

The MARC Formats: An Overview

KATHARINE D. MORTON

Abstract: The USMARC Archival and Manuscripts Control (AMC) format provides archivists with the means to communicate and exchange information about holdings in machine-readable form. This article gives a brief history of the USMARC formats developed by the Library of Congress for bibliographic data and authority records and outlines the basic structure of a MARC record. It explains the role USMARC formats have played in the creation of national and international standards for the exchange of bibliographic information and in the growth of bibliographic utilities. The review process for proposed changes to the formats is described.

About the author: Katharine D. Morton is head of Manuscripts and Archives at Yale University Library. She is a member of the Connecticut State Historical Records Advisory Board and the Archives, Manuscripts, and Special Collections Task Force of the Research Libraries Group. She received an M.A. from the University of Wisconsin, an M.L.S. from Southern Connecticut State University, and a certificate from the Case Western Reserve University Advanced Workshop in University Archives. This article is a revised version of a paper presented at the 48th annual meeting of the Society of American Archivists, 3 September 1984, Washington, D.C.

THE USMARC FORMATS HAVE PERFORMED a vital function for libraries and other bibliographic agencies as standards for the representation of bibliographic and authority information in machine-readable form. While the formats were initially perceived to be of little use to archivists, the situation has changed dramatically with the introduction of the MARC AMC format, which provides a new vehicle for exchanging and communicating information about archival holdings in machine-readable form. To be understood and used properly, MARC AMC should be placed in the broader context of the family of MARC formats. This article gives a brief history of the formats and an explanation of the structure of a MARC record. It outlines the role MARC has played in the development of standards and in the growth of the bibliographic utilities. Finally, it discusses how formats are modified and improved and suggests how archivists may participate in this process. A bibliography is appended.

MARC is an acronym for machine-readable catalog. There are MARC formats for seven types of bibliographic materials: books, serials, maps, visual materials, music, archives and manuscripts, and machine-readable records. A format focuses on a type of material and identifies the elements necessary to describe that material, both physically and intellectually. Since each type of material has its own special needs, some data elements appear in one format but not in others.

The data elements of the MARC records are like a store of building materials, carefully prepared and labeled for future use. When a bibliographic record is created in conformity with

MARC construction standards, the pieces of information can be readily identified, retrieved, and manipulated by a computer in a variety of ways to generate a number of desirable products. The MARC formats are frameworks for communication, vehicles designed to carry bibliographic information from one unit or agency to another. When a record appears in a MARC format, the agency knows precisely what it is receiving and what it must do to use that record and the information it contains for its own purposes.

In the mid-1960s efforts got underway at the Library of Congress to make cataloging information available nationally—not only in hard copy, as had been done by the library for many years, but also in a machine-readable form that could be manipulated by other libraries to meet internal needs. In a pilot project funded by the Council on Library Resources, the Library of Congress examined the feasibility of distributing cataloging data in machine-readable form to sixteen selected user libraries. These libraries received weekly tapes and tested the utility of the data in a variety of internal processing applications. The MARC pilot project, as it was called, lasted from January 1966 to June 1968. This successful experiment led to the development of a new, improved communications format for the distribution of bibliographic records, MARC II. The MARC Distribution Service was established as a permanent part of the Library of Congress in order to expand the distribution of cataloging records on magnetic tape.¹

From its inception, the MARC II format was intended as a structure that could be used for all types of bibliograph-

¹For a description of the project, see Henriette D. Avrams, *The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources* (Washington, D.C.: Library of Congress, 1968).

ic data. Practical constraints of time and staffing made it impossible initially to identify and analyze data elements for the full range of materials. It was decided instead to design a format structure that would accommodate all types of bibliographic information and to develop formats for each type of material, one at a time, beginning with books.

Other considerations affected the design of the MARC II format. Because it was a vehicle for the exchange of bibliographic information between systems with independent computer facilities, manipulation of the format by an assortment of electronic data processing equipment had to be possible. The MARC format was also designed to support a range of library processing functions, such as bibliographic display, acquisitions, and information retrieval. It was not intended to mandate formats used by any internal processing system; the assumption was that each system would devise internal formats to suit its own needs. Since MARC records would be used to create a variety of products, including catalog cards, computer output microform, bibliographies, displays on a CRT screen, and accession lists, the for-

mat could not be tailored to the generation of a single product to the detriment of others.²

The MARC format for books first appeared in August 1968 under the title *Subscriber's Guide to the MARC Distribution Service*.³ MARC formats for serials, maps, films, manuscripts, and music were published by 1976.⁴ The practice of issuing separate pamphlets for each type of material was discontinued in 1980 when all previous editions were superseded by a single looseleaf publication, *MARC Formats for Bibliographic Data*, which presents the specifications for all types of material in an integrated, composite format.⁵

MARC Formats for Bibliographic Data is updated periodically, and several major updates have recently appeared. Update number nine contains the MARC format for machine-readable records. Update number ten includes the visual materials format, which is a revised and expanded films format that accommodates prints and photographs. This update also contains the revised MARC format for manuscripts, called the Archives and Manuscripts Control (AMC) format. In addition, update ten is a

²For an account of the development of the MARC II format, see Henriette D. Avrams, John F. Knapp, and Lucia J. Rather, *The MARC II Format: A Communications Format for Bibliographic Data* (Washington, D.C.: Library of Congress, 1968).

³Library of Congress, MARC Development Office, *Subscriber's Guide to the MARC Distribution Service* (Washington, D.C.: Library of Congress, 1968). The first, second, and third editions, 1968-1969, appeared under this title. The latest edition, published in 1972, is *Books: A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Books*, 5th ed.

⁴Library of Congress, Information Systems Office, *Serials: A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Serials*, preliminary edition (Washington, D.C.: Library of Congress, 1970); Library of Congress, Information Systems Office, *Maps: A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Maps* (Washington, D.C.: Library of Congress, 1970); Library of Congress, MARC Development Office, *Films: A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Motion Pictures, Filmstrips, and Other Pictorial Media Intended for Projection* (Washington, D.C.: Library of Congress, 1971); Library of Congress, MARC Development Office, *Manuscripts: A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Single Manuscripts or Manuscript Collections* (Washington, D.C.: Library of Congress, 1973); and Library of Congress, MARC Development Office, *Music: A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Music Scores and Musical and Non-musical Sound Recordings*, draft (Washington, D.C.: Library of Congress, 1973).

⁵Library of Congress, Automated Systems Office, *MARC Formats for Bibliographic Data* (Washington, D.C.: Library of Congress, 1980).

cumulation of all previous formats: books, serials, maps, music, visual materials, archives and manuscripts, and machine-readable records. The Library of Congress is also working on a second edition of *MARC Formats for Bibliographic Data*, which will incorporate all previous changes and contain additional introductory material.

In addition to this family of formats for bibliographic records, there is a MARC format for authority records. These records specify the correct headings to be used for names, subjects, or series, based on the *A.L.A. Cataloging Rules for Author and Title Entries*, the *Anglo-American Cataloguing Rules*, and the *Library of Congress Subject Headings*.⁶ Authority records are vital in maintaining consistency of entry in a catalog or data base. If the name of a person or corporate body is entered in a variety of forms on different occasions, those entries will be scattered, and some may be overlooked by the patron. Authority records are kept to lessen such inconsistencies; this format provides the specifications for authority records maintained in machine-readable form.

A MARC format has also been developed for records of holdings and location information. The existence of a standard format facilitates the exchange of holdings and location information in machine-readable form. This is especially important for holdings information on serial publications,⁷ but the format is

designed to accommodate all types of material. This work was a joint project of the Library of Congress and the South-eastern Association of Research Libraries. The final draft of the format was issued in February 1985.⁸

The basic elements of a MARC record are the structure, the content designation, and the data itself. The structure is the overall framework for the machine-readable record, its container, if you will. Since a machine-readable bibliographic record is essentially a string of characters on magnetic tape or other medium, content designators are necessary to identify and to set apart the data elements in the record. The data elements can then be manipulated by a range of electronic data processing equipment for a variety of processing functions. In the MARC records, the content designators are called tags, indicators, and subfield codes. The content of those data elements in the MARC record that comprise a traditional bibliographic record is defined by standards outside the formats. These standards may be in the form of manuals, such as the *Anglo-American Cataloguing Rules*, or identifying numbers, such as those assigned in the *National Union Catalog of Manuscript Collections*.⁹ For example, the MARC format does not specify how to determine a main entry, just that one is needed. The content of other data elements that are represented by codes in the MARC format is defined within the USMARC formats. There are lists of codes for country

⁶*A.L.A. Cataloging Rules for Author and Title Entries*, 2d ed. (Chicago: American Library Association, 1949); *Anglo-American Cataloguing Rules*, 2d ed. (Chicago: American Library Association, 1978); and Library of Congress, Subject Cataloging Division, *Library of Congress Subject Headings*, 9th ed. (Washington, D.C.: Library of Congress, 1979). There are quarterly and annual supplements to the latter volume.

⁷Serials are defined in library cataloging terms as publications in any medium issued in successive parts bearing numerical or chronological designations and intended to be continued indefinitely. Newspapers, newsletters, and professional journals are serials.

⁸Copies may be obtained from the Cataloging Distribution Service, Library of Congress.

⁹Library of Congress, Manuscripts Section, *National Union Catalog of Manuscript Collections* (Washington, D.C.: Library of Congress, 1959/61-).

Variable Fields of MARC Formats for Bibliographic Data	
Functional Blocks by Tag Number	
00X	Control fields
0XX	Variable fields, general information
1XX	Main entry
2XX	Title & title paragraph
3XX	Physical description
4XX	Series notes
5XX	Bibliographic notes
6XX	Subject entries (name or topical)
7XX	Added entries other than subject or series
8XX	Series added entries
9XX	Local use

Table 1

of publication, geographic areas, and languages, for example, in the appendixes of the *MARC Formats for Bibliographic Data*.

Structurally, a MARC record consists of three main parts: the leader, the record directory, and the fields. The leader consists of the first twenty-four characters of the record. It describes the record that follows and provides the information necessary for the machine-processing of the record. This information includes the total length, type, bibliographic level, and status of the record.

The second major part, the record directory, serves as an index to the exact location of each field, or section, in the record. This directory comprises a series of fixed-length, twelve-character entries

that describe each field in the record. This information consists of the tag, or numeric identifier of the field; the starting character position, or location, of the field within the record; and the length of the field. The record directory makes possible a random arrangement of fields on the storage medium and facilitates the addition of new fields to the record.

The fields, the third part of the MARC record, organize and contain the basic descriptive information. There are eleven functional groups, or blocks, of fields for bibliographic records (see Table 1).¹⁰ The length of the fields may vary according to the nature of the data they contain. This is an important feature, since bibliographic records vary greatly in size. A

¹⁰Some of the library terminology used in Table 1 may be unfamiliar to archivists. As the name implies, a main entry is the principal access point under which the bibliographic record is presented, most commonly the name of an author or creating agency. Series notes (400s) and series added entries (800s) are unrelated to record series but instead refer to serial publications such as journals, newspapers, and monographs published as part of a series. The subject entries (600s) and added entries (700s) represent terms in addition to the main entry under which the record is indexed in the data base. These may be names of individuals, names of corporate bodies, topical headings, or geographic names. This table gives only a basic outline of the MARC fields; within each block, scores of fields are identified and defined in the *MARC Formats for Bibliographic Data*.

three-digit tag, stored in the record directory, identifies each field. The fields for authority records have the same numeric-block structure as the bibliographic records. Because authority records are functionally quite distinct from bibliographic records, the fields are defined very differently. For example, rather than main entry, title, and physical description, the terms "headings," "see and see also references," and "tracings," are used.¹¹

The USMARC formats are part of a larger network of communication standards and have played a major role, both nationally and internationally, in the development of standards for the exchange of bibliographic information. Building largely on the development of MARC formats in England and the United States, the International Organization for Standardization adopted in 1973 a format structure for the exchange of bibliographic information from one system to another (ISO 2709).¹² Although not itself a bibliographic format, ISO 2709 provides the outline or framework for the design of operational bibliographic formats, such as the USMARC formats. In the United States, the American National Standards Institute (ANSI) has established a standard (Z39.2) for the structure of a record.¹³ Like the ISO standard, it provides a structure from which any number of implementations can be designed. Thus, USMARC formats are in conformity with both the national and international standards for the exchange of information on magnetic tape.

The USMARC formats have served as models for bibliographic exchange formats in other nations. Since the issuance of the MARC II format by the Library of Congress, over twenty countries have developed and implemented MARC formats. Although there are considerable differences in the national MARC formats, virtually all follow the structure of the ISO standard.

The existence of these national MARC formats has necessitated the development of an international exchange format for MARC records. UNIMARC, the acronym for universal MARC, is a format developed under the auspices of the International Federation of Library Associations.¹⁴ Built on the ISO standard, UNIMARC provides a vehicle for the exchange of information among countries with their own MARC formats. With UNIMARC, an agency does not have to develop conversion programs for each of the MARC formats it handles.

The USMARC formats are also linked to work on International Standard Bibliographic Description (ISBD). ISBD is an effort to standardize practices in the content and arrangement of a bibliographic record. With ISBD, data elements in a record can be recognized more easily despite the language of their content. ISBD also facilitates the automated manipulation of the data. UNIMARC incorporated these standards, and it is hoped that the adoption of ISBD by more countries will increase the similarities in the records they produce.

In the United States, the Library of Congress has led in the development of

¹¹All of these terms refer to headings or to the record of headings that appear for items in a catalog or an online file.

¹²International Organization for Standardization Documentation, *Format for Bibliographic Information Exchange on Magnetic Tape*, ISO 2709-1973.

¹³American National Standards Institute, *American National Standard Format for Bibliographic Information Exchange on Magnetic Tape*, ANSI Z 39.2-1971.

¹⁴International Federation of Library Associations and Institutions, *UNIMARC: Universal MARC Format*, 2d ed. rev. (London: International Federation of Library Associations, 1980).

standards that define the data elements to be included in a record that is shared with other institutions or contributed to a national data base. Known as the National Level Bibliographical Record, this basic record is the minimally acceptable MARC record when full cataloging is not possible or desirable.¹⁵

The USMARC formats have had a significant effect on national and international library activities through the development of standards. The distribution of MARC bibliographic records by the Library of Congress has had an even greater impact on library technical processing operations in the United States. For recently published materials, cataloging in many libraries is now a process of modifying an existing record for local use rather than creating an original one.

The availability of MARC bibliographic records has spurred the growth of regional library networks and bibliographic utilities such as OCLC (originally the Ohio College Library Center, now the Online Computer Library Center), RLG (the Research Libraries Group), and WLN (Washington Library Network). Each utility acts as a central processing facility that gathers and transmits bibliographic records for member institutions. The MARC tapes received from the Library of Congress are an important component of the utility's operations. The tapes are loaded centrally into the utility's data base. A member library locates a bibliographic record in the central data base, modifies the record online to suit its internal needs, and orders catalog cards to its specifications from the utility. If the item has not already been cataloged, the member library enters the record it creates into the data

base so that it can be used by other libraries as well for derived cataloging. Since the data base reflects the holdings of the members, it provides the information needed for interlibrary loan activity. Thus growth of the central data base is of primary importance to both technical services and public service functions.

In addition to the creation of MARC records for recently published materials, the retrospective conversion of older bibliographic records to machine-readable form is highly desirable. For acquisitions and cataloging, it is useful to have records for older titles that are still actively added to libraries. Online access to these records can facilitate interlibrary loan and other public service activities as well. The Library of Congress's pioneering efforts in retrospective conversion greatly expanded the number of MARC records available for distribution. Other cooperative "recon" projects, such as the CONSER project for serials, have made significant additions to the pool of shared MARC records. Manuscript and archival repositories have recently become involved in a similar project: three repositories have completed the entry of their records into the Research Libraries Information Network (RLIN) with Title IIC funding from the Office of Education, and twelve others are currently undertaking retrospective conversion projects into the same data base with funding from the National Endowment for the Humanities and the Pew Trust.

The MARC formats have greatly facilitated the integration of bibliographic information for all types of material in centralized data bases. While this has advantages for the technical processing of materials, it has an even greater impact

¹⁵Library of Congress, Processing Services, *National Level Bibliographic Record* (Washington, D.C.: Library of Congress, 1980-). NLBR documents based on the MARC books, films, music, maps, and serials formats have been issued, and second editions of each are planned. A document for machine-readable data files is in preparation.

on the use of the materials by the researcher. Materials that may have been described in physically separate, unfamiliar, or overlooked files can be retrieved through an integrated approach to bibliographic materials, regardless of type.

The extensive use of USMARC formats in a variety of situations has led to the review and modification of the formats. The Network Development and MARC Standards Office of the Library of Congress oversees the review of proposed changes. It is assisted by a USMARC Advisory Group, which includes the American Library Association's Committee on Representation in Machine-Readable Form of Bibliographic Information (MARBI) and representatives of the national libraries, bibliographic utilities, special library groups, and others. The Society of American Archivists, through its Committee on Archival Information Exchange (CAIE), has been part of the USMARC Advisory Group since 1983. Institutions using MARC AMC can submit requests for changes to CAIE for recommendation to the MARC Standards Office. The Cataloging Distribution Service of the Library of Congress makes available, on a yearly subscription basis, a loose-leaf service called *USMARC Formats: Proposed Changes*. This documents suggested changes, the process of review, and the final decisions on the USMARC bibliographic, authorities, and holdings formats.

Many proposed changes in the MARC formats result from particular needs of users or from particular characteristics of certain types of material. The major revision represented by the MARC AMC format is a prime example of this process. Other proposals for change have resulted from a broader analysis of the format and its uses. The Library of Congress sponsored an extensive review of the MARC formats in 1980/81. A detailed list of problem areas within the formats was prepared for further study.¹⁶ One such area was mixed format materials—items or entities that could logically be placed in two or more formats. (An example of mixed format material is a map that is also a serial.) The selection of one format over the other could preclude the use of data elements that are needed but have not been defined as applicable to that kind of material. The review process prompted discussions on the feasibility of a total integration of all of the bibliographic formats created to date and the definition of data elements that could be used in any combination for any type of material.¹⁷

It was clear to those involved in the review process that a statement of principles of MARC format content designation was needed before meaningful revision could take place, and a draft document was prepared for discussion.¹⁸ Further work resulted in the compilation of a set of underlying principles for the MARC formats, many of which are not explicitly stated or clearly articulated in the MARC publications.¹⁹ A statement of

¹⁶Working under contract with the Library of Congress, David Weisbrod of the Yale University Library Systems Office prepared thorough enumeration of the aspects of the MARC formats that merited review.

¹⁷John C. Attig, "The Concept of a MARC Format," *Information Technology and Libraries* 2 (March 1983): 9–17; D. Kaye Gapen, "MARC Format Simplification," *Journal of Library Automation* 14 (December 1981): 286–292.

¹⁸Library of Congress, Network Development Office, *Principles of MARC Format Content Designation*, draft, (Washington, D.C.: Library of Congress, 1981).

¹⁹John C. Attig, "The US/MARC Formats—Underlying Principles," *Information Technology and Libraries* 1 (June 1982): 169–174.

principles will be included in the forthcoming second edition of the *MARC Formats for Bibliographic Data*. Work has continued at the Library of Congress on format review, and a format integration document has been prepared for discussions with the bibliographic utilities whose members would be affected most directly by substantive modifications to the MARC formats.

The environment in which MARC records are used in the United States has changed dramatically in the past decade. The one-way distribution of Library of Congress cataloging records that characterized the initial use of MARC has been replaced by an interactive and interdependent utilization. Many bibliographic agencies are participating in the creation of MARC records, as well as in the exchange of bibliographic information in machine-readable form. Now archivists have the opportunity to use the MARC AMC format to communicate information about their holdings in international, national, regional, and local data bases. They can explore applications of the format and share their discoveries with colleagues. And they can choose to participate actively in the process of revision in order to maintain the format as a flexible, functional tool for the archival community.

APPENDIX

BIBLIOGRAPHY

- Attig, John C. "The Concept of a MARC Format." *Information Technology and Libraries* 2 (March 1983): 9-17.
- . "The US/MARC Formats—Underlying Principles." Reports and Working Papers, *Information Technology and Libraries* 1 (June 1982): 169-174.
- Avram, Henriette D. *MARC, Its History and Implications*. Washington, D.C.: Library of Congress, 1975.
- . *The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources*. Washington, D.C.: Library of Congress, 1968.
- . *RECON Pilot Project: Final Report on a Project Sponsored by the Library of Congress, the Council on Library Resources, Inc., and the U.S. Department of Health, Education, and Welfare, Office of Education*. Washington, D.C.: Library of Congress, 1972.
- Avram, Henriette D., John F. Knapp, and Lucia J. Rather. *The MARC II Format: A Communications Format for Bibliographic Data*. Washington, D.C.: Library of Congress, 1968.
- Avram, Henriette D., Sally H. McCallum, and Mary S. Price. "Organizations Contributing to Development of Library Standards." *Library Trends* 31 (Fall 1982): 197-223.
- Avram, Henriette D., and Sally H. McCallum. "UNIMARC." *IFLA Journal* 8 (1982): 50-54.
- Crawford, Walt. "Library Standards for Data Structures and Element Identification: U.S. MARC in Theory and Practice." *Library Trends* 31 (Fall 1982): 265-281.
- Gapen, D. Kaye. "MARC Format Simplification." *Journal of Library Automation* 14 (December 1981): 286-292.
- Hagler, Ronald, and Peter Simmons. *The Bibliographic Record and Information Technology*. Chicago: American Library Association, 1982.
- International Federation of Library Associations. Committee on Cataloging. *ISBD(M): International Standard Bibliographic Description for Monographic Publications*. London: 1974.

_____. *ISBD(S): International Standard Bibliographic Description for Serials, Recommended by the Joint Working Group on the International Standard Bibliographic Description for Serials Set Up by the IFLA Committee on Cataloging and the IFLA Committee on Serial Publications*. London: 1974.

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*. Munchen-Pullach, Berlin: Verlag Dokumentation, 1972.

Library of Congress. Automated Systems Office. *MARC Formats for Bibliographic Data*. Washington, D.C.: Library of Congress, 1980. Looseleaf publication with quarterly updates.

Library of Congress. RECON Working Task Force. *Conversion of Retrospective Catalog Records to Machine-Readable Form: A Study of the Feasibility of a National Bibliographic Service*. Washington, D.C.: Library of Congress, 1969.

_____. *National Aspects of Creating and Using MARC/RECON Records*. John C. Rather and Henriette D. Avrams, eds. Washington, D.C.: Library of Congress, 1973.