browsing European subject gateways via a common classification system (DDC), in *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting Sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC, p. 25-33.

Landry, Patrice. 2001. The MACS Project: Multilingual Access to Subject (LCSH, RAMEAU, SWD). *International Cataloguing and Bibliographic Control* vol.30, no 3, 46-49.

Landry, Patrice. (2004). Multilingual Subject Access: The Linking Approach of MACS. *Cataloging & Classification Quarterly* Vol. 31, nos 3-4, 177-191.

Lee, Edmund. 2005. Building interoperability for United Kingdom historic environment information resources. *D-Lib Magazine* Vol. 11, no. 6. Available from: http://www.dlib.org./dlib/june05/lee/06lee.html

Leth, Pia ; Ingrid Berg. Subject indexing in Sweden: The creating of a national system based on international standards in a country that often wanted to go its own way. 70th IFLA General Conference and Council, August 22-27, 2004. http://www.ifla.org/IV/ifla70/papers/041e-Leth_Berg.pdf

Lopes, Maria Inês et Julianne Beall, ed.(1999). *Principles Underlying Subject Heading Languages (SHLS).* Working Group on Principles Underlying Subject Heading Languages, IFLA Section on Classification and Indexing, München: K.G. Saur.

MacEwan, Andrew. (2000). Crossing language Barriers in Europe: Linking LCSH to Other Subject Heading Languages. *Cataloging & Classification Quarterly* Vol. 29, nos 1-2, 199-207.

Mann, Thomas. 2006. A critical review [of] "The Changing nature of the catalog and its integration with other discovery tools. Final report. March 17, 2006. Prepared for the Library of Congress by Karen Calhoun". Prepared for AFSCME 2910, The Library of Congress Professional Guild. April 3, 2006.
http://www.guild2910.org/AFSCMECalhounReviewREV.pdf

Mann, Thomas. 2005. Will Google's keyword searching eliminate the need for LC cataloging and classification? Available from:
http://www.guild2910.org/searching.htm

McCulloch, Emma and Ali Shiri and Dennis Nicholson. 2004. Challenges and issues in terminology mapping: a digital library perspective. E-LIS: E-prints in Library and Information Science. (Deposited on 17 January, 2005).
http://eprints.rclis.org/archive/00002994/

McIlwaine, Ia C. 2003. Trends in knowledge organization research. *Knowledge Organization* Vol. 30, no 2, 75-85.

Nicholson, Dennis. 2002. Subject-Based interoperability: Issues from the HighLevel Thesaurs (HILT) Project. 68th IFLA Council and General Conference, August 18-24, 2002. Available from: http://www.ifla.org/IV/ifla68/papers/006-122e.pdf

Nicholson, Dennis and Susannah Wake. 2003. HILT:subject retrieval in a distributed environment, in *Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting Sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC, p.61-67.

Vizine-Goetz, Diane; Hickey, Carol; Houghton, Andrew and Thompson, Roger. 2004. Vocabulary mapping for terminology services. *Journal of Digital Information* Vol. 4, issue 4. Available from (http://jodi.esc.soton.ac.uk/articles/v04/i04/Vizine-Goetz/

Zeng, Marcia Lei ; Lois Mai Chan. 2004. Trends and Issue in Establishing Interoperability Among Knowledge Organization Systems. *Journal of the American Society for Information Science and Technology*, Vol. 55, no.5, 377-395