

# Using MARC classification format for UDC and mappings to other KO systems for an enriched authority file

*Rosa San Segundo*

*Carlos III University of Madrid, Spain*

**ABSTRACT:** The USMARC classification format, developed in the early 1990s for the DDC and LCC systems, is also amenable for other classification systems. This paper presents a proposal for using the MARC classification format for UDC. There are advantages in using this format for the UDC data in an authority file, e.g., for the MRF records and records for combined notations as well. There has been a trend in library catalogues for subject interoperability between traditional classification systems such as the UDC, DDC, LCC and subject headings. An example with great impact is WebDewey, which offers interlinking between classification numbers, the alphabetical index of the tables and LCSH. Another example is the electronic version of LCC Plus, also including links to LCSH. Subject gateways built upon library authority files can support the interoperability between classification systems and subject headings. These gateways can be the backbone of a more universal access through hypertextual navigation structures supported by classification systems including UDC. To our knowledge, the MARC classification format has not yet been applied to the UDC and in this paper we are going to propose a solution supported by some examples.

**KEYWORDS:** Classification format, MARC

## 1. Introduction to MARC classification format application

The difficulty in information retrieval by means of systematic classifications represents a great limitation in current library systems, since they are unable to adequately integrate classification-related data. Their difficulty in performing searches of the elements constituting a notation or a compound classification number is due to the fact that they are not easily processable, as current library catalogues are not designed to contain data in such a complex form. The use of MARC classification format enables a chain of symbols of a notation to be recorded in a MARC classification field and thus to enable retrieval of elements.

MARC formats allow online representation and communication of bibliographical data. They have also been created for classification-related authority records. MARC authorities are valid in bibliographical records and catalogues for authority records, whether they are proper names, subject headings or others.

The Library of Congress began developing MARC authority formats in order to manage authority control in its catalogues. MARC links classification numbers with the subject headings assigned in the authority file, and with the alphabetical terminology contained in the tables of the classification systems in use. In addition, the updating, maintenance and modification of online classification systems are articulated in a more straightforward way.

The USMARC classification format was developed between 1988 and 1990 in the USA. It was conceived to host the DDC and LCC systems, although its creators also had in mind that it could

---

be applied to other classification systems such as UDC. Following the USMARC example, IFLA UNIMARC Committee set up a working group to develop a classification format compatible with UNIMARC and UDC was included in its plans.

The application of MARC classification format offers the possibility of improving retrieval by the UDC, as the notation can be broken down in authority records. Currently, such a retrieval cannot be fully undertaken in online library catalogues, as the distinction cannot be made between a main number and a subordinated one, and the different auxiliary numbers cannot be differentiated, which is why searches cannot be performed using the elements constituting a notation. MARC has not been applied yet to UDC notations.

## **2. The MARC classification format as a Classification Authority File**

The application of MARC classification enables information retrieval via the notations of a classification system. It would operate to build a Classification Authorities File: the classification data would provide the authority for classification records and may be retrieved in a broken down way similar to other MARC authorities formats. Once it has been introduced into an authority, a UDC notation would operate just like any other authority file. Hence, its recovery would mean a better performance of UDC, since in OPACs it is essential to have a file with systematic classification linked with a proper classification authority file, as well as a friendly interface enabling visualisation and navigation through hierarchies and facilitating a search tool (pre- and post-coordinated) using both classification numbers and natural language terms.

The use of classification formats would provide an improvement in classification tables maintenance and in authority control management; great help in classification tasks; access and use of different classification systems or subject headings, facilitating interoperability in record authorities and including the application of expert technology in catalogues; articulation of an online topographic catalogue; possibilities for the creation of other classification system tools, such as multilingual indexes for the classification system; and improvement in access through subjects in online catalogues (Guenther, 1994).

## **3. Application of MARC classification format to UDC**

The representation of electronically classified, structured and organised information into hierarchies, its processing by means of traditional classification systems and its online retrieval in catalogues is of great use. The Library of Congress, a pioneer in implementing MARC formats, is attempting to convert the whole LCC to MARC format, a process which begun with the HF and HG tables. It is also addressing the conversion of the DDC; this conversion is more viable and has been based on the twentieth edition (ESS, Editorial Support System). Nevertheless, the application of the MARC classification format to UDC has not been undertaken yet, and this is the proposal that is formulated here.

In order to manage, maintain and exchange the UDC classification data, a format will be needed to accommodate and differentiate the following elements (McIlwaine, 2003): (a) the notation, with a provision for different types of numbers (main number(s), special auxiliary number(s)); (b) the relationship between elements within a UDC compound number; (c) broader hierarchies; (d) references, scope notes, application notes, and the previous record of notation terms and indexes. Additional data required for the proper management of the classification are also

important: the UDC edition being used, the edition language, notes, the edition type and others. The application of the MARC classification format would provide the possibility of information retrieval in a properly assigned classification, since a systematic catalogue is a well-created catalogue.

The application of MARC classification enables a proper information retrieval via the notations of a classification system, since these notations would be included in a Classification Authority File. The classification data will provide the authority for classification records and may be recovered with greater accuracy. Their application to UDC entails taking greater advantage of the systematic catalogue through UDC.

#### **4. Application and types of subject interoperability**

A priority objective for library catalogues has been subject interoperability and, especially, interoperability between traditional classification systems (UDC, DDC and LCC). Such interoperability is a significant improvement for those systems that have been published with associated subject headings, such as DDC and LCC.

The classification system with subject headings with the greatest impact has been the DDC in its electronic version (edition 22), which has the LCSH associated with it. The LCC Plus also has the LCSH linked to it in its electronic version, and had a great impact.

The creation of subject gateways becomes the instrument for the creation of interoperability between classification systems, subject headings and thematic descriptions. These gateways provide a fundamental structure that will support digital libraries, which always point towards the universal virtual library or the universal bibliographical catalogue projected by Otlet and La Fontaine.

The interoperability of content is a very broad issue, and may be established between different classification systems, different languages within the same classification system, alphabetic and systematic systems and others. But, moreover, interoperability between subject headings and classification systems may also articulate interconnection between different controlled vocabularies in different languages.

When retrieving information, users should not be aware of the *modus operandi* of the articulation of the different indexing languages. The ideal way would be that the user formulates a single search, instead of formulating it in different ways and in different languages. Therefore, the ideal thing would be to make different controlled vocabularies and classification systems to be interoperable.

As an authority file, the MARC classification format is capable to articulate four types of interoperability:

##### **4.1. Interoperability between subject headings in the same language**

Regarding interoperability between subject headings in the same language, the most noteworthy proposal is the one implemented between the Library of Congress Subject Headings (LCSH; see at <http://www.loc.gov/cds/lcsh.html>) and the Medicine Subject Headings (MeSH). It integrates online catalogue subject headings which establish relations between authority files, data bases and metadata. By using the MARC format, relationships have been created between authority files of

over 10,000 records, although the entire synthetic LCSH structure cannot be done completely, as it has to be manually done in order to finish it. This case of interoperability has been methodologically exported for other subjects.

The so-called metathesaurus of the National Medicine Library (UMLS, Unified Medical Language System) of the USA has made over thirty subject heading vocabularies, thesaurus terms and classification systems interoperable, using for the latter processing techniques based on lexical units. It is based on the creation of a specialised lexis which includes over 180,000 entries containing verbs, nouns, adjectives, etc. on Biomedicine. It also has a semantic network including 132 semantic types; this semantic network establishes the categorisation of all the components of the metathesaurus, with 53 links between the semantic types or network nodes and the relationships (links) between them.

#### **4.2. Interoperability between multilingual subject headings**

Multilingual subject headings are intended to be set up by a system of linguistic equivalences, establishing relations between terms in different languages. This task of comparing terms is not an easy one. If the comparison is merely done between the terms of the subject heading lists in different languages, it is called "terminological comparison" and it is addressed from linguistic problems. If the comparison is based on the establishment of equivalence between equivalent authority records in different languages, it is called "semantic equivalence", and here semantic problems are addressed. And if the equivalence is established via application, this is "syntactic equivalence", and addresses technology-linked aspects.

The European multilingual project MACS - Multilingual Access to Interoperability Subjects (Landry, 2000), was created in 1997 at the Conference of European National Libraries (CENL) in order to attempt to address the lack of multilingual access to European databases. Switzerland, as a multilingual country, was especially interested in the articulation of this project. It establishes interoperability between three subject heading lists: the Library of Congress Subject Headings (LCSH), RAMEAU, and SWD, enabling access to a common list of the three SHL (Subject Headings Languages). The framework of this interoperability is articulated via a system of links among the different catalogues. The search can be formulated in the four national libraries, either in one or in all of them: the Swiss National Library, the French National Library, the German Library and the British Library. Searches can be undertaken either in one subject heading list (SHL) or in all the SHLs. This also entails the possibility of formulating a search in the catalogue of a library in another language, in addition to visualising the retrieved records which can be undertaken via some of the MARC formats (USMARC, UNIMARC) and also MAB and in three languages (English, French and German), in collaboration with and financed by the four National Libraries. Currently, MACS is restricted to authorised headings and it is only applicable as an SHL dictionary. The future incorporation of links to other authority recorded elements may bring with it the undertaking of a multilingual virtual file.

Regarding multilingual subject headings, the multilingual database on French monuments heritage, MERIMEE (<http://www.numerique.culture.fr/mpf/pub-fr/index.html>), is also noteworthy. It encompasses religious, civil, school, military and industrial architecture and is aimed at articulating interoperability between controlled vocabularies in different languages. It covers three areas: the inventory undertaken by regional services, old inventories included in PREDOC and historic monuments under protection since 1913, including the decree that registered and classified them. Thus the MERIMEE database is linked to five databases: the THESAURUS, PALISY, MEMOIRE and ARCHIDOC databases (<http://www.culture.gouv.fr/culture/inventai/presenta/bddinv.htm>).

Another European project on interoperability between multilingual subject headings is the HEREIN European Project, the European Information Network of Cultural Heritage Policies, which was begun by the Council of Europe and is financed by the European Union. It was sponsored by six countries: Spain, France, Ireland, the United Kingdom, Norway and Hungary, and Belgium later on. The multilingual thesaurus included in the HEREIN project (<http://thesaurus.european-heritage.net/>) aims to offer a terminological standard on national policies which address the subject of architectural and archaeological heritage.

#### **4.3. Interoperability between subject headings and classification systems**

Subject interoperability between subject headings in the same and different languages may be extended and lead to interoperability between subject headings and classification systems in the same language. The most noteworthy is DDC for Windows 22, which articulates edition 22 of DDC tables and offers the advantage of being used in a large number of libraries in 135 countries, translated into 30 languages and used in the USA by 95% of public and school libraries in addition to a large number of university and specialised libraries. Its alphabetical index was one of the instruments which placed DDC within the framework of modern classification systems.

Interoperability is established between the classification numbers and the alphabetical headings of the tables, to their notations, and to the Library of Congress Subject Headings (LCSH). The latter has already been applied to the two latest editions of the tables.

It is also worth mentioning the interoperability between LCC and LCSH (Library of Congress Subject Headings). The product of this conjunction is called LCC Plus (see <http://www.loc.gov/catdir/cpsol/lcco/>), and it is available in the Web Classification, an interface which is being developed in the Library of Congress.

Regarding interoperability with UDC, there is a new product based on the UDC abbreviated edition and its correlation and translation to the General Finnish Subject Headings.

In Spain, interoperability between UDC and the Subject Headings List of the National Library of Madrid from the authority files is being undertaken by a project sponsored by AENOR (<http://www.aenor.es>).

#### **4.4. Interoperability between classification systems**

Subject interoperability may be extended and embrace the one established between classification systems. One of the most noteworthy of the interoperability projects among classification systems is the British teaching thesaurus project, HILT (High Level Thesaurus Project - <http://hilt.cdlr.strath.ac.uk/index.html>). Although being a project originated in Great Britain, it is also aimed to encompass Australia, Canada and the USA. It makes different controlled vocabularies interoperable, and attempts to facilitate both subject search and navigation. The interoperability it presents encompasses LCSH, DDC and the UNESCO Thesaurus, UDC and AAT (Art and Architecture Thesaurus).

Another relevant proposal is the DARPA project on interoperability or comparison through metadata between controlled entry languages and retrieval languages or, in other words, controlled vocabularies and metadata vocabularies.

The European RENARDUS project (<http://www.renardus.org>) addresses interoperability between specific classification systems, controlled vocabularies in different languages and their convergence. This programme is based on technical models and computer tools, which is why interoperability is essentially based on computer instruments.

Another project that should also be mentioned is the Polish interoperability project which converges several controlled vocabularies such as the Subject Headings Language (SHL), the Thesaurus of Common Topics (TCT), UDC and the Polish Thematic Classification (PTC).

Likewise, the American Mathematics Society (AMS) is working on interoperability between Mathematics Subject Classification (MSC) and DDC with the 510 class related to Mathematics, a project developed at the New York State University in Albany.

Finally, we must not forget the interoperability developed between the Swedish Classification System (SAB) and DDC. The project is financed by the Royal Library of Sweden and the table conversion is published by the library and can be found online (<http://www.kb.se/Bus/SAB/sabheadings.htm>). It is focused on articulating a new collaboration framework on heritage in these countries. Its primary objective is to enable the exchange of information on heritage policies, as well as collaboration in working groups with forums, associations and other heritage areas such as archives, libraries and museums. It makes reports available on heritage policies in these countries addressing access, protection and conservation. Regarding this point, a thesaurus was drawn up, including the key words in different documentation such as heritage policy reports and others. It was not based on any already existing thesaurus, but it attempted to shape a specialised multilingual thesaurus in three languages (English, French and Spanish) and it may be extended to other languages. It methodologically establishes hierarchical, equivalent and associative relationships.

### **5. Proposal for applying the MARC Classification Format to the UDC Master Reference File (MRF) with subject interoperability**

The advantage of a classification format for UDC also entails the possibility of ensuring that the Master Reference File (MRF) from where every UDC translation, version and adaptation emanates, converges as an authority in the MARC classification format. The classification format applied to UDC means that the authority of numbers and notations may also be managed by using the MRF and it is integrated as such.

084	8#\$audc\$cInternational medium edition
<b>153</b>	##\$a512.5\$hMathematics and natural sciences\$hAlgebra\$jGeneral algebra
084	8#\$audc\$cInternational medium edition
<b>153</b>	##\$a642.12\$hHousekeeping. Home economics. Domestic science\$hFood. Cooking. Dishes. Meals\$hMeals and mealtimes. Tableware\$jMorning meal. Breakfast

Figure 1 – Examples of coding UDC numbers with MARC Classification format

For a synthetic classification such as UDC, the most important advantage of a classification format would be the possibility of using already implemented new notations and searching each compound number in a broken-down format, in addition to using natural language terms to retrieve classification numbers. Classification numbers would provide the authority for classification records, which could be saved and maintained as a separate file, as a database similar to other databases built upon MARC authority formats. This means that once they have been introduced into an authority file, classification numbers may be linked to the bibliographic records in both the cataloguing and retrieval processes.

This authority file may be used in classification tasks to assign notations to bibliographic records; it would serve to automatically validate classification numbers; and it would provide a system with mechanisms to validate their correct assignment, global updates and reclassification of bibliographic records (Woods, 1996). This authority file may be printed, maintained and exchanged or shared between libraries. It may act as a monitor for the authorities of the UDC notations which do not appear in the tables. The creation of a classification record for a notation would provide an authority of that notation and guidance for its future use.

In order to obtain the maximum performance from UDC in an OPAC, it is essential to have automated UDC tables, a UDC authority file and an interface which enables visualisation and navigation by hierarchies and which facilitates a search tool (simple and post-coordinated) using both classification numbers and natural language terms. Consequently, the MARC classification format is very useful for the maintenance, management and printing of UDC notations, and to provide a system which is fitted with mechanisms to validate their correct assignment. To a large extent, the MARC classification format facilitates the tasks of the classification work as it authorises the conversion of classification numbers to other classification systems.

Incorporating classification data into online systems in the ways suggested above promises to be of great potential for UDC, but has yet to be tackled.

### **Bibliography**

All URL valids as of 06.02.2010

Guenther, R.S. (1994) The Library of Congress Classification in the USMARC Format. *Knowledge organization*, 21(4), pp.199-202.

Landry, P. (2000) The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD). In *IFLA 66<sup>th</sup> Council and General Conference*. Jerusalem, Israel 13-18 August 2000. Available at: <http://archive.ifla.org/IV/ifla66/papers/165-181e.pdf>.

McIlwaine, I. C. (2003) *La Clasificación Decimal Universal. Guía de uso / Adaptada por Rosa San Segundo*. Madrid: AENOR.

Robbio, A.; Marini, A.; Maguolo, D. (1999-2001) *Mathematics Subject Classification MSC and Dewey Decimal Classification DDC*. Available at <http://www.math.unipd.it/~biblio/math/engddc.htm>.

San Segundo, R. (1996) *Sistemas de Organización del conocimiento. La organización del conocimiento en las bibliotecas españolas* Madrid : Universidad Carlos III ; Boletín Oficial del Estado, 1996, ISBN 84-340-0886-6.

San Segundo, R. (2005) *Interoperability between classification systems using metadata*. In *Advances in Metadata Research*. Paramus, NJ: Rinton Press, 2006, pp. 140-147.

Tennis, J. (2002) *Layered information modeling and a knowledge organization paradigm: inherent classification and the design of interoperable systems*. Paper presented at "Knowledge Technologies Conference 2002", Seattle, March 11-14, 2002. Available at: [www.knowledgetechnologies.net/proceedings/presentations/tennis/LayeredFinalPaperSubmitted5.1.2002.doc](http://www.knowledgetechnologies.net/proceedings/presentations/tennis/LayeredFinalPaperSubmitted5.1.2002.doc).

Woods, E.W. (1996) *Requirements for a Format for Classification Data: Final Report – July 1996*. Available at: <http://archive.ifla.org/VII/s29/projects/rep0796.htm>.

Zeng, M. L.; Chan, L.M. (2004) Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55 (5), pp.377-395.