

Cataloging Systems: 1973

Applications Status

Maurice Freedman

I would like to begin with some acknowledgments. I am not a systems librarian, and I haven't had any hands-on experience with machines. Brian Aveney, John Knapp, and Mike Malinconico are the people who have helped educate me to the realities of hardware. I am especially indebted to Jerry Pennington, the assistant director of technical services at the Hennepin County Library and systems analyst there, without whom nothing would have worked.

INTRODUCTION

The applications status of cataloging in 1973 is quite difficult to characterize by such evaluative terms as good, bad, healthy, hopeful, of concern-it is all of these things, depending upon one's focus. In the past five years there have been substantial controversy and concern primarily over the automation of cataloging, perhaps the most complex area to have been seriously dealt with by systems librarians. In some respects it is ironic that the successes finally started to come at the same time the most reactionary statements were being made. An effort will be made to identify some genuine causes of concern, highlight some of the successes, and point out some of the directions which must be pursued. Let us begin with some general observations.

Perhaps the most obvious thing which can be said is that the days of the massive megabuck research enterprises are, for practical purposes, behind us. The granting agencies just do not have the same quantities of funds available to subsidize research they had in the past. As well, some of the research results, obtained at great expense, need not be duplicated

Director of Technical Services, Hennepin (Minnesota) County Library System

and are available to all. In the next five years, it will in the main take hard dollars directly from library budgets to pay for cooperative or individual efforts. The past five years of automation have seen some substantial commitments: the Stanford University Libraries through the period of 1967-74 will have spent grant money totaling close to \$3 million on the BALLOTS-SPIRES projects, while The New York Public Library, unable to obtain outside funding, spent about \$4 million of its own funds to develop its various bibliographical control systems, just to mention a couple of the heavyweights. Finding the money in the period of 1973-78 will take cost justification, greater ingenuity, tenacity and, in all probability, more self-sacrifice than was needed in the past quinquennium.

Continuing in a general mode, that the computer can be successfully used as an aid to the cataloging enterprise has been amply demonstrated. There have been spectacular successes proving that the computer technology is viable and that the blue sky is no longer beyond the horizon but is unquestionably the atmosphere we live in and breathe.

Notably the Ohio College Library Center has performed the feat of providing an on-line cataloging service for libraries all over Ohio and elsewhere for libraries in Georgia, Pennsylvania, and New Hampshire. Printing the catalog cards at night for titles requested or cataloged during the day, OCLC is providing a cataloging service apparently less expensive and far more efficient than its participants can provide for themselves. By the summer of 1973, an OCLC user will have on-line access to a data base of almost 1 million records.

Although less glamorous but at least as significant, The New York Public Library has solved the problem of authority control through the use of computer technology, thereby permitting automatic control of all subject and name authority terms which occur within bibliographic records, as well as cross-references, scope notes, and other authority components intrinsic to professional cataloging. Those of you who have received letters from LC explaining why changing such terms as "aeroplanes" to "airplanes" is too expensive should appreciate that a single input transaction to NYPL's authority control system can accomplish this kind of change, thereby liberating catalogers from the tin can of outmoded terminology to which they've been tied.

The most widely felt effects of the successful application of computers to cataloging are on two cataloging products: the book catalog and the catalog card. Commercially and noncommercially produced book catalogs can be found in such disparate places as Ventura, California, and New York City. Such commercial organizations as Science Press and Autographics and, in the noncommercial sphere, the Orange County

(California) Library and The New York Public Library have produced book catalogs based on a variety of principles, but all are successfully continuing to produce them with regularity. Critics of automation should take note that in 1968 it was doubtful whether a book catalog exceeding 100,000 titles was possible. The extraordinary *University of California Union Catalog Supplement, 1963-67* with its 750,000 titles and 2.6 million entries demonstrates the current capability, not potential, of the digital computer. Book catalogs can provide the least expensive, most efficient, and most timely means of access to a library's resources for the greatest number of people.

Line-printed catalog cards produced by commercial or institutional organizations are found in libraries all over the country. Such competing firms as Josten's Catalog Card Corporation, Richard Abel, and Baker and Taylor are using line printers to provide card sets to their customers, who at least in part are incidentally purchasing books they job. The University of Chicago, an original MARC Project participant, has printed over 2 million cards for its own use, and OCLC, a nonprofit corporation, has a fantastic card service. It is clear that computer-generated cards have enjoyed a sound fiscal and technological base.

In 1973, the question is entirely decided that the computer can be used as a tool to facilitate, change, and improve cataloging techniques, processes, and services. In 1968, this point might have been questioned: hardware considerations notwithstanding, few could have envisioned an Ohio College Library Center or the *University of California Union Catalog Supplement* as realizable in the next five years. This is not to say that there have been large expenditures for automation projects where the results have been primarily of research value, or where the results are not yet known.

MARC AND LC SERVICES

From the present perspective, the work titled *The MARC II Format*, written by Library of Congress staff members Henriette Avram, John Knapp (now with Richard Abel Co.), and Lucia Rather, has been the single most important event in the automation of cataloging. With a national standard for the communication of bibliographical information, the feasibility of sharing machine-readable cataloging data was permanently established. Since the inception of the MARC distribution service in 1968, the Library of Congress has disseminated over 300,000 machine-readable English language monographic cataloging records in the MARC

II Format, and currently includes Cataloging-in-Publication data. The direct or indirect dependence of the library community on the MARC service is a reality. Whether libraries get card sets from commercial firms, OCLC, or their own line printer (as the University of Chicago and Columbia University do), or whether they, as does the New York Public Library's 82-library branch system, get the majority of their catalog records from MARC for ultimate reproduction in a book catalog, the MARC service is a necessary component in the nation's cataloging service. Information Design, Incorporated, founded a searching and cataloging system which was originally based wholly on the MARC service. This firm takes the MARC tapes, extracts and builds title, series, and card number indexes from them; it also has a Videocomp create a set of card images on microfilm from which the full set of overprinted cards can be made.

The Library of Congress, which has supplied a variety of bibliographic services to the country since 1901 when it began to distribute catalog cards, has not yet been able to accomplish all of the things in automation it set out to do. In the 1960s, LC, through its various services, found it increasingly difficult to meet the mushrooming demands of the national library community. The demand for catalog cards created an opportunity for card services that commercial suppliers leaped to fill, and prompted LC's Processing Department to recognize that automation was the best approach toward solving LC's card distribution problem. LC's Inability to fully implement Phase 11 and produce cards on demand for all records in machine-readable data has had several serious implications for the library community, some negative and some positive.

MARC service subscribers have been seriously affected. The Library of Congress has had the capability since September 1971 of producing photocomposed catalog cards of such quality that the average person could not distinguish them from the hot-type printed card. But LC is still following a procedure which, to the outsider not fully appreciative of LC's internal complexities and problems, seems paradoxical: the English monograph records are first printed using the traditional printing methods, and only after the finished printed card is approved by a proof editor does the conversion to a MARC record begin. It is painfully clear to both LC and the MARC subscribers that eliminating the hot-type process would probably eliminate the delay in distributing MARC records.

Furthermore, libraries purchasing either card sets or card services, such as NEWS sets or proof slips, either pay for the service with extended waits or, in the case of the NEWS set, with exorbitant costs. The NEWS-set service currently costs about \$325 per month, and provides a library or

firm with the English-language catalog card output of LC on a daily basis, and on the average two to three weeks in advance of the offset-produced proof-slip service, which includes most of LC's foreign- language and English-language cataloging. The success of Phase II of the Card Division's system is a necessary precondition for lowering costs and hastening deliveries of the basic card services.

On the positive side, resourcefulness, ingenuity, and creativity characterized the library community's response to the problem. In the commercial area, Information Dynamics Corporation and Information Design, Incorporated, have adapted LC catalog products by combining microfilm and computer technology to provide services which enabled libraries through micrographics to produce their own catalog cards. In addition they include multiple indexes to the LC cataloging data, thus instantaneously improving over the manual single-access files of proof slips maintained by individual libraries, and eliminating the file maintenance problems associated with traditional card files.

As mentioned earlier, the high-speed line printer has come into full bloom as a card-printing device. Through consultation with the American Library Association, IBM implemented the ALA print train which contains not only upper- and lowercase characters, but also all the diacriticals required by various Roman alphabet languages. In 1973 computer-generated catalog cards are routinely produced in all sectors of the nation, the MARC records being a chief input source for these suppliers.

At least one commercial book catalog firm, Science Press, is using MARC records as an input source, primarily for public libraries. The Oklahoma State Department of Libraries' MARC-0 service provides a MARC-based SDI service, as well as listings of MARC records. The Richard Abel Company and the University of Massachusetts use the MARC service as a source for providing catalog cards and processed books. Still other uses of the MARC service are made. Some of these will be considered in the discussion which follows.

SOME PROJECTS OF SPECIAL SIGNIFICANCE

The Ohio College Library Center has developed the nation's best card delivery service. One of the major preconditions for OCLC was the difficulty and cost to libraries in the sixties to adequately solve their basic cataloging problems: namely, how can the library quickly and cheaply get authoritative cataloging information? A nonprofit corporation was

formed in Columbus, Ohio, and as of June 1972, OCLC had 49 members in Ohio plus experimental users at Dartmouth College, the Cooperative College Library Center in Atlanta, Georgia, the Pittsburgh Regional Library Center, and the Union Library Catalog of Pennsylvania with terminals at Drexel, Temple, and Pennsylvania Universities. OCLC under the direction of Frederick Kilgour and backed by massive funding from the State of Ohio and grant sources established a center for computer-based technical services which up to the present has produced arranged catalog card sets on demand and on-line union catalog information.

Although already described in a variety of places, some of the basic features of the system are worth examining and evaluating. OCLC has a Xerox Sigma 5 computer with its own operating system, an OCLC developed system specially geared to efficiently deliver products and services. The OCLC file of cataloging data has been growing at the rate of 30,000 titles per month and is creating one of OCLC's severest problems: the capability to store new cataloging data. In the first third of 1973, the secondary memory will be doubled, but it will be only adequate for one year. It will also be difficult for OCLC to take advantage of any new equipment manufactured by other firms because of the dual problems of rewriting the portions of the operating system required to operate the new devices, and building the black boxes required to physically connect them to the Sigma 5. (OCLC once solved similar problems when it successfully connected the IBM 1403 N-1 line printer to the Sigma 5.)

OCLC has two input streams to its cataloging data base. The first is the input of MARC records from LC. The other is the addition of original cataloging on-line when no MARC record exists for the given title. Two separate problems arise from this. It is difficult to control the quality of the record which is originally input. The first record of a title becomes the permanent one with the exception that a MARC record can bump an originally cataloged one. Furthermore any user may decide to add a record, irrespective of the existence in the file of a record describing the same bibliographical entity, thus further exacerbating the storage problem. Unfortunately, the sharing of cataloging of various levels of quality has real limitations and problems. Another way of characterizing this deficiency in the system is to note that there is no built-in control of name and subject authorities entering the system via OCLC input.

The other problem, which has a tremendous impact on cost, is the storage of all LC MARC records on-line. As of January 1973, 62 percent of the LC MARC records had never been used by OCLC members. Of course, it is difficult to know which 38 percent will be used, but some

alternatives might be sought. (The University of Chicago announced at one time that it found that it was only practical to save the current six months of LC MARC output. Further discussion of this problem will occur later).

There are three methods of access to the data base of OCLC: by LC card number, author-title (the first three letters of the author's surname and first three letters of the first significant word of the title), and title only (by keying in the first three letters of the title's first significant word and the first letters of the next three words of the title). The implementation of this truncated search method is one of the great accomplishments of OCLC. In a matter of seconds the entire data base is searched and the record (or records), if found, displayed.

Among other things which happen each evening, in Columbus, Ohio, catalog card sets are printed in a single arrangement for all of the records for which a given library requested cards. During the day holdings records are instantly created for each institution cataloging a given record. So OCLC is truly creating a powerful on-line Ohio union catalog, as well as an excellent card service. And at least two networks have plans to replicate OCLC. In 1968 most librarians were not really believing this would be possible.

The other major MARC-based on-line system is BALLOTS (Bibliographic Automation of Large Library Operations using a Timesharing System), a project which began in 1967 at the Stanford University Libraries. The coordination of two Stanford automation projects, BALLOTS and SPIRES (Stanford Public Information Retrieval System) has allowed the BALLOTS system to extensively use and incorporate several features of the SPIRES management and searching software. BALLOTS-MARC is the first of eleven modules of a comprehensive BALLOTS system to be fully implemented by the end of 1974. (This discussion, as with OCLC, is restricted primarily to that which is operational. Since I've been here, I've been informed that module 2, a process information file, is operational; module 3, which includes a file of permanently cataloged records, is going to be operational next week). BALLOTS-MARC offers its sole user at this time, Stanford University, neither authority control nor a large data base. But it does provide a fantastic search capability. It presently contains only a small portion of the LC-MARC data base and provisions for original input will come in subsequent modules. BALLOTS-MARC is used both in the acquisitions and processing modes, as well as in cataloging. When a MARC record is found, BALLOTS-MARC has the capability of producing a purchase order, a vendor invoice, seven possible different file slips, two book labels,

and an average of ten different catalog cards. The BALLOTS-MARC capability is particularly effective in the area of information retrieval. The search method is interactive and negotiable. There is on-line access by virtually any form of, or combination of, name(s); by one or more title words; or by combinations of name and title words; and by LC card number. If more than one title meets the search request, then the number of titles meeting it is displayed. The user then has the options of either expanding or decreasing the information in the search request, or displaying each individual citation. This ability to readily manipulate the search request is unprecedented in the use of searching general library catalog files. It is questionable whether other automated cataloging systems will in the near future be able to afford to duplicate Stanford's search capability on a large file.

An interesting sidelight of the Stanford research was its conclusion that capturing diacriticals in machine-readable form and displaying them on CRTs was not cost-effective. Research libraries which regard these marks as an absolute necessity might ponder Stanford's results.

The other automation projects which should be discussed in some detail in a 1973 review are the New York Public Library bibliographical system, the *University of California Union Catalog Supplement*, and the Hennepin County Library book catalog project. Each of these is especially significant for different reasons.

The New York Public Library's automation effort produced the first wholly computer-controlled bibliographical control system. NYPL was forced to automate because of the deteriorating condition of its card catalog and the continuing increase in the cost of manually maintaining the catalog. The NYPL system currently has as its major features: a provision for MARC input and original input, maintenance capability, authority control, sophisticated sorting and filing capability, and the power to regularly produce a photocomposition driver tape for printing book catalogs. The three basic files the NYPL system maintains are: (1) an authority file of names, subjects, series, and conventional and uniform titles, which includes the filing form of all authorities as well as catalogers' notes, cross-references, and other such information one normally associates with manual authority files; (2) a bibliographical file of records in a MARC-like format which contains the full catalog record with MARC tags, indicators, and subfield codes, and also pointers from each authority term to the control number of that term on the authority file; and (3) the MARC service data base, which is periodically searched as LC card number queries are made of it.

One of the refinements of the NYPL authority control system is the

automatic authority search performed for all new entries added to the file. It works as follows: all new terms are passed against the existing authority file and either are identical or not identical to terms on the file. If identical, the term is automatically accepted. If it matches a cross-reference, it automatically gets switched to the valid term. If not identical to a term or cross-reference, it is sent back to the cataloger for maintenance or approval. NYPL built in the feature of automatically accepting all authority terms entering the system from MARC records.

The system of bibliographical control NYPL has developed is the most comprehensive mechanized bibliographical control system in large scale at this time. This writer has seen printouts from a prototype authority control system which has operated at Richard Abel Company. It is understood that in the State of Washington the MECCA system has automatic authority features and that Science Press advertises automatic authority control with its MARC-based book catalog service.

In the process of developing its authority list, NYPL created an invaluable file of data which in either hard copy, machine-readable form, or microform can be an invaluable aid to catalogers around the country. It will continue to have such value until LC can make its authority information available.

NYPL has done extensive work in the area of automatic filing. Most filing forms are automatically generated, but provision is made for the insertion of special filing forms for which algorithms could not be utilized. For example, the Revolutionary and Civil Wars file automatically by date and not by spelling. (Altogether there are 54 chronological subdivisions automatically recognized). But, numbers in filing positions in titles have to have sorting forms manually inserted.

The NYPL bibliographical system's chief deficiency is that it is a batch system. There is the laborious process of sending proof copy back and forth to the computer, but this is true of all batch input systems which, of course, cannot provide for interactive cataloging.

The *University of California Union Catalog Supplement, 1963-1967* (also known as UCUCS) represents a notably different strategy of library automation. Ralph Shoffner, Project Director, Institute of Library Research (ILR), University of California, Berkeley (and now with Richard Abel Co.), developed a strategy which had as its target the conversion of a catalog card into a machine-readable record with the MARC 11 structure, but with less specificity than required by MARC for catalog records. For printing purposes, the *Supplement*, a book catalog, did not need to distinguish personal name main entries from corporate or other name main entries; hence all nontitle main entries were assigned the MARC tag 100, which normally is used only for personal name main

entries. Other such simplifications were introduced with the additional consideration that an error rate of one error in each five entries would be permissible. It was the ILR concept that machine methods could be introduced at subsequent dates for the purpose of refining MARC tagging and coding, and automatically identifying and/or correcting errors. The obvious corollary is that the initial product has many more errors than one associates with research library catalog products, and the data base is of less value to potential customers than if it had been more precisely tagged.

Underlying the project was the demand that costs be kept at a minimum. The concept of automatic field recognition (AFR) was introduced as the tool for the conversion of 750,000 titles comprising the catalog. AFR, as implemented at Berkeley, bypassed the batch input process which usually begins with a human editor explicitly identifying and tagging the various data elements on a catalog card or work sheet. Instead of the computer having the incoming information properly labeled, AFR automatically tagged and coded the data elements by identifying certain cataloging elements defined as constants. For this to work, the input operator had to follow certain conventions (e.g., always begin with the call number), but then did type directly from the catalog card. The Berkeley AFR program, although far cruder than the one developed at LC for the RECON project, did allow for a cheap conversion of a catalog record to the MARC II format with codes and delimiters sufficiently specific for the production of the UCUCS catalog. (The RECON project at LC utilized a format recognition program geared to produce a pure and precise MARC 11 record. RECON tested the feasibility of the RETrospective CONVersion of existing catalog records, working directly from the cards for a relatively small number of titles. LC has no immediate plans for a large-scale retrospective conversion.)

In addition, a maintenance program called FIX was developed which had the dual characteristic of allowing one to change something as specific as a single character within a given subfield or as general as deleting an entire field or record, and then building a new MARC record. Sort keys were then built for all of the filing fields and a tape (actually several tapes) was produced which ultimately drove a Videocomp to photocompose the page masters.

Looking at UCUCS itself, one will find a host of typos and problems. On the other hand, it is inconceivable that a clean catalog could have been created in the short period of time for the relatively low cost: 750,000 titles, 42,000 pages, 47 volumes, and 250 class-A bound sets cost a total of 1.25 million dollars, or \$1.16 per title. The machine-readable data base generated in the process should prove to be of value to other

libraries or commercial firms despite its condition. UCUCS's usefulness as an acquisitions and cataloging tool for the various University of California libraries is obvious, and assuredly should manifest itself at some point in the literature.

Previously mentioned was the fact that NYPL had expended \$4 million for research and development. The Institute of Library Research of course expended a significant amount of money on research and development, as well. But the Hennepin County Library (HCL), a public library serving suburban Minneapolis residents, developed an automation program predicated on the transferability of software and the universality of the MARC record. Unable to allocate substantial amounts of money for research and development, HCL negotiated the transfer of the ILR software. Implementing and modifying AFR and FIX to meet local cataloging requirements, Hennepin converted its shelflist of approximately 85,000 titles to the MARC 11 format. Having budgeted \$1.50 per title, just for the conversion, but spending considerably less, HCL, using FIX, tried to create a clean data base. (Several years ago, Richard DeGennaro had published an article in *Library Trends* saying that the price of getting a clean MARC record was \$1.50. You will recall that the University of California had a catalog of 750,000 titles for \$1.16 per record.) Controlling the flow of printouts was one of the biggest problems. The printouts for each record and the various pages of corrections ran into the hundreds of thousands of pages, which had to be stored, retrieved, and replaced on an orderly basis. This problem in part has been solved by dumping all approved records on upper and lowercase COM microfiche, thus obviating the need to refer back to the hard copy. The other major problem was the fact that HCL's production deadline of a June 1972 book catalog seemed like an impossibility because the remaining ILR programs for sort-key generation and printing had not yet been fully tested and debugged. This of course made transfer to Minnesota wholly impractical. With one full-time programmer and one systems analyst, HCL had to wait. In the meantime, the New York Public Library was in the process of announcing the publication of its *Research Libraries* and *Mid-Manhattan Library* book catalogs and the successful implementation of its bibliographic control systems.

Through a series of negotiations, NYPL agreed to make minor modifications to its software to enable it to accept an HCL/ILR/MARC record and thus produce the Videcomp driver tape for printing the 2,800 page catalog. For the first edition of the Hennepin County Library Catalog, NYPL actually processed about 80,000 records through its system. In the process, HCL, using the NYPL authority file, and

incidentally getting the benefit of all matches, cross-references, scope notes, etc., was able to clean up many of its variations in entry with simple authority transaction. (This was a key element; we are a suburban library that has evolved from a very small library. The procedures involved over a period of thirty years were almost infinite in variety.) The resulting catalog of course would not have been possible working through the ILR system. On the other hand, passing 80,000 records through the system at once with little time for review because the publication deadline had already passed had at least two negative effects: (1) many filing and authority errors did get through, and (2) a minor error occurred in the call number field (any call number over 12 characters in length was converted to fiction).

At present, the Hennepin County Library is in the process of transferring the entire NYPL system as it is constituted for NYPL's branch libraries to Minnesota. HCL's current plans are to use the MARC input portion of the NYPL system, but to continue to create MARC records through AFR and FIX for original cataloging rather than the more complex and expensive input utilized by NYPL. This should once accomplished, establish the transferability of entire complex computer systems and interrelatedness of large systems through the MARC format. In reality this task is proving tougher than anticipated. Hennepin County will have to add a second programmer to concentrate entirely on the transfer. Lastly, because of the remote processing, the supplements are taking far too long to produce.

In terms of its originality, the HCL project introduced little that was new—basically it used the products of other original library automation efforts. Its willingness to use and modify the software developed elsewhere will enable HCL to have the capability of the New York Public Library's system and the Institute of Library Research system without the concomitant development costs.

It should be noted that the Institute of Library Research and the New York Public Library both gave their software to the Hennepin County Library without a fee. The chief implication of successful software transfer is that a library with access to a computer and a small but qualified systems staff can, in principle, use and maintain a system developed at great expense, but by comparison require only minor expenditures of staff and money to operate. It seems to this observer that from a cost standpoint this is certainly one of the most practical approaches libraries embarking on automation programs can consider. One critical concept implicit in this is accepting the wheel someone else has invented, albeit less perfect than the one you could not afford to invent on your own. Another is that library institutions other than ILR

and the New York Public Library try to be as generous and cooperative in terms of software exchange. The third concept is the power of MARC, without which none of this could have been possible.

In the area of practicality, note must be taken of the automation program of the University of Massachusetts Library processing center. A complete technical services system was developed which includes on-line order input and computer generated catalog kits. The University of Massachusetts system has few of the refinements and subtleties of the Ohio College Library Center, the New York Public Library, or BALLOTS; on the other hand it purchases and delivers to participating Massachusetts colleges the book selected by the college with a set of ready-to-file cards, a book pocket, and appropriate labels for a cost of \$1. Established strictly as a processing operation, the simplicity of the University of Massachusetts system offers a most inexpensive processing service to its many member academic libraries.

Computer-output microfilm (COM), a magnetic tape-to-film process utilizing a cathode ray tube (CRT), will have an ever-increasing impact on library automation. COM catalogs have been created in such disparate places as El Centro Junior College, Dallas, Texas, and Tulsa Junior College, Tulsa, Oklahoma. Each of these libraries took its machine-readable cataloging records in magnetic tape form and had a COM service bureau convert those data into microfilm and microfiche, respectively.

The most exciting application of the microform technology to cataloging is at the Georgia Institute of Technology's 'project LENDS. The library's card catalog was filmed, using traditional microfilming methods, and converted to microfiche. New cataloging data was captured in machine-readable form and published in a COM microfiche format supplement. The supplement is periodically replaced with a newer cumulative supplement. Altogether Georgia Tech currently has 50 copies of its main card catalog and its most recent cumulative supplement available throughout its processing areas, in various department libraries and offices, and even at a remote campus. The net effect of this project is the successful and inexpensive access it provides to Tech's collections. This use of microfiche is the cheapest method yet of breaking away from the location of the centralized library catalog without losing the valuable information it holds.

In relation to COM, another far more sophisticated device has been alluded to several times, the electronic photocomposition device. Videocomp, Fototronic, and Photon are brand names for the devices which at high speed set the cold type for the LC Phase 11 catalog and card plates, the Information Design microfilm, and the page masters for most of the

book catalogs being printed these days. The per-page cost of these typesetting devices has plummeted in the past few years. One can now have a book catalog of graphic art quality, as opposed to line printer quality, without incurring excessive costs.

SUMMARY AND CONCLUSION

Let us review what has been discussed up to this point. Several general points were made at the outset asserting that the use of the computer technology in the last five years demonstrated its viability and usefulness when applied to catalog processes and products. Line-printer cards and computer-generated book catalogs can be found all over America. It is entirely beyond question that computers can be used for cost-effective cataloging processes. As well, the various systems discussed demonstrate that for practical purposes the technology imposes few constraints: both on-line and batch systems have been found useful and operative. The chief cost problems seem to be associated with the effort to provide on-line access to huge quantities of data, the requirements of ultrasophisticated search capabilities, and the difficulties of getting clean data using batch input methods. We have noted that it is possible to transfer complex programs; and that a field recognition program geared toward producing a generalized MARC record is capable of minimizing the cost of creating a MARC record.

The national dependence on the Library of Congress' success in automating its technical processes was also emphasized. Automation efforts around the country are critically dependent on LC's ability to provide prompt machine-readable cataloging records and will be measurably enriched when LC successfully implements its plans for the distribution of MARC name and subject authority data. Through the New York Public Library's success in the area of authority file development and control, it is clear that at least the technology is already available for this process,

The Institute of Library Research's publication of the LC subject supplements in a cumulative edition has been of great value. Nonetheless it is still distressing to note that the cut-off date for data for the most recent edition of the LC list of Subject Headings was June 1964.

It is difficult to predict the next five years. Clearly there will be more and more automation of cataloging: more institutions will be engaging in it, and more will be receiving products of such efforts. I will close with my hopes for the future.

I hope that the success of OCLC is not converted into an OCLC replication movement which will cut off the development of alternative network systems. Such other systems should include at least one which begins its implementation with some form of rigorous bibliographical control built into its initial stage, and one which would attempt to utilize standard operating systems and standard programming languages. Certainly this latter goal, if successful, should contribute substantially to the successful transfer and maintenance of the system.

In a lighter vein, I hope that the LC card number will continue to be used by LC and other American libraries for identifying catalog records. Invented at the turn of the century to identify printed catalog cards, and having successfully withstood and repulsed the future shock of the check digit, the LC card number appears to be the principle form of access to even the most sophisticated data base systems of today.

I hope that a central data bank will be created, preferably under LC's control, to which all networks can have on-line access for titles which their member libraries do not have holdings. Obviously the burden of an individual network's carrying large numbers of titles not held by its members is costly, and should not be necessary.

I hope that the next five years see far greater utilization of the same software in different libraries. Libraries not subsumed by networks can have a high quality system with a limited expenditure. As well, all libraries should try to be as cooperative and generous as the New York Public Library and the University of California have been.

I hope that automation librarians investigate the possible application of the COM technology. COM is cheap to produce and cheaper. To duplicate; it is easy to use; and it has a fast turnaround time.

I hope that systems librarians will neither have to waste their time defending library automation in the future, nor, at the same time, make outrageous claims or promises. There was altogether too much of the latter in the 1960s and of the former at the outset of the 1970s. The sequence is not coincidental. (I might add that I was a little distressed this morning to hear hierarchical notions invoked regarding systems analysts and programmers, who might in some way be "better and smarter than librarians," intrinsically finding solutions that librarians couldn't find. I had hoped that we were beyond putting down librarians for not being systems people

I hope that the Library of Congress will be able to expeditiously deliver the products and services which it has promised and which the library community so desperately needs. It should also be recognized that until LC *does* meet the library community's needs, libraries must more aggressively and openly make these needs felt, as well as cooperatively develop realistic alternatives.

And toward that end, I hope that all of the libraries, institutions, and businesses creating data bases make as a Major order of business the development of a program which can convert their internal record into a MARC 11 record. This seems to be the real key to shared cataloging.

Finally, I thank you for giving me your attention for so long a time. My last hope is that this talk has been of some value.

DISCUSSION

Butler: Thanks very much, Mitch. That was a lot of material in a short period of time. We'll take questions from the floor for a few minutes, and we'll start the open forum later.

Bill Larsen (North Suburban Library System): Mitch, you mentioned the problem that you had with your supplements. You have also mentioned the potential of COM. Have you considered putting out your supplements in COM?

Freedman: Yes, I think there's a long way to go with taking COM microfiche and putting them out for general public use. We are using COM internally, for staff, but I have some doubts about the technology in terms of display devices, storage and access and retrieval of fiche in a public library situation. The problem isn't the speed of turnaround and printing. The problem is that we're sending material to New York; they run it through one part of their system; it comes back to us; we have to edit it and send it back again to run; and every time there's a hang-up, we've got a week in the mail. That's the primary reason the supplements have been slow in coming.

Judith Hopkins: Since any library which automates its procedures tends to automate in terms of its existing procedures and when it considers its future needs, I'd be interested to know, in line with your use of software developed by other libraries, what changes in cataloging procedures were required at Hennepin County? Do you feel these changes were beneficial to your cataloging in general, or made because of the adoption of the software but didn't improve your cataloging?

Freedman: No changes at all were necessary to our cataloging in terms of the conversion to a MARC record. We started with a unit record card which had already been in existence in our shelflist, and as far as what we're doing right now, we're using unit record cards for MARC

cataloging procedures. [I think it's like a red herring, though, because we're following the Anglo-American Cataloging Rules, using the LC List of Subject Headings, and the manual procedures used in taking the intellectual cataloging record to the application of those rules and putting it on some kind of hard-form document for an input operator are lower level in terms of significance, I think, than had we not been following standard cataloging rules. With the NYPL system, the authority control file has actually helped our cataloging. We've had a variety of authorities, and we adopted the position that we would not recatalog the whole collection. A public library turns over in five to ten years anyway, with the exception of basic reference works. It has helped us without our having to make major changes. When we install the NYPL system, it is my hope that our catalogers will interact with it and write out transactions to change authorities on the file. This all required new procedures, not necessarily changes in present operations. They are doing authority work every time they catalog a book.

Mary Kay Daniels (Card Division, Library of Congress): I would like to state that LC is trying to help you, Mitch, as well as all our subscribers throughout the library world. I have quite a few comments that I would like to make, because I felt there were some statements that do need elaboration.

Insofar as the card distribution service is concerned, MARC titles are indeed made available on demand, in effect. Some of them *are* handled through inventory but on a demand basis; in fact, this demand is handled on a daily basis. Every order that we receive, if it is available in a card format generated from a MARC record prior to being printed through the Government Printing Office, is made available to our subscribers through that method-first. Whichever card is ready first, that's the card that is received.

Freedman: Isn't it true that if the MARC record has not been received, the record is pulled out of the original linotype stock?

Daniels: Yes. If what you are alluding to is that there are two job streams going on at the moment for processing new card copy for English language materials, you are right. This is primarily occurring for one, and only one, reason at this time: because the printing and cutting capabilities within the Card Division that are linked to the photocomposed card copy cannot handle the volume that will be induced by the requirement for providing new card copy for these titles. It is expected that this equipment will be available in the Division by December 1973.

At that time these two job streams should no longer exist. As of now, all newly cataloged film copy and sheet map materials are not going through this duplicative process. They are going directly through photocomposition, and your depository cards, your NEWS set cards, are photocomposed cards.

Then you also mentioned the time involved in getting bibliographic materials, and I thought I would like to bring your attention to some of the current timings. The rates of turnaround have increased dramatically in the last four years, as most of our subscribers, I think, would readily admit. At present we are on a seven-day turnaround time for cards which are in inventory, and this covers about 75 to 80 percent of the titles that we receive. Interestingly enough, 50 percent of these receipts are in MARC, which shows you something of the volume of how MARC is helping us, as well as everyone else. Insofar as reprinting, anything that's in MARC we can now do on a one-day turnaround basis, using this photocomposition and offset printing of these titles by Card Division staff. Anything that's not in MARC, that therefore must be reprinted for inventory through GPO, is now on a five-day turnaround. Really, any longer timings that may occur are normally caused by orders falling into the category of preassigned card numbers for which cataloging has not been completed, into the category of reprinting cards via Copyflo for titles not in MARC that are not maintained in inventory, or into the category of orders requiring a search by author and title to determine the card number for a title prior to an order being filled.

Freedman: I welcome LC getting back into it, because I think the LC card with its authoritative data is superior to the lineprinter card.

Daniels: You also talked about some of the relative timings involved in procuring various types of cards that are available. I'd like to point out that I don't think you can say that one type of card is consistently two or three weeks behind or ahead of another type of card. The distribution of all of these types of cards is built on cycles, and the only thing you can positively predict is the sequence in which these cycles start. Anything along the line can interrupt an individual card within an individual cycle, or an entire cycle. The normal sequence of the cycle is that after all of the new copy is printed, depository set cards are pulled, are alphabetized by title, and should be the first type of card mailed out. But, concurrent with alphabetizing of the depository cards, there is the pulling of cards for NEWS sets; so that this cycle is started immediately after the first step of the depository card set cycle. For those of you who might not be familiar

with the term, "NEWS set" is used to designate English-language monograph titles. Then, proof sheet copy, which is really no longer a proof copy but is actually an offset copy of the printed card, is usually the third category of materials sent out. So, in terms of 1-2-3, it's normally depository-NEWS set-proof sheet. I've often seen in the literature various studies of the timeliness of proof sheets vs. MARC tapes, but really that is not the most valid comparison of LC data going out in a card image vs. a tape record form.

There is one other thing that perhaps is not widely known and that should be more widely known. The Card Division is now really the sole subscriber to the MARC Services. In essence, we get all the tapes produced by the MARC Development Office. We are doing the cumulation, selection, and copying necessary for the MARC subscribers. As a part of this process we are attempting to overcome some of the problems that deal with duplicate LC card numbers, particularly as you get into having more than one interactive data base, such as serials and books, where you might wish to recatalog a monograph as a serial, and you would want to delete it from one data base and add it to another. Unless carefully managed, a MARC subscriber getting two tapes on the same day with this type of information might well be misled.

Butler: Mary Kay, that has been very helpful. Mitch mentioned briefly, in passing, feeling the necessity for the completion of Phase II and didn't really explain Phase II. Maybe you could summarize? What are the remaining steps?

Daniels: Perhaps we should look at Phase II the way initially announced, the way presently in operation, and the way it should be or will be in the immediate future. Initially, the grand concept was that all the titles would be in machine-readable form, everything would be handled on demand, and would be prepared like clockwork. However, everything is not in machine-readable form, nor is there enough time in a 24-hour day to produce all titles on demand in individual sets, given the current configuration of equipment. As a result there have been reevaluations of what is the best way to handle all the card copy we do handle. Consequently, we now have two new copy card distribution systems, as I noted before, and four card reproduction systems: (1) offset printing in GPO of cards not in MARC, (2) photocomposition and offset printing within the Card Division of titles in MARC (except that modified records are not printed for distribution through photocomposition, but are printed for distribution through the GPO branch, in order that the complete information that appears on LC's Official Catalog card may be given).

Both of these first two reproduction systems are for "popular" titles, that is, for titles which the Card Division maintains in inventory. Other titles, whether in MARC or not, are considered "nonpopular" and are not maintained in inventory but are reprinted on a demand basis. To provide a little further background on this practice of popular vs. nonpopular titles, we based this split of inventory vs. demand items on ordering statistics we have maintained in the automated card procedure since 1969. So we now have approximately four-and-one-half years of data. Once a year, usually in the fall, we realign what is in inventory based on these statistics. We are presently using the algorithm that if a card has not been ordered three or more times over a period of two years, we delete it from inventory. If a card has been ordered four or more times over a period of one year, and is not already in inventory, it is added to inventory. The two remaining card reproduction systems used thus cover the redistribution of nonpopular items by: (3) photocomposition and offset printing within the Card Division of titles in MARC that are not maintained in inventory, and (4) the photographing of a master record card, and printing of Copyflo cards for titles not in MARC that are not maintained in inventory.

As to where we are going in the immediate future, the most pressing concern is increasing the amount of equipment on hand needed to process the new copy printing requests for all records being entered into MARC. We can now do all the integral parts of Phase 11, including generating new copy, reprint copy and demand copy. Insofar as new copy, we can produce not only the unit card used for depository sets and individual card orders, but also overprinted cards for LC's official and public catalogs. We can also produce all the book catalog format cards required for those catalogs which are produced manually, including main entry, subject heading, added entry, see reference and added entry in reference formats. In December of this year we do expect that we will have the additional equipment, and all the MARC that can be processed through the system, will be processed through the system.

Freedman: For those of you who want to know more about LC's automation program, there's an excellent article in the Spring 1972 issue of LRTS, which covers all phases of what LC is doing, what they promise to do, with the deadlines.

Daniels: I'd like to make just one more comment. We have talked briefly about subject headings and book catalogs. In May of this year the Card Division was given the go-ahead to actually use our film data base to produce the film book catalog. Right behind that, we will also be responsible for doing the subject headings' 8th edition.

Bruce Warren (Autographics): I don't mean this to be a loaded question, but once you've successfully transferred software from ILR and NYPL to the computer center at Hennepin, do you perceive any operational, political, perhaps communication problems, with Hennepin itself being responsible for the production of the book catalogs, especially in the supplement area?

Freedman: Do I see any problems internally for doing it?

Warren: Right.

Freedman: The only problem is, we have to do it. That's all. We do it or we don't do it, and making it work is what we've lived with since we started. We started with a dependence on ILR, and ILR, like most other automation projects in this world, wasn't proceeding on the time basis that it had planned. We were stuck waiting for the rest of the software, and we were just lucky that NYPL had a system that produced a book catalog and that took a MARC record, so we were able to match up with them. We had to make it work; we could have gone to a commercial firm as well, which we would have done, too. The commercial firms are now using MARC input to their systems, and that's the real key. It's freedom for the client as well as for the vendor. The vendor will have lower input costs, but the client should have the sense to require a MARC tape back from the vendor so it's not tied to the internal processing format of a given book catalog vendor. It can take its tape and go elsewhere if it has a chance for better service or better price, but it's liberating for both of them.

Unidentified speaker: Is the computer a part of the Hennepin County Library System itself?

Freedman: In October, Hennepin County is installing a two megabyte device. Right now it has one megabyte; we've never had any problem getting time or adequate core for anything we wanted to do.

Margaret Williamson (University of Houston): In the NYPL authority file, do they also contain and generate all cross-references needed?

Freedman: Let me put it this way: their authority file controls authorities. Their catalogs have input, either from the LC List of Subject Headings or authority work or original subject headings. Whatever they have on there, where our terms went in and got a match, either a match to a cross-reference or to the valid term, we got everything associated with that authority term from NYPL. We got scope notes that are unparalleled

