

IFLA Satellite Pre-Conference of the Classification and Indexing Section
 “Looking at the Past and Preparing for the Future”
 Florence, Italy, 20-21 August 2009



20 years SWD – German subject authority data prepared for the future

Yvonne Jahns

Deutsche Nationalbibliothek

Abstract

The German subject headings authority file - SWD - provides a terminologically controlled vocabulary, covering all fields of knowledge. The subject headings are determined by the German Rules for the Subject Catalogue. The authority file is produced and updated daily by participating libraries from around Germany, Austria and Switzerland. Over the last twenty years, it grew to an online-accessible database with about 550.000 headings. They are linked to other thesauri, also to French and English equivalents and with notations of the Dewey Decimal Classification. Thus, it allows multilingual access and searching in dispersed, heterogeneously indexed catalogues. The vocabulary is not only used for cataloguing library materials, but also web-resources and objects in archives and museums.

1. Introduction

As subject indexers we look mostly into the future: how could we enhance our indexing languages, how could we respond to requests for information when knowledge itself is constantly changing, how could we handle these millions of new publications, to make them quickly available to information seekers, how could we improve our indexing rules, finding automatic methods to save human resources, how could we find solutions for the semantic web etc. Librarians at the forefront of organizing knowledge – our last section satellite meeting on subject retrieval, held in Dublin (Ohio) in 2001 was conceived in this light.¹

Let us back-pedal and look at the past, and overlook the last 20 years - and we can realise, that we are well prepared for the future. German librarians created a capital indexing language. This vocabulary is called *Schlagwortnormdatei (SWD)*, meaning nothing else than subject headings authority file. As all of us know, controlled vocabularies are essential for information retrieval tools because of the complexity and the ambiguousness of natural language which our information or knowledge is expressed through.

¹ McIlwaine, I.C. et al : 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, München, Saur 2003

Kalebasse ist Kürbisflasche, also die Flasche, die aus einem Flaschenkürbis gemacht wurde bzw. aus der Frucht des Flaschenkürbisses, also der Kalebassenfrucht, Callabash, callabash, Lagenaria ...



http://upload.wikimedia.org/wikipedia/commons/thumb/a/a2/Pots_for_sale_in_Kenya.jpg

Getting subject access is a common day to day phenomenon for scientists, students, every user of search engines, online-catalogues, web-directories, even for people doing translations or creating instruction manuals. All of them need a special combination of linguistic and content information. In order to provide people with topics they want, we describe these with controlled subject headings.

For better understanding such artificial languages and of course encouraging international cooperation and data exchange, below the principles underlying our subject heading language (SHL) are described.

2. History and partners of the subject authority file

Our *SWD* authority file is a terminologically controlled vocabulary. We often say thesaurus, but strictly speaking, if it is one or not depends on the definition of thesauri like in *ISO code 2788*, if it is one or not. The word "thesaurus" is derived from 16th-century Latin, in turn from Latin *thesaurus*, from ancient Greek *thesauros*, meaning a collection of things which are of high importance or value.² Since such definitions are always a controversial issue (due to the lack of certain hierarchical conceptual relationships in our vocabulary) I will not expound this problem and will avoid the term thesaurus here. Undoubtedly it is a kind of knowledge repository.

SWD started as a list of about 120,000 standardised subject headings from some libraries – the German National Library (*DNB*) and a few Bavarian libraries - in the 1980s, resulting from their subject (card) cataloguing practise. At the same time librarians approved the first German Rules of the Subject Catalogue (abbreviated in German *RSWK*). There was a long tradition of verbal indexing in German libraries, but no common rules until this time. The beginning of IT in libraries in the 1970s allowed data exchange and required harmonized rules and of course authority files. Lists were no longer effective. In 1986 *DNB* started cataloguing by *RSWK* and *SWD* and others followed. Since then, *SWD* subject headings are determined by these subject indexing rules and guarantee consistency of indexing.

² <http://en.wikipedia.org/wiki/Thesaurus>

During this time, efforts in developing subject heading languages were made all over Europe. Interoperability projects and mergers of existing vocabularies were on the agenda.³

Today our rules are applied by a huge number of libraries, scientific and public, many of them participating in producing and maintaining subject headings. The German National Library cooperates with all major libraries and also library networks or service centres in Germany, Austria and Switzerland, among them special institutions like art museums' libraries or church libraries. Actually the participants and users do not only come from the library sector, but also from museums, archives or broadcast stations or others from the media sector.

The cataloguing rules are now accepted principles and the authority files are harmonized tools and information sources for subject retrieval. Simultaneously, our rules come under criticism not to be web-oriented enough. We do consider these arguments carefully, but one could certainly write another full paper on this issue.

3. The database

3.1. Content and structure of authority records

SWD covers all fields of knowledge, it is a universal vocabulary, very similar to *LCSH* or *RAMEAU*. There are now 550,000 headings and even more references providing an efficient information retrieval.

The file consists of different categories, the main entities are: geographical names, topical headings (concepts and objects), terms for time and form, names of institutions (corporate bodies), titles of works.

(Names of persons are stored in an extra authority file, *PND*, used also for cataloguing author's names.)

SWD 2009:

550.000 headings

160,000	topical headings (concepts), also events, time periods and forms
190,000	geographical names
130,000	names of corporations
70,000	titles of works

As you can see in the figure, the greater part of terms are individual terms - names of places, events or titles of works and so on. The general nature of the headings was changing from 1987 until today, now topical terms grow softer. In other words, we can keep a hybrid nature of our *SWD* file: on one hand there is a thesaurus of hierarchically

³ Landry, Patrice: The Evolution of Subject Heading Languages in Europe and their Impact on Subject Access Interoperability, in: New Perspectives on Subject Indexing and Classification – Essay in Honour of Mada Heiner-Freiling, Leipzig 2008, S. 249

structured terms as semantic units for conveying concepts and on the other hand a name authority file.

Subject headings can be single words or complex phrases, usually nouns, since nouns are the most concrete part of speech. Commonly known designations are preferred, usually names and terms in German. Encyclopaedias, dictionaries or other thesauri have to be consulted before creating a heading, to ascertain the most commonly used name or term. This source will be given as a reference. Unlike a dictionary, an authority record does not define words, and definitions can be found at the referenced sources.

An example of an authority record is shown below:

SWD	
link	http://d-nb.info/gnd/4165900-4
SWD-ID	4165900-4
Schlagwort (descriptor)	 s Kürbis
Quelle (source)	Lexikon der Biologie : allg. Biologie, Pflanzen und Tiere
Entitätencode (sub entity code)	snz
Synonyme (synonyms)	Cucurbita pepo Gartenkürbis
SWD-Systematik	24.3;32.4
DDC-Notationen	583.63 635.62 641.3562
Oberbegriff (broader term)	-
Unterbegriff (narrower term)	Ölkürbis Riesenkürbis
Verwandter Begriff (related term)	-
Theman in (subject in)	125 Publikationen (DNB catalogue)
	<ol style="list-style-type: none"> 1. <i>Kürbis & Co.</i> <i>Bath : Parragon, [2008]</i> 2. <i>Kürbis</i> <i>Augsburg : Weltbild, 2008</i> 3. <i>Kürbis</i> <i>München : Compact-Verl., 2008</i> 4. <i>Kürbis-Kochbuch</i> <i>Husum : Husum, 2008</i> 5. <i>Kürbis, Kraut & Co.</i> <i>München : Zabert Sandmann, 2008, 1. Aufl.</i> 6. ...

What are the possible see-references and relations between terms? These are the well-known categories of thesauri, divided into three types:

- equivalent,

- hierarchical,
- associative relations.

Synonyms, alternative spellings, variant forms of names, former names, broader and narrower terms, related terms, etc. are indicated as you can see in the example above.

We of course have a close look at the IFLA-Functional Requirements for Authority and Subject Authority Data and check how we can upgrade our vocabulary to guarantee international data exchange.

Besides the main entities we distinguish between 40 sub-entities, indicated by a special code, e.g. the letters *snz* indicating the sub entity for “biological or chemical nomenclature” or *sip* meaning the sub-entity “authority names for individual product names and brand marks”. Thus we can cluster the vocabulary.

Authority records are structured additionally by ISO country codes, ISO language codes and subject groups. This systematic allows us to view sector extracts or to get partial views of the vocabulary.

Subject headings are highly specific, according to the rule of finding always the narrowest term, but accurate and comprehensive as well for description of a topic. Compared to other well-known thesauri like *MeSH* or *AGROVOC*, *SWD* is well-developed and meets the needs of specialised institutions despite of being a universal vocabulary.

The vocabulary is highly scientific, too. The conceptual dynamics of all disciplines is represented.

Most of the topical headings belong to the disciplines of technology, biology, chemistry, medicine and economics, representing the enormous literary warrant in these fields, as for instance compared to fields such as library science or terms for handicraft works. Yet, we also have terms like knitting, papier-mâché or origami since German librarians index all kinds of documents with *SWD* and also the German National Library collects and catalogues a huge number of non-scientific publications.

3.2. Maintenance procedure and construction principles

Our vocabulary is not at all static. New publications need new terms to describe contents adequately. The database is updated daily. Beside the construction of new headings, existing terms have to be maintained. The use of habitual language has to be monitored. Constant updates represent the current development of research and science. We follow innovative terms and language usage just as the editorial staff of an encyclopaedia is watching the latest news. In my opinion, it is important that headings are created not only by people who are responsible for developing the vocabulary, but also by indexers, who apply the terms.

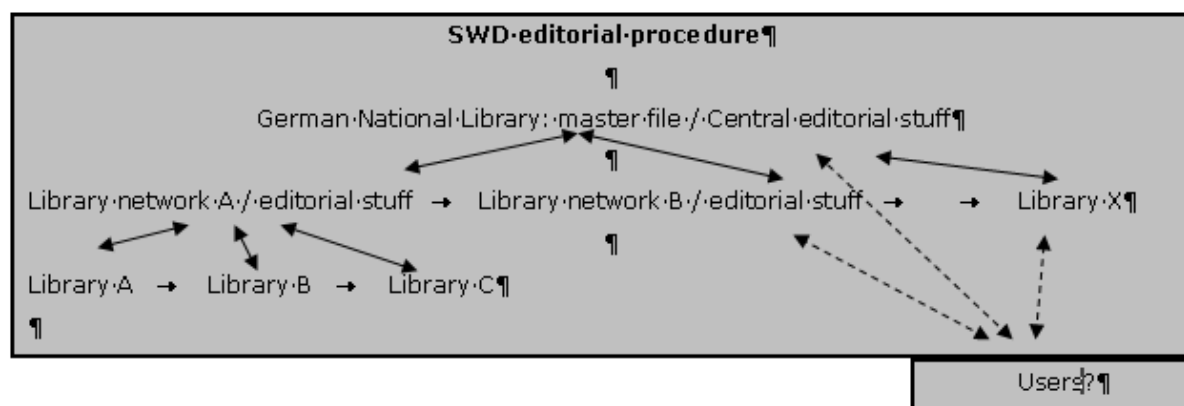
These are the major construction principles of our vocabulary: ⁴

⁴More details have already been published in the country report of the seminal IFLA-study “Principles Underlying Subject Heading Languages”, our section of classification & indexing did in the 1990s, see Lopes,

- Terminology is controlled. Uniform headings are built with synonyms. Sometimes definitions are formulated. Each term is placed in its context, allowing a user to distinguish between homonyms like "bureau" - the office and "bureau" - the furniture.
- Each heading is stored in an authority record with a (permanent and unique) ID number.
- Terms follow the literary warrant.
- Terms are audience (user) oriented.
- Complex and compound subjects are expressed by linking component parts of subject headings via syntagmatic relationships (syntax principle).

The National Library hosts a central authority database, maintenance is co-operative. All participants are able to send data, correct⁵ and communicate online.

Not only two or three editors create the controlled vocabulary on behalf of other cataloguers, but dozens of indexers and editors around the country create new records in their familiar disciplines, reflecting the development of current research. In which respect is this different from the *Wikipedia* principle? In fact, it is much the same procedure, creating high synergy by many of experts. Only a few editors make final revisions. In case there are discrepancies, all editors can send an email to the others, with a direct link to the authority record itself. A group of about ten editors from all participating consortia meet twice a year for personal discussion.



One of our future plans is to be more open for new heading entries generated by users themselves. If they fail in subject retrieval, they can send us (the editorial staff) a proposal.

We also learn from social tagging systems which terms users are searching for and which kinds of search behaviours exist.

Maria Inês / Beall, Julianne et al: Principles Underlying Subject Heading Languages (SHLs), München, Saur 1999.

⁵ Different editorial levels are set up, depending from editorial experience of the participant / completeness of the authority record regarding the cataloguing rules

In my view, linguistic components should be integrated into our catalogues or search engines to avoid “no results” due to the use of a plural search term, which in German are quite often non-trivial modifications of the noun (*Kindergarten / Kindergärten*). Integrating this dictionary feature directly into the authority file could be a solution saving the time to add this feature into every online catalogue. If we do so, we will also get better success using subject headings in automatic indexing processes.

3.3. Data formats

Standard communication formats for subject headings are the same as for bibliographic records:

- *MAB2*, the automated library exchange format in Germany.
- *MABxml*, using a XML structure, suitable for XML based technologies like OAI, SRW/U and XSLT.
- *Pica+*, basically as an internal format in some libraries.
- *MARC21 authority*, following the decisions of the Committee for Library Standards in December 2004 to internationalize the German standards uniform use of *MARC 21* as the exchange format for all libraries in Germany and Austria.⁶

Converting subject headings to appropriate semantic web formats like *SKOS* will be a great challenge. Expressing structure and content of our heading language with *RDF-Schema* or *OWL* allows the construction of machine-processable statements.⁷ Within the European project *TELplus*⁸ a first RDF-version of the *SWD* was built in 2008.

3.4. Access tools

SWD is available via a Z39.50 Gateway and also on DVD-ROM along with the Name Authority file (*PND*) and the Authority file of corporate bodies (*GKD*), furthermore on other storage media via National bibliographic services.

Headings are of course searchable via DNB online catalogue <https://portal.d-nb.de/>. There are plans to make our subject terminology available as a public domain data set under the license of Creative Commons <http://creativecommons.org/licenses/by-sa/3.0/deed.de>

There is also a structured display, that *DNB* created within a project, where one can find a comfortable navigation among the relations of the terms, <http://melvil.d-nb.de/swd>.

⁶ for *SWD-MARC* authority see http://www.d-nb.de/standardisierung/pdf/konkordanz_2.pdf

⁷ Svensson, Lars G.: Unified Access – A Semantic Web Based Model for Multilingual Navigation in Heterogeneous Data Sources, in: New Perspectives on Subject Indexing and Classification – Essay in Honour of Mada Heiner-Freiling, Leipzig, 2008, S. 241

⁸ <http://www.theeuropeanlibrary.org/telplus/>

s Bildband	
IDN	041453956
UF	s Abbildungen <Formschlagwort>
UF	s Ansichtspostkarte <Formschlagwort>
UF	s Photographien <Bildband>
UF	s Bildbände
UF	s Bildatlas
UF	s Bilddokumente
UF	s Bilder <Formschlagwort>
UF	s Bildmaterial <Formschlagwort>
UF	s Ansichten <Formschlagwort>
UF	s Bildsammlung
UF	s Abbildungswerk
UF	s Illustrationswerk
BT	s Buch
NT	s Atlas
NT	s Karikatur
RT	 s Bildliche Darstellung

•	OB3 s Schrifttum ▼▶
○	OB2 s Druckwerk ▼▶
▪	OB1 s Buch
▪	 s Altes Buch ▼
▪	 s Ausländisches Buch
▪	 s Beschäftigungsbuch ▲
▪	 s Bestseller
▪	 s Besucherbuch ▶
▪	 s Bibliophiles Buch ▼▶
▪	 s Bildband ▶
○	UB1 s Atlas

Other library networks offer a similar web access, e.g. the south-west network offers it free searches under <http://swb.bsz-bw.de/DB=2.104/>

	[Sachschlagwort]
Schlagwort:	Welp Suche nach hierarchisch untergeordneten Begriffen?
PPN:	213144417
SWD-Nummer:	4507272-3
Quelle:	B 1996
Erläuterung:	Verknüpfte z.B. mit Hundehaltung bzw. einz. Hunderasse, mit Fuchs oder Wolf
SWD-Systematiknummer:	32.6
DDC-Notation:	636.707
Synonym:	Hund / Jungtiere Junger Hund

4. Use and applications

So what is it good for? There are four main traditional applications of *SWD*:

- having a source vocabulary for subject indexing
- ensuring consistency of cataloguing among various library networks
- having topical search keys (access points)
- a navigation tool for information retrieval in OPACs.

The first aims we have already touched. Adding subject headings one-by-one to a string means building a kind of a very short abstract of the contents of publications. The German national library does this for nearly all new titles of the publishers' book trade, following the rules of syntactic indexing of *RSWK*.

In contrast to the Anglo-American cataloguing practise, German bibliographic records do not contain subject headings as text strings, but links to authority records via ID numbers. If an authority record has been changed, title records will be updated automatically. Also in this way, searching synonyms is always possible. As a contrast, German subject headings are not as rigorously pre-combined as in the *LCSH*, therefore words can be freely combined. *SWD* is a multi-faceted, post-coordinated vocabulary. Searching such single-concept terms is very effective. *SWD* contains only the source vocabulary. Cataloguers can select terms, combine them and form heading strings as they wish.

Link zu diesem Datensatz	http://d-nb.info/984075259
Titel	Lava und ich : ein Welp, ein Soldat und ein Stück Menschlichkeit im Irak-Krieg / Jay Kopelman ; Melinda Roth. Aus dem Amerikan. von Sigrid Eicher
Personen	Kopelman, Jay ; Roth, Melinda
Ausgabe	1. Aufl.
Verleger	Stuttgart : Kosmos
Erscheinungsjahr	2007
Umfang/Format	189 S. ; 22 cm
Einheitssachtitel	From Baghdad, with love <dt.>
ISBN	978-3-440-11099-7 3-440-11099-0
Einband/Preis	Pp. : EUR 16.95, sfr 31.30 Pp. : EUR 16.95, sfr 31.30
Schlagwörter	USA ; Soldat ; Irakkrieg ; Welp ; Geschichte 2004-2005 ; Erlebnisbericht
DDC-Notation	956.70443092 [DDC22ger]
Sachgruppe	950 Geschichte Asiens

Single subject headings or single words are searchable as well as their synonyms. A few German library online catalogues also allow searching and browsing of the whole subject headings phrases.

Guaranteeing consistency of the cataloguing practise means minimizing the cataloguing efforts and costs for every single library. Using the same source vocabulary allows network cataloguing and most efficient data exchange.

Apart from indexing library holdings, e.g. books, newspapers, electronic resources, DVD, microforms, websites, *SWD* is used more and more for describing museums' exhibits or photographs.

We will now have a short look at the other goal – to navigate in online catalogues via subject headings. As we have seen in the beginning, the main goal of our indexing work is to help information seekers to find the relevant information. Most library catalogues today integrate subject headings into the basic search index, in order to enhance the search vocabulary. The system automatically searches for all synonyms of the users' search term to find more hits. Beyond this, only a few online catalogues offer navigation via term structures, which means clicking on word nets or clouds, asking for associative terms (and documents).⁹ This is a great challenge for the design of user interfaces. Libraries should agree on standards also in this area. Investments in subject authorities may not remain an end in itself. User orientation should be maximized and we should look for new authority applications. There is for example a whole market in the field of business knowledge organization.

Generally, authority records, particularly name authorities could not only be applied in the way of traditional cataloguing as we did the last twenty years. New databases, digitization and open access to data networks and all the possibilities of the web produced exponentially increasing data amounts. Such a volume and growth of digital data makes subject access, retrieval precision and selection more necessarily then ever before. As demonstrated by several current projects, authorities offer the right infrastructure. They allow linking information and resources, using the same documentation vocabulary and the same identifier for names and topics. This is the advantage of our documentary languages over today's indexing realised in search engines.

Networked authority data could be much more profitable, if the vocabularies provided worldwide would be mapped.

5. Mappings and links

Most libraries have resources indexed by different vocabularies and classification schemes. Therefore query links for searching heterogeneously indexed collections are needed. High-quality (precise) results of these crosswalks are based on (manual or automatic) mappings between vocabularies. Much research was done in recent years in

⁹ Queens Library does it with word clouds showing related terms, spelling variations and translations, <http://aqua.queenslibrary.org/>

this field. Web services or so-called terminology services provide data to facilitate subject interoperability.¹⁰

The vision of a virtual international authority file (VIAF) was born in the *VIAF* project of *OCLC*, the Library of Congress, Bibliothèque nationale de France and *DNB*, started in 2003¹¹. A first prototype was established with names of persons, but combined authority data now include also geographic entities. This extension matches various names of both governments and natural geographic units as rivers, mountains etc.¹²

5.1. Crosswalks to other thesauri

German librarians and documentalists were very creative in building subject headings languages.¹³ Often invented only for local application, they wanted to meet special user's needs. There are different controlled vocabularies for different disciplines. That is why mapping projects started to support cross searching, mushrooming simultaneously with all the virtual libraries which have been founded the last ten years.¹⁴ *SWD* was part of several mapping initiatives, last in the project *KoMoHe*¹⁵. There cross-concordances were produced, meaning manually created links that determine equivalence, hierarchy, and associative relations between terms from *SWD* and other controlled vocabulary. Favoured expressions of any documentary language can be related to those of another either by congruence, 1:1 equivalence or 1:many equivalence, which includes "AND" or "OR" relationships.

SHL 1 soziale Entwicklung = *SHL 2* Gesellschaft „AND“ Entwicklung

Vocabularies were analyzed in terms of topical and syntactical overlap before the mapping started. All mappings were created by researchers or terminology experts. Essential for a successful mapping is an understanding of the meaning and semantics of the terms and the internal relations within the vocabularies of concern.

The vocabularies mapped to *SWD* or those *SWD* was mapped to are the following:

- *AGROVOC Thesaurus (AGROVOC)*: A multilingual vocabulary in the agricultural domain (39,000 terms).
- *INFODATA Thesaurus (INFODATA)*: A vocabulary in the information science domain (1,000 terms).

¹⁰ A good overview is given by Nichol森, Dennis: A Common Research and Development Agenda for Subject Interoperability Services? <http://pro.tsv.fi/stks/signum/200805/4.pdf>

¹¹ <http://www.oclc.org/research/projects/viaf/>

¹² Hengel, Christel: The Virtual International Authority File (VIAF), in: New Perspectives of Subject Indexing and Classification – Essays in Honour of Magda Heiner-Freiling, Leipzig 2008, S.269

¹³ Kunz, Martin: Subject retrieval in distributed resources: a short review of recent developments <http://www.ifla.org.sg/TV/ifla68/papers/007-122e.pdf>

¹⁴ 64 crosswalks with more than 500,000 relations were established alone funded by the German Ministry for Education and Research.

¹⁵ http://www.gesis.org/en/research/information_technology/komohe.htm

- *Medical Subject Headings (MeSH)*: A vocabulary in the medicine domain (23,000 terms).
- *Psyn dex Terms (PSYNDEX)*: A vocabulary in the psychological domain (5,400 terms).
- *Standard Thesaurus Wirtschaft (STW)*: A vocabulary in the economics domain (5,700 terms).
- *Thesaurus Bildung (BILDUNG)*: A vocabulary in the pedagogic domain (50,000 terms).
- *Thesaurus Sozialwissenschaften (THESOZ)*: A vocabulary in the social science domain (7,700 terms).

Although the usefulness of terminology mappings is generally acknowledged, yet the actual effectiveness is so far rarely evaluated. In an information portal with many different databases, the question becomes crucial whether cross-concordances can enable a distributed search. A few studies have shown that they work successfully in bridging language gaps.¹⁶

DNB wants to use all these existing mappings to enhance *SWD* search vocabulary, to provide seamless searches across *SWD*-indexed catalogues and other databases in the future. One benefit of such mappings that should not be underestimated is quality control, resulting systematic inspection of the vocabulary.

5.2. Multilingual linkage (LCSH, Rameau)

As we have seen above, the corresponding English and French authority files are the Library of Congress Subject Headings (*LCSH*) and the Répertoire d'autorité-matière encyclopédique et alphabétique unifié (*RAMEAU*), respectively. To overcome linguistic barriers and to finally provide a truly multilingual search, a project called *MACS* (Multilingual Access to Subjects) was started in 1998, initiated by *CENL*. A working group from four libraries was set up, the British Library, the Bibliothèque Nationale de France, *DNB* and the National Library of Switzerland. Equivalence links are established between the three subject headings languages, *LCSH*, *RAMEAU* and *SWD*. The first study took a subset of headings from the three authorities in the fields of Sports and Theatre.¹⁷ Different mapping methods were explored and a prototype of a linking tool was developed. This database called *LMI* (Link Management Interface) still exists. It is a web application, independent of any partner library system.

¹⁶ Mayr, Philip / Petras, Vivian: Cross-concordances: terminology mapping and its effectiveness for information retrieval http://www.ifla.org.sg/IV/ifla74/papers/129-Mayr_Petras-en.pdf

¹⁷ Clavel-Merrin, Genevieve: The need for co-operation in creating and maintaining multilingual subject authority files <http://www.ifla.queenslibrary.org/IV/ifla65/papers/080-155e.htm>

LCSH	RAMEAU	SWD
Women	Femmes	Frau
Women – Employment	Femmes – Travail	Frau “AND” Berufstätigkeit
Women authors	Femmes écrivains	Schriftstellerin

Because all the tests were successful, the *MACS* organization was refined and linking mechanisms were improved. The Swiss and German National Libraries want to finish the linking work for the 50,000 most-used headings in 2009, which is done manually by subject experts. Then, we will finally be able to share our resources with other libraries, using *LCSH*, *RAMEAU* or *SWD* and moreover to improve access to our collections in the language of the user’s choice. Furthermore, users will be able to search several catalogues simultaneously. Queries will be translated into the other subject headings. This is particularly important for European search environments like *TEL*¹⁸.

5.3. Linking subject headings with DDC

Mapping subject headings to notations of the *DDC* is the aim of the *CrissCross* project, a cooperation of the German National Library and the Cologne University of Applied Sciences.¹⁹ In doing so, it faces the challenge of finding a crosswalk between two different knowledge organization systems. *SWD* works as initial vocabulary, *DDC* is the target language. Appropriate Dewey numbers are added directly to the *SWD* records.

descriptor	Kürbis	
synonyms	Cucurbita pepo Gartenkürbis	
DDC notations	583.63	(Botanik - Cucurbitales)
	635.62	(Gartengemüse - Pâtissonkürbisse und Kürbisse)
	641.3562	(Lebensmittel – Zucchini,...)
	641.6562	(Kochen mit Kürbissen)

Mappings are carried out as specifically as possible. Due the polysemy of many subject headings, one can be assigned to several Dewey classes. The degree of correspondence or determination between a term and a class is specified by a number from 4 to 1. This measure of determinacy shall be used for ranking mechanisms in future retrieval. The project is underway and nearly 70,000 subject headings are mapped to the classification so far. With this, we have an additional structural element within the *SWD* as well as an enhanced access vocabulary for the *DDC*. Both can be used repeatedly in retrieval scenarios.

¹⁸ <http://search.theeuropeanlibrary.org/portal/de/index.html>

¹⁹ <http://linux2.fbi.fh-koeln.de/crisscross/index.html>

5.4. Cooperation with online-encyclopaedias

In cases where authority files are effective, provided that they not only arrange controlled terms, but are broadly accepted, links to / from other online sources can expand their area of application.

We started referencing name authorities to Wikipedia articles and we want to go on with this for topical terms. Thus, indexers and users can look up definitions of terms and *Wikipedia* users receive a list of related publications (via our library catalogue). For the same purpose *DNB* started cooperating with Germany's most important encyclopaedia and *SWD* framework *Brockhaus*²⁰, but as you may know, the publisher went out of business and their online encyclopaedia will not be maintained anymore. We have to face a general crisis of printed encyclopaedias, raising challenges also for future authority work.

6. Future plans - a common German authority file for persons, institutions and subjects: GND

Beside *SWD* there exist two other major German authority files, one for names of persons, *Personennormdatei (PND)*, and one for names of institutions / corporate bodies, *Gemeinsame Körperschaftsdatei (GKD)*. The first one is used for alphabetic and subject cataloguing as well, that is, for describing whether a publication is from a or about a person. The second one is mostly used for formal descriptions, in case a publication is published by an institution. *GKD* was the first authority file project, going back to the 1970s, when computerized cataloguing and also *RAK* started. *PND* and *GKD* follow the rules of alphabetic cataloguing, therefore rules for names differ from the *SWD*. Turning to *RDA* and a closer cooperation of alphabetic and subject cataloguing, our aim is to bring together heading construction rules and merge our authority files.

This merged file will be called *Gemeinsame Normdatei (GND)*, common authority file, and will hopefully be completed within the next years. At the moment we create or adopt a data format closely related to *MARC21 authority*.

7. Closing

Organizing knowledge retrieval in a controlled way is a multidisciplinary task requiring more than building and applying a controlled vocabulary. It is an exciting task for information professionals, even more so in today's web retrieval, where people can look up information from any place in the world. Overcoming language barriers will be one of the most important tasks, especially for our European initiatives like the search engine *Europeana* or gateways like *TEL*.

Authority data can used today either in standardised library cataloguing or indexing databases or websites. Libraries should realise the new potential of linking multimedia

²⁰ <http://www.brockhaus.de/>

web resources of different providers, bringing together books, films and other documentations.

In spite of its advantages, authority work remains to be an expensive and coordination-intensive work, but there is no doubt that the benefits outweigh all efforts.