

UNDERSTANDING MARC : ANOTHER LOOK

Roy Chang

Cataloging Coordinator

Rene Raatjes

*Cataloging Librarian
The University Library
Western Illinois University
Macomb, Illinois, U.S.A.*

Abstract

MARC format has been widely used and discussed in our profession. However, there appear to have a wide spread misunderstanding of its real structure and attributes. This article discuss the needs for us to understand it a little more. Also, it presents the general misconceptions about MARC, the compatibility of MARC, the structure of MARC, standardization and data communication, and some major issues related to MARC format. In this library automation age, MARC is a key element in library services, and it deserves us to take another look.

MARC has been a familiar term in the library field for more than a decade. Most librarians know that MARC stands for Machine-Readable Cataloging, and many have a general, if somewhat vague, understanding that the MARC structure forms the basis for the manipulation and communication of bibliographic data by means of a computer. Any library using OCLC, RLIN, or one of the other bibliographic utilities to catalog its books, has had to come

to terms with the MARC format, at least on a practical working level. In recent years, however, the influence of MARC has extended beyond the cataloging unit. As more libraries move toward the use of circulation control systems, online public access catalogs (OPAC), and various local automated systems, the need to have bibliographic records conform to the "MARC standard" has grown increasingly important. Librarians are becoming aware that MARC records furnish the major source of data for building online catalogs and are beginning to realize that their choice of library automation system is dependent on the ability of that system to support the MARC standard.¹ The time has arrived when a fuller understanding of the MARC format can place librarians in a more secure position from which to participate in database decision-making and maintenance. Never has it been more important to take another, closer look at MARC.

Popular Misconceptions About MARC

Although the MARC record format has been widely used and discussed, most librarians have only a limited understanding of it. Even if their library uses a bibliographic database system, it has not been necessary for librarians to know exactly how records are processed or in what shape they are stored in the machine or on archival tapes. Public service librarians have viewed MARC as just another catalog format replacing the old 3" x 5" catalog card by a record display on a computer screen. For most catalogers, the extent of their required knowledge of MARC has been the input format which appears on their computer screen and which they often erroneously believe to be identical with the MARC record format. The example below shows a display of the input format designed by OCLC to accommodate the data elements required to make a complete bibliographic record. It uses codes known as tags and indicators to identify various elements of bibliographic data

(referred to as fields). Within most fields, it uses subfield codes known as delimiters, to further narrow the identification of data elements:

```

010      gb84-1258
020      0063182785 (pbk)
050 1    HQ814
100 10   Walczak, Yvette
245 10   Divorce : b the child's point of view / c Yvette
          Walczak with Sheila Burns.
260 0    London : a San Francisco : b Harper & Row, c 1984
.....

```

A more sophisticated but still incomplete understanding of MARC recognizes that the MARC format also contains fixed fields which accommodate data such as language of text, presence of a bibliography, source of cataloging, etc. In an OCLC display these appear at the top of the record, as shown below:

```

Type: a  Bib lvl: m  Govt pub:   Lang: eng  Source: d  Illus:
Rep:      Enc lvl:   Conf pub: 0  Ctry: enk  Dat tp: s  M/F/B : 10
Desc: a   Int lvl:   Dates: 1984
010      gb84-1258
...

```

Another misconception involves the understanding of the term "full MARC record", which is often thought to be one that displays all the tags required by the data in the record, rather than one which represents the standard USMARC format.

In actuality, few librarians have ever seen the real standard record format known as LCMARC or USMARC, since it was designed for the computer, not for the human eye. The display formats shown above are only the visible part of the MARC record. The other parts of the real MARC record are the leader and the directory, both invisible in online systems, but "vital to

communication and some forms of processing” You may never see a leader or a directory on line, but those elements make USMARC processing efficient and flexible.² In the real MARC record, the tags are not included with the indicators attached to a particular field. Nor does the USMARC leader match the same fields as the fixed fields of the familiar input format. The visible input formats used by the various bibliographic utilities (OCLC MARC, UTLAS MARC, etc.) all have the same general structure as USMARC, but vary from each other in their use of extended non-USMARC fields. For example, OCLC MARC uses an 049 tagged field to show item holdings, while RLIN MARC uses 95X tagged fields for the same purpose.

Most bibliographic database systems use the formats shown above to display a bibliographic record in place of the conventional 3" x 5" card format. Some systems, however, offer library users the option of viewing the same record online in a 3" x 5" card format or, as in the case of OCLC, may also produce cards offline for use in the library's card catalog. As accustomed as librarians are to the traditional catalog card, they are becoming increasingly comfortable with the tagged format, which offers more room and flexibility in displaying and storing a bibliographic record.

Why Bother Trying to Understand MARC

Despite some of their misconceptions, librarians have for years managed to make practical use of MARC. So why bother trying to learn what the MARC format really is? Walt Crawford, in the introduction to his book, *MARC for Library Use*, summarizes the reasons that today's librarians need to increase their understanding of the MARC record:

Many librarians create and use MARC records without ever understanding the nature of MARC itself. While no such understanding is

required for cataloging, librarians need to know more about MARC as their uses of computers expand. A thorough understanding of MARC will help when dealing with vendors of services, when considering online catalogs and other automated systems, and when considering possible local development of automated systems.³

In recent years more and more libraries are implementing some type of automated system. Because of this trend, the need to understand the MARC record format takes on new importance. Most librarians have heard enough about MARC to know that it represents the standard for machine-readable record formats. As computer applications become more common in libraries and the opportunity to share bibliographic records increases, librarians are becoming more conscious of the need for standardization of bibliographic records. We need to be assured that our automated systems are in accord with whatever standards MARC has established. When faced with the responsibility of choosing an automated system, the librarian must assess the capability of that system to accept, store, and process MARC records. Since most vendors claim that their systems are MARC compatible, it is important that librarians know enough about MARC to be able to verify such claims.

The Structure of MARC

A complete description of the MARC structure is beyond the scope of this paper. Readers who are interested in the detailed specifications of MARC should refer to the MARC documentations of Library of Congress or of national bibliographic systems such as RLIN, WLN, UTLAS, or OCLC and to W. Crawford's *MARC for Library Use*. However, a general description of the real structure of MARC may be helpful in visualizing the overall picture.

The MARC format is divided into three main parts: the leader,

the directory, and the variable fields. The variable fields are, in turn, subdivided into two groups: the variable control fields and the variable data fields. The following example illustrates the overall format of a MARC record:

| LEADER | RECORD DIRECTORY | VARIABLE FIELDS |
|-----------------|------------------|-----------------|
| ----- | ----- | |
| 0123456789..... | 23 | |

The MARC record begins with a 24-character field commonly referred to as the "leader" of the record. The leader contains specifications necessary for the recognition and basic computer processing of the entire record. The first 5 characters of the leader specify the length of the record. Since the leader is a numeric field, the number indicated in the field will be right-justified, e. g., 00845. The 5-numeric-character field implies that a MARC record can be as long as 99999 bytes. Users of the MARC record do not see this data on the screen. For example, what the user sees in the beginning position of an OCLC record is the record identification number. This should not be confused with the record length, which occupies the beginning position in a USMARC record. Following the designation of record length are three 1-character data fields for coding the record status (new, revised, deleted, etc.), the type of record (language material, music, map, etc.), and the bibliographic level (monograph, serial, etc.). These three fields along with character 17, the encoding level (degree of completeness of record) can be seen online in the fixed fields position of an OCLC record.

Following the 24-character leader is the directory. The directory can be thought of as the "road map to the record."⁴ It consists of a separate entry for each field in the record and shows the tag (3 numeric characters) which identifies the type of field, the length of the field (4 numeric characters), and the starting character position (5 numeric characters). These entries are used by the computer to efficiently locate any tag no matter what system produced the

record. Since the length of the field is specified by 4 numeric characters, it is possible to have a field length of up to 9999 bytes. The tags in a MARC record are located in the directory, not in the variable fields as we are accustomed to see them on the computer screen.

Following the directory are the variable fields. Although relatively few variable fields have been defined, in theory a variable field could be designated for every number between 001 and 999. In other words, 999 tags could conceivably exist within a MARC record. Theoretically, the tags could also be represented by non-numeric characters, such as AAA, AAB, AAC, etc. or the combination of numeric and alphabetic characters, such as 1AA, 2TL, 3ED, etc.

As indicated earlier, variable fields are subdivided into variable control fields and variable data fields. The tags for variable control fields are those beginning with 00, i.e., 001-009. These control fields do not use indicators and subfield codes. Again, there are structural differences between the MARC record and that of a bibliographic database such as OCLC. For example, in USMARC, the 001 field contains the Library of Congress card number, which is used as a control number by the system. In OCLC MARC, however, 001 contains the OCLC record number. This is also a control number, but it is a system-supplied number, entirely distinct from the LC card number. Similarly, most of the fixed fields data shown on the OCLC screen will, in the MARC record, be stored in the 008 variable control field.

Following the variable control fields are the variable data fields. The variable data fields begin with 2-character indicators, which serve a variety of purposes, according to the specific field. For example, one of the indicators in the 245 (title) field specifies whether or not the title should be traced. These indicators are followed by one or more delimiters which precede and identify various elements of the data within the field. As mentioned earlier,

the tag itself is not part of the variable data field. A variable data field for an imprint may look as follows:

(260) 0 &a New York : &b D. McKay Co., &c c1976.

The 260 tag appears only in the directory. The "0" represents the 2 indicators (a zero and a blank), and "&a", "&b", and "&c" represent the delimiters.

MARC Compatibility

Sometimes, "for good reasons" a vendor may advise a library to use a sub-standard MARC record format. The ostensible basis for this argument is that it would take too much space to store the full MARC record, and that nothing would be lost by doing away with the leader, the directory or other data elements which are not visible in the human-readable catalog. If a local system can process OCLC MARC records and produce an online catalog that appears to contain all the data elements needed for the purpose of retrieval and precise identification of a bibliographic record, then does it matter how the records are formatted and stored in the machine or on magnetic tape?

The answer is a definite "Yes".

MARC is a standardized bibliographic record format. A standardized record format can facilitate communication among systems, whether they are linked into an online network or communicate through offline media such as magnetic tapes or magnetic disks. As data sharing and inter-system communications increase, a realization of the importance of the standard record format—MARC—will become better focused. W. Crawford discusses the importance of adhering to MARC compatibility:

MARC compatibility means flexibility, and allows a library to move

towards an integrated system.....MARC compatible systems are designed for the future..... MARC provides a common ground for sharing data; without compatibility, a library is foreclosing such sharing.⁵

D.S. McPherson offers related advice:

When a library evaluates an automated system, concern about record formats may take a back seat to other criteria such as system features and purchase price. In the long term, however, use of a system that does not meet existing standards may prove extremely costly.⁶

We sometimes hear that a library is using a local system to download its own bibliographical records from a national bibliographic database system, by connecting the system to the printing port of the terminal and sending each of its records off the printing port. It is true that the local system can capture all data elements of each record as it appears on the computer screen or print-out. But that record is not the same as the one on the national system archival tape. Therefore, the local system needs to employ an additional program (separate from the program that processes the standard MARC record) in order to be able to process the records downloaded from the terminal.

We also hear at times that certain libraries are using micro-computer packages such as dBASE III + or RBASE 5000 to catalog special materials, and that they are creating their records in MARC format. It is quite possible that such database management systems could be used to produce a true MARC record. However, it would require extensive programming efforts to achieve so complex a record format, because while both dBASE III + and RBASE 5000 can handle fixed field records quite easily, they cannot deal effectively with variable length records. In fact, any relational database management system would be unsuitable for handling variable length records since the relational record characteristically places records in a flat talbe form. Although samples of the output

record format show that MARC tags are used to indicate data fields such as author, title, and subject, the data in these fields is often truncated when space requirements exceed the flat table limits.

Some of the system vendors point to their tagged display format as evidence that their system can process and output MARC records. Such assertions are not uncommon but should not be taken at face value. As such times, the librarian must insist on asking the essential question: Can the local system reproduce a complete MARC record from the records stored in the system, if in the future the records had to be transferred to another system?

It is easy to understand why the Technical Standards for Library Automation Committee (TESLA) of ALA became concerned several years ago about the MARC compatibility of automated library systems being marketed. As a result, it launched a compatibility survey of various vendor's products. The survey indicated that there was a generally strong vendor commitment to the MARC format. "However, there were enough nonstandard practices reported to indicate that MARC compatibility cannot be assumed and the customers should question prospective vendors carefully in a number of areas."⁷

Standardization and Data Communication

The codes and data fields within the MARC structure are not the same as the MARC structure itself. To be MARC compatible, both the structure and the interior codes must follow a set of standards. The MARC structure standard was set by the American National Standards Institute and is known as ANSI Z39.2-1979.⁸ Although the ANSI standard did not specify the standard for tags, indicators, and data-element identifiers (delimiters), those defined by LC have been accepted as standard practice, for example, 100 for main entry-personal name, 245 for title, etc.

The need for strict adherence to the standard is related to the

nature of computers. The computer simply cannot recognize any departure from a standard code. The slightest variation would mean a total loss of information. For example, if one system used a 300 tag to identify the physical description field and another system used a 301 tag, it would be impossible to communicate the information between systems. To be able to transfer a computer record from one database to another, the record format and the subfield codes must be compatible. Without compatibility, a whole set of new programs would have to be developed to process the records being transferred from one system to another.

A data processing standard of such rigidity is not without difficulties. The need to adhere to a standard means that there will be less flexibility to accommodate local needs. Bibliographic records for individual libraries have an abundance of local characteristics. The various bibliographic data processing centers also need room for local accommodations. For this reason, systems such as OCLC or RLIN have provided additional tags and codes not included in the standard MARC format. In the OCLC system, for example, some of the extended tags include 049 (local holdings), 090 (locally assigned LC-type call number), and 949 (local notes). Many additional subfield codes have also been provided within each system.

The OCLC 049 field contains several system-defined subfields as well as two subfields, &1 and &0, which can be locally defined by the individual library. Similarly, OCLC designates the entire 949 field for local system usage. Both indicators and subfield codes within the 949 field are meant to be defined locally. It is conceivable that the entire bibliographic record could be redefined to fit into this local field. However, one should remember that the 949 field is intended for local usage only. Over-extension of locally-defined codes would, of course, diminish the communicability of the record.

The OCLC system will accept data entered into the local usage

fields, but it will simply leave them there as part of the record; it will not do anything with them. Other tags or subfields not defined by USMARC or OCLC MARC will be rejected by the system.

Because each bibliographic utility has adopted a somewhat different set of extended tags and subfield codes, bibliographic communication among these systems requires additional processing. For a discussion of alternative ways to resolve conflict in communication among systems, see R. Renaud's "Resolving Conflict in MARC Exchange."⁹

Computer programs for reading a MARC record are much easier to develop than those for constructing a MARC record. Many vendors will use MARC records produced by LC, OCLC, and RLIN, etc. as input data, but they will not reconstruct MARC records for other systems to use. In some cases, the local system will use a totally different structure to store records. For example, the LCS system in Illinois has a much simplified record format, although its structure still maintains the framework of leader, directory, and variable fields. A specially constructed program would, again, be required to process these records.

Issues Concerning MARC

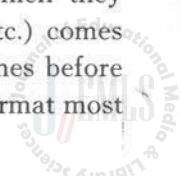
There is no question that MARC is a highly complex record format. To develop a program for reading the MARC record would not be an easy task for a novice programmer. The easiest record for a programmer to work with and for a computer to process is one containing limited data elements which can be entered into fixed rather than variable fields. For example, a record consisting of library staff names and telephone numbers requires a very simple format. Both the name and the number can be treated as fixed field data.

On the other hand, a book record with title and author data is a difficult record format, even though it also involves only two data

elements: the title and the author. A book title can vary greatly in length from very short to very long. Similarly, there can be a single author, no author, or more than one author indicated. The length of the author's name is also widely variable. Since a large portion of the MARC record consists of variable fields, it is easy to foresee the difficulties of developing processing programs. The presence of numerous subfields, especially the complex subfields in the 049 field of OCLC MARC present added difficulties for processing. To complicate matters even further, many of the tags and subfields can also be used repeatedly within the MARC structure. All in all, the large number of tags, indicators, and subfield codes combined with variable-length data elements make the MARC format a highly complex record structure to deal with.

Is there a good reason for the complexity of MARC? We may reply that it is the nature of the bibliographic record that makes the machine record structure so complex; and it is the users' information needs that, in turn, dictate the nature of the bibliographic record. We librarians require a record format that will accommodate all needed data elements related to a bibliographic record. MARC, with its flexibility to accommodate multiple authors, multiple subjects, and all types of subject headings, is a format designed to fill this need.

It is inaccurate to maintain that in developing a machine-readable record "We put the card catalog in electronic form....."¹⁰ The MARC record is not limited to the traditional access points found on a catalog card, but allows the record to be manipulated in numerous additional ways. For example, catalog records may be accessed by LC card number, ISBN, key words, etc., none of which is accessible in the card catalog. It is true that MARC arranges the variable data elements in the approximate order in which they appear on a catalog card, e.g., the call number (050, etc.) comes before the main entry (100, etc.), and the main entry comes before the title (245), and so on. The designer of the MARC format most



likely reasoned that librarians are accustomed to this order of displaying bibliographic elements. However, the input and output format of a record can be independent of the format that is stored in the computer. Name entries, whether main entries (1XX) or added entries (7XX) can be accessed together, despite their location in different fields. Similarly, series data, whether they occupy 4XX fields or 8XX fields, share identical access. The order of these tags does not prohibit the programmer from reordering them when they are processed. However, once the meaning of these tags, indicators, and subfield codes are set, they should be standardized so that all systems can easily process each others records.

How to utilize the data elements within a MARC record is really up to the local system. As far as computers are concerned, any field can be selected to be indexed for quick retrieval. Any of the data elements can be extracted from the MARC structure for building any type of data model: network, hierarchical, or relational. The flexibility of the MARC record allows for the creation of specialized types of data files. For example, a subject authority file can be created by selecting data entered in the fields tagged 6XX.

The complaint that "The MARC record does not provide adequate subject access to the very materials it has been used to access"¹¹ reflects a misunderstanding of MARC's potential. The MARC format provides the fields for any number of subject headings and sub-headings. It is up to others to use them for providing adequate access. The local cataloger must accept the responsibility for inputting whatever subject entries are deemed necessary for adequate access to any bibliographic record. The MARC format itself cannot be blamed for the failure to make use of its capabilities.

Conclusion

MARC records have long served as the key data source for

library automation systems. These systems must not only be able to meet the bibliographic information needs of the library's users, including the librarians themselves, but they must also embody the standardization of format which is a prerequisite to data sharing and system communication. Although librarians are making increasing use of MARC for both technical and public services, their conception of its real nature remains cloudy. Because MARC is made for "machine-eyes", it is not easy for those librarians who are not yet computer-literate to fully understand its nature and its potential. In the automation age MARC is essential to the library profession. If librarians persist in continuous ignorance of the MARC format, the future electronic catalog could be totally at the mercy of system designers and data processing personnel. The main principles of library service might be severely compromised in favor of convenience of data processing. As G. Patton suggests, the best results will come from the active participation of system experts who understand library functions and librarians who have a fundamental understanding of computer systems.¹² The key to the excellence of future library operations is library automation. Many library systems that are being developed today will have a long-term effect on library services. One of the key elements in system development is the MARC record. To be able to take an active role in the decision-making related to library automation systems, one must have a good understanding of the nature and possibilities of the MARC format. Librarians as a profession must accept the challenge of understanding MARC and how it relates to the future of library service.

Notes

1. Robert A. Walton & F. R. Bridge, "Automated System Marketplace 1987: Maturity and Competition", *Library Journal*, 113 : 6 (April 1, 1988), p. 39.
2. Walt Crawford, *MARC for Library Use: Understanding the USMARC*

Formats (White Plains, N.Y.: Knowledge Industries Publications, 1984), p. 7.

3. *Ibid.*, p. 1.
4. Michele I. Dalehite, "MARC Format on Tape: A Tutorial," in *From Tape to Product*, ed. Barry B. Baker (Ann Arbor, Mich.: Pierian Press, 1985), p. 25.
5. Crawford, p. 52.
6. Dorothy S. McPherson, "MARC Compatibility: A TESLA Survey of Vendors," *Information Technology and Libraries*, 4 (September 1985), p. 241.
7. *Ibid.*
8. For a discussion of the "American National Standards for Bibliographic Information Interchange on Magnetic Tape," see Crawford, p. 29.
9. Robert Renaud, "Resolving Conflict in MARC Exchange," *Information Technology and Libraries*, 3 (September 1984), pp. 255-261.
10. Kevin Hegarty, "Myths of Library Automation," *Library Journal*, 110 (Oct. 1, 1985), p. 471.
11. *Ibid.*, p. 48.
12. Glenn Patton, "Interaction: Letters to the Editor," *Library Resources & Technical Services*, 32 : 1 (Jan. 1988), p. 9.