

Frederick G. Kilgour
Director
The Ohio College Library Center
Columbus, Ohio

SUMMARY OF MARC

MARC is a label attached to an increasing variety of formats containing machine-readable cataloging information. However, MARC is first and foremost a format for interchange of bibliographic information, and BNB MARC and MARC II derive from the proposed "USA Standard for a Format for Bibliographic Information Interchange on Magnetic Tape."¹ If a format is to bear the label MARC it should have the capability of generating a standard communication format derived from the USA standard.

Although this standard in no way prescribes data contained within a MARC format, those data are unfortunately beginning to acquire the label. Hence, codes and cataloging information are also beginning to be known as MARC. Nevertheless, this paper will attempt to distinguish between format and content.

Applications of machine-readable cataloging information must surmount various categories of obstacles to achieve routine operation, and until the present time relatively little has been achieved if one views the library community as a whole. However, from the viewpoint of research and development the attainment is considerable and successful. Although the LC MARC project has been designed for the use of LC cataloging information, primarily for catalog card production, the data on MARC II tapes have been used not only for catalog card production, but also for bookform catalog production and a variety of other listings, such as acquisitions lists and selective dissemination of information (SDI) reports. It has also been used in the design of on-line catalogs.

Clearly, MARC II and BNB MARC are successes. They communicate bibliographic information in a useful format, and the bibliographic information itself is being used. Nevertheless, applications of the MARC format and its contents have revealed a few areas—surprisingly few—where evolutionary improvement may be expected. On the whole the MARC formats are astoundingly well designed.

It cannot be too often emphasized that the primary characteristic of MARC communication formats is physical; the USA standard prescribes only physical characteristics. The MARC communication format is analogous to the 75 by 125 mm. catalog card employed by the unit-card system with its standardized format for main entry, body of title, physical description, notes, and tracings. Like the

standardized unit card, the MARC communications format is designed primarily to facilitate interchange of cataloging information.²

PROCESSING PROBLEMS

The subcommittee that designed the USA standard attempted to produce a format for communications that would also be efficient for processing. On the basis of present experience, it appears that the committee almost, but not quite, attained this goal. The bringing together of directory information as the second logical group of data in the standard format has proved inefficient for space and processing. Relocation of tag and length of field information as a leader for each tagged element of data, thereby making possible omission of starting character position, greatly facilitates additions or subtractions to the record, for it is no longer necessary to recalculate affected starting character positions. This change is relatively minor, but it does reduce processing costs.

A beneficial, general principle for processing bibliographic data is that whenever possible, processing should be done only once and results that may be reused should be saved. At the present time, MARC II does not provide for storage of sort keys (whereas BNB MARC does provide some sort data) or for information useful in formatting call numbers. The Library of Congress is currently working toward an amendment to the MARC II format that would provide for retention and communication of generated sort keys. Similarly it would be useful to provide a technique for communicating LC call numbers in the 050 field in a partially formatted status. The Dewey classification number in the 082 field presents no problem.

A whole series of questions related to the minimum amount of data required for a useful MARC record still remain unresolved. For catalog construction these records may go into bookform catalogs, card catalogs, or on-line computerized catalogs. Each of these catalogs has different requirements. For use in a bookform catalog the sort key should be in sufficient detail to enable machine alphabetizing of all entries in one grand sort, but the cataloging data can be somewhat less complete than that required in a card catalog. For example, tracings serve no purpose in a bookform catalog, and there is somewhat less of a requirement for determining an exact position of a main entry since it is possible for the eye to glance over fifty to a hundred entries at a time.

A record to be used to produce a card to go into a card catalog requires a less precise sort key, since the cards will only be roughly alphabetized for filing. However, determination of the main entry must be more precise, for it is more important in a card catalog that the card be in its correct location. Also, it is necessary that at least one card in one catalog contains tracings that reveal the location of other cards describing the same book.

Although there has been little experience with on-line catalogs, it seems apparent that a sort key probably will not be used to arrange records in a file and that only the most rudimentary of sort keys would be required for possible alphabetizing of a score or more records for displaying on a cathode ray tube terminal. Since the user would almost always be working with small displayed catalogs containing only tens of items, it seems unnecessary to elaborate on bibliographic description beyond the originals on title pages and at the heads of articles.

From the previous paragraph it is clear that MARC records for on-line catalogs require considerably less data than do MARC records for production of catalog cards or bookform catalogs. Hence, the question arises as to how complete any given MARC record should be and still be usefully interchangeable. Clearly, MARC records for an on-line catalog can be more simple than those from which catalog cards or bookform catalogs would be produced. Therefore, it would burden those constructing on-line catalogs to require that their cataloging data also be adequate for manufacture of catalog cards and bookform catalogs. Rather, it appears that it would be more useful to define a minimum record that could be used as a base by another user to construct a more complicated record required for the second user's purposes.

Horseback opinions expressed in committee meetings will not be sufficient to determine such a minimum standard. Rather, it will be necessary to design in detail a MARC system based on minimal data. Once such a system has been designed and tested, the minimum standard can be established.

Although MARC II provides for abridged titles in accordance with Rule 133B of the *Anglo-American Cataloging Rules*,³ it does not provide for production of abridged titles as added entries when the abridgement does not include the first words of the title. A significant portion of libraries employ abridged titles or striking titles that do not include first words of the title, thus the addition of this capability to the MARC II system would enhance it for many users.

PROBLEMS WITH CALL NUMBERS

Cataloging codes in the twentieth century have largely neglected subject added entries and appear to have entirely neglected call numbers. This absence of rules for the construction of call numbers, associated location stamps, and holdings information generates the most difficult problems to solve in the computerized production of catalog cards. In the OCLC's (Ohio College Library Center) catalog production system it has been necessary to add subfield codes only in call number fields. OCLC employs "\$*" to indicate that a call number has been partially formatted, "\$+" to identify a location

stamp, "\$-" to identify holdings information, and a "\$," to indicate that a class number in a call number field is not followed by a book number.

Because of the seemingly uncontrolled variety of information included in Library of Congress call numbers, it can be said that LC call numbers are pseudo free text. Therefore, it is not possible to parse LC call numbers with confidence. Three examples of bizarre call numbers follow:

<u>Card Number</u>	<u>Call Number</u>
64-62399	\$aJ82\$bD792 May 8, 1964
74-604824	\$aQE75\$b.P9 no. 280 etc.
70-449376	\$aJX197\$b.A2 E/ECE/665

If call number data are to approach free text as LC call numbers now do, it will be most helpful in the mechanized formatting of such call numbers if a separator code is placed at the end of each logical segment of the number.

Call number problems in MARC II data are further enhanced by the Library of Congress's practice of placing in the 050 field—Library of Congress call number—whatever the Library of Congress uses as a call number. It is suggested that the term Library of Congress call number should be defined as a Library of Congress classification number to which a derived book number has been added thereby yielding a call number. It is this type of call number that is useful to other libraries.

However, MARC II records emanating from the Library of Congress contain three other types of information in this field. First, there is the case of a Library of Congress call number as just defined, but followed by a \$a and an alternate class number. This condition occurs most frequently, but not necessarily entirely, in the case of LC PZ3 and PZ4 call numbers. Second, there is the case of only a classification number appearing, and third, the word "Law" is a frequent occupant of the 050 field.

None of these three types of data are Library of Congress call numbers as ordinarily defined, and none can be used as they exist in the Library of Congress call number field as call numbers on cards for other libraries. Therefore, in processing the 050 field it is necessary to determine if this LC call number field actually contains an LC call number, or what part of the data is an LC call number. The OCLC catalog production system puts out a unit card whenever case two or three occurs.

On the MARC II weekly tape numbered V2N002 that was distributed on April 2, 1970, 12 percent of the records bore either case one, two, or three types of data. However, if it were only 1 percent of the

records that had non-Library of Congress call numbers in the 050 field, it would still be necessary to expend needless processing time on every record to determine the character of the data in the 050 field. It is suggested that the MARC II record would be improved by designating a new separate field for each of the three cases presented in the previous paragraph, thereby enhancing the efficiency of processing.

CATALOGING DATA

There are no apparent problems associated with the computer processing of MARC II cataloging data. To be sure, the error rate in cataloging data is uncomfortably high, and it has been rumored that an analogous circumstance exists in BNB MARC records. Presumably the present frequency of errors in both LC MARC and BNB MARC will diminish when the two institutions begin routine applications of their records for their own purposes.

MARC II contains Library of Congress cataloging, and as Richard E. Coward has pointed out,⁴ this cataloging does not always conform to Anglo-American cataloging rules because of the inability of the Library of Congress for administrative reasons to adopt these rules in their entirety. Therefore, MARC II cataloging data do not conform to either American or international cataloging standards.

The existence of nonstandard cataloging data in MARC II raises a problem that should be resolved. It is inconceivable that the Anglo-American rules can be discarded at this time. However, it is not inconceivable that the Library of Congress could prepare cataloging data for MARC II that would conform with the Anglo-American rules, or at the very least could indicate in a record a cataloging element or elements not in conformity with the rules. Resolution of this problem would enhance the international acceptance of a MARC format, as Coward has already shown.

Since BNB MARC and MARC II were created, experience makes it abundantly clear that the MARC-type format for machine-readable cataloging information is a success. In the case of MARC II, there appear to be only three relatively small problems associated with computer processing. First, there is the awkwardness of the directory being located in front of all the data instead of being spread out through the data with each pertinent piece of directory information being located at the beginning of each field. Second, there is the problem of the generation and retention of sort keys on which the Library of Congress is now working. Third, there is the problem of formatting call numbers. This problem exists because of the lack of detailed rules for setting up call numbers, and it is suggested that such rules be drawn up.

MARC II cataloging information does not conform with the

Anglo-American rules. This nonconformity can generate problems for libraries that have established the policy of conforming to Anglo-American rules in their entirety. Hence the problem should be resolved, and there are presumably several avenues that could lead to an effective resolution.

The MARC format is a standardized format and like all standards can be expected to evolve. It is gratifying that at this early moment in the history of MARC there appears to be minimal evolutionary pressure to attain full success in a classical cataloging environment.

REFERENCES

1. United States of America Standards Institute. Z39 Committee. Subcommittee on Standardization in the Field of Library Work and Documentation. "USA Standard for a Format for Bibliographic Information Interchange on Magnetic Tape," *Journal of Library Automation*, 2:53-65, June 1969.
2. Kilgour, Frederick G. "Standardization for Interchange of Cataloging Records—MARC II. In *Proceedings of the Third International Congress on Medical Librarianship*. (In press.)
3. *Anglo-American Cataloging Rules*. Prepared by the American Library Association, the Library of Congress, the Library Association, and the Canadian Library Association. North American Text. Chicago, ALA, 1967, p. 195.
4. Coward, Richard E. "MARC International," *Journal of Library Automation*, 2:181-86, Dec. 1969.