Rebecca S. Guenther
Network Development and MARC Standards
OfficeLibrary of Congress, Washington, DC 20540

The Library of Congress Classification in the USMARC Format

Rebecca Guenther is Senior MARC Standards Specialist in the Library of Congress. She received her M.S. in Library Science from Simmons College, Boston, MA. Current responsibilities: work on natl. & int. library automation standards, including USMARC bibliographic, authority, classification, holdings, and community information formats, and USMARC code lists for languages, countries and geographic areas.

Guenther, R.S.: **The Library of Congress Classification in the USMARC Format**
Knowl.Org. 21(1994)No.4, p.199-202, 9 refs.
The paper reviews the development of the USMARC Format for Classification Data, a standard for communication of classification data in machine-readable form. It considers the uses for online classification schedules, both for technical services and reference functions and gives an overview of the format specification details of data elements used and of the structure of the records. The paper describes an experiment conducted at the Library of Congress to test the format as well as the development of the classification database encompassing the Library of Congress Classification (LCC) schedules. Features of the classification system are given. The Library of Congress will complete its conversion of the LCC in mid-1995. (Author)

1. Introduction

The Library of Congress recently completed the development of a machine-readable format for classification data to allow for the communication of classification records between systems and to provide a standard for the storage of classification data in the computer. The USMARC Format for Classification Data joins the family of MARC (Machine-Readable Cataloging) formats: bibliographic, authority, holdings, and newly developed community information formats (1).

Implementation poses great challenges for institutions, particularly for those responsible for the maintenance of library classification schemes.

2. The USMARC Formats

The USMARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form. They are communication formats, primarily designed to provide specifications for the exchange of bibliographic and related information for sharing between systems within the United States. Developers of the USMARC formats have attempted to maintain some compatibility with other national and international formats (e.g., CANMARC, UNIMARC). The Network Development and MARC Standards Office of the Library of Congress is responsible for developing and maintaining the USMARC formats, which now consist of the following: USMARC Format for Bibliographic Data, USMARC Format for Authority Data, USMARC Format for Holdings Data, USMARC Format for Classification Data, and the USMARC Format for Community Information (expected publication date: May 1993). Development and maintenance is accomplished in consultation with user communities.

Content designation in the USMARC formats are codes and conventions used to explicitly identify data elements in a record. The goal is to characterize the data elements with sufficient precision to support manipulation of the data for various functions (2).

Functions supported include display, both formatting in an online display and in producing a printed or other type of product, and online information retrieval. How an institution displays the data is not specifically covered in the USMARC formats, although they do provide for display constants, which are terms, punctuation, or spacing that are system generated for display. For instance, the hyphen separating the beginning and ending numbers of a classification number span is generated as a display constant by the system based on the structure of the classification number field of the record. Institutions may display the data as desired. In the USMARC formats the data is organized into fields, each identified by a three-character numeric tag.

MARC is the generic term for Machine-Readable Cataloging; USMARC, the format that is used in the United States) has proven to be flexible, efficient, and easy to maintain for library automation systems (3).

It is used worldwide for storing, sharing, and manipulating bibliographic information by computer. The USMARC Format for Classification Data joins the other established formats with a similar structure and goals.

3. The USMARC Format for Classification Data

The USMARC Format for Classification Data contains specifications for converting classification data into machine-readable form for communicating and storing classification data. USMARC classification records provide the authority for classification numbers and their captions. Each classification number or caption is given its own USMARC classification record, which may include not only the number and caption, but also any notes or other information associated with it. With USMARC classification records, the institution responsible for maintaining the classification scheme would likely create and distribute the majority of classification records, rather than the shared environment currently used for bibliographic records, where many institutions may contribute

to a shared database. A Joint Working Group on a classification format has been formed by the International Federation of Library Associations (IFLA), Sections on Information Technology and Classification and Indexing. Work is underway to establish a UNIMARC Format for Classification.

4. Uses for USMARC Classification

The following are uses of interest primarily to technical services applications:

Catalogers' workstation.

An online classification schedule could be incorporated into a catalogers' workstation, which might also include an online catalog of bibliographic and authority records, cataloging and inputting documentation, and perhaps holdings information. Online access to classification numbers may save time for the classifier, who could perform a keyword search to locate quickly the possible classification numbers through captions, notes, or index terms. Synthesized classification numbers could be provided (i.e., those numbers built from tables within the classification schedule to provide a more explicit number for the subject), so that the work need not be duplicated.

Printing schedules and providing timely updates.

Producing the Library of Congress Classification from an online file could greatly facilitate the cumbersome process that the Library of Congress currently uses to publish revised editions. The LCC was developed over a period of time by different people. It was designed as a shelf location and browsing device and has been maintained as such (4). Many of the now forty-eight schedules are in need of revision, and have been kept up-to-date by LC Classification Additions and Changes. This method can prove cumbersome, since it requires using several classification volumes (i.e., a base volume and any number of updates), rather than a single revised edition.

Linking to bibliographic records.

Maintaining machine-readable classification data in the same format as bibliographic data has many advantages. By maintaining an online classification schedule, an institution may be able to validate classification numbers assigned to bibliographic records; perform global updates of a bibliographic file when a classification number is added or changed; and facilitate reclassification projects, either from one classification scheme to another, or within a single classification scheme.

Linking to subject/name authority records.

A link may also be provided to subject and name authority records, which could assist the cataloger in assignment of appropriate subject headings and classification numbers. It also could enhance subject access to the schedules by using a controlled thesaurus, as is used for bibliographic records.

The following are uses that will benefit reference applications as well as technical services:

Providing subject access to schedules

In previous research, scholars and researchers have recognized the possibilities of classification data providing subject access to library material (5). Explorations on the use of the LCC for online subject access have been conducted (6). Online classification data can enhance retrieval of bibliographic records by providing a different type of subject access, through a classed catalog approach, rather than through controlled subject headings. It enables the library user to see the interrelationships between topics and classification numbers, and facilitates browsing from more general to more specific topics and numbers. In addition, more precise searching of bibliographic records is possible through classification numbers for certain types of searches. For example, a well-known individual author or work may have its own classification number in the LCC and a search by class number may retrieve a listing of bibliographic records on that author or work more efficiently than a traditional author or title search. Having the classification data accessible online will facilitate this process. Subject access will also be enhanced by providing a wealth of subject terms in the classification schedules that may not be available in the bibliographic records and their controlled subject vocabulary.

Basis for an online shelflist.

The Library of Congress has had numerous requests to automate its shelflist. If this project is undertaken, the online classification schedules would certainly provide a basis for that type of effort, or for another institution using LCC. An automated shelflist would be a valuable resource for bibliographic material.

5. USMARC Classification Format Overview

The USMARC Format for Classification Data (7) allows for the explicit identification of defined classification data elements so that the data may be manipulated for a variety of functions. The implementation of the format, or system developed to use the data, will determine how well the intended functions are met. The format is designed to provide an authority for a number and its caption, with any associated notes or instructions for how to apply it. USMARC records are created for classification numbers from the schedule, from numbers specified in a table providing subdivisions, and for general explanatory index terms not associated with a particular classification number. (For a further review of the classification format fields, see (8).)

Required Classification Fields

A classification data record must have at the minimum the following fields: 008 (Fixed-Length Data Elements, which contain processing information about the record

and number), 084 (Classification Scheme and Edition, which specifies the classification scheme; the format was developed to accommodate any classification scheme); and 153 (Classification Number), which includes the number, the caption thatdescribes the subject represented by the number, and the superordinate level of the caption hierarchy. In rare cases, field 154 (General Explanatory Index Term) with field 753 (Index Term-Uncontrolled) areused instead offield 153. Other associated fields, such as tracings, notes, index term fields, and number building fields may also be contained in a USMARC record.

USMARC Classification Format Field Blocks

The following are the groups of fields in the USMARC Classification Format.

Leader
Directory
Control information, numbers, codes (010_084)
Classification numbers and terms (153, 154)
Complex see references (253)
Complex see also references (353)
Invalid number tracings (453)
Valid number tracings (553)
Note fields (680_685)
Index term fields (700_754)
Number building fields (761_768)

6. USMARC Classification Format Development

The Library of Congress recognized the need for classification in machine-readable form, and decided to develop a USMARC format for classification for communicating and storing classification data. The USMARC Format for Classification Data began as an extension of the USMARC Authority Format. However, it was found that there was not as much overlap between the two types of data as expected, and it was rewritten as a separate format. The format was developed between 1988 and 1990 in close consultation with the two major classification schemes in use in the United States, Library of Congress Classification and Dewey Decimal Classification. It was approved provisionally by the American Library Association's Machine-Readable Bibliographic Information Committee (MARBI). Its provisional status means that experimentation is needed before full approval.

7. Library of Congress USMARC Classification Experiment

The Network Development and MARC Standards Office of the Library of Congress embarked on a classification experiment to put to use with actual records the new USMARC format that had been developed. Its goals were to experiment with the adequacy of the format as a carrier for classification data; to determine the likely editing of the schedules necessary for full conversion; to determine specifications for an online system to use the data; to gain an understanding of the financial and staff requirements for a larger effort; and to look at ways to enhance access to classification numbers and terms.The Office chose a PC-based MARC record making system called Minaret, which has been used throughout the Library for AmericanMemoryprojects. Its flexible software allows for user-controlledindexing and display forms and a user-defined format. Thus, a new format such as classification,previouslyunknown to cataloging systems could be defined and records created using it. Minaret canoutput records in standard USMARC for use by other systems using their own software. Input isgenerally easy and quick, and most classification records tend to be short. It was decided to begin the experiment using the LCC schedule H-HJ (Social Sciences: Economics).

Shortly after beginning to input classification records in the H-HJ schedule, it became apparent that the display of full USMARC records was not adequate to effectively use online classification data. Because classification data requires that the user see the hierarchical relationship between classification numbersand terms to make full use of the data, the Library contracted with the developer of Minaret, Cactus Software, to obtain an enhanced version of the software. This enhancement provided a browse display, which shows classification data in a form similar to the page of an LCC classification schedule.

8. Minaret Classification Database

As part of the Classification Format Experiment, USMARC classification records were input for all classification numbers and captions in the H-HJ schedule, consisting of about 15,000 records from the schedule itself and about 6000 records from tables. The classification database has been used on a daily basis by a cataloger from the Business and Economics cataloging section instead of a printed schedule. In addition, the cataloger is enhancing the database with subject terminology as needed for subject access to the schedules. Because of the expressed interest of the law library community in online classification (partially for its potential use in reclassification projects), the subclass KJE (Law of Europe) was converted by an outside contractor and the resulting database was shown at the American Association of Law Libraries annual conference in San Francisco in July 1992.

MARC records have been made available to institutions wishing to experiment with the format, however, these records do not include the software to display them, like a classification schedule.

9. Conversion Efforts for LC Classification

As a result of the success of the classification experiment and the desire expressed by many institutions to incorporate machine-readable classification data into their systems, the Library of Congress' Cataloging Directorate formed a technical committee to look at how a full conversion of LCC might be accomplished. The Technical Committee on LC Classification in Machine-Readable Form recommended that all schedules be converted into the USMARC classification format.

As of late 1994, about two-third of the LCC schedules had been converted to machine-readable form, both by outside contractors and by internal Library of Congress staff. It is expected that the entire conversion will be completed by mid-1995. The first LCC schedule to be printed from the MARC records (H: Social Sciences) was completed in late 1994. Other printed schedules will follow in early 1995. The Cataloging Distribution Service will also distribute the records as a MARC distribution service and is exploring a CD-ROM product.

10. Characteristics of the Online Classification Database

USMARC classification records are contained in one master database which spans all subject areas. The following characteristics are apparent when viewing the online classification database:

An online browse display shows classification numbers with captions hierarchically arranged on the screen much like the page of a classification schedule. (Hierarchies are shown as indentions.) Numbers have been assigned to all captions. Some numbers are suppressed in the enhanced browse display.

As one browses down the screen, the previous hierarchy appears in a box at the top of the screen. This shows the user the relationship between the current caption and the previous captions.

All notes, references and subarrangements appear in the enhanced browse display. The user can jump to a number referred to in a reference by using the "enter" key. It is not necessary to view the full MARC record, although this may be done by pressing a function key.

If a table is to be applied to a number for subdivisions, the table can be windowed onto the bottom of the screen by pressing the "enter" key. For example, a book entitled *European Parliament: Rules of Procedure* would use the base classificatino number KJE5390 for European Parliament, which is made more specific by applying a form subdivision from Table 9 for "Rules governing the organization: monographs". The table can be "windowed" onto the bottom of the screen while viewing the hierarchical arrangement for European Parliament. The user can also search captions in the tables to find the appropriate number to be added onto a base number.

The user can search by various methods: by classification number, by caption, by index term (from the index at the back of the classification schedule), or by a combined index of both caption and index term. This allows the user to have a rich set of terminology to access the classification number, and ultimately works on that specific subject.

By using this wide range of subject terms from various sources, the user can search by variant terms, including terms in other languages (particularly in the law schedules which typically include French and German terminology in the classification captions and index), acronyms and initialisms.

11. Conclusions

What began as an experiment to test the USMARC Classification Format has grown into a full conversion of the Library of Congress Classification schedules. Once all schedules are converted, it will be possible to search across classes, something that has been cumbersome up until now. For instance, one's research could include legislation about a certain industry as well as economic information about that industry. The ability to search in both the law (class K) schedules as well as the economics schedule (class H) will enable the user to have access to a wider range of material. In addition, incorporating classification data into online library systems using USMARC will allow for links between subject authorities, name authorities, and bibliographic data. One could determine the classification number for a specific topic and link to the library's collection in the bibliographic files through that classification number, allowing for a classed catalog approach to research material. Thus, one could see the subject arrangement that would be found on the shelf and the hierarchical relationship between topics. Maintenance and reclassification could largely be done by a computer, a development of particular interest to law library collections, for which it is desirable to reclassify material into the newly developed LCC law (class K) schedules. Incorporating classification data in online systems could result in the addition of rich subject terminology for online searching. If such potential is realized by systems implementing online classification data, classification data will become a powerful research tool.

References

(1) USMARC Format for Classification Data: Including Guidelines for Content Designation. Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1991.

(2) The USMARC Formats: Background and Principles Washington, D.C.: Network Development and MARC Standards Office 1989. p.4.

(3) Crawford, W.: MARC for Library Use. 2nd ed. Boston: G.K. Hall, 1989. p.7a

(4) Chan, L.M.: The Library of Congress Classification System in an Online Environment. Cat. & Classif. Quart. 11(1990)No.1, p.9.

(5) Markey, K., Demeyer, A.: Dewey Decimal Classification Online Project: Evaluation of a Library Schedule and Index Integrated into the Subject Searching Capabilities of an Online Catalog: Final Report to the Council on Library Resources. Dublin, OH: OCLC Online Computer Library Center 1986.

(6) Chan, L.M.: Library of Congress Classification as an Online Retrieval Tool: Potentials and Limitations. Inform. Teclmol. & Libraries 5(1986)No.3.

(7) USMARC Format for Classification Data: Including Guidelines for Content Designation. Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1991.

(8) Guenther, R: The Development and Implementation of the USMARC Format for Classification Data. Inform. Technol. & Libraries 11(1992)No.2.

(9) Technical Committee on LC Classification in Machine-Readable Form: Final Report. Washington: Library of Congress, September 1992.