

An Approach to Developing Computer Catalogs

A method of developing computer catalogs is proposed which does not require unit card conversion but rather the accumulation of data from operating programs. It is proposed that the bibliographic and finding functions of the catalog be separated, with the latter being the first automated. Such automation is seen as being advantageous on a cost basis.

CARD CATALOGS WILL INEVITABLY BE REPLACED with a computer system. If the costs of card catalogs continue to rise and eventually meet the declining costs of computer usage, then at some point in this decade, economic pressures and other factors will force libraries to abandon card systems for computer systems.¹

It is not difficult to support this hypothesis. One need only examine the operating costs of almost any library system over the past few years. The costs of creating and maintaining card catalogs represent a significant portion of most academic library budgets; as catalogs increase in size along with collections and multiply with decentralization, these costs increase out of proportion to the growth of the library's total budget. In dealing with this phenomenon, libraries have introduced many improvements in catalog methodology, but few are yet preparing for a major conversion to a computer-based catalog.

Robin W. MacDonald is coordinator of technical services and J. McRee Elrod is head of the cataloging divisions, University of British Columbia Library, Vancouver, B.C.

One major reason many libraries are hesitant to prepare for this anticipated change is the cost of converting existing catalogs to machine-readable form. Although the cost of conversion is significant, it seems less when compared with the eventual costs of storing information, of computer time to service the inquiries, of the many remote inquiry terminals needed for access and the other related costs of developing and maintaining the necessary software systems. Emphasis on the cost of conversion have tended to obscure other more important issues.

Any consideration of the eventual replacement of card catalogs should begin with the purposes of a catalog. A basic purpose is to put the user in touch with information, either by providing holdings information or by providing a detailed bibliographic description of specific works. The first is a finding function, to find what works are available, or to find a known work. The second is a bibliographic function, to provide an exact description of items held by the library.²

Traditionally these two functions are combined in card catalogs and they reside together fairly well. The relative

importance of these two functions, the difference between them, and the possibility of separating them, are significant when considering an approach to developing computer catalogs. A full bibliographic description is needed for one, yet a brief author and title will usually suffice for the other. Studies have indicated that many users, and most catalog queries, need finding information about possible or specific works.³

With card catalogs, the cost of providing one function along with the other has not represented a significant additional cost; once entered into the information store, the continuing cost to the library remains more or less constant regardless of the amount of information entered.

With computer systems the amount of information stored represents a cost variable: the quantity of information entered will not only affect storage costs on a continuing basis but will also indirectly affect the costs of accessing the catalog. For each inquiry the amount of information processed will affect the time for accessing a given entry, the time to read, transmit and display the information for the user, and will also affect the requirements of the remote communication device used. All these factors contribute to the cost of using the system. The cost of using card catalogs, on the other hand, is generally not a concern of the library, but rather a cost the user must bear. With computer catalogs, the cost of use will be borne by the library budget.

Another reason for examining the costs of card catalog systems in contrast to computer systems is the cost of decentralization, where collections are divided among a variety of branch libraries, resulting in increased costs. Studies of catalog costs at the University of British Columbia have shown that nearly 40 percent of the cards produced, sorted, and filed into the public catalogs are required for branch library

catalogs, and since a union catalog is maintained these are all duplicate card sets. With computer systems, much of this redundant catalog data found in duplicate catalog files can be eliminated by centralized data store with decentralized access, and this access need not be restricted to library locations.

The increasing difficulty of using large card catalogs augments the need for computer systems with their greater flexibility. Card catalogs, generally difficult and costly to revise, are particularly inflexible for revising the structure of subject headings. Subject access is important to many users, but some parts of large card catalogs become nearly unusable when as many as a thousand entries fall under one heading. It becomes expensive and difficult to introduce changes locally or to follow revisions made by the Library of Congress in the subject-heading structure. Some relief for this can be provided by divided catalogs with subject guide cards; but each time a change or rearrangement is introduced, the complexity of manual filing may increase, and the user is still left with an inflexible situation as well as a subarrangement of entries that doesn't help find what is wanted. Search options would be more numerous in a computerized catalog, and revision easier. Most importantly, they will be ultimately less costly than people.

These pros and cons have been discussed before, but always with the complete unit-card-by-unit-card conversion in mind. Are there no alternatives to perpetuating the card catalog on magnetic tape? Perhaps, as a first step, it was necessary to explore that possibility. That exploration produced much valuable information, stimulated much thought, and established an important standard in the MARC format. Do libraries actually need a complete bibliographic tool at the immediate disposal of every patron? Or do libraries need better finding devices to satisfy the

most frequent demands made of the catalog? We believe there is a way to obtain a better finding device as a preliminary step in catalog conversion, without compromising the traditional bibliographic standard and without excluding the possibility of providing a computerized bibliographic tool at some point in the future.

If the two principal functions of the card catalog are separated, and dealt with as separate but interrelated requirements, the finding function could be served by a computer-based system, while the bibliographic function could continue to be served by a card file. Taking this approach, it is possible to begin by providing a computer inventory record: a file that would be linked to the circulation system, and would answer queries regarding what the library has, where it is located, and if it is available. There would be only one point of access, probably by call number, which users would have to find by using the existing card catalog. The information required for this initial file would be similar to the information now used for most automated circulation systems. The file could in many cases be established from circulation records where these are in machine-readable form. Additional sources of information for such a machine file might be automated acquisitions and serials systems.

The systems design for beginning an on-line cataloging system would be fairly critical; it would be important that the software, access methods, and basic design of the system allow the beginning file to become the central file to an eventual catalog system. Specifically, it would be necessary to be able to "build" additional files in connection with the central file, to provide additional approaches. Systems already do exist that provide multiple access points using separate "index" files linked to a "master" file. It is also important that the system be on-line because the introduction of

this mass of additional data would outstrip the present batch-processing print-out methods.

Once the central file is established, the next step would be to determine the order in which the various access points other than call number should be developed. The choice would depend upon the particular library, and could begin with an area not already covered by the existing card catalog, such as an index to the classification system used. This, perhaps, is not the easiest alternative, but it is attractive in that the central file would likely be in call-number sequence, and an additional subject access device not commonly found in North American libraries would be provided. Alternately, the standard access points of author, title, or subject could be considered. The conversion method in each case would depend somewhat on the files available for the conversion process, but essentially the various authority files would be converted. Each "heading" would be entered once only and the related call numbers entered and linked to the central file. Where divided catalogs are used, it would be possible to begin converting at the start of a particular file, and gradually replace the card-catalog access with a computer system.

For many academic libraries, subject access would probably be the first choice in terms of the number of users that would benefit. A subject file might be established by using the LC subject headings on magnetic tape as a beginning source file, the seventh edition being currently available in a MARC format, with the prospect of the supplements also soon becoming available. Replacement of that part of the card catalog immediately necessitates consideration of those users who want bibliographic detail; it would then be necessary to provide bibliographic information in some form.

One approach would be to provide a

card shelf list. This, of course, would compel users who need a bibliographic description to use one system to find the item, and a second system to obtain further bibliographic detail. At the same time there would have to be an improvement in services to users of the finding process. This improvement would become a faster, easier-to-use, locating tool, accessible from many locations, easily and quickly revised and updated.⁴ It would eliminate some redundancy in existing card systems and provide additional benefits to users and the library not possible with card systems.

This approach obviously requires selection of information to appear in the automated system, and of provisions made for those alphabets, symbols, etc., that cannot be reproduced on the standard computer auxiliary equipment nor the library network of inquiry devices. These are not easy decisions; much is unknown about catalog usage and which elements of information are most useful. Those items now romanized on LC cards for nonroman alphabet materials might be used in the finding device: call number, author, title, and imprint date. These elements are the minimum required for filing in the traditional card catalog—hence their romanization.

Not only must user needs be considered, but also the needs of the library's processing staff. In order to identify an item as an added copy or volume, the minimum bibliographic data usually required is author, title, imprint, collation, and series. This amount of bibliographic detail might not be included in the machine system; thus, the addition of a copy or volume might require the consultation of the classed card file by a library assistant, but no notation would need to be made there. In a large system, this would mean that a copy of the classed file might be needed in both public and work areas. Occasionally adding a slightly different edition as a copy by guess, it would be possible to

add copies without reference to the card file. This could be done by including the number of pages in the data entered on the machine system. Rarely would two different editions be published in the same year, with the same number of pages (other than the simultaneous two-country publication of the same title, which is already ignored for added copy purposes in many libraries). The card copy would, in any event, match copy one. The number of pages is also relevant on the machine file for thickness of books when dealing with stack management on an automated basis.

The approach outlined here would change many processing operations, both during and after development. Subprofessionals, who formerly compared Library of Congress cards with books, might be comparing MARC information on a display device with a book in hand and selecting data for the machine catalog; professionals, who formerly prepared worksheets for typists, instead might be entering information directly on a machine system. A shift in the skills required in processing would occur, along with a possible shift in the professional/subprofessional ratio. Fewer, more highly skilled subprofessionals might be required once the library is no longer producing and filing cards in such large quantities.⁵

The minority of users requiring full bibliographic information in such a system would still be consulting a card file (probably by call number). Not only would this file be the only source of complete bibliographic information, but it would also be the only place in which nonroman scripts would be shown in their original form. Oriental, Greek, Slavic, and Sanskrit materials would appear in the machine system in romanization only.

This raises the issue of the significance of the MARC developments relative to this approach. Although individual libraries should compromise on the

availability of a standard bibliographic description, MARC developments should not necessarily make this kind of compromise as well. Rather, a full conversion should be made at one location and local library systems would expect to extract data needed for the local system. The availability of the full bibliographic data should continue to be a national concern, and the frequently-proposed National Network would conceivably maintain this information in a central store, or in regional centers.

Those users requiring bibliographic details could connect through the local library system to access a data file located at some distant center. The manner of storing the full bibliographic data might include some form of graphic representation, rather than digital, for some languages, characters or symbols that may never be compatible with available, general-purpose computing devices.⁶

Although data stored graphically could not be manipulated as could bibliographic data in digital form, it could be called up for display by a machine system. Thus the bibliographic function needed by some users, a service weakened earlier, could be recovered and probably improved.

The costs of conversion, storage, processing, etc., are not individually less on this basis, but collectively, the total cost would be a great deal less than if each and every library maintained the full data. Without pursuing all the implications or requirements of a National Network on this basis, considerable study of user needs should be undertaken before massive retrospective conversions are made.⁷ Also, a realistic evaluation should be made of the technical and economic limitations for computer reproduction of the many characters and symbols needed to represent Oriental, Slavic, Sanskrit, and other nonroman alphabets. Even now, libraries are unable to influence the manufacture of

a good circulation terminal. The unpleasant truth must be faced: libraries simply don't have the influence to ensure that the necessary tools will be available when needed. If specialized devices are essential, then their acquisition should be examined before undertaking any long-range plans. Also needed are large budgetary increases over a period of time to cover the cost of computer time, terminals, networks, and other costs.

This information system would develop from a manual system through a mixed system to an automated system. During its development the system might be thought of as existing for a time on each of three plateaus. The first plateau continues parts of the existing catalog—author/title and subject files. The circulation system, serial record, and shelf list have merged into one on-line machine system, accessed by call number. Patrons would determine the call number from the card catalog, then use it to query a keyboard terminal to determine location and availability. The system responds, possibly with a visual display, giving volume, copy, location, and status information. The system efficiently conveys the addition of volumes and copies, the placing of titles on reserve, the loss or loan of items, or the removal of material for binding.

At the second plateau the traditional card catalog has been replaced. The call number has served to link information entered from a variety of sources. From the authority file or author/title file all names (personal, corporate, and series) have been entered along with the call number associated with them. From the subject catalog all subject entries have been entered along with their call numbers. Titles have entered the system through a combination of circulation and acquisition history tapes. (Earlier an index to the classification might have provided a different type of subject control. Terminals, consisting of typewrit-

er-like keyboards with visual display screens, may be queried by entering author, title, series, or subject. Call numbers with brief identifying data and holding and status information as detailed in the description of the first plateau would be displayed.

Those desiring complete bibliographic data would consult a card file by call number. Here is filed one card for each bibliographic item in the collection. For roman alphabet materials this might be a computer-produced card based on MARC tape; for nonroman script materials, a Library of Congress card; and locally produced cards for originally cataloged material. The method for establishing this file might be by cards electrostatically printed from microfilm of the former shelf list. Copy and location information would no longer be entered in the public or work-area card files.

For large or decentralized library systems, the bibliographic file might be kept centrally in card form and available at many locations on a microform, probably cartridge microfilm. At least one copy of this file would be kept in hard-copy form, the location and arrangement depending upon the particular library's requirements.

The final plateau would replace this file by direct access to a central data bank. The existence of material unique to the local collection, or reluctance to see the last vestige of the three by five inch card catalog vanish, might allow at least one of the classed card files to be maintained. The filing of but one card per bibliographic item, and that by the rapid mode of call number, would aid the processing staff as would the ease for making changes in the system.⁸

The final product, then, would function much like the computerized catalog as usually conceived. It would not, however, be achieved by unit card conversion. Nor would the data be stored in bibliographic units, but rather in a va-

riety of files (such as name authority, subject authority, and title) linked by call numbers. The local library would escape the high cost of storing and manipulating the great mass of data required to translate present card catalogs into machine form.

Some of the current concern about catalog-filing problems could be eliminated by developing various access files, so that entries are inserted using human judgment rather than programming logic. The basic file initially would be created from preordered entries, i.e., the order of the particular file being converted. Then as new entries are added to these authority files, they would be inserted by catalogers in their proper place, perhaps by instructing the system to file the new entry between two others. At any rate, it is unlikely that programming of complex filing rules would be essential; however, it may be desirable to have the system "place" new entries initially and have the placement confirmed by a filer or cataloger, with the ability to shift the logical position of an entry. Since accessing the file would be controlled by a software system, the logical or physical placement of entries may not be as rigid or as important as it is with card systems; on the other hand, there may not be a need to depart from the existing structure of catalogs. An understandable, logical arrangement should be the main objective.

This proposal requires study and development, particularly in regard to the needs of libraries for National Union Catalogs, regional and national networks, shared cataloging, and MARC developments. The idea should be pursued, however, as an alternative to the seeming impossibility of each library ever having a full MARC catalog file.

REFERENCES

1. Dolby states that "unless the storage, transmission, and retrieval of information in library archives is automated, there is no hope

- of keeping pace with the exponential growth of libraries." J. L. Dolby, et al., *The Cost of Maintaining and Updating Library Card Catalogs: Final Report* (U.S. Office of Education, Bureau of Resources [Washington, D.C.: 1969]).
2. These two functions have been frequently distinguished, but it has usually been assumed that they must be combined in a computerized catalog. Cf. Robert Hayes and Joseph Becker, *Handbook of Data Processing for Libraries* (New York: Becker & Hayes, 1970). On p. 587 the authors assert: "A catalog is usually regarded as being a very different thing from an index . . . catalogs providing very full and complete data and indexes providing very brief data." The idea of automating the index function of the catalog seems not to have occurred to the authors.
 3. Catalog-use studies were made at the University of British Columbia in preparation for the division of the dictionary catalog into author/title and subject files, and again in preparation for the imprint date filing of subject entries behind their guide cards.
 4. Peter Simmons has shown the advantage of such ease of updating in the reserve book procedure, suggesting "faster responses to changing conditions with less human effort" in the adjusting of loan periods for books under heavy demand. "Reserve Collections," *Canadian Library Journal* 29:86 (Mar.-Apr. 1972).
 5. The average card set at UBC is now fourteen. Each year, \$35,000 is spent maintaining the library's official author/title file alone.
 6. Conference on Microform Utilization, the Academic Library Environment, Denver, 1970, *Microform Utilization: The Academic Library Environment* (Denver, Colo.: University of Denver, 1971) p. 55.
 7. Ben Ami Lipetz, *User Requirements in Identifying Desired Works in a Large Library* (New Haven: Yale University Library, 1970).
 8. Time studies at UBC have shown that alphabetic sorting and filing is more time-consuming and therefore more expensive than call-number sorting and filing.