

Classification Issues: Report

IFLA General Conference, Oslo, Norway, 2005

Classification and Indexing Section

Nancy J. Williamson

During two meetings of the Section in Oslo, reports were presented on five projects and working groups. The 'final draft' of "Guidelines for the Construction of Multilingual Thesauri" had been posted on IFLANET for world wide review and "many positive comments and suggestions had been received". Based on the comments McIlwaine would prepare the final English version in the Fall of 2005 and publication is expected in early 2006. A Working Group on a Virtual Clearing House of Available Subject Access Resources had been created during the Berlin Conference in 2003. Work has been progressing steadily and a sample of about 20 subject access tools has been prepared for testing. In 2005/06 the work will focus on identifying subject access to resources which are used either by national libraries or used at the national level in such tools as national union catalogues, or national bibliographies but not maintained by a national library. In a second phase an effort will be made to include other important resources. Subgroups of the Working Group have been formed to deal with specific issues and with regions of the world. A Working Group on Guidelines for Subject Access by National Bibliographic Agencies was established at Berlin in 2003 and re-launched at the Oslo Conference to be chaired by the newly elected chair of the Section, Patrice Landry. The tasks of this Working Group are to "collect and study the indexing policies and practices of the national bibliographic agencies, to analyze the data, and draft guidelines." The plan is to issue the guidelines in 2007. The Section is also taking a leading role in a new Division Working Group on the Functional Requirements for Subject Authority Records" (FRSAR). Its terms of reference are "to build a conceptual model of Group 3 entities within the FRBR framework as they relate to the aboutness of works; to provide a clearly defined structured frame of reference for relating the data that are recorded in subject authority records; to meet the needs of the users of those records; and to assist in an assessment of the potential

international sharing and use of subject authority data both within the library sector and beyond."

Finally there is a Cross-Section Working Group on a Multilingual Dictionary of Cataloguing Terms and Concepts which was struck in 2002. Two members of the Standing Committee will coordinate the Sections' involvement in it.

During the Oslo Conference, the Section participated in a half-day session on 'Classification and Indexing with Cataloguing'. Three of the papers presented were on various aspects of the subject approach to information. All three were international in scope and are available on the Internet from the Oslo Conference Programme at <http://www.ifla.org/IV/ifla71.programme.html>.

Marie Balikova of the National Library of the Czech Republic presented a paper on "Multilingual Subject Access to Catalogues of National Libraries (MSAC)." MSAC is an online subject access system being developed to facilitate collaborative catalogue access between the National Library of the Czech Republic and the national libraries of six other countries – Slovakia, Slovenia, Croatia, Macedonia, Lithuania and Latvia. The project is in its early stages and is intended to provide users of online catalogues and internet services of these institutions with an indexing and retrieval tool enabling them to enter a query in their own language and retrieve from all the catalogues.

Since the existing catalogues had been indexed with different vocabularies and indexing tools, the project is complex. It was necessary to deal with a number of essential factors including the standardization of tools, browsing and searching methods, and the interoperability problems of library catalogues and digital collections of heterogeneous institutions. The goal is a "one-stop, seamless" searching apparatus. Two approaches were suggested – mapping and the use of a switching language. Authority control is considered essential for the sake of consistency and the multilingual system must meet the

needs of users who might have different requirements; for example to search multilingual collections, using queries in one language, or to locate documents in a number of languages, via an interface in the language of their choice. Translation might be required or a summary in a language other than the language of the documents. This would require sophisticated technology, staff with special language skills and plenty of money. Several verbal subject systems were investigated and two were selected for a possible merger and the required cooperation. Analyses of the subject analysis process in the online environment were carried out on both verbal terms and classification systems. This produced a list of desirable features for such a system. As a result of the investigation, the *Universal Decimal Classification* (*UDC*) was selected as the only feasible system with the necessary features to serve as a common indexing tool. The article cites the merits of *UDC* as its international use, meaningful notation, ability to deal with facets and suitability as a possible mapping device for the controlled terms of the participating languages. *UDC* was also cited as having advantages through its numerical notation, language independent coding and its flexibility and support of detailed expressions in both English and Czech, *UDC* allows for adaptation of citation order to meet local requirements. However in this international environment context, consistency of application would be essential. The process of mapping between Czech verbal expressions and *UDC* numbers is described and the process is being carried out intellectually. Eight special language indexes will emerge.

MSAC is being developed in two phases. The first phase includes "the development of [a] Czech topical authority file and the integration of language variants of the participating libraries into the Czech subject authority file." In the second phase, "the combinations 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." This could be automatic but intellectual checking is assumed. Access is to be provided via Z39.50 protocol. Libraries not supported by this protocol will be able to join the system with some adjustments.

With reference to the future, the authors note the fact that "the Internet European multilingual community uses more than 30 languages with different repertoires and encodings." MSAC has implications for the creation of a multilingual subject retrieval

tool to introduce a mapping scheme into existing systems – a tool that is considered to be an essential element of the European Library Service. They also note that MSAC at its very beginning "is dependent on voluntary work of teams from the participating libraries" and communication is almost entirely dependent on e-mail. Up to the time of writing, there had been no external financial support for the project. A new prospect for the future development of the MSAC project is probable through joining the TEL-ME-MOR project (The European Library Modular Extensions for Mediating Online Resources). It is funded by the European Commission under its Information Society Technologies (IST) Programme and the ten new member states of the European Union have been invited to join.

In a paper entitled "The Répertoire de vedettes-matière and RAMEAU: Two Indexing Languages in French: A Necessary Luxury?" Jo-Anne Bélair, Université Laval, and Françoise Bourdon and Michel Mingam, Bibliothèque nationale de France, demonstrated the necessity of having two indexing languages in the same language – French. The needs of the users are linguistic and cultural on the one hand and administrative and managerial on the other. The two indexing languages exist in an atmosphere of collaboration supported by an official agreement and day to day cooperation.

In a brief history of each language, the authors explain that they come from different origins and have developed in different ways. The RVM (Canada) derives from the *Library of Congress Subject Headings* (*LCSH*). Translated, beginning in 1946 and first published in 1962 it exists in three formats – Web, CD-ROM and a downloadable format. It has been a work in progress, in 1994 having obtained permission to use "as English equivalents to its subject headings, the terms contained in" the *Medical Subject Heading* (*MeSH*) and the *Art & Architecture Thesaurus* (*AAT*). RVM has also begun to include Library of Congress classification numbers. It began as an in-house tool but its application has expanded. In 1974 the National Library of Canada adopted RVM as the national standard for subject indexing in French. In the same year it was adopted by the Bibliothèque publique d'information du Centre Georges-Pompidou (Paris), and other institutions followed. Today it is applied in all types of French language libraries in Québec, as well as in a number of institutions in Europe.

In the early 1980s the National Library of France and the French university libraries adopted a unique

list of subject headings which would become known as the *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (Rameau). It was inspired by the Laval RVM and originates from it. Rameau language develops from user demand both internal and external to the Bibliothèque nationale de France. Its national expansion has been accompanied by the development of an institutional infrastructure provided by the Centre national Rameau. The Centre includes a user and partner network organization and supplies products to a number of lists, journals and guides. A Rameau website (<http://rameau.bnf.fr>) was established in 2000, whereby users can make direct suggestions for creating and modifying headings, and training aids are offered. Rameau has been adopted in other countries including Poland and Romania.

In explaining the relationship between the two indexing languages, the authors acknowledge that while the languages have common roots they have diverged somewhat over the years. Both are based on natural language and have developed a "local or national dimension – a trait that is unavoidable." French language in Quebec is not exactly the same as the French language in France and vice versa. Also the cultures are different. Thus the way patrons search library catalogues is different. It follows that there is a requirement to bend in the direction of the particular library clientele. Hence, there is a legitimate reason for the existence of the two lists. There are also administrative and managerial differences. RVM evolves in a bilingual context and uses catalogue copy prepared by other institutions and machine translates *LCSH* headings into French whereas Rameau used *LCSH* as one source among others and diverges "not only from the American headings themselves but from certain semantic fields, if they are too different from the French fields." The authors state that the need to be autonomous does not mean that the two indexing languages ignore each other. The agreement uniting the two was renewed in 2004. It provides a framework for cooperation while still respecting the autonomy of each. Also, it provides for reciprocal citations in each language's headings and products. There is reciprocal exchange of products. The access to data, and the sharing of thoughts and experiences are beneficial to both. In the day to day work there is "systematic consultation" of both databases whence changes and modifications are being made. The "exchange of thoughts about the best way to translate into French new scientific concepts for which there are no French language sources" and "improvement of language on ei-

ther side of the Atlantic" benefits both partners. The authors feel that the existence of two somewhat different tools in the same language "is a necessary luxury" and a model of "true international collaboration, respectful of local requirements."

The third paper by Magda Heiner-Freiling, of the Deutsche Bibliothek and Patrice Landry of the Swiss National Library focuses on "The Use of the *Dewey Decimal Classification (DDC)* for the Organization of the National Bibliographies: Switzerland and Germany/Austria." It discusses the reasons that led to the use of *DDC* for this purpose, and deals with the difficulties of adopting a common approach using the Second Dewey Summary (100 divisions). It also describes the challenges for the three libraries of incorporating German requirements into the *Deutsche Nationalbibliografie*. Their concerns have been brought about by the introduction of national bibliographies to the web environment and the need to "undertake studies on how to bring electronic documents under bibliographic control." The issues were intensified by the need to arrange national bibliographies in a "user-friendly" way and to ensure efficient subject access to records. In addition there had been "a long-standing IFLA and UNESCO recommendation to arrange bibliography according to an international classification scheme." These events provided the background for the decision by the three libraries to work together to produce a common *Dewey Decimal Classification (DDC)* arrangement for the national bibliographies of Austria, Germany and Switzerland.

The position of each library is described. The Swiss National Library undertook a major reorganization of its operations and structures which included a study to select a classification scheme for its collection – a study which also looked at the national bibliography. For the Swiss national bibliography, this evaluation led to the replacement of an indexing system based on UDC with the Dewey Second Summary (100 Divisions) with modifications for Swiss history and geography. In 2004 Die Deutsche Bibliothek (DDB) decided "to use a new scheme based on *DDC*" for the new *Deutsche Nationalbibliografie (DNB)* and the Austrian national bibliography followed suit in the same year. These moves can be attributed to two factors. The first was the impending publication of a German translation of *DDC*. Secondly, since 1982 the three libraries had been using two different systems, a situation that placed constraints on the exchange of data for use by the bibliographies. There was growing international acceptance of *DDC* and the change to *DDC* would im-

prove bibliographic services and facilitate cooperation with foreign national libraries.

Originally, it was decided that the Dewey Second Summary (100 Divisions) would be a suitable level of access for the bibliography and that the standard *DDC* organization should be used. In many cases there was a good match between *DDC* and the former subject categories, in all fields of science and in subjects where literary warrant was high. However this was not suitable throughout. Among the areas cited as problems are some of the places where *UDC* and *DDC* differ (e.g. the 800s). With respect to music, a deeper level of division than the Dewey outline was needed (specifically in 782 and 784-788 sections on printed and recorded music). In 2002 and 2003 an analysis of the distribution of documents with the 100 Divisions was conducted. Documents were not spread evenly in all classes and while some classes could be merged, others could benefit for the use of more specific numbers. Also, there were discipline based problems, for example in the melding of all branches of engineering.

As well, there was a lack of correspondence with modern disciplines such as engineering, electronics, mechanical engineering and environmental technology which had been separate subject categories prior to 2004. Another troublesome discipline was archaeology. Moreover the Second Summary level introduced in 2004 proved to be too broad and would not satisfy users' bibliographic needs. In some cases, in classes containing a variety of topics, subtopics were invisible (e.g. 790 Sports, games and entertainment). As a result, beginning in 2006 the *DNB* will use full *DDC* notations.

However, the problem would not be insurmountable. The authors concluded that

- the cooperation was an interesting and rewarding experience. Ultimately each national
- library could adjust and adapt the classification for its own needs and the work done was a good basis for further developments. Further, it strengthened support for the German version of *DDC* to be published in 2006.

Universal Decimal Classification

Extensions and Corrections to the UDC (E&C) no. 27 was published in November 2005. As the Editor in Chief mentions in her report: "this year has seen many changes in work relating to the *UDC* and next year promises even more." The business office has

been reopened in The Hague under the capable direction of Dr. Perry Moree. Also, and the financial footing of the Consortium has "improved greatly." A number of changes in personnel have taken place or will do so shortly. The Chair of the Consortium, Alan Stevens has stepped down and Geoffrey Robinson a long-term and extremely hard working, knowledgeable and devoted member of the Editorial Advisory Committee has completed his work with this issue of *E&C*. A number of other changes have taken place on the Advisory Committee – serving members have departed and new members added. The Editor in Chief herself is planning to step down sometime in 2006 and a new editor will be chosen.

Professor Ia McIlwaine has been Editor in Chief since 1993 and her 13 years in that position have seen great changes in *UDC* and a significant move forward to becoming a classification suited to the 21st century. Professor McIlwaine has been willing to be experimental and has modernized *UDC* in many ways, bringing to bear on it great diligence and a sound knowledge of classification theory. A scan of *E&C* over those 13 years provides a story in itself. While it is not possible to discuss all of the many changes here, it is important to mention some of the most significant ones. In general numerous classes have been overhauled and updated. Perhaps the work most symbolic of the direction in which *UDC* would like to go new table for Religion which, now fully-faceted, has removed the problem of bias. Bias in religion is a major weakness still present in other classification schemes. The common auxiliaries have been cleaned up and strengthened and the new common auxiliary table for "relations, processes and operations" is a valuable addition. An experimental project is the proposal for class 61 Medicine still in progress. Another major contribution is the extended place table published as a supplement to *E&C* 27. This table is probably the most extensive geographic table in existence and has the potential to be adopted by other schemes. During Professor McIlwaine's tenure of office the *UDC* has become available online providing access to the whole classification. It is difficult to sum up the entire contribution of this Editor in Chief, but it is safe to say that she has made the *UDC* come alive in a way that it did not exist before. In doing so, she has been cognizant of the needs of *UDC* users and has laid a solid foundation for *UDC* development and revision that should stand it in good stead for years to come. Although the work of revision will never be finished, she is leaving *UDC* in an excellent position to con-

tinue to develop and serve users in the future. Any new editor will find Professor McIlwaine's record to be a hard act to follow.

In addition to the reports of the Chair of the Consortium and the Editor in Chief, *E&C* 27 includes a number of articles and some revised tables. Articles from Hungary and Poland describe *UDC* activities in those countries. Hungary is about to publish a new printed edition and a version of the database will be published in a Hungarian-English version bilingual CD. In Poland *UDC* has had long history and will celebrate the 100th anniversary of its implementation in that country in 2007. Its use dominated into the 1990's when computer technology introduce subject headings and keywords into the electronic systems. Despite the inability to use *UDC* in their computerized library systems, it continued to be used and they began to search for new ways to access *UDC* class numbers. Today the Polish libraries find that *UDC* "has a chance to become the second information language used in electronic library catalogues beside subject headings." A review of OPACs shows that 30% of libraries offer *UDC* as an alternative to subject headings in searching. *UDC* is also used as a method of browsing databases of reference resources in the National Library. In some libraries it is the preferred language of search. *UDC* currently has three types of uses in Polish libraries – to connect *UDC* codes and subject headings, so that alternatives can be used in searching; to connect *UDC* with a thesaurus permitting access to broader, narrower and related terms; and to develop *UDC* authority records including in them word equivalents and reference links. The *UDC* has survived along with subject headings and is the sole system in libraries not yet automated. There are plans to publish "new, revised" *UDC* schedules" in 2006 for use from the beginning of 2007. One of the

most important inclusions in this issue is the "Guidelines for the revision and maintenance of the *UDC*." These guidelines are "concerned with the *construction* of the classification not with its application." They are intended to provide guidance to persons involved in the revision of the *UDC* and its Master Reference File (MRF). An added touch is a page of photographs of some of the participants in discussions of *UDC* at the IFLA Annual Conference in OSLO in August 2005.

The major reason for the annual publication of *E&C* is to announce the changes and proposed changes to the *UDC* tables. Included in this issue is one set of proposals related to the ongoing experiment with class 61 Medicine. Proposed schedules for the Glandular, Endocrine and Respiratory systems are given. The first phase of this project will probably come to an end in 2006 with the publication schedules for the Digestive and Reproductive systems. Following that a reassessment and revision of the proposal will be carried out to upgrade and rationalize differences resulting from the development of the tables over a period many years. Finally, included here are "Revised *UDC* Tables." completed in the past year. These fully approved tables include additions and changes to the common auxiliary table 1e Common Auxillaries of Place, the largest sections of which deal with "(497.7) Macedonia (1991-). Former Yugoslav Republic of Macedonia" and "(54) States of the Indian subcontinent." Also included are major revisions for "33 Economics. Economic science," "64 Home economics. Domestic science. Housekeeping," "684 Furniture and allied industries," "94 History of the British Isles" and, in a separate volume, the Extended Place Tables. As always the Editor in Chief invites comments on proposals and revised tables to enable her to respond the needs of users.