into a manual system, but it is hard to see how savings can be effected by a computer at this point unless we can get machine readable input ready-made from a source like MARC II tapes. If it is cheaper to process these tapes to find the descriptive and subject catalog information for individual items as they come to a library, rather than get it from printed copy or originate it, then we will use them. Unfortunately, present indications are that it may cost more.

Storage is one of the cheapest things we have today. Even amortizing land and building cost, we can keep reference books a year on the shelf for 20¢ apiece. At 5 M bits each, this is 4¢ per million bits per year. The card catalog is somewhat more expensive at 48¢ per year per million bits. Add an annual increment of eight percent or so and costs of conventional storage are still bearable. Not so the costs of on-line storage, which is the only conceivable form of computer storage for this type of material.

The numbers are not much better for present day library reference staff work in information retrieval. Reference work and catalog service may account for about ten percent of the personnel budget of a large library. Half of the time of these professionals may go into information retrieval. Assuming that they will have to spend a good deal of time training customers in using the computer, we may be able to save half the present budget to put toward the machine. This will not go far. In fact, my conclusion is that computerized information retrieval will require practically all new money. Major new financial support will be needed for large scale information retrieval, SDI and other individualized computer based services which we in the libraries want to provide.

Librarians have always been quick to adopt new technology, for instance for catalog card production, for micro storage, for quick, expendable copies. Computers are no exception. They are urgently needed now for inventory control. If we can afford anything that computers have to offer, it is this.

Using computer technology—Frustrations abound

by HENRIETTE D. AVRAM

Library of Congress
Washington, D.C.

The automation of libraries is a fairly recent entry to the growing number of areas of applications for computers. Is this an indication that librarians have been resisting advancing technology or could it be that the process of controlling large stores of information is so complex and the hardware, software, and brainware still too limited to cope with this complexity? Might it also be that computer specialists, underestimating the challenges, have evinced little interest in the library problem?

My experience in the library world suggests that these states and conditions have all combined with negative effect. The function of a library is to provide reference service to users and to make readily available the contents of its collections. The efficient performance of this function is directly related to the successful and timely completion of processing, i.e., the selection acquisition, cataloging, classification, and shelving of a book. The rapidly increasing number of books and periodicals places the greatest strain in this area and thus pinpoints the prime candidate for mechanization.

Before discussing one of the major automation activities at the Library of Congress and its associated problems, some facts about LC are in order to set the environmental background. The Library of Congress has in its collections about 55.5 million items: books, serials, maps, music, prints and photographs, manuscripts, etc. Approximately 75 million records contain the control information and bibliographic description of this collection. Its largest file, the Official Catalog, contains some 14.5 million records. An inventory of files showed that there are about 1,260 different files which are used in the Library's operations. Under Title II-C of the Higher Education Act of 1965, the Library has been charged with the additional responsibility of acquiring and cataloging all works, published anywhere in the world, important to scholarship. The materials flowing into the Library include items written in 70 different languages, represented by 20 distinct alphabets.

One of the basic functions of librarians is the recording and organizing of bibliographic data to facilitate access to and use of the books and other materials contained in the collections of libraries. Although bibliographic data may be recorded and stored in a variety of ways, the card catalog record has been the preponderant medium used by libraries in the United States. The bibliographic information on the catalog record is basically of two kinds: (1) a description of a book in terms of author, title, etc., and (2) some kind of notation to be used in locating the book on the shelves. The locating notation also usually comprises a means for arranging together materials on the same and related subjects. A catalog record distinguishes in a unique place one book from all the other books represented in the catalog. The catalog
card, with its basic information, can be used again and again to provide multiple access capability—usually author, title, subject—and forms the basis of what is known as the unit card system. Essentially librarians are attempting to organize and make readily available the intellectual output (books) of other humans in all disciplines. This involves the application of a rather complicated set of rules to the output of very unpredictable human beings, the result being that it is safe to say that almost every rule will find its exception manifested.

Since the Library of Congress is the major source of bibliographic information for the American library community, it was natural to conduct an experiment at LC to test the feasibility and utility of centrally producing cataloging data and distributing these data to users. Project MARC (for Machine-Readable Cataloging) was in operation for 19 months in test and pilot phases involving sixteen cooperating libraries. The project was successful and a full operational system providing selected machine-readable cataloging data for all interested libraries will begin early in 1969. During the pilot period, recommendations for improvement were received from the participants, a cost model was maintained, and the procedures for preparing bibliographic data for conversion to machine-readable form and the processing of these data were improved. The format for the interchange of the record was evaluated by staff members of many organizations: the Library of Congress, the National Library of Medicine, the National Agricultural Library, the United States of America Standards Institute Code for Information Interchange, and the proposed standard for Magnetic Tape Labels and File Structure.

The library community, although operating in a very imperfect world in terms of having both second and third generation computers, configurations progressing from minimal to maximum (when is a 1401 a 1401?), and I/O devices not capable of handling the necessary character sets, has forged ahead to adopt standards. This is a significant step forward.

The introduction of computers to libraries poses special problems in file organization and hardware while providing new opportunities for multiple access to information. We are faced with deciding how information can best be structured and stored for effective retrieval. Imposed on top of all classic functions performed by librarians, i.e., acquisitions, cataloging, classification, reference, is the function of searching. The search argument varies with the inquiry. It ranges from data on an order slip to the information on the title page of a book, to the Library of Congress catalog number, to a name in an authority file. The questions of file structure—where in the file to search and when to stop searching—are related to discovering the criteria for the determination of identity. The human mind has certain categories of analytic capability which cannot yet, if ever, be captured by machine. Therefore, we must create ploys which cause the machine to approach, in effect, the desired objectives.

Studies at the Library of Congress show that the storage requirements for 1972 is $4 \times 10^9$ characters. Interesting developments in hardware technology in the next five years should partially resolve the problems of large random access stores at acceptable costs. If we can approach an efficient solution for organizing information and consequently retrieving from the files, one nagging question that remains is how best to convert the files and in what order of priority. Because libraries cannot limit coverage in time and discipline, files reflecting the past must in time be converted to machine-readable form. Many conversion strategies have been proposed and the final decision must be based upon reasonable grounds as to use and cost. The conversion of bibliographic information requires specifications for the representation of this information in machine-readable form, i.e., decisions regarding data elements that need explicit identification and the definition of a character set for input, storage, and display. The character set needed to encode bibliographic data is essentially infinite because it is open-ended. Not only are we concerned with many languages in a multiplicity of alphabets, but in addition, any

\(\text{ASCII} \), the standard for Recorded Magnetic Tape for Information Interchange, and the proposed standard for Magnetic Tape Labels and File Structure.
author can use any character at will. The obstacles then become challenges seeking creative solutions.

Librarians and computer scientists have rarely communicated well with one another, and this lack of communication results from the fact that each group is too parochially oriented to its own field. Both groups are actually striving toward precision but each sees precision in a different way. The librarian is concerned with precision in the definition of the record, for he must be precise in this definition in order to uniquely represent a book for retrieval. The computer person is interested in precision in method, i.e., an exact description of a process, so that his program will perform efficiently and produce the output required. Machine people have a tendency to minimize the librarian's problems of precision and exhibit a general reluctance to become interested in the data except as it affects the computer application. Without a complete understanding of the complexity of the data, the capabilities of the computer are oversold, thus later causing what might be termed a credibility gap. Librarians, on the other hand, must recognize the potential and the limitations of the new technology and provide the necessary guidance for the efficient use of communication and information manipulation devices.

Success will not come overnight but will depend upon the combined efforts of the most talented people that can be found in many disciplines.

Computers in service to libraries of the future

by HOWARD W. DILLON
Harvard University
Cambridge, Massachusetts

Automation activities in libraries have been undertaken with accelerating frequency over the past ten years. It is no longer uncommon to find successful projects in almost every type and size of library throughout the country. Libraries have demonstrated that they can develop and operate ordering and processing systems to control financial and bibliographic information at the time a new item is added to the collection. Book catalogs and other holdings lists are produced and distributed in many formats. Automated circulation control systems, particularly the data collection type, are widely accepted and functioning successfully.

This portion of the panel discussion will describe a few projects with which the speaker is familiar. The projects selected have been chosen because they represent major attacks on problems which must be solved in order for all libraries to move forward with automation.

A recurring problem for systems of library processing developed over the past years has been the reliability of the data first entered into the system at the time a purchase order is placed. In off-line systems, the need to edit or replace the information used at the time of the order with better data obtained when the item purchased is in hand has been a critical update problem. Few systems, therefore, attempted to carry computer processing from ordering through to the completion of cataloging as one integrated system. On-line processing capabilities and the development of a nationally distributed, timely bibliographic record by the Library of Congress in the MARC II communications format, make it more likely that integrated technical processing systems for libraries can be made operational.

Second, given a communications format for the sharing of bibliographic data, libraries with common processing requirements are undertaking to share in the cooperative development and design of processing systems. Standardization and compatibility have been hallmarks of library systems for many years. While not always perfectly achieved, librarians have traditionally demonstrated their concern with the sharing of bibliographic data and collections. The advent of computer processing has not altered that basic philosophy. Rather, it provided an opportunity to realize the goal of compatibility with greater perfection.

Systems to be considered in this presentation will include:

SYMBIOSIS—SYstem for Medical and BIOlogical Sciences Information Searching

NELINET—The New England Library Information Network

The Integrated, Computer-Based, Bibliographical Data System for a Large University Library, being developed by the University of Chicago Library, including the subsequently coordinated activities of Columbia and Stanford university libraries in this acquisitions and cataloging system.
The Washington State University Library Technical Services System, which is a project to develop an on-line processing capability for that library.

The presentation will review the objectives of these projects, summarize their accomplishments to date, and discuss hardware or operating system software problems encountered.