Library Standards for Data Structures and Element Identification: U.S. MARC in Theory and Practice

WALT CRAWFORD

Introduction

The most prominent standards for data structures and element identification in U.S. library automation are the U.S. MARC formats. This paper considers some aspects of building, maintaining and using U.S. MARC and other such standards. The first section considers the costs and methods of establishing and maintaining U.S. MARC, and some of its benefits. The second relates U.S. MARC to the underlying standard ANSI Z39.2-1979. The third considers the relationship between the standard, content, and processing, and includes some other examples of data structure and element identification standards. The fourth section considers levels of compatibility within U.S. MARC and standards in general. The final section gives a few notes on U.S. MARC in practice.

Some Definitions

Data structures provide explicit frameworks for data and (sometimes) data element identification. Without knowledge of the data structure, it is impossible to deal effectively with the data; in a machine context, it is frequently impossible even to read the data. Data structures may provide for subfields (or positions) within fields, within records, within either files or result sets. The discussion here is limited to the record and more detailed levels.

Walt Crawford is Manager, Product Batch, Applications Software Development Group, Computer Systems & Services Division, The Research Libraries Group, Inc.
Data element identification as a term is self-explanatory. Data element identification can be of four varieties, which may be intermixed in a given record or data structure:

1. **Positional or implicit**, where the structure defines certain positions within a record (or within a field) as having specific meaning. Examples in MARC include leader positions, fixed-field elements, and indicator values (the latter two being positional elements within explicit higher-level structural elements).

2. **Explicit by code**, where the structure provides for coded content designation requiring external knowledge of the coded values. Most MARC data element identification is of this type, including fields and subfields.

3. **Explicit by label**, where the label is intended to be self-explanatory. This type is frequently called "self-contained identification." ISSN usage within printed materials is of this type. The label "ISSN" identifies the data element.

4. **Self-identifying**, where the element requires no identification.

Content designation is used to differentiate those elements of a record which designate content from the content itself. The term is generally synonymous with explicit data element identification. In U.S. MARC, tags, indicators and subfields are all elements of content designation. The term is widely used in dealing with U.S. MARC.

Standards, as used in this paper, include not only those established by the American National Standards Institute (ANSI), International Organization for Standardization (ISO) or other standards agencies, but also other explicitly stated sets of rules, particularly those used in common by more than one agency.

The U.S. MARC formats, sometimes called LC MARC or simply MARC II, are the series of formats for bibliographic and authority data established and published by the Library of Congress (LC). The primary concern of this paper is the MARC Formats for Bibliographic Data (MFBD).²

**COSTS AND BENEFITS OF U.S. MARC**

Standards for data structures and element identification are like other standards. They cost money to develop, establish, maintain, and use; that cost should be justified by benefits. When a standard becomes too expensive to maintain or use, it should cease to exist.
U.S. MARC

U.S. MARC, as specified by MFBD, is both an elaborate data element identification standard and a data structure standard. The data structure is a set of choices within the framework of ANSI Z39.2-1979, "Bibliographic Information Interchange on Magnetic Tape." "MARC" is frequently used to refer either to the data structure or to the data element identification standard. U.S. MARC is dynamic; the Library of Congress carries out a continuous maintenance program in cooperation with ALA, bibliographic services, and other interested parties.

Revision and Costs

U.S. MARC was originally developed by LC to distribute LC cataloging data to others. The story of that timely effort, headed by Henriette Avram, has been told elsewhere. Revision is an ongoing process, involving staff at LC, guidance and comment from many interested parties, and quarterly meetings to review proposed changes. Two of these meetings are held by representatives of ALA's Resources and Technical Services Division, Library and Information Technology Association, and Reference and Adult Services Division who compose the committee on Machine Readable Form of Bibliographic Information (MARBI). These are held at the ALA Annual Conference and Midwinter Meeting, and involve the committee, LC staff, representatives from the bibliographic services, and others. The other two meetings are held at LC in the spring and fall, and involve most of the same participants.

The direct cost of maintaining U.S. MARC is probably at least $250,000 per year, and that is counting only LC staff costs; direct costs to the bibliographic services, ALA divisions, and the National Library of Canada for the quarterly meeting; and costs of documenting and implementing changes. U.S. MARC is the most expensive content designation standard in the library community—unless AACR2 is considered a designation as well as content standard—and it is also the most beneficial.

Benefits of U.S. MARC

The most obvious benefit of U.S. MARC is the successful sharing of cataloging data. This started as a one-way sharing: LC, which had long provided printed catalog cards, began to provide information which users could manipulate for their own needs. The development cost of
U.S. MARC has probably been repaid through the benefits of the MARC Distribution Service alone.

The commonality of the communications format and data element identification standards also supported the growth of data sharing beyond LC. The development of OCLC and, later, of other bibliographic services has permanently altered the nature of technical processing in thousands of libraries. While each bibliographic service serves special needs in special ways, all have in common the data element identification standards of U.S. MARC (either directly or through mnemonic mapping), and all receive and distribute data in some compatible superset of U.S. MARC. The tens of millions of U.S. MARC-formatted records created and used through the resources of the services have certainly justified MARC revision costs many times over.

U.S. MARC formats are also widely used with local extensions, for a variety of local processes in libraries and library-related organizations. As early as 1972, the University of California at Berkeley began to use an extended U.S. MARC format for serials payment information. Research Libraries Group's Research Libraries Information Network II uses extended U.S. MARC for all batch processing, including production of acquisitions forms and special reports, as well as catalog cards. Library vendors also use U.S. MARC to provide a variety of services to even the smallest libraries.

The commonality of format has also allowed some sharing of computer software; as computers get cheaper and programmers get more expensive, this sharing seems likely to spread.

U.S. MARC has added a large and explicit vocabulary to library jargon, that of tags and subfields. While this may be a mixed blessing, it does allow rapid, unambiguous communication.

U.S. MARC AND ANSI Z39.2-1979

The data structure underlying the U.S. MARC formats is a formal ANSI standard—ANSI Z39.2-1979. That standard specifies the length of the record leader and significance of most of its positions. It further specifies the position and makeup of the record directory, the nature of field and subfield delimiters, and the length of tags. Finally, it specifies that only characters are used in records, never binary or other coded forms of data.

The current ANSI Z39.2 is generalized beyond that point, allowing for a wide range of implementations, or potentially for a single imple-
U.S. MARC

mentation with a range of record characteristics so wide that it would be quite difficult to process the records. The choices made for the underly-ing structure were all in the direction of flexibility and extendability, with the only real restriction being that tags are three characters long. That is not much of a restriction, since numeric and lowercase alphabetic characters may be used. There can be 46,656 distinct tags, and the structure allows for a wide range of element identification below the tag level.

As a data structure standard, ANSI Z39.2 is too flexible to use on its own in an efficient processing system. Potentially, each record could differ in: (1) number of indicators per field; (2) number of characters making up a subfield code; (3) length of "length of field" in directory entries; (4) length of "offset" in directory entries; and (5) length of the "implementation-defined portion" of directory entries, allowing sub-records or other structural extensions.

The generality of ANSI Z39.2-1979 allows implementation of sophisticated record structures. It is, however, a relatively difficult standard to process. The standard does not require that a given imple-mentation include all of the options, and U.S. MARC does not. Limiting the options allows more efficient implementation.

The present U.S. MARC formats include only a single option for each record-controlling variable in ANSI Z39.2-1979. There are always two indicators per variable data field, subfield codes are always two characters long, "length of field" is always four characters long, "offset" is always five characters long, and there is no implementation-defined portion. These choices are a compromise between flexibility and practi-cality. They help to keep U.S. MARC efficient.

EFFECTS OF U.S. MARC ON CONTENT AND PROCESSING

Any information-carrying medium affects the information to be carried. A standard for data structure and element identification affects the data which can be included, how they are entered and maintained, how they can be processed, and how they can be retrieved. The follow-ing breakdown is one way of evaluating a standard for data structure and element identification. U.S. MARC is a comprehensive standard, so all elements of the breakdown are applicable.
Data Entry and Maintenance

Does the standard require a sophisticated data entry system? Does data entry require expert operators? How much manual coding of content designation is required? Can data be verified by computer? Is it easy to edit existing records?

U.S. MARC data can be entered using a simple system (for instance, Basic Fix/FIX at the University of California), but data entry then becomes tedious and difficult, with little or no verification and difficult editing. Systems allowing easy modification of existing records, good verification, and good editing require considerable sophistication.

U.S. MARC requires a good deal of human coding of content designation, but actual data entry does not require great expertise. Verification can be quite refined, up to and including full authority control, depending on the system. The format lends itself to editing.

Content Restrictions

Can the standard handle a wide variety of data? Must data be abbreviated to suit the format? Can a variety of relationships be stored? Is there room for future needs? Can simple records be entered? Can very complex records be entered?

U.S. MARC shines on almost all these counts, though ANSI Z39.2-1979 is potentially even stronger. Data abbreviation is almost never required by the standard, though most implementations must set some tighter limits. The format allows up to 9999 characters in a field, and 99,999 in a record. Realistically, no record can be over 32,000 characters (and some systems are far more restrictive: 8192 is a common limit), and few editing systems can handle a field of more than 1500 to 1800 characters. (LC has, in fact, distributed one or two MARC records with “505” fields which are too long to be processed by some systems.)

Simple records coexist with complex records in U.S. MARC. The formats have room for future needs, and reserve room for local extensions. Few standards for data structure and data element identification have the versatility and lack of content restriction found in U.S. MARC and other comparable MARC formats.

U.S. MARC does not support extensive internal data structuring: it is not possible to include a “record within a record,” or to provide full content designation for more than one level of a multilevel record. U.S. MARC does handle a variety of data relationships and has room to handle more. A general technique has been adopted in U.S. MARC which expresses such relationships, where full content designation of
related bibliographic entities might be required.\textsuperscript{7} This technique avoids "nested records" or "subrecords," using instead record number linkages to separate U.S. MARC records.

Storage and Processing

Is there a computer-language bias? How much data overhead is involved? Can records be processed quickly? How large must the machine be? Can generalized, efficient software be written for record processing?

U.S. MARC requires strong string-handling languages, such as PL/1. Data overhead is fairly high. Record processing is fast, particularly for a format with such extended capabilities. It is possible to process U.S. MARC records on some microcomputers, but most U.S. MARC processing is done on large systems.

Because all data are stored as characters, and because data element identification techniques are consistent throughout (except for control fields), U.S. MARC allows generalized data-processing techniques where the task of extracting data is independent of the particular data to be extracted. Table-driven software works well for U.S. MARC, allowing programs and systems which can be maintained and modified quickly and safely. (This is true specifically for record analysis and retrieval; it is not as true for data entry or editing.)

Data Retrieval and Manipulation

Does the standard lend itself to multifunction records? Is selective (partial) retrieval possible and meaningful? Does the standard support varied and sophisticated retrieval? Does it support sorting and other manipulation?

U.S. MARC and various extended U.S. MARC formats allow for multifunction records. The format allows meaningful selective retrieval: if an application requires only the main entry, short title and physical description, these tags and subfields can be retrieved \textit{without regard to any other tags or subfields in the record}. The high degree of content designation supports sophisticated retrieval techniques. While U.S. MARC records cannot be sorted without use of a sort key, the records do provide some support for sorting.
Summation and Other Examples

While not a "cheap" standard, U.S. MARC is unusually versatile and flexible, and, once built, the records can be processed efficiently and easily. Two examples follow of other data structure and element identification standards, one from the publishing and book trade and one from the abstracting and indexing community.

BISAC Order and Invoice Formats

The Book Industry Systems Advisory Committee (BISAC), a voluntary collaboration of publishers, booksellers, wholesalers, and librarians, was founded in 1975 "for the purpose of improving the interchange of technical information pertaining to the ordering, handling, and movement of published materials." BISAC has developed and published several standardized formats, designed for inexpensive data entry and processing, at some expense in flexibility.

The Purchase Order Tape Communications Format and the Invoice Communications Format are both made up of eighty character fields. Records are connected by a common area repeated in each field, and each field begins with a two-character identifier. All data element identification within a field is implicit (positional).

The BISAC formats are not designed to handle "difficult" situations, but are suited to the bulk of book trade transactions. They allow simple, cheap, "fill-in-the-blanks" data entry. They allow easy verification, and are single-purpose. Data must be abbreviated or coded in many cases.

Unlike U.S. MARC, the BISAC formats are single-purpose, efficient, and relatively simple. They could be characterized as "COBOL-biased," but could be processed easily using any business-oriented language. These are typical business formats, flexible enough for most transactions, simple enough for efficient use, but somewhat lacking in flexibility and versatility.

Chemical Abstracts Standard Distribution Format

There is no commonly used standard format in the abstracting and indexing community. Most producers have their own format, tailored to their own needs. These are probably cheaper than any common format could be, from the producer's perspective.

One such format, more versatile than many, is the Standard Distribution Format (SDF) used by Chemical Abstracts for its distribution services. SDF is well documented, flexible, and allows room for future data element identification. It is less flexible than U.S. MARC, and does
not allow equally generalized software, but it is a versatile, well-designed format, designed to minimize storage requirements.

Records are variable in length with a directory similar to MARC but with binary length and offset values, and with short data elements actually stored in the directory. A record may have up to 255 data elements (there is no lower level of element identification than the field), and may be up to 3520 characters long. SDF is strongly biased toward IBM Assembler; all control elements and fields begin at doubleword (eight-byte) boundaries, and the documentation includes IBM Assembler subroutines for data retrieval. Data overhead is unusually low for a directory-based format. There is heavy use of binary and bit-string data both in the directory and in fields. As a result, fully generalized data handling techniques are not suitable.

SDF appears cheaper to key, edit and store than U.S. MARC. It is a fairly versatile format, and is presumably cost-effective for Chemical Abstracts. SDF is a single-producer standard—more commonly termed a “format”—and it does serve typical standards purposes within the Chemical Abstracts family of distribution services.

COMPARABILITY BETWEEN AND WITHIN STANDARDS

Compatibility is a frequently used term; the phrase “MARC compatible” is frequently abused. MARC compatibility is an important topic for library automation and is being addressed by a working group of the Technical Standards for Library Automation Committee (TESLA) of ALA/LITA. While the issue of MARC compatibility cannot be settled here, some discussion of levels of compatibility may be useful.

Identity: Precise Compatibility

The highest level of compatibility is identity. Identity normally results from common implementation of a single comprehensive standard by more than one agency. Identity implies that all processes working on one case will work the same on other cases. It requires that character set, record structure, content designation, data element identification, coded values, and rules for content be the same in all cases.

An implementation of U.S. MARC would be identical to U.S. MARC if it included all (and only) data elements contained in MFBD, stored in ALA Extended ASCII (American Standard Code for Informa-
tion Interchange), using the structural definitions given in MFBD, and using ISBD punctuation and AACR II cataloging rules, as used in MFBD.

Reversibility: Full Compatibility

Two standards are fully compatible if records using either can be algorithmically transformed to the other, and back again, without any loss of information whatsoever. This level of compatibility is full reversibility.

Alternate Character Set Usage—U.S. MARC is defined using an extended ASCII character set. Most large library-related computers use as a primary character set the IBM-defined Extended Binary Coded Decimal Interchange Code (EBCDIC) character set. Most users of MARC data define “extended EBCDIC” character sets, translate MARC data from ASCII to EBCDIC on receipt, and translate data from EBCDIC to ASCII on transmission.11

Alternate Storage Mechanisms—MARC as a structural standard is well suited to sequential processing, but not to direct access (as in a bibliographic service or online catalog). Most online implementations transform MARC data into a different structure for online use, transforming data back to the MARC structure for sequential use or transmission.

Alternate Communications Technique—When using telecommunications to pass MARC data, a structure which uses fewer characters is highly desirable. Given the current U.S. MARC standard, it is possible to strip part of the leader and all of the directory, attaching each tag to its field; such a record can be transformed back to standard U.S. MARC by a simple program with no loss of data.

The OCLC MARC format and the RLIN MARC format are not compatible at this level. Both are supersets of U.S. MARC, and fall into the third level of compatibility, described next.

Superset Compatibility

One format is a superset of another (second) format if records in the second format can be algorithmically transformed into proper records of the first format, and later transformed back into the second format, without any loss of information at any level. In such a case, all records in the first, or superset, format should be algorithmically convertible into proper records of the second, or subset, format, but some information may be lost in the process. This level of compatibility differs from
simple convertibility in two respects: the converted record becomes a proper record of the superset (or at least a proper portion of a record), and the subset record can be reconverted without loss.

There are at least three types of supersets which can occur in dealing with U.S. MARC. A given superset could include elements of all three types.

1. **Content extensions.** The OCLC, RLIN and UTLAS communications formats maintain precise structural equivalence to U.S. MARC, and include all U.S. MARC content designation and data element identification. They also define additional data elements, such as local holdings fields or acquisitions fields and subfields.

   These formats are Extended U.S. MARC formats. The U.S. MARC record can be algorithmically extracted without loss of designation or content, by a simple algorithm. Software which can process U.S. MARC will process the extended record, though it will not recognize the extended fields.

2. **Structural extensions.** A database management system could incorporate all content and content designation from U.S. MARC within an expanded structure. If it is possible to build the bibliographic core of a record in such a system directly from a U.S. MARC record, by program, and to rebuild the U.S. MARC record by program without loss of information, the structural extension represents a proper superset of U.S. MARC. Since programs to process U.S. MARC would probably fail in attempting to process the extended structure, the superset would not be called "Extended MARC."

3. **Character set extensions.** Until such time as non-Roman character sets are actually defined for the MARC formats, an implementation using them is a superset of U.S. MARC.

The preceding list deals with extensions to MARC. At a slightly lower level of MARC compatibility are proper subsets of U.S. MARC—formats which can be algorithmically converted to processable U.S. MARC records, and back again, but which cannot store full U.S. MARC records without loss of information.

That level edges over into the next level down, depending on what one considers to be a "proper" U.S. MARC record. A local format using a full "008" field, full indicators and subfields, but only allowing a subset of tags, would be a proper subset of U.S. MARC and could with some justification be labeled a "MARC subset."

On the other hand, a format which was all uppercase, had no stored subfields, only allowed up to thirty characters for author and sixty for title, had no coded values, and did not allow any other content designa-
tion, would really belong in the next lower level, since the resulting “U.S. MARC” record would be almost useless.

Convertibility: Unidirectional Compatibility

A format is convertible from another format if it is possible to convert records from the second format into the first by program, but not to reverse the process without loss of information. This is unidirectional compatibility, and is what is commonly referred to as “MARC compatible.”

Unidirectional compatibility is the broadest level, and the most open to abuse. “MARC” is a selling point in library automation, and there is a tendency to stretch a point in calling something “MARC compatible.” Some fairly clear levels of convertible formats can be stated, based on extent of reversibility and data storage.

1. **Reversible with loss of content designation.** Such a format allows for all textual data within U.S. MARC to be stored in a meaningful manner, but may omit subfields, indicators and other forms of content designation. The “restored” MARC record would be less useful than a full U.S. MARC record, but would be intelligible, and might be restorable to full U.S. MARC information with some manual or intellectual assistance.

2. **Reversible with loss of content.** Here, “reversible” is really a sales pitch. All-capital formats, formats with short maximum lengths for data elements, and formats allowing only certain data elements to be stored all fall into this category. Theoretically, if the program is sufficiently tailored, you can “reverse” almost anything, down to and including a format where the full bibliographic record appears on an 80-column punched card. Except where the loss of content is along clear subset lines (for instance, dropping all added entries), such implementations more properly belong in the next category.

3. **Nonreversible.** A format labeled as such is an honest attempt to incorporate MARC data into a simpler record. In this case, the claim is that U.S. MARC records can be read and data derived to build local records. There is nothing wrong with nonreversible formats, probably the most frequent non-MARC use of MARC records. What is wrong is blithely labeling such formats “MARC compatible,” when the only “compatibility” is that the data elements in the record can be algorithmically derived from U.S. MARC records.
Pseudo-compatibility

The distinction between pseudo-compatibility and incompatibility is a subtle one, based on appearance or lack of information. Pseudo-compatible and incompatible formats are alike in that records in either format cannot be algorithmically converted into the other format without unrecoverable loss of information.

U.S. MARC IN PRACTICE: SOME NOTES

Mnemonics and Tags

Some early system designers felt that MARC tags were too foreign to library practice, and that subfielding would be difficult. Some systems were designed using mnemonics for fields and subfields, with the system providing the tagging. One example is the BALLOTS system, which totally excluded numeric tags.

Use of mnemonics for fields and subfields grew more difficult as the format expanded, and became cumbersome with the serials format. Use of mnemonics for subfields was effectively impossible by the middle to late seventies.

Surprisingly, the MARC tags and subfields have become common usage in the library field. This is partly due to their brevity and precision: “111” is shorter than “Main Entry—Conference or Meeting Name,” and “x11” is a brief way of saying “Conference or Meeting Name, whether Main, Subject, Series, or Other Added Entry.”

Newer systems and revisions of older systems (such as RLIN II) tend to use tags and subfields, using mnemonics only for fixed fields and local extensions of U.S. MARC. The language of MARC has become part of library jargon.

Complexity of the Formats

As early as MARC I and as recently as 1981, it has been claimed that the complexity of content designation in U.S. MARC makes it too expensive to enter data, that the cost of assigning fields and subfields is a substantial addition to the cost of cataloging.

It was recently suggested that libraries could save money and get by just as well if all “name” fields were collapsed to a single field (i.e., 100, 600, 700, and 800) and most subfields were eliminated. Technical processing staff at a variety of libraries consistently said that the extra time
required for proper content designation is a small part of the overall flow of original cataloging. It is, of course, nonexistent in online copy cataloging, which provides 90 percent or more of the cataloging for all but the largest libraries.

Retrieval and manipulation depend heavily on the existing level of content designation. Personal name searching can use a different technique than corporate or conference name searching.

**MARC as a Communications-Only Structure**

MARC was designed for tape transmission of bibliographic records—originally, for transmission from the Library of Congress to others. At the time, many assumed that any user or builder of such records would use them in a different structure, translating to or from MARC only at system boundaries.

An informal survey at the LITA ISAS (Information Science and Automation Section) Programmer's Discussion Group (involving fifteen institutions) showed that eight of the fifteen institutions made active use of the MARC structure as an *internal processing format*. This is not surprising, speaking from the perspective of personal experience at UC-Berkeley and now at RLG; MARC, with minor extensions but no substantive changes, is an effective, efficient and flexible batch-processing format.

The greatest advantage of using MARC with extensions but without major revisions is, of course, that the processing software is resistant to change in data element definitions. The MARC directory is efficient for individual item retrieval within a batch environment. If you are looking for one set of fields, you don’t need to know what other fields have been added.

Use of the structure is not restricted to catalog card production. At RLG/RLIN, all acquisitions product generation (e.g., orders, claims, cancellations) is based on MARC structured records. UC-Berkeley has been driving its acquisitions and product generation systems from MARC structured records for nearly a decade now, at low cost and with high flexibility.

The MARC structure is principally useful for batch processing. Interactive online use usually requires transformation to some database structure.
The LC-Centrism of U.S. MARC

U.S. MARC has been "LC-centric," concentrating on the needs of the Library of Congress. Most U.S. MARC development since the early years has avoided additional LC-centrism. In recent years, there has been movement away from existing LC-centrism—studies have been done, the MARC review process is focusing on the issue, and the Library of Congress itself is working to lessen the bias. In the course of this movement, it has become clear that many libraries want some LC-centrism. In some cases, they want to distinguish between data elements actually assigned by LC and those assigned by others. Moves to generalize the format will leave in some LC-centrism, at the request of other libraries, even when the Library of Congress would prefer to see the bias eliminated.

U.S. MARC and Analytics

An analytics technique for U.S. MARC was approved at the ALA Midwinter 1981 MARBI meeting. It adds field 773—"In"—to the existing formats (and adds some other supporting codes). While not the sophisticated structural solution which was originally proposed, the current solution is easy to implement, easy to use, and allows full extendability for complex situations.

U.S. MARC and Structured Data

ANSI Z39.2-1979 allows implementations which store several levels of fully content-designated bibliographic entries within a single record (for instance, a set of maps with entries for each map). This "subrecord technique" was considered as a possible change to U.S. MARC.

The decision to use linkages to separate U.S. MARC records was made largely because of the expense and difficulty of subrecords. All users of systems reflecting U.S. MARC, whether they ever used subrecords or not, would pay a high ongoing price for the technique, in addition to the extremely high price of initial implementation.

Some intended uses of the subrecord technique could not have been handled. A record of more than 32,000 characters, including all overhead, is essentially unprocessable on any current business computer.

*The development of analytics capability in U.S. MARC has been a complex and difficult one stretching over many years and showing, in its final phases, the historic changes in attitudes toward MARC. It would be impossible to give even a brief version of that development here.
Intricate multipart records would have exceeded this limit frequently enough to cause a continuing problem.

The Three-by-Five-Inch Card Orientation of U.S. MARC

U.S. MARC does have features which specifically serve card-oriented needs, and does retain the concept of "main entry." Many of those features are required for any unified single-record display.

U.S. MARC uses the same subfielding and other content designation for main entries and equivalent added entries. The same level of information may be provided, and systems which do not use a "main entry" are well served by U.S. MARC. While U.S. MARC supports the three-by-five-inch card, it does not do so to the detriment of other uses.

Conclusion

Standards for data structure and element identification can range from small and simple standards to those as large and complex as U.S. MARC. Any such standard should be judged by its cost/benefit ratio. U.S. MARC has benefited libraries far beyond its costs.

Thousands of libraries of all sizes use tens of millions of U.S. MARC and extended U.S. MARC records, directly and indirectly. U.S. MARC has served these libraries well, and continues to do so. U.S. MARC continues to evolve, making the formats more useful while retaining the worth of existing records and processing systems. The evolution is sometimes slow and painful, but is done with concern for the past and present, as well as for the future.

The future is longer than the past, but is based on that past. Existing U.S. MARC-formatted records are as relevant to future library needs as existing books are to future readers; those who would scrap either must be able to justify the change.

The future may bring a format, not evolved from present U.S. MARC, which is so superior as to make mass conversion worthwhile. No such format has yet appeared, and no convincing case has been made to this point for radical change in U.S. MARC. Those who attack U.S. MARC should bear the burden of proof: showing a superior alternative, and showing its economic validity. The case for MARC III (?) may be made, but it has not been made yet.
U.S. MARC

References

15. Crawford, "Programmers Discussion Group Meets."