

# Principles of Format Design

Henriette D. AVRAM and Lucia J. RATHER: MARC Development  
Office, Library of Congress

*This paper is a summary of several working papers prepared for the International Federation of Library Associations (IFLA) Working Group on Content Designators. The first working paper, January 1973, discussed the obstacles confronting the Working Group, stated the scope of responsibility for the Working Group, and gave definitions of the terms, tags, indicator and data element identifiers, as well as a statement of the function of each.<sup>1</sup>*

*The first paper was submitted to the Working Group for comments and was subsequently modified (revised April 1973) to reflect those comments that were applicable to the scope of the Working Group and to the definition and function of content designators. The present paper makes the basic assumption that there will be a SUPERMARC and discusses principles of format design.*

*This series of papers is being published in the interest of alerting the library community to international activities. All individual working papers are submitted to the MARBI interdivisional committee of ALA by the chairman of the IFLA Working Group for comments by that committee.*

## INTRODUCTION

In order to have this paper stand alone, the scope and the definition and functions of the content designators as agreed to by the Working Group are summarized below:

1. The scope of responsibility for the IFLA Working Group is to arrive at a standard list of content designators for different forms of material for the international interchange of bibliographic data.
2. The definition and function of each content designator are given as:
  - a. A tag is a string of characters used to identify or name the main content of an associated data field. The designation of main content does not require that a data field contain all possible data elements all the time.
  - b. An indicator is a character associated with a tag to supply additional information about the data field or parameters for the processing of the data field. There may be more than one indicator per data field.

- c. A data element identifier is a code consisting of one or more characters used to identify individual data elements within a data field. The data element identifier precedes the data element which it identifies.
- d. A fixed field is one in which every occurrence of the field has a length of the same fixed value regardless of changes in the contents of the fixed field from occurrence to occurrence. The content of the fixed field can actually be data content, or a code representing data content, or a code representing information about the record.

#### BASIC ASSUMPTION—SUPERMARC

There appears to be little doubt that the format used for international exchange will not be the format presently in use in any national system. The first working paper addressed the obstacles that preclude complete agreement on any single national format, and a study of the matrix of the content designators assigned by various national agencies substantiates the above conclusion. Consequently, we are concerned with the development of a SUPERMARC whereby national agencies would translate their local format into that of the SUPERMARC format and conversely, each agency would accept the SUPERMARC format and translate it into a format for local processing.<sup>2,3</sup> SUPERMARC, therefore, is an international exchange format with the principal function that of transferring data across national boundaries. It is *not* a processing format (although if desired, it could be used as such) and in no way dictates the record organization, character bit configuration, coding schemes, etc., to be used within processing agencies.

The SUPERMARC format, however, should conform to certain conventions, namely the format structure should be ISO 2709 and the character representation should be an eight-bit extension of ISO 646.\* The latter convention means that data cannot be in any other configuration than a character-by-character representation.

SUPERMARC assumes not only agreement on the value of content designators but, equally as important, on the level of application of these content designators. Whatever the agreed upon level of content designation is, those agencies with formats more detailed will be able to translate to SUPERMARC but will be in the position of having to upgrade all records entered into their local system from other agencies. Likewise, local formats consisting of less detailed content designation than SUPERMARC must upgrade to the SUPERMARC level for communication purposes.

Where the actual content of the record is concerned, i.e., the fields and/or data elements to be included, it is highly probable that the decision of the Content Designator Working Group will be that data, if in-

\* ISO/TC 46/SC4 WG1 is presently engaged in the definition of extended characters for Roman, Cyrillic, and Greek alphabets and mathematics and control symbols.

cluded in the record, are assigned SUPERMARC content designators, but that not all data will always be present. This permits the flexibility required to bypass some of the substantive problems of different cataloging rules and cataloging systems. For example, one agency may supply printer and place of printing while another may not. It may be assumed, however, that all agencies will conform to the specifications prescribed by the ISBD and other such standard descriptions as they become available.

## PRINCIPLES OF FORMAT DESIGN

Prior to any deliberation regarding the actual value of content designators, the Working Group realized it must agree on a set of basic principles for the design of the international format. The first working paper set forth, in the form of questions, some of the issues that must be taken into account in arriving at the principles. Several members of the Working Group expressed their opinions and these were considered in the formulation of the principles. The principles were discussed at the Grenoble meeting in August 1973. Five of the principles were adopted and the sixth was deferred for further analysis based on working papers to be written by some of the members. The sixth principle was adopted at the Brussels meeting in February 1974.

The six basic principles are stated below with a discussion following each principle:

1. *The international format should be designed to handle all media.* It would be ideal if at this time all forms of material had been fully analyzed. This is currently not the case. Agreement on data fields and the assignment of content designators can realistically only be accomplished if there is a foundation upon which to build.

Therefore, the forms of material have been limited to those listed below because, to the best of our knowledge, these are the only forms where either experience has been gained in the actual conversion to machine-readable form or in-depth analysis has been performed to define the elements of information for the material.

*Books:* all monographic printed language materials.

*Serials:* all printed language materials in serial form.

*Maps:* printed maps, single maps, serial maps, and map collections.

*Films:* all media intended for projection in monographic or serial form.

*Music and Sound Recordings:* music scores and music and nonmusic sound recordings.

At the meeting in Brussels, the decision was made to use the ISBD as the foundation for the definition of functional areas for the formats. Since at the present time an ISBD exists only for monographs and serials, these materials will receive first priority by the IFLA Working Group.

Still under consideration is the question whether manuscripts should be included in the forms of material within the scope of the

Working Group. Pictorial representations and computer mediums have not as yet been analyzed. When these forms have been analyzed, they should be added to the generalized list.

2. *The international format should accept single-level and multilevel structures.* There is a requirement to express the relationship of one bibliographic entity to another. This relationship may take many forms. A hierarchical relation is expressed for works which are part of a larger bibliographic entity (such as the chapter of a book, a single volume of a multivolume set, a book within a series). A linear relation is expressed for works which are related to other works such as a book in translation. This discussion is concerned with hierarchical relationships and the need to describe this relationship in machine-readable records. There are a number of ways in which hierarchical relationships may be expressed. One method is to place the information on the related work in a single field within the record. For example, the different volumes of a multivolume set may be carried in a contents field. When a book is in a series, the series may be carried in a series field. This may be termed using a single-level record to show a hierarchical relationship. Another method is to use a multilevel record made up of subrecords.<sup>†</sup>

The concept of a subrecord directory and a subrecord relationship field was discussed in Appendix II to the ANSI standard Z39.2-1971.<sup>4</sup> The appendix illustrated a possible method of handling subrecords and expressing relationships within a bibliographic record but was not part of the American standard. Similarly, in 1968 the Library of Congress published as part of its MARC II format a proposal to provide for the bibliographic descriptions of more than one item in a single record, and represented this capability as "levels" of bibliographic description.<sup>5</sup> The international standard (ISO 2709) defines a subrecord technique without an explicit statement of a method to describe relationships.<sup>6</sup>

More recently, a level structure was proposed in a document by John E. Linford,<sup>7</sup> and an informal paper by Richard Coward<sup>8</sup> gave the following example of a level structure:

<i>Level</i>	<i>Record</i>
Collection	1 subrecord
Sub-collection	1 subrecord
Document	1 subrecord
Analytical	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">1 subrecord</div> <div style="text-align: center;"> <div style="border-top: 1px solid black; width: 100%; margin-bottom: 2px;"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <div style="text-align: center;">1 subrecord</div> <div style="text-align: center;">1 subrecord</div> <div style="text-align: center;">1 subrecord</div> </div> </div> </div>

<sup>†</sup> A subrecord is a "group of fields within a bibliographic record which may be treated as a logical entity." When a bibliographic record describes more than one bibliographic unit, the descriptions of the individual bibliographic units may be treated as subrecords.

Several national agencies have expressed concern regarding the efficiency of the ISO 2709 subrecord technique and have suggested that a modification be made to the subrecord statement.

There are alternative techniques which could be incorporated in the international exchange format to build in level capability. Methods have been suggested that would cause a revision (specifically the number of characters in each directory entry) to the ISO standard; other alternatives might not. Regardless of the final technique agreed upon, national agencies should maintain the authority to record their cataloging data to reflect their catalog practices, i.e., either describing the items related to an item cataloged as fields within a single-level record or as subrecords of a multilevel record.

3. *Tags should identify a field by type of entry as well as function by assigning specific values to the character positions.* Assigning values to the characters of the tags allows the flexibility to derive more than a single kind of information from the tag. For example, it should be possible by an inspection of the tags to retrieve all personal names from a machine-readable record regardless of the function of the name in the record, i.e., principal author, secondary author, name used as subject, etc.
4. *Indicators should be tag dependent and used as consistently as possible across all fields.* Indicators should be tag dependent because they provide both descriptive and processing information about a data field. If the value assigned to an indicator is used as consistently as possible across all fields, where the situation warrants this equality, the machine coding is simplified to process different functional fields containing the same type of entry.
5. *Data element identifiers should be tag dependent, but, as far as possible, common data elements should be identified by the same data element identifiers across fields.* The principle has been adopted that the format will handle all types of media and consequently the projected number of unique tags may be quite large. In addition, since all types of media are not yet fully analyzed, the number of unique fields is an unknown factor. While it is undeniable that making data element identifiers tag independent would be desirable, the limited number of alphabetic, numeric, and symbolic characters would restrict the number of data elements to the number of unique characters. This constraint on future expansion seems to be more important than any advantages gained from making data element identifiers tag independent.

If data element identifiers are tag dependent, then additional refinements could be added in one of two ways: (1) the principle of identifying common data elements by the same identifiers across fields could be followed as far as possible, *or* (2) the identifiers could be given a value to aid in filing. The two refinements appear to be mutu-

ally exclusive since a data element in one field may have a different filing value from the same data element in another field. Since the first refinement should be useful for many types of processing, and the second would be useful only in filing, the former seems to be the better option.

6. *The fields in a bibliographic record are primarily related to broad categories of information relating to "subject," "description," "intellectual responsibility," etc., and should be grouped according to these fundamental categories.* The first working paper discussed as an obstacle the lack of agreement on the organization of data content in machine-readable records in different bibliographic communities. A subsequent paper consisting of comments made by staff of the Library of Congress on the proposed EUDISED format discussed in greater detail the analytic versus traditional arrangement.<sup>9, †</sup> The majority of the national formats designed to date are arranged by using the function as the primary grouping and the type of entry as the secondary grouping. Several working papers produced by committee members supported the arrangement by function on the grounds that it followed the traditional order of elements in the bibliographic record and therefore simplified input procedures. Grouping of the fields first by function and then by type of entry was agreed to at the Brussels meeting.

#### REFERENCES

1. Henriette D. Avram and Kay D. Guiles, "Content Designators for Machine Readable Records," *Journal of Library Automation* 5:207-16 (Dec. 1972).
2. R. E. Coward, "MARC: National and International Cooperation," in International Seminar on the MARC Format and the Exchange of Bibliographic Data in Machine-Readable Form, Berlin, 1971, *The Exchange of Bibliographic Data and the MARC Format* (Munich: Pullach, 1972), p. 17-23.
3. Roderick M. Duchesne, "MARC: National and International Cooperation," in International Seminar on the MARC Format and the Exchange of Bibliographic Data in Machine-Readable Form, Berlin, 1971, *The Exchange of Bibliographic Data and the MARC Format* (Munich: Pullach, 1972), p.37-56.
4. American National Standards Institute, *American National Standard for Bibliographic Information Interchange on Magnetic Tape* (Washington, D.C.: 1971) (ANSI Z39.2-1971). Appendix, p.15-34.
5. Henriette D. Avram, John F. Knapp, and Lucia J. Rather, *The MARC II Format; A Communications Format for Bibliographic Data* (Washington, D.C.: Library of Congress, 1968), Appendix IV, p.147-49.
6. International Organization for Standardization, *Documentation—Format for Bibliographic Information Interchange on Magnetic Tape*. 1st ed. International standard ISO 2709-1973(E). 4p.

† In an analytic tagging scheme, the first character of the tag describes the type of entry and subsequent characters describe function; in a traditional tagging scheme, the first character describes function and subsequent characters describe type of entry.

7. Council for Cultural Cooperation. Ad Hoc Committee for Educational Documentation and Information. Working Party on EUDISED Formats and Standards, 3d Meeting, Luxembourg, 26-27 April 1973, *Draft EUDISED Format* (Second Revision). Prepared by John E. Linford.
8. Paper sent from Richard Coward to Henriette D. Avram, "Notes on MARC Sub-record Directory Mechanism."
9. Henriette D. Avram, "Comments on Draft EUDISED Format (Second Revision)," unpublished paper.