Communications

MARC Format Simplification

D. Kaye GAPEN: University of Alabama, University.

This is a summary of a paper written on the consideration of the feasibility as well as the benefits, disadvantages, and consequences of simplification of the MARC formats for bibliographic records.[1] The original paper was commissioned in June 1981, by the ARL Task Force on Bibliographic Control as one facet in exploring the perceived high costs of cataloging and adhering to MARC formats in ARL libraries. The conclusions and recommendations, however, are entirely those of the author and the opinions and judgments stated here result from a wide-ranging canvas of technical services people, computer people, and/or library administrators. Because the MARC format has so many uses, the paper is divided into five perspectives from which the MARC format can be viewed: history, standards, and codes; present purposes; library operations; computer operations; and online catalogs.

The Library of Congress has already begun a review of the MARC format and has distributed a draft document.[2] The general thrust of that review is a close examination of the MARC format in an attempt to begin to lay the foundation on which revised MARC formats can firmly stand—particularly in regard to content designation (tags, indicators, and subfield codes used to identify and characterize the data explicitly).

As that review deals with the very specific, this paper aims generally at attempting to paint with broad strokes a picture of today's MARC in its many relationships, benefits, costs, and what the impact would be to the whole from any change to the part.

PERSPECTIVE: MARC HISTORY, STANDARDS, AND CODES

Relationships

The original MARC format document established conventions for encoding data for monographs. Though it was understood that early applications were going to relate to the production of catalog cards, the MARC designers looked ahead to an increasing emphasis on data retrieval applications. Other design considerations included, for example, the necessity for providing for complex computer filing, allowance for a variety of data processing equipment, and an attempt to provide for some analytical work (more specific description of contents notes or other types of analysis).

Later the single MARC II format was transformed into a series of formats, and as time passed, those formats became inextricably tied to other developments at the national and international levels: The International Standard Bibliographic Descriptions, the Anglo-American Cataloguing Rules, 2d ed., UNIMARC, the National Level Bibliographic Records, and the National and International Communications Standards; e.g., ANSI Z39.2–1979 and ISO 2709.

Benefits

The benefits of the MARC formats and other standards and codes have been substantial both philosophically and pragmatically. The sharing of cataloging records through the computer-based, online networks have been shown in a variety of cost studies to have contained the rate of rise of per unit cost. A further benefit of the MARC formats is the momentum its creation gave to the steady movement toward standardization which can benefit individ-

ual libraries in a number of ways: first, bibliographic information can be exchanged among libraries and countries. Second, in recent years we have moved steadily toward creating an environment in which the Library of Congress would become one of many authoritative libraries thus enhancing the shareability of records.

Costs

The early costs of the development and implementation of the MARC formats were borne by LC (aided by Council on Library Resources funds). LC continues to bear most of the costs of MARC formats, such as new MARBI proposals, duplication and distribution of documentation, and so forth. Direct investment of library dollars came through the purchase of the MARC tapes and the development of systems to receive, process, and output data in MARC formats.

Impact of Change

Throughout the years of its use, the MARC format content designation and content rules have been augmented or modified. In the beginning, however, databases were small and changes could be absorbed more readily. The number and complexity of the formats have increased, as have the interrelationships of the MARC formats with other standards and codes resulting in a present environment in which the impact of change is felt more strenuously.

PERSPECTIVE: PRESENT RELATIONSHIPS AND CONSTRAINTS
Relationships

Today's close interrelationships between the MARC formats and other codes and standards affect both library and computer operations. Though, for example, the general International Standard Bibliographic Description was implemented by the library community prior to the adoption of AACR2, the second edition of the rules has firmly incorporated the ISBDs. When this format description system is combined with the machine-based MARC formats, some ISBD information will be supplied by humans and some generated by programmed machine manipulations.

As a second example, in the last couple of years, the Library of Congress has spearheaded the development of National Level Bibliographic Record(s) which define the specific data elements that should be included by any organization creating cataloging records which may also be shared with other organizations or be acceptable for contribution to a national database. As the logical idea of a national database comes to fruition, it is necessary for the MARC format to provide for greater specificity in the coding of originating library, modifying library, and so forth.

Benefits

The benefits of the use of the MARC format continue to lie in the ease with which bibliographic information can be shared and the concomitant beneficial impact on cost control. In addition, the MARC format supports a host of other standards and codes and the benefit from these relationships has been consistency in and fostering of standards development. In the bibliographic arena, the more that standards are developed—locally, regionally, nationally, and internationally—the more we will be able to transmit and share bibliographic data, thus controlling the costs of original cataloging. On the other hand, we also "pay" when we standardize.

Cost

The two costs associated with increased standardization are additional time and thus cost required to meet standards, and the increased expense of maintaining local practices which may often be idiosyncratic. In relation to the latter, while many local idiosyncrasies are often unnecessary and counterproductive, there are generally some which have become an integral part of a large catalog database or upon which a major procedural activity is based. But, to benefit from compliance with standards, increasingly we will move away from local practices.

In terms of the time required to adhere to the MARC format, it is possible to continue to utilize the format (or participate in systems that use it) and yet control the amount of complexity with which one has to deal. Both AACR2 and National Level Biblio-

graphic Record documents allow for "levels of description" which provide for more or less description; and various online networks allow, in a similar manner, for limited input standards.

As we view the array of standards and codes which together make up today's bibliographic scene, we can see that each of the separate elements is consistent within itself, is understandable, and counts for only a portion of the costs associated with the cataloging process. The combination of elements, however, begins an accretion of complexity that for most requires an effort of organization and education in order to control work flow and meet standards.

Impact of Change

Because the MARC format is closely interwoven with a number of national and international codes and standards, changes to the format would have implications far beyond the local library. At the very least, discussions would have to involve a host of individuals and groups, all at different stages of development and implementation based upon the present MARC format.

PERSPECTIVE: LIBRARY OPERATIONS
Relationships

In the library-operations perspective, any operations related to the MARC format have to be viewed as only one of many elements which must be interfaced with daily work flow. Let us look, for example, at the amount of time which might be expended in a typical large academic library by cataloging personnel in training and ongoing work activities required in MARC-related operations.

In those libraries which obtain access to cataloging databases as members of networks, contact with the MARC format is filtered through the standards, requirements, MARC implementation design, documentation and other related training facilities of the network. Libraries which maintain their own databases do the same kind of filtering, though staff may have somewhat more control of the user cordiality of the interface. The shared networking environment, however, generally seems to imply more standards and requirements be-

cause of the attempt to guarantee as much "shareability" as possible.

Libraries participating in OCLC, for example, must train staff in the following codes: AACR1; AACR2; standard subject heading codes; standard classification codes; OCLC/MARC formats for each type of material being cataloged; OCLC bibliographic input standards; OCLC Level I and Level K input standards; OCLC systems users guides; in some instances, input standards documents for regional or special-interest cooperatives; local library interpretations, procedures, and standards.

Any close review of the time library staff expend in the use of these tools for either training or ongoing operations reveals that MARC per se requires only a limited proportion of a typical library staff person's day. While training may be intensive at either the beginning of a person's job or at the beginning of work with a new type/format of material, this portion of the cataloging unit cost is small.

Benefits, Costs

In the cataloging activity, the benefits from the use of the MARC formats are at least two: first, the MARC format as part of an online cataloging system permits the machine-production of catalog cards at a major savings over manual production. Second, access to a shared cataloging database permits the use of "clerical" catalogers at an estimated unit cost saving per book of twenty dollars when compared to "original" cataloging.[3] Third, depending upon the information available in the cataloging record, the time required for decision making during the cataloging process can be decreased significantly.

Impact of Change

It was the general consensus of the technical services people I contacted that simplification of the formats through the consistent assignment of tags would make training and introduction to new formats somewhat easier, but that any savings of time would probably be trivial. There was no consensus that either simplification or shortening would result in any significant time or cost savings.

To a certain extent, the use of the very

specific MARC formats has made the descriptive cataloging process (and the training to undertake it) clearer in that the logical relationships and description of the data elements are so clearly exposed through the assignment of tags and other codes. Also, once initial familiarity with the format(s) is achieved, ongoing use becomes second nature. It is also possible for cataloging staff to control the complexity with which they will deal through the use of less than "full," but still nationally acceptable levels of cataloging and, hence, MARC coding.

Finally, most technical services people believe that cataloging and maintenance activities in libraries have always been complex, requiring long and detailed procedures and intricate work flow. While membership in networks requires new skills and knowledge, it is the sum of the whole rather than the difficulty of any single portion which affects unit costs today. Changing the MARC format through either simplification or shortening would have only a slight effect on the total technical services operation and costs.

PERSPECTIVE:
THE COMPUTER
OPERATIONS ENVIRONMENT
Relationships

In looking at computer operations, there are at least two major subdivisions: operations that serve only one client (e.g., a library system serving itself) or operations that serve many clients (e.g., RLIN or Blackwell/North America). The constraints differ for each operation and are further complicated by whether or not the computer operation must be able to produce as well as accept bibliographic records in a MARC format.

Each computer facility, for example, can have distinct operating software depending upon the type and mix of computing equipment used. In addition, each computing facility translates the MARC-formatted records into an internal processing format which may differ extensively from MARC. Too, further tailoring may be done for batch processing as opposed to online operations and computer operations which serve a single user may not have to re-create records in the MARC format and may even

more radically redesign the MARC-formatted records for internal use.

As changes to the MARC format occur over the years, each computer system will write additional software to incorporate those changes into the then existing system. In some instances, it may be too difficult to attempt to convert old databases to reflect changes in MARC coding, and there will then exist an "old" database and a "new" database for that particular MARC field or subfield. Since changes have occurred in many fields, most databases are an amalgam of new and old interpretations (this is true in relation to cataloging codes, too) of MARC coding, and original internal software design may reflect the same type of patchwork quilt.

Operating these computer systems is complicated, in addition, by the fact that a wide range of user library needs and desires must be accommodated. Indeed, a report prepared by Hank Epstein for the Conference to Explore Machine-Readable Bibliographic Interchange (CEMBI) revealed after an exhaustive review of the use of MARC data elements that there was no data element not used by someone![4]

Benefits

Benefits that accrue to computing operations as a result of the MARC format include the use of what was called "a pretty decent general communications format," which facilitates communications, card/COM production, and online information retrieval. As a communications format it is as coherent as any other structure for carrying bibliographic data. Because the format allows for a very specific level of detail in description, computing operations can supply a variety of products to fill a variety of needs.

Costs

While specific cost information was not available for inclusion in this paper, discussion does reveal some widely held generalizations. First, the MARC format does not seem to be any more complex or costly to use than other variable field communications formats. Beginning programmers are generally introduced first to the internal communications format of their particular

computing system, and when they come to the MARC tags rapidly become familiar with the coding through experience. Indeed, if the programmers know the structure of and have a specification for the format, they can work with that format even though they may be unfamiliar with it from the users' point of view. Thus, the format itself, and training in its use does not seem to be significantly costly.

Second, every change in the MARC format requires some programming effort and may or may not require concomitant changes in the database. The consensus of the computer people with which I spoke was that the sophistication and specificity of the MARC formats was a good thing, but the inconsistencies among formats is problematical. The benefits of consistency can be important, but to justify changes financially, the major changes should be done at one time. Indeed, most individuals doubted whether or not there was sufficient capital in these straitened times to be able to implement consistently a major MARC format change—and this is from the perspective of both the operations serving one and many users.

Impact of Change

Without a philosophical and practical framework (or benchmark) against which to compare the benefits and costs of alternative solutions to MARC format maintenance issues and without a better and more comprehensive description of the requirements of the internal processing formats of the computer operations, it is difficult to assess clearly the costs and benefits of MARC format changes. It does seem to be the case presently that, once established, computer operations can deal with the complexity and specificity of the MARC format without undue ongoing financial investment.

The strength of the MARC format for computer operations lies in its specificity. For the batch processing environment especially, the MARC format is a reasonably efficient format and one that facilitates development. Its inefficiencies are not drastic and its specificity buys valuable flexibility. Severe cuts or major simplifications would be a mistake since discontinuing specificity is a one-way street—once it is gone, it can-

not be retrieved. The ability of the machine to assist in editing is weakened by the loss of specificity and it then becomes more difficult to edit out poor data. Simplification through consistency, rather than shortening, would produce the most beneficial impact—though it must be done carefully to be cost beneficial.

PERSPECTIVE: ONLINE CATALOGS
Relationship

The major difficulties facing us when we attempt to discuss the relationship of the MARC format to online catalogs is that, first, we know so little about how people think when they use our card catalogs; and, second, we have so little experience with how those thought and use patterns might change when the online catalog replaces the card catalog. Another aspect of online library system development is the combination of subsystems such as acquisitions, serials control, or authority control with the online catalog and the implications of such a combination for system design, the internal processing format, and compatibility with the MARC format.

The index design of most large online catalogs or information retrieval systems today relies upon precoordinated search keys in order to facilitate the large sorting activities that have to occur. The second indicator in the 700 field, for example, is designed for the purpose of formulating search keys, filing added entries or for selecting alternative secondary added entries. This type of specificity is necessary for both card production and online retrieval. Taken together, all of these considerations make most systems and library technical people hesitate to recommend any major changes to the MARC format at this time.

Benefits

At this time, therefore, in terms of information retrieval, there does not seem to be any major force toward either simplifying or shortening the MARC format to facilitate retrieval. This becomes an even more cogent sentiment when we consider that major development efforts have already been begun in the areas of online catalog access and information retrieval. Delays in these development efforts now caused by

changes in the MARC formats could be enormously wasteful of the time and effort already invested, and could postpone urgently needed implementation of new, easily maintainable online systems.

Costs

There is no firm cost data to guide us in considering the impact of MARC format changes in the information retrieval environment. Generally accepted assumptions are, however, that because of our lack of knowledge and experience in this area, it is simply too risky and potentially costly to experiment.

Impact of Change

Overall, without more experience in this area, it is the general opinion that the fullest level of descriptive specificity of the MARC format might be required to design and implement online catalogs/information retrieval systems which can be responsive to the needs of a variety of users and levels of information. Interaction with other subsystems and formats is also incomplete, thus clouding our vision of the impact of change over the breadth of the library community.

SUMMARY AND CONCLUSIONS

The original purpose of the MARC format is still a cogent and necessary one—that of allowing for a great variety of individual library needs for products, practices, and policies via a standardizing communications format. Both catalog card production and online retrieval necessitate the same level of specificity, though particular tags, indicators, and subfield codes may vary.

As we look toward a variety of authoritative cataloging sources the MARC format, in addition to a specific coding of bibliographic information, might also have to specify descriptions of cataloging actions so that the greatest degree of "shareability" might exist. Some of this related authority-type information will either be carried as part of the MARC format or in some manner as linked records.

The computer operations that utilize the MARC formats exist under the constraints of a variety of internal processing formats and design constraints. For each internal processing system, however, the specificity of the MARC format offers flexibility and efficiency for a number of different processes and products.

Taken by itself, the MARC format is no more difficult to work with than any other standard or technique for both librarians and computer people. While it might be useful for librarians to implement training aids such as online documentation, access to library manuals (particularly that of the Library of Congress), and so forth, the benefits of aids such as these are trivial since the coding can be learned rather quickly through experience. For computing people, on the other hand, changes in the formats can be very expensive and disruptive. There is general agreement, moreover, that over the long term we have got to be able to maintain the MARC format in response to experience with retrieval and other theoretical and technical advances. The main thrust of maintenance in the computing realm is consistency across formats, but approaching this type of simplification requires a number of preliminary steps if it is to be implemented effectively.

We need to develop a vocabulary for jointly discussing the elements of the problem. In addition, a major review needs to be undertaken of the internal processing formats and design constraints of the major computer operations—both to serve as a benchmark for measuring the impact of format changes, and as a guideline for newly developing systems to assist in avoiding mistakes in the development of new computer operations.

Someone needs to be thinking about and designing the ultimate, comprehensive MARC format—not to be implemented, but to serve as a springboard for discussion and for consideration of system design. We need to establish limitations on what we will handle with the MARC formats and where we will begin to rely on underlying formats instead. The development of a comprehensive MARC conceptualization would also provide a protocol for undertaking the improvement of MARC and would serve as a benchmark against which local systems could be compared.

At the very least, the steps described here would facilitate the consideration and implementation of making the formats consistent across types of material—a goal which is seen by all to be highly desirable.

We need a format which is consistent, easily maintainable without being uncontrollably disruptive, and responsive to changing needs which are likely to accelerate as we gain experience with online systems.

Rather than recommending or supporting the implementation of specific changes to the MARC format, it is essential that the library community begin to establish the framework and benchmarks necessary to maintain the MARC formats over the long term as well as to guide short-term considerations. ARL and others can play an important role in undertaking and encouraging a broader approach to this pressing problem. Such an approach will not only reduce the risk of decision making, but will also assist in the development of the cost/benefit data needed to enhance consideration of format changes.

REFERENCES

1. D. Kaye Gapen, *Simplification of the MARC Format: Feasibility, Benefits, Disadvantages, Consequences* (Washington, D.C.: Association of Research Libraries, 1981), 22p.
2. "Principles of MARC Format Content Designation," draft (Washington, D.C.: Library of Congress, 1981), 66p.
3. Ichiko T. Morita and D. Kaye Gapen, "A Cost Analysis of the Ohio College Library Center On-Line Shared Cataloging System in the Ohio State University Libraries," *Library Resources & Technical Services* 21:286–302 (Summer 1977).
4. Council on Library Resources Bibliographic Interchange Committee, *Bibliographic Interchange Report*, no.1 (Washington, D.C.: The Council, 1981).

Comparing Fiche and Film: A Test of Speed

Terence CROWLEY: Division of Library Science, San José State University, San Jose, California.

INTRODUCTION

For more than a decade librarians have been responding to budget pressures by altering the format of their library catalogs from labor-intensive card formats to computer-produced book and micro-formats. Studies at Bath,[1] Toronto,[2] Texas,[3] Eugene,[4] Los Angeles,[5] and Berkeley,[6] have compared the forms of catalogs in a variety of ways ranging from broad-scale user surveys to circumscribed estimates of the speed of searching and the incidence of queuing. The American Library Association published a state-of-the-art report[7] as well as a guide to commercial computer-output microfilm (COM) catalogs pragmatically subtitled *How to Choose; When to Buy.*[8]

In general, COM catalogs are shown to be more economical and faster to produce and to keep current, to require less space, and to be suitable for distribution to multiple locations. Primary disadvantages cited are hardware malfunctions, increased need for patron instruction, user resistance (particularly due to eyestrain), and some machine queuing.

The most common types of library COM catalogs today are motorized reel microfilm and microfiche, each with advantages and disadvantages. Microfilm offers file-sequence integrity and thus is less subject to user abuse, i.e., theft, misfiling, and damage; in motorized readers with "captive" reels it is said to be easier to use. Disadvantages include substantially greater initial cost for motorized readers; limits on the capacity of captive reels necessitating multiple units for large files; inexact indexing in the most widespread commercial reader, and eyestrain resulting from high speed film movement.

Microfiche offers a more nearly random retrieval, much less expensive and more versatile readers, and unlimited file size. Conversely, the file integrity of fiche is lower and the need for patron assistance in use of machines is said to be greater than for self-contained motorized film readers.

THE PROBLEM

One of the important considerations not fully researched is that of speed of searching. The Toronto study included a self-timed "look-up" test of thirty-two items "not in alphabetical order" given to thirty-six volunteers, of whom thirty finished the test. The researchers found the results "inconclusive" but noted that seven of the ten librarians found film searching the fastest method. "Average" time reported for searching in card catalogs was 37.3 min-