

LEVELS OF MACHINE READABLE RECORDS

RECON Working Task Force: Henriette D. AVRAM, Chairman; Richard DE GENNARO; Josephine S. PULSIFER; John C. RATHER; Joseph A. ROSENTHAL and Allen B. VEANER.

This study of the feasibility of determining levels or subsets of the established MARC II format concludes that only two levels are necessary and desirable for national purposes: 1) the full MARC II format for distribution purposes; and 2) a less complex subset to be used by libraries reporting holdings to the National Union Catalog.

INTRODUCTION

In March 1969, the Advisory Committee to the RECON Working Task Force, after approving publication of the initial RECON report (1), endorsed investigation of a number of questions raised in that report as well as consideration of certain issues not covered in the initial survey. The basic tasks to be undertaken have been described in another article in this issue (2). With further support for RECON from the Council on Library Resources, Inc., the Working Task Force has met several times to explore some of these problems. This article reports the conclusions reached with respect to one task: the feasibility of determining a level or subset of the established MARC II format that would still allow a library using it to be part of a future national network.

DEFINITION OF "LEVEL"

During the initial RECON study the Working Task Force, for discussion purposes, considered levels of encoding detail of machine readable catalog records in relation to the conditions under which conversion might occur. A level was distinguished by differences in 1) the bibliographic

completeness of a record, and 2) the extent to which its contents were separately designated. With respect to the latter point, the RECON report stated:

"A machine format for recording of bibliographic data and the identification of these data for machine manipulation is composed of a basic structure (physical representation), content designators (tags, delimiters, subfield codes), and contents (data elements in fixed and variable fields). Although the basic structure should remain constant, the contents and their designation are subject to variation. For example, a name entry could be designated merely as a name instead of being distinguished as a personal name or corporate name. When a distinction is made, a personal name entry can be further refined as a single surname, multiple surname, or forename. Likewise, if a personal name entry contains date of birth and/or death, relationship to the work (editor, compiler, etc.), or title, these data elements can be identified or can be treated as part of the name entry without any unique identification. Thus individual data elements can be identified at various levels of completeness." (3)

Appendix F of the RECON report tentatively defined three levels:

"*Level 1* involves the encoding of bibliographic items according to the practices followed at the Library of Congress for currently cataloged items, i.e., the MARC II format. A distinguishing feature of level 1 is the inclusion of certain content designators and data elements which, in some instances, can be specified only with the physical item in hand.

"*Level 2* supplies the same degree of detail as in level 1 insofar as it can be ascertained through an already supplied bibliographic record. . . .

"*Level 3* would be distinguished by the fact that only part of the bibliographic data in the original catalog record would be transcribed. In addition, content designators might be restricted . . ." (4).

At the outset of the present study, however, it was recognized that incomplete bibliographic description is not acceptable in records for national use. In addition, it seemed that the question of having a level below level 2 really arose from a desire to define a machine readable record with a lesser degree of content designation rather than one with less complete bibliographic data. It was decided, therefore, to concentrate the study effort on this task, and the original formulation of level 3 was discarded.

On further consideration, it was realized also that the distinguishing feature between levels 1 and 2 was not significant. Omission of data elements that cannot be determined unless the book is in hand may simplify an individual record but does not simplify the content designators in the format because these elements are often present in other

records. Thus, as far as content designation is concerned, levels 1 and 2 (as originally defined) were in fact the same.

Once this similarity became apparent, it was recognized that the specification of levels really depended on the functions of machine readable catalog records from the standpoint of national use.

FUNCTIONS AND LEVELS

On the basis of present knowledge, it seems that machine readable records will serve two primary functions for national use. The first involves the distribution of cataloging information in machine readable form for use by library networks, library systems, and individual libraries; the second involves the recording of bibliographic data in a national union catalog to reflect the holdings of libraries in the United States and Canada. In this report, the first is called the distribution function; the second is called the national union catalog (NUC) function. Each of these functions can be related to a distinct level of machine readable record.

The Distribution Function

The distribution function can best be satisfied by a detailed record in a communications format from which an individual library can extract the subset of data useful in its application. At the present stage of library automation, it is impossible to define rigorously all of the potential uses of machine readable catalog records. Thus, there is no way to predict which data elements may not be needed or to rank them according to their value to a wide variety of users under different circumstances.

To confirm the wide variation in treatment of the MARC II format, an analysis was made of the use of MARC content designators by eight

Table 1. Use of MARC Content Designators by 8 Library Systems or Networks

Number of libraries	Fixed fields (19)	Number of items		
		Tags (63)	Indicators (126)	Subfield codes (181)
8	—	26	—	1
7	—	6	—	88
6	—	3	2	45
5	1	5	7	15
4	6	3	9	9
3	7	2	92	11
2	4	4	16	9
1	1	7	—	3
None	—	7	—	—

Note: Only six libraries supplied information on fixed fields.

library systems and emerging networks. The data from this analysis were synthesized for presentation in two tables. Table 1 shows the acceptance of content designators in terms of the absolute number of libraries using them. It should be read as shown by the following examples: 1) 26 of the 63 MARC tags are used by all eight libraries; 2) 92 of the 126 indicators are used by only three libraries. Table 2 shows the acceptance of content designators in relative terms. Thus, if only three libraries were using a particular tag and all used the associated subfield codes, the acceptance of those subfield codes was calculated as 100 percent. In both Tables 1 and 2, the columns on indicators and subfield codes include responses only from those libraries that were definitely using the tag with which a given indicator or subfield code was associated. The analysis excludes tags for which no immediate implementation is planned by the MARC Distribution Service.

Table 2. *Percentage of Acceptance of MARC Content Designators by 8 Library Systems or Networks*

Percent of libraries	Number of items			
	Fixed fields	Tags	Indicators	Subfield codes
	(19)	(63)	(126)	(181)
100	—	26	—	10
75-99	1	9	2	134
50-74	13	8	16	32
25-49	4	6	108	5
1-24	1	7	—	—
0	—	7	—	—

The major findings of this analysis may be summarized as follows:

1) Of 19 fixed fields, 14 were used by at least half of the libraries and all were used by at least one library.

2) Of 63 tags, 43 were used by at least half of the libraries and 26 were used by all of them. Seven tags were not used by any of the libraries studied, but these tags cover items that will appear in machine records produced by the National Library of Medicine, the National Agricultural Library, and the British National Bibliography.

3) Of 126 indicators, only 18 were used by at least half of the libraries. The highest degree of acceptance was the use of the same two indicators by six libraries. On the other hand, each indicator was used by at least two libraries.

4) Of 181 subfield codes, 176 were used by at least half of the libraries that were using the related tags. Each subfield code was used by at least a quarter of the libraries that could express a relevant opinion.

The foregoing analysis confirmed the view that a nationally distributed record should be as rich in content designation as possible. Failure to provide this detail would result in many libraries having to enrich the record to satisfy local needs, a process more costly than deleting items selectively. Therefore, as of now, the present MARC II format constitutes the level required to satisfy the national distribution function.

The National Union Catalog Function

As noted above, the NUC function relates to the use of machine readable records to build a national union catalog. At first thought, it might appear that this function overlaps the distribution function. As far as Library of Congress cataloging is concerned, this view is correct. It is valid also with respect to cooperative cataloging entries issued by the Library as part of the card service. However, the two functions are quite distinct as far as regular reports to NUC are concerned.

The essential difference between the two categories of catalog records is that those issued as LC cards have been completely checked against the Library's authority files and edited for consistency, whereas only the main and added entries of NUC reports have been checked for compatibility. The impact of this difference can be judged from the fact that an attempt to distribute NUC reports as proof slips several years ago was abandoned because the response to this service did not justify its continuance.

Distributing NUC reports in machine readable form would add another dimension to the problem of processing them, because, to be flexible enough for wide acceptance, NUC reports would have to be entirely compatible with those issued by the MARC Distribution Service. Since compatibility would involve more detailed content designation than many libraries might put into their records for local use, libraries would have to be willing to provide this detail in NUC reports, or the level of NUC reports would have to be upgraded centrally. As the certification of the bibliographic data and the content designators would entail a major workload for the Library of Congress, it does not seem practical to pursue this goal at present.

It is possible, however, to define a subset of content designators to cover the eventuality that outside libraries may be able to report their holdings to NUC in machine readable form. A MARC subset can be determined for the NUC function because this function involves processing records in a multiplicity of places to be used centrally for specifically definable purposes. The distribution function, on the other hand, involves the preparation of records at a central source to be used for a wide variety of purposes in a multiplicity of places. The difference is vital when it comes to stating the requirements for the two types of records.

The specifications of a machine readable record to fulfill the NUC function depend on the nature and functions of the national union catalog itself. The content designators for such a record will be defined in a separate investigation now being conducted by the Working Task Force. The present study was considered to be completed once the feasibility of defining a level of machine readable record for that purpose was established.

CONCLUSION

The findings of this study of the feasibility of defining levels of machine readable bibliographic records are as follows:

1) The level of a record must be adequate for the purposes it will serve.

2) In terms of national use, a machine readable record may function as a means of distributing cataloging information and as a means of reporting holdings to a national union catalog.

3) To satisfy the needs of diverse installations and applications, records for general distribution should be in the full MARC II format.

4) Records that satisfy the NUC function are not necessarily identical with those that satisfy the distribution function.

5) It is feasible to define the characteristics of a machine readable NUC report at a lower level than the full MARC II format.

REFERENCES

1. RECON Working Task Force: *Conversion of Retrospective Catalog Records to Machine Readable Form* (Washington, D. C.: Library of Congress, 1969).
2. Avram, Henriette D. "The RECON Pilot Project. A Progress Report." *Journal of Library Automation*, 3 (June 1970), 10-22.
3. RECON, op. cit., p. 43.
4. Ibid., p. 164.