I shall describe three applications of MARC tapes. Two of these are in progress at Stanford University and the third is a commercial service furnished by an organization with which several Stanford librarians are associated in an advisory capacity.

I. Technical Processing in the Large Research Library

The first application is research oriented and is part of a large scale project to perform technical processing functions through on-line computer services in a large university. A prototype technical processing system was operated at Stanford during 1969, and the results published in a recently issued report. Enough was learned from the prototype to justify development of a production system, an effort which is now in progress. Central to this effort is the concept of making MARC and other files searchable at a CRT. In the prototype system, MARC could be searched by:

- author
- words in the title
- words in corporate body headings
- publication date
- LC card number

Words could be truncated if desired. This software supported interactive searches: simple searches (one access point) or compound searches (several access points connected by Boolean operators). Typewriter terminals were used for the prototype.

What I will next be describing is a set of external specifications for the production system; note that these specifications do not refer to a system which is operational.

Once it was proven feasible to search centrally distributed data bases by terminals, it was next desired to construct local data bases to cover records not within the scope of the MARC Distribution Service. An objective was to design a CRT screen format for input of bibliographic records. Several existing input systems were examined and a set of design criteria postulated.

1. The number of transactions between the CRT terminal and the main frame should be minimized.
2. The incidence of data overflow in various fields should be minimized.
3. A significant amount of input data should be pre-edited before new records are added to the data base.
4. The terminal should display the roman alphabet in upper and lower case, but *not* diacritical marks or special characters.

5. The need for coding and tagging should be minimized.

To meet the first three of these requirements, considerable statistical knowledge of bibliographic records was required.

To minimize the number of terminal/computer transactions, it was necessary to input and process as many data elements in one step as possible, i.e., as many on one screen as possible. The number of data elements which can be contained in a single screen display is a function of the screen or buffer capacity and the lengths of data elements. The practicality of working with a screen containing many data elements is also dependent upon knowledge of the frequency of occurrence of data elements. Our analyses enabled us to rank order the principal data elements in a bibliographic record both by frequency of occurrence and by length, so that an operator could enter a nearly complete bibliographic record on a single CRT screen approximately 90% of the time. I say "nearly complete bibliographic record" because at the time of ordering, complete bibliographic data are not available for many records. Statistical analysis of MARC was indispensable to this design work.

A prime advantage of input on a formatted CRT screen is that little or no tagging is required because everything on the screen is already mapped to a set of data elements. For this type of design one needs to know more than the minimum, maximum or average length of data elements of variable length. For each data element to be input on a screen with a 90% chance of avoiding overflow, we must know in advance the field size which will accommodate 90% of the occurrences of that data element. A series of MARC tapes was analyzed for this purpose and the 90% point calculated for each variable length data element. In the event of overflow, our requirements specify software to automatically call up a second screen to accommodate those 10% of the data elements which exceed the maximum length which fits on one screen. The second screen will repeat all of that data element thus far keyboarded and provide the balance of the screen for additional input. The need for overflow screens and special software will be reduced as programmable CRT's with large buffer capacity become available at prices libraries can afford.

We also will require pre-editing of certain fixed fields to assist the terminal operator and increase the probability of creating accurate data. For instance, date of publication must be 4 characters and equal to or greater than 1450. Several dozen error conditions will be detected and reported back to the operator. Needless to say, we make no attempt to anticipate spelling errors.
II. Regional MARC Network

Development work on the production system has, unfortunately, been slowed owing to the diminution of research funds. However, we expect to receive our first four CRT terminals within the next few months and will be ready to begin support of the second application, which is oriented towards the utilization of MARC in a regional network.

Stanford is presently working with four other academic libraries in the San Francisco Bay Area in the conduct of a feasibility study to determine whether a MARC data base can be used by libraries outside of Stanford for acquisition and cataloging. We know this type of networking is technically feasible — the job here is to assess economic feasibility and political viability. To assist in this task, two persons from each institution are working about half time performing local cost analyses and gathering data to indicate the potential utility of MARC for a regional network. Statistics indicate that Stanford's work load is about equal to the combined load of the remaining four libraries. Here are some sample annual data:

<table>
<thead>
<tr>
<th></th>
<th>Stanford University Libraries</th>
<th>4 Bay Area Academic Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Total purchase orders</td>
<td>32,000</td>
<td>37,000</td>
</tr>
<tr>
<td>Expected no. of MARC searches</td>
<td>23,000</td>
<td>24,000</td>
</tr>
<tr>
<td>Expected no. of successful MARC hits (=no. of p.o.'s which can be generated without further keying)</td>
<td>14,000</td>
<td>14,000</td>
</tr>
<tr>
<td>No. of process slips or workslips produced (multiple copies)</td>
<td>81,000</td>
<td>91,000</td>
</tr>
<tr>
<td>Standing Search Requests (SSR's)</td>
<td>5,600</td>
<td>5,600</td>
</tr>
<tr>
<td>SSR's: matches found</td>
<td>3,600</td>
<td>3,800</td>
</tr>
</tbody>
</table>

It is proposed to provide on-line searching of current MARC, including SSR's plus complete packs of printed cards sorted for immediate filing in each library's catalog.

Machine cost estimates have been prepared for the following categories of work: (1) on-line transactions, (2) batch transactions, (3) overhead, such as data base storage, rental of terminal, batch utilities, etc. All cost estimates are based upon currently prevailing rates for Stanford's IBM 360/67.

A. On-line transactions

It is estimated that on-line searches will range in cost from about $.02 through $.08 and $.17, depending upon search complexity and amount of communication between the terminal and the main frame. Input of new bibliographic records will cost about $.05. All figures are for machine time only; labor is excluded.
B. Batch output transactions

This includes purchase orders, cards, process slips, spine labels, matches to standing search requests, and the cost of purging no longer needed SSR's. All are estimated at $.02 each.

C. Overhead

Each CRT terminal rents for $225/month at current prices. Terminal connect time is $3.50/hour. Communication line costs will range from zero for Stanford, which has its own lines, to $170/month for the heaviest, most distant user. Disc storage looms as the largest cost — about $1200 per pack (IBM 2314) per month. About two packs are needed to maintain current MARC records, and their indexes for the preceding six months. However, it should be noted that this fixed cost is independent of the number of users — it costs just as much to store MARC for Stanford as it does to store it for a dozen or more libraries in the region. The same cannot be said of maintaining individual depository sets of LC cards, or reproducing, overprinting, proofing and sorting cards — all services which can be performed centrally.

III. CardSet

I shall now describe a new application of MARC which is being developed by a commercial firm in consultation with libraries at Stanford. This is CardSet, a MARC package which offers the complete MARC data base and is marketed by Information Design, Inc., of Menlo Park, California.

Output of CardSet is a complete, headed set of cards for each MARC record selected by the user. Cards are produced in Times Roman typography (the same type used for LC printed cards), and have LC call numbers, headings, and tracings in place. The end product is obtained through a two stage process: MARC tapes are converted to microfilm card images, and the images are converted to cards through xerography. Two separate, cartridge-mounted reels of film are produced: one is a sequential file containing full LC/MARC bibliographic data; the second is an index to the first. Both are 16mm microfilms. The LC/MARC file may be designated the "master data base". It grows sequentially, each weekly tape's COM equivalent merely being added to the tail end of the previous film. (In actual practice, no splicing is done — a complete, new cumulated update of the film is generated, and the previous version discarded). The index film is cumulated and completely replaced twice a month.

The CardSet system provides rapid access to MARC records through a two-step, lookup process. First, the index film is consulted; the index film is used in a low cost reader or reader/printer. A user may look up an item in the index by Library of Congress card number, by title, or by names of series. From the index entry, he obtains a unique address which gives the location of the corresponding full bibliographic data on the second film. If a work has not yet been cataloged but has a
preassigned LC number, the "address" is the abbreviation NYC, for "not yet cataloged". The film of the master data base is loaded into a Xerox Microprinter, and with the aid of a rapid location device (the Rap/ID), the addressed frame containing the full bibliographic data is brought to the screen.

On the screen, it can be seen that each "page" of COM film in the master database contains six card images. This feature makes possible the unique ability of the CardSet system to generate complete sets of cards with LC call numbers and headings already overprinted. Each of the six card images has been properly formatted and headed in the COM unit. When the image is properly centered and the Microprinter is loaded with pre-punched card stock, it is possible at the push of a button to make a full set of cards ready for filing in a library catalog. If an entry requires more than six cards, CardSet produces the number of necessary additional frames (of six cards each) to accommodate the right number of cards.

Introductory price of the CardSet service is $3,000 per year, inclusive of the master data base, twice monthly update service to the indexes, and rental of the reader/printer for the indexes, but not the Xerox Microprinter for card production. Use of the Rap/ID frame locator device is included with the Microprinter. At a volume of 20,000 exposures per years, the direct cost (exclusive of labor) per set of six cards is estimated at $.12 each per set, or 2¢ per card. Satellite index stations are available at a rental of $1200 per year; this arrangement serves decentralized library systems or regional networks. All MARC records produced to date will be made available to subscribers, and all records remain the property of the buyer.

IV. Conclusions

Inherent in the development and implementation of large systems are manifold human and managerial problems. These problems, which are described in a separate paper (2), include: articulation of quantitative goals, self-evaluation of current systems, assuring of adequacy of communication between users and system designers, user involvement in the design, documentation of the design, scheduling, estimation of operating costs, fund raising, external communication, and, in the case of networks, marketing. Problems of personnel aspects of library automation are described in (3), and an analysis of major decision points in library automation in (4).

The utilization of MARC in large-scale systems, particularly on-line applications, is exceedingly complex and expensive to develop. To operate large computer based bibliographic systems is probably economically infeasible for any one institution, with the possible exception of the few largest. Even so, because neighboring institutions predictably buy and service a subset of the library materials acquired by the larger institutions, it makes economic and political sense for users in a region to band together in the interest of sharing the fixed overhead costs of large systems.
The United States has a long tradition of local autonomy in library management; in the U.S. no network can be created which imposes conditions upon its users. Both cooperation and standardization must be negotiated; they cannot be achieved by administrative fiat. Thus, the challenge in creating the network is not technical but political.

However, as is clearly stated in the 14th annual report of the Council on Library Resources (5), the development of automation and the continuation of local autonomy in bibliographic operations are mutually exclusive. The decline of research and operational budgets for libraries in the U.S. may well force librarians to choose between standardized products and services or seriously degraded ones. As an alternative, a commercial service may be called upon not only to package bibliographic data bases but possibly also to engage in facilities management, i.e., to run libraries on a "profit" making basis. Whether this can be done while maintaining or increasing the level of service remains to be seen.

References and Bibliography


