Almost sixteen years ago I was graduated from what would be described in the parlance of TV as "another" library school. The curriculum to which I was exposed was probably little different from what the sister schools throughout the United States were dispensing at that time: cataloging and classification, book selection, library administration, history of libraries and books, etc. When, after commencement, I went out into the world and began applying the knowledge I had acquired at library school, I cataloged and I classified and I performed all of the other conventional functions in a manner little different from the professional routine which my colleagues had been following for decades.

Few people anywhere were talking about computers and data processing as we know them today, for the very good reason that the entire subject field was in process of being created. About this time, too, one heard nothing about such phenomena as "information explosions," "data storage and retrieval," or "coordinate indexing," because the coiners of these now popular expressions had still to mint them. All of us were—with few exceptions—blissfully unaware of the profound social and professional revolution on whose brink we stood. I for one, with my well-established, personal lack of "mechanical aptitude," never dreamed that it would be my professional destiny to become involved so heavily with the gadgetry and jargon of data processing.

There were, however, some organizations including a handful of libraries, that were already deeply involved, if not with electronic data processing, with the growing information handling problem. Among these pioneers was the Army Medical Library, previously known as the Library of the Surgeon General of the U. S. Army, and which, since 1956, has been called the National Library of Medicine. Having mentioned the topic of my presentation, I wish to make clear

Mr. Taine is Chief, Bibliographic Services Division, National Library of Medicine.
that in this paper I shall not be describing data processing activities carried on throughout the National Library of Medicine. Instead, I shall be restricting my discussion to the application of these new techniques to the Library's indexing program, where they have been given most attention to date.

This constraint is really not as narrow as it sounds. When an application is carefully selected and designed, data processing may become a powerful adjunct to librarianship as it has already become in business and technology. Although the amplitude of appropriate applications will vary with the organization or the library, the computer system will remain merely a foundation upon which specific library programs may be superimposed. In the case of the National Library of Medicine, we have chosen to begin utilizing these new tools in the indexing program; tomorrow we may decide to tackle the serial record or the preparation of the catalog; in the more distant future we may decide to take on the devising of a system for the total control of the Library collection, including inventory, circulation, inter-library loan, and perhaps even miniaturization.

Fifteen years ago, a Committee of Consultants for the Study of Indexes to Medical Literature Published by the Army Medical Library—an activity in which the Library had been engaged for about seventy years—was appointed. A parallel research project was set up at Johns Hopkins University's Welch Medical Library, which included among its charges the study of the possibility of using machine methods. It is interesting to note retrospectively, the names of some of the eminent Committee and Project members. Among the librarians who served on the Committee were Verner Clapp, Ralph Shaw, and Mortimer Taube. Eugene Garfield, now the Director of the Institute for Scientific Information, started his career with the Research Project; while he was with the Project, more than ten years ago, Dr. Garfield devised a method of preparing printed indexes through the use of punched card equipment.

As a result of the Committee's activities, certain decisions were made in 1950 that radically affected the Library's publications program. The venerable Index-Catalogue, which had been published for three quarters of a century by the traditional letter-press method, was suspended. In its place were issued two separate publications, the annual Printed Catalog and the monthly Current List of Medical Literature, both prepared by the then novel method of shingling individual 3"x 5" slips or cards into columnar or page formats. Elsewhere in the Library during the same period, new photoduplication equipment and techniques such as xerography and home-grown "Photoclerk" devices were also being utilized.
Index Mechanization Project

Serious limitations in the method of preparing the Current List of Medical Literature began to be noted and attention was devoted to exploring ways of improving the ongoing system. In 1956, following the establishment of the National Library of Medicine in the Public Health Service, the search was intensified. By the end of 1957 a preliminary plan had been devised, and financial support was sought to mount a research project to study and implement new methods of index preparation. The Council on Library Resources responded with a grant of $73,800 to cover the work to be performed during the fiscal years 1959 and 1960. Early in the course of the NLM Index Mechanization Project's existence, in April 1959, the Library also contracted with an expert in the field of biomedical applications of computers to "conduct a study to investigate the feasibility of using electronic digital computers for the publication of the Index Medicus and also as a basis for the construction of an efficient reference and bibliographic service." These are some of the more immediate antecedents to our mechanization and computer systems.

The present system used in the publication of the Index Medicus, Cumulated Index Medicus, and Bibliography of Medical Reviews emerged from the Project in 1960. A detailed report on the Library's experience in conducting the Index Mechanization Project was published in January 1961 in the Bulletin of The Medical Library Association and, after about 4,000 copies were distributed, it was recently officially declared "out of print." Since the contents of this report are available, I shall restrict myself here to a summary of the mechanized index preparation system that evolved from the Project, draw some conclusions based on our experience that may be of interest, and move on to a more detailed discussion of the computer system now under development by the Library.

The primary objective of the Project was to improve the printed index composition methods by introducing mechanized techniques and, by so doing, to enhance the currency and expand the coverage of the publication. A second aim that was not fated to be accomplished successfully at that time pertained to the feasibility of using the basic publication system also as a bibliographic storage and retrieval system. It was concluded "... that a retrieval system could not be successfully grafted on to a publication system, which deals in large measure with the problems of composition. ... In the context of a very large system, it would be more suitable to start with the design of a retrieval system, and then to proceed with the publication. ..." This is, in fact, the most significant lesson we learned from the Project, and we hope to put it to good use in the design of MEDLARS.
Since January 1960, the Library has been publishing the monthly issues of the new *Index Medicus* by means of the new system which utilizes punched paper tape typewriters, punched card tabulating equipment, and the Listomatic Camera. At the heart of the system lie two major factors: one, the choice of basic format, the other, the employment of a high-speed precision camera (the Listomatic Camera) which photographs text imprinted along the top two-thirds of an inch of a punched card.

The format of the *Index Medicus* is based on the unit citation, in contrast to the fragmented citation used in its predecessor publication, the *Current List of Medical Literature*. A typical unit entry consists of the following elements: the names of all authors cited in the original publication, the title of the article either in the original English or in English translation, the abbreviated journal title, the volume number, inclusive pagination, date of issue, and a symbol indicating the language of the article if other than English.

With the Listomatic Camera, one, two, or three lines of copy, precisely positioned along the top of a punched card, are photographed at the rate of 230 cards (up to 690 lines) per minute. A special punch is "read" by the camera informing it as to the number of lines on the card to which the aperture must be adjusted. Other portions of the card can be punched for other purposes, e.g., the provision of various sorting indicia.

The following characteristics are basic to the system. The great majority of unit citations can be compressed into entries of three lines or less. The unit citations can be mechanically and repetitively transferred to punched cards. The punched cards can then be coded to allow rapid arrangement by machine. The arranged cards can then be quickly passed through the camera, and the negative film can be cut into column lengths and assembled in page form for the direct production of a lithographic printing plate.

The preparation of the *Index Medicus* proceeds as follows. First the individual journal article is analyzed for its subject content. The analyst assigns appropriate subject headings, following a subject-heading authority list to assure consistency. Foreign language titles are translated. The information is recorded on a simple "data sheet" which is clipped to the beginning page of the journal article to which it refers.

The data sheet, together with the journal itself, is next passed on to a subprofessional assistant who, working with a code dictionary, assigns codes for the main subject headings and certain other data. The purpose of this operation is to convert the indexer's contribution into a form suitable for mechanical processing.

From here the encoded data sheet and journal are sent to the "input-typing station" where a typist, operating a Model AA Justo-writer recorder, types the unit citation in its full form and, in the
process, produces a perforated paper tape together with a readable proof copy. After the proofreading, the tape, clipped to the journal, moves on to the "key-punch station," and a set of cards is punched for each article. A typical situation would perhaps call for five cards: three subject cards, each punched with its individual subject-heading code, and two author cards, each punched for the name of a particular author. Additional punches indicate language, year of publication, and the number of lines to be photographed on the card. At this point the issue of the journal is released and added to the Library's collections; the perforated paper tape and the corresponding set of cards are sent to the "output-typing station."

The "output typist" inserts the perforated tape in the reading element of a Model JU Justowriter reproducer and the punched cards in a special card-holding platen of the machine; the machine automatically imprints the text of the unit citation across the top of each of the cards. After a final inspection the processed cards are stored until the end of the collection period each month. The cards for the month are sorted and arranged by multiple passes through the IBM 083 sorter which handles the cards at the rate of 1,000 per minute.

Three auxiliary master sets of cards are required. The first set contains all subject headings as they appear in the authority list. The second set contains the entire cross-reference structure of "See" and "See also" references. The third set is a special Listomatic "program set." The latter, consisting of about eight cards for each page of the Index, furnishes instructions for a special device on the collator which counts the varying number of lines on the cards until the maximum number per column is reached and merges column-spacer cards, page indicators, and running-head indicators. These "non-entry" sets, punched and imprinted in advance, can be used over and over in the preparation of successive issues.

By means of the IBM 087 collator, the arranged entries are checked against the master subject-heading set. Subject headings needed for the issue are selected, while other subject headings are rejected. The selected subject headings are then checked against the master cross-reference set. Cross references needed are selected, and cross-references to headings not appearing in the particular issue are rejected; the selected subject headings and cross-references are merged and incorporated with the main file of previously arranged entry cards. The entire file—entries, main headings, and cross-references—is now in the desired order for publication; after this file has been merged with the Listomatic program set, the material is ready for the camera.

The entire file is photographed by the Listomatic Camera on 400-foot rolls of film which, after having been processed, are cut into column lengths and "stripped" into page form. The time required for
all these operations, beginning with the arrangement of the entries, is ten working days for an issue of about 50,000 entries. The publication is printed by offset reproduction from the page negatives.

The National Library of Medicine publishes twelve regular monthly issues of the *Index Medicus*, each containing only new material. At the end of the year the entries in the twelve issues are mechanically consolidated into one alphabet by the same procedure as that just described. The entire file of cards again passes through the Listomatic Camera. The cumulation requires about six weeks of machine time. The “manuscript” of the cumulated *Index*, on film, is turned over to the American Medical Association which publishes it as the annual *Cumulated Index Medicus*.

After administering the new system for almost one year, we attempted a fairly detailed appraisal of its effectiveness for inclusion in the *Index Mechanization Project Report*. Now, let us see how the system appears to me after more than three years’ experience with it. There is no doubt in my mind that the new mechanical composition system was a distinct improvement over the older *Current List* shingling method. Thanks to the Listomatic system, we have been able to realize a significant increase in coverage of the medical literature from an average annual level in the neighborhood of 110,000 articles to 125,000 in the first year of operation, 140,000 in 1961, and about 146,000 items in 1962. While providing more rapid, current publication, relative unit costs have also gone down. As you can see, we have been quite pleased with the new system. But how about others?

One of the objectives of the Index Mechanization Project was “To provide a demonstration of methods that might aid scientific indexes in general, and assist in similar enterprises in other fields.” The project succeeded in the objective of demonstrating the potential usefulness of such a system, but no significant trend towards emulation developed. Although we at the NLM have sporadically run a number of publications for other government organizations (inspired largely by the availability of our Camera) and commercial service bureaus have executed another handful of bibliographic products, on the whole, these applications have generally been half-hearted. This relative scarcity of any imitators is a highly significant aspect of the system.

This is probably as it should be. Now that the Index Mechanization Project has gone into history we are able to look back and to realize that those were years of general transition between an era of primitive, non-automated operational systems and increasing experimentation and investigation on the one hand, and the beginning of the exploitation of more sophisticated data processing techniques which we are witnessing today. The vast majority of libraries and other information-oriented agencies will be moving rather directly, abruptly, and traumatically from the Stone Age into the Space Age; the transition
stage that our present system represents will, therefore, probably not play a major role in future developments in library technology. In our own case, however, the more gradual transformation has been of inestimable benefit.

MEDLARS

Now let us turn our attention to discussing the NLM's present MEDLARS Project, and I shall attempt to recount the highlights of its history, objectives, and system characteristics. MEDLARS stands for Medical Literature Analysis and Retrieval System.

In the Fall and Winter of 1960, it became clear to the Library that a new program was needed to meet the bibliographical needs of the medical community. The following general objectives for the Library's indexing program emerged:

Increased coverage of the current substantive medical literature of the world up to totality.
More rapid processing to accelerate the availability of the information contained in the system.
Provision of broader accessibility to the bibliographic items by analyzing to a greater depth and along additional avenues of approach such as language and geographic origins of the document.
The capture and delivery of pinpointed and prescribed segments of the total file in a variety of patterns of selection and arrangement.

Underlying these aims is the fundamental principle of exploiting a single inputting of information to produce a multiplicity of output products. The object is to utilize the initial screening, identification, analysis, description, and processing required in the normal course of bibliographic activity, so as to provide variations as well as replicas of the original citations for different purposes. This is scarcely a revolutionary concept to librarians accustomed to the unit catalog card entry. In MEDLARS, the digitalized information about authors, titles, subject headings, and other bibliographical data may well be considered a supercharged analog of the unit card.

To achieve these goals, the Library again attempted to create an integrated bibliographical system possessing a publication and retrieval capability, both of a high order of sophistication and magnitude. One of the conclusions reached in the course of the earlier Index Mechanization Project was the inadequacy of punched cards as the medium for a system as complex and large as this. Since a more
powerful system was required, it was natural for the Library to investigate the feasibility of utilizing an electronic digital computer as the keystone of the new project.

At the beginning of 1961, systems specifications for MEDLARS were written and distributed to appropriate organizations active in related fields of operation. Accompanying the "specs" were invitations to submit proposals for a "planned approach to the development, testing, and implementation of an effective capacity within the Library for the storage, retrieval and publication of typed information from medical... literature." Among the specific primary objectives of MEDLARS were the following:

1. Improve the quality of and enlarge the Index Medicus, at the same time reducing the through-put time required to prepare the monthly edition for printing to five working days.
2. Provide for machine compilation of bibliographic listings in a variety of rather broadly defined subject areas within the general field of medicine. Examples of the scope of this type of service might be such topics as "diseases of the cardiovascular system" or "cancer." These listings would be made available on a recurring, periodically scheduled basis, essentially as a current awareness type service.
3. Make possible the addition to the Index Medicus, and other MEDLARS products, of citations representing monographs and serial titles, as well as journal articles and research project reports.
4. Make possible the prompt and efficient servicing of relatively complex demand requests for bibliographic information.
5. Increase the depth of indexing per article, both by adding more specific terms to the subject heading authority list and by increasing the number of terms used by the indexer in the analysis of the item.
6. Increase the coverage of the literature from an annual indexing of about 145,000 items to about 250,000 by the end of the decade.
7. Reduce the duplication of literature-screening activities now being carried on by many organizations.
8. Keep statistics and perform analyses of the MEDLARS operations, to provide useful system intelligence.
9. Provide for future expansion to add other objectives not entirely clearly defined or feasible in the earlier phases of the project.

The last point refers to certain secondary objectives which are scheduled for somewhat later implementation. The Library has
underplayed these future applications for fear of jeopardizing the successful implementation of the primary objectives that we have just described. It is interesting to note the history of the secondary objectives starting with the brief mention in the original MEDLARS letter of invitation of "other functions of the Library, where feasible (that) may be integrated into the system at a future date" down to the present where the Library now has deemed it necessary to commit actual resources to the serious study of these concepts. For the moment, it should be emphasized that these "blue sky" projects are not an integral part of the MEDLARS that is scheduled for implementation in 1964. Included in these secondary objectives are the following:

1. The storage and retrieval of graphic text images in addition to the bibliographic citations that are already in the basic MEDLARS.

2. A decentralized national network of mechanized search centers utilizing the MEDLARS input data centrally processed at the National Library of Medicine.

3. Conversion of the Library's entire Serial Record File to the computer in order to facilitate the control of posting, claiming, and general information relating to the medical serial literature entering the collection.

Other potential applications are being uncovered in the course of the present investigation.

About seventy-five firms ultimately received the "bid packages" and, before the deadline was reached on April 24, 1961, twenty-five proposals had been received in the Library. These proposals varied considerably in their method of presentation, size, appearance, lucidity of expression, imagination, and comprehension. The Information Systems Operation of General Electric Company's Defense Systems Department, located near the Library in Bethesda, was finally selected and, on August 14, 1961, the MEDLARS project was officially launched.

Primarily for purposes of more effective management control, the entire MEDLARS project was broken down into three major phases: Preliminary Design, Detailed Design, and Implementation.

The first phase consisted of the study and analysis of the Library's stated requirements and objectives in parallel with the development of an embryonic system concept. Estimates of projected work loads and request demands played an important part in this stage as did investigations of suitable equipment for use in the different subsystems of MEDLARS.

Phase I was completed by the end of January 1962, and Phase II, the Detailed Design, commenced on February 1, 1962. This part of
the Project covered the design and engineering of the system merely outlined during Phase I, to the level of detail necessary to transform it into an operational reality. During this phase, which ran until the summer of 1963, a tremendous effort was required. The detailed procurement specifications for every item of equipment had to be written for use in their purchase or development. All the major and minor computer programs had to be written and tested. Plans had to be developed for the training of the entire present staff complement in the Bibliographic Services Division and for the considerable number of additional personnel, including several top level key people required to fill the various new positions created by the new system. Detailed planning for the preparation of the computer site to provide additional mechanical facilities was also essential for the operation of the equipment configuration.

Before the completion of Phase II, the third and final implementation stage was kicked off. Phase III actually got underway officially at the end of January 1963. During the Implementation Phase, the equipment will be procured, installed, and tested. On March 14 the computer was delivered to the Library and installed in a specially prepared area. The training of all the members of the NLM staff will be completed with the desired objective achieved of having developed the in-house capability of operating and even further expanding the new system.

It would be useful at this point to give some idea of the quantitative loads that the system is being designed to handle. In 1964, the first year of operation, about seventy journal issues, representing about 650 articles plus approximately twenty monographs will enter the system daily; by 1969 this is expected to rise to about 100 periodical issues daily or approximately 1,000 indexed articles plus an additional seventy monographs. Ten demand requests per day are anticipated initially; ninety, later. About fifty recurring bibliographies are provided for. From the output viewpoint, the load in terms of characters composed is expected to be about 100,000 daily at the outset for demand bibliographies alone; close to a million characters per day by 1969. The Index Medicus will add a load of almost 9 million characters each month, rising to twelve and a half million in five years; an additional annual load of more than a hundred million characters will stem from the Cumulated Index Medicus in 1964, and about 150 million characters five years later. These figures do not include an average of about 800,000 characters per week that will be required for the composition of the recurring bibliographies.

To get a better idea of how MEDLARS will operate, it will be helpful to understand the nature of the unit record which is like the life blood of the entire system. Each of the major functional components of MEDLARS interprets the elements of the unit record according to its own requirements. The indexer begins the construction of
the unit record by deriving and posting subject and other tags, translating or transliterating the article title, and supplying descriptive indexing data. The typist reduces the alpha-numeric form in which the information arrives at her station to the digitalized codes that are machine usable. To the computer, the record represents a processing load for compression, for storage, for search, and for output processing. At the output end of the system, the unit records manifest themselves in the sheer quantity of characters they comprise and the manifold forms of entry which they are capable of assuming. To the user, hopefully, the unit record is the bibliographic pot of gold at the end of the rainbow. The following elements constitute the unit record: (1) names of authors, (2) title of article in English, (3) title of article in vernacular, (4) journal title abbreviation, (5) volume and page information, (6) date of issue, (7) language, (8) subject tags, (9) geographical tags, (10) form tags, (11) check tags, (12) date of entry.

Most of the elements, singly or in combination, with the exception of the title of the article itself and volume and page information, are susceptible to search.

Now we might turn our attention to the subsystem components of MEDLARS. The operations of the system can be divided logically into three parts, as follows: Input and conversion subsystem, Manipulation subsystem, and Output subsystem.

Input and Conversion Subsystem

The Input and Conversion Subsystem performs two basic functions: (1) It prepares the information to be read into the computer, and (2) It converts this information into computer language.

As journals are received by the Library, they are sorted and distributed to a staff of indexers. Book material will be distributed to a staff of catalogers who will perform somewhat similar functions. The analysts select the articles containing pertinent information and prepare a data sheet for each article containing those elements of the unit record that are appropriate to the particular citation. They translate titles of foreign material and analyze the information contained in the material by posting one or more of the headings listed in the Library's Medical Subject Headings. Pertinent check tags are also noted on the data sheet. Check tags are search access points other than subject headings, such as form or geographic headings, not contained in Medical Subject Headings. Another interesting innovation is the provision for the analyst to suggest new terms for consideration as ultimate additions to the subject heading list. These unofficial terms are called "provisional headings." The result will be a unit record for each article, monograph, book, or serial title which consists of a citation plus associated headings.
The data sheets and associated journals are used by a staff of input typists, and the data are converted to a form which can be fed into the computer. These operators type on machines that simultaneously produce typewritten copy and punched paper tape. There is an average of about ten articles per journal piece indexed. For each piece, the journal information is typed only once and is sequentially followed by the other unit record elements for all articles in the journal.

The proofreading staff checks the typed copy prepared by the input typists and indicates necessary corrections. A separate "correction" paper tape containing all the correction and control information is then prepared and reproofed. A complete unit record, any element of a unit record, or individual words can be added, deleted, or changed. The original reel of paper tape containing many unit records and the correction tape are then read into the computer. Any errors detected by the computer programming are printed out on the computer printer.

In addition to preparing unit records for entry into the computer, the Input and Conversion Subsystem also processes requests for the retrieval of certain selected parts of the stored information. These requests will be of two broad types: those calling for the preparation of periodically printed publications, such as Index Medicus, and those calling for the retrieval of bibliographic reference information to satisfy requests received by the Library from its consumers, that is, recurring bibliographies and demand bibliographies, respectively.

A search request is completely defined by the following items:

1. Retrieval criteria
2. Limits on the number of unit records to be printed
3. Manner in which specificity of retrieval criteria may be expanded or contracted
4. Arrangement in which unit records retrieved are to be printed
5. Selection of specific elements of unit records to be printed
6. Format of the total printed output, i.e., the physical arrangement of words on the page
7. Typography of the printed output

Those elements of the unit record that may be examined during a search in order to determine whether the unit record qualifies for retrieval include author name (s), journal title abbreviation, language, appended subject headings, geographical headings, form (article, monograph, etc.) headings, and date of acquisition of entry. Such elements may be used singly or in any combination of logical products and sums to specify the retrieval criteria of a single search request.
As requests for bibliographic information are received by the Library, they will be sorted and distributed to a staff of bibliographic specialists called Searchers. These Searchers translate the requests into the language and format required by the system. In the same manner that analysts select those subject headings that adequately classify an entry, the bibliographic Searchers select the headings that identify the entry for retrieval. The retrieval criteria elements are listed and organized into the form of logical expressions following the rules of Boolean algebra.

The demand bibliographic requests are typed by the input typists in the same manner as are the citations from medical literature. The typed copy is proofread, corrections made, and the paper tape read into the computer. The information retrieved by the computer to satisfy these search requests is screened by the Searchers to make sure that it satisfies the original requests before being transmitted to the requestor.

In 1963, the existing mechanized system will continue to publish Index Medicus and associated publications. The unit records generated will, however, be fed into the computer’s file for use in meeting future demand bibliographic requests. During the transition period, output information will continue to be sorted and collated, and photographed by the Listomatic Camera. When MEDLARS finally and completely supersedes the existing system, it will have a tremendous store of references on hand to call upon in filling individual demand bibliographic requests.

**Manipulation Subsystem**

The Manipulation Subsystem revolves around the high-speed digital computer. It accepts the unit records on paper tape, checks them for the presence and correctness of those elements for which such checks are possible, does some pre-processing to facilitate and speed the subsequent processing, and stores the unit records on magnetic tape. In response to search requests, it searches its cumulation of unit records for those that qualify for retrieval, edits them, and composes them for output on magnetic tape.

In the performance of the tasks, the Manipulation Subsystem will utilize computing equipment consisting of the Minneapolis-Honeywell 800 computer and associated tabulating equipment.

The journal article records; other data are stored on reels of magnetic tape. These tape files are prepared, updated, and used by various computer programs which manipulate the files to provide the desired outputs.
A highlight of the Manipulation Subsystem is the use of two separate citation files. In one file, the Compressed Citation File, the citations are carried essentially in coded form to enable demand searches to be made faster and more efficiently. In the other file, the Processed Citation File, the citations have been enlarged to provide the complete form of entry required for Index Medicus and other recurring bibliographic publications in a more efficient manner.

Output Subsystem

The Output Subsystem transforms the magnetic-tape product of the Manipulation Subsystem into the specific form desired.

Two basic output techniques will be available in MEDLARS. For routine man-machine communication, some of the recurring bibliographies and most of the demand searches, a high-speed computer printer will provide single font copy in continuous paper, or on 3" × 5" card forms. The Index Medicus, Cumulated Index Medicus, Bibliography of Medical Reviews, most of the recurring bibliographies and a small number of demand searches will, however, be printed out in a more interesting manner.

During the preliminary design phase, considerable effort went into the study of the publication composition requirements of the system. To satisfy these quantitative and qualitative requirements it was necessary to combine relatively high speed with equally superior levels of typographic quality. The investigations of equipment possessing this kind of capability led to a dead end, and the decision was reached to attempt to develop a satisfactory machine for this specific purpose.

Following the appraisal of competitive bid proposals, a sub-contract was awarded to Photon, Inc. in August 1962 for the delivery to the Library of a Graphic Arts Composing Equipment (GRACE) that will be capable of transforming the magnetic tape output of the Manipulation Subsystem into exposed film from which, after processing, a printer will be able to produce the Index Medicus and related publications. GRACE accepts electrical signals from a magnetic-tape transport operating fully under her control. These signals arrive in bursts of coded characters which represent a full line of legible type across the multi-columnar page. This type is made up of a combination of 226 different alphanumeric symbols which are set on film or paper at a rate in excess of 440 characters per second.

Once a month, for the publication of Index Medicus, the unit records in storage will be expanded, sorted, edited, and rewritten on magnetic tape. Once a year the material will be merged for the annual
Cumulated Index Medicus; even with the help of GRACE the annual Cumulated Index Medicus is estimated to require about two weeks for composition.

The demand search requests will be processed by a staff of search specialists and converted to a form suitable for input to the computer on paper tape. The tape will be read into the computer, which will then search its files for the citations that meet the request's specific requirements.

For the recurring bibliographies, the search specialists will periodically prepare the search and output format parameters for entry into the computer. As new citations thereafter enter the computer, they will be tagged with the recurring bibliography identification for future, periodic retrieval.

All of these functions will be performed speedily. The monthly issue of Index Medicus will require but five days for completion of the film "manuscript," and recurring bibliographies will be processed in one or two days. Demand bibliographies will be produced in a similar period, except for priority requests which will be given same-day service.

One of the keystones of the entire system is the subject heading authority list, Medical Subject Headings. In preparation for the first issue of the new Index Medicus the 1st Edition of Medical Subject Headings was carefully compiled in 1959. This list was published and used as the basis for all 1960-62 issues of Index Medicus and Cumulated Index Medicus that have appeared. To prepare for the computerized retrieval system, the full scale revision of this list was initiated. The huge task of converting 4,400 main headings and sixty-seven standard topical subheadings contained in the 1959 list into a 5,700 main heading system with no topical subheadings was completed in October 1962 and published in January 1963. The cross-reference structure was augmented by a completely categorized arrangement of the main headings in addition to the purely alphabetically-ordered listing.

An innovation in the updating and publication arrangements for Medical Subject Headings is the provision for the revision of the lists to be made annually hereafter and for the entire alphabetical and categorized listings to be published as a second part of the regular January issue of Index Medicus. All subscribers to the Index Medicus will thus automatically receive the authority list, as they did this year, and have it available for reference in the use of the index.

Another important change in the subject approach revolves around the increased depth of indexing to which I earlier referred. In the past, the Index Medicus subject section included an average of about two entries per article. For the retrieval system, this level appears to be entirely inadequate and the goal has been raised to an average of eight to ten subject tags per item. However, some simple
arithmetic quickly indicated that it would be impracticable to publish all of the citations under these additional headings in the Index Medicus. The solution that was finally adopted was a compromise between the conflicting requirements of the publication and retrieval systems. The publication average was upped 50 per cent, from an average of two to three subject entries and the remainder of the selected headings, designated as "non-IM tags," are retained in the computer system for search purposes. It should be made absolutely clear that each heading per se is not "IM" or "non-IM"; rather, the indexer, on the basis of the individual article in hand, decides on the disposition and treatment of the subject tags.

When MEDLARS becomes operational early in 1964, the National Library of Medicine expects to realize substantial advantages over the present, limited mechanized system. This does not mean that all of the Library's functions in this area will be mechanized. Certain tasks, such as indexing, cataloging, preparation of search requests, and proofreading will be performed by people while machines will perform such operations as storage, retrieval, and composition of bibliographic citations.

In the development of a computer information retrieval system such as MEDLARS, there is a natural tendency to stress the data processing aspects at the expense of these humanly performed functions. Since it is indisputable that the mechanized portion of the system requires a tremendous effort to engineer, it becomes necessary to exercise care and discipline to avoid relegating the other parts to a kind of second-class status. This could be fatal. Difficult and complex as the data processing problems that confront us are, I have the conviction that the success or failure of MEDLARS will be more directly related to the non-mechanized elements.

Discussion

Ralph T. Esterquest

The impact of MEDLARS on the medical library world is not that of the familiar metaphor—the pebble dropped in the pond, casting concentric ripples that reach many points on the shore. The impact

Mr. Esterquest is Librarian, Harvard Medical Library.
is no pebble for sure. It is a mighty rock. The waves it will cause will surge and splash for a long time to come.

MEDLARS is the great bibliographic break-through of our generation.

Let me first remind you of certain characteristics of medical libraries—especially medical school libraries—that seem relevant to our subject.

1. The materials of a medical library are subject to a rapid rate of obsolescence. The heavily-used item is this year’s book and this month’s periodical.

2. Periodicals seem more important than books. Although books are indeed used, the emphasis is definitely on the periodical article. The typical medical library will, this year, place on its shelves almost as many new periodical volumes as monograph volumes.

3. Many users of a medical library want information in a hurry. It’s not that he’s a surgeon, elbow-deep in a chest operation, calling for a book to read up on what to do next. But frequently, in a teaching hospital, an unanticipated observation at eleven in the morning suggests a highly useful consultation at two o’clock, and a bit of reading meanwhile can increase substantially the effectiveness of such a conference, or teaching situation.

Add to these characteristics the literature explosion. The total number of medical periodicals is increasing all the time. There were 3,900 in the World Medical Periodicals list in 1953; 5,800 in the 1961 edition. The medical scientist is badgered by a public that wants him to find a cure for cancer. He’s given $100 million by Congress and told to get busy and take care of heart disease. The medical scientist tries to cope with the literature and puts the heat on the librarian. “Do something!” he cries.

At least the National Library of Medicine is doing something. MEDLARS.

The NLM has had a long history of publishing indexes and catalogs, and has a tradition of leadership in bibliographic control. The quality of this national leadership has never been better than it is today. Medical librarians are not surprised to see their National Library undertake MEDLARS.

This exciting project is notable on many counts. One important count is MEDLARS’ dual purpose: the publication of an index; and the compilation of tailor-made bibliographies. In respect to index publishing, MEDLARS a) greatly speeds up the time schedule for compilation; b) increases accuracy; c) increases the depth or thoroughness of subject analysis; d) greatly increases coverage, from 130,000 articles annually (now) to 250,000 in 1969; and e) increases scope by the inclusion of catalog entries of monographs as well as periodical articles. In respect to the compilation of demand
LIVER DISEASES (C4)
Some remarks on chemical injury of the liver. Its relationship to disease. Cameron R. J Forensic Med 8:114-5, Jul-Sep 61
Liver damage in treatment with monoamine oxidase inhibitors. Ronnberg-Halvorsen B:
Streut-Lakaridts 58:2582-92, 15 Sep 61 (Sw)
245 p. illus. WM 100 B746m
statistics
339 p. illus. WM 32 A2 215a
450 p. WM 444 G962b (Jap)

LOCOMOTION (G1)
LUMBOSACRAL REGION (A1)
Lessons of the lumbosacral region in industry. Mathews RM:
Rev Venes Sanc Atl Antic Seg 28:546-9, Sep 61 (Sp)

LUNG (A4)
The problem of dormant fetal organs: the kidneys, lungs and the gut. Jakoby S: Bred Neesat 3:343-56, Dec 61
The influence of age and environment factors on the behavior of reaggregated embryonic lung cells in culture. Grover JW:
Exp Cell Res 26:344-59, Mar 62
Heart Bull 11:54-7, May-Jun 62
Pulmonary extracoronal endomy. Siltanen P:
Duodecem 76:264-9, 1962 (Fin)
Respiratory and metabolic related changes in acid-base
Effects of hypophysectomy on hepatic circulation. Orrego H, et al:
Rev Med Chile 90:144-53, Feb 62

LIVER DISEASES (C4)
Some remarks on chemical injury of the liver. Its relationship to disease. Cameron R. J Forensic Med 8:114-5, Jul-Sep 61
Liver damage in treatment with monoamine oxidase inhibitors. Ronnberg-Halvorsen B:
Streut-Lakaridts 58:2582-92, 15 Sep 61 (Sw)
245 p. illus. WM 100 B746m
statistics
339 p. illus. WM 32 A2 215a
450 p. WM 444 G962b (Jap)

LOCOMOTION (G1)
LUMBOSACRAL REGION (A1)
Lessons of the lumbosacral region in industry. Mathews RM:
Rev Venes Sanc Atl Antic Seg 28:546-9, Sep 61 (Sp)

LUNG (A4)
The problem of dormant fetal organs: the kidneys, lungs and the gut. Jakoby S: Bred Neesat 3:343-56, Dec 61
The influence of age and environment factors on the behavior of reaggregated embryonic lung cells in culture. Grover JW:
Exp Cell Res 26:344-59, Mar 62
Heart Bull 11:54-7, May-Jun 62
Pulmonary extracoronal endomy. Siltanen P:
Duodecem 76:264-9, 1962 (Fin)
Respiratory and metabolic related changes in acid-base
Effects of hypophysectomy on hepatic circulation. Orrego H, et al:
Rev Med Chile 90:144-53, Feb 62

LIVER DISEASES (C4)
Some remarks on chemical injury of the liver. Its relationship to disease. Cameron R. J Forensic Med 8:114-5, Jul-Sep 61
Liver damage in treatment with monoamine oxidase inhibitors. Ronnberg-Halvorsen B:
Streut-Lakaridts 58:2582-92, 15 Sep 61 (Sw)
245 p. illus. WM 100 B746m
statistics
339 p. illus. WM 32 A2 215a
450 p. WM 444 G962b (Jap)

LOCOMOTION (G1)
LUMBOSACRAL REGION (A1)
Lessons of the lumbosacral region in industry. Mathews RM:
Rev Venes Sanc Atl Antic Seg 28:546-9, Sep 61 (Sp)

LUNG (A4)
The problem of dormant fetal organs: the kidneys, lungs and the gut. Jakoby S: Bred Neesat 3:343-56, Dec 61
The influence of age and environment factors on the behavior of reaggregated embryonic lung cells in culture. Grover JW:
Exp Cell Res 26:344-59, Mar 62
Heart Bull 11:54-7, May-Jun 62
Pulmonary extracoronal endomy. Siltanen P:
Duodecem 76:264-9, 1962 (Fin)
Respiratory and metabolic related changes in acid-base
Effects of hypophysectomy on hepatic circulation. Orrego H, et al:
Rev Med Chile 90:144-53, Feb 62
bibliographies, MEDLARS draws upon a vast store of bibliographic information, selects relevant items through the sophisticated manipulation of thousands of subject and other tags, and prints out its tailor-made product at great speed.

It is important to remember about MEDLARS, however, that its end product is not a nugget of scientific information, it is a bibliographic citation.

A basic principle of MEDLARS is that the input item is a "unit record," or unit citation. This is analogous to the familiar unit card in cataloging. It's as though thousands upon thousands of unit cards—everyone different—were thrown haphazardly into a vast tub, and then, because each unit card had many different, and invisible, and magic handles, you the magician can pull out any group or combination you could name. Out they pop, all in perfect order—exactly as you wished it, when you rubbed your Aladdin's lamp.

In order that you may see how well-ordered the unit citations actually will be when out-put by MEDLARS, Fig. 1 shows a sample page of the new Index Medicus (as of January 1964). The subject index (Fig. 2) is the main section, followed in each monthly issue by the author index. The typographic excellence of these pages testify to the sophistication of the MEDLARS hardware, specifically GRACE (Graphic Arts Composing Equipment).

An important aspect of MEDLARS is the attention that has been given and is being given to the subject heading apparatus. Only a short time ago, the National Library of Medicine developed a revised up-to-date subject heading list. Many medical libraries adopted this list and have been using it as a major part of their subject authority list. Now, with MEDLARS on the scene, extensive further revision has been undertaken, topical sub-headings have been eliminated, and many new, more specific headings have been added. One reason given for these wholesale changes has to do with the fact that, in recent monthly issues of Index Medicus, the number of citations under some subject headings had become too great for easy scanning; thus new, more specific headings were created to break up the long list. Wouldn't more sub-divisions have been a better answer? Has this something to do with the searching technique of MEDLARS? I should hope that Mr. Taine would tell us why the new Medical Subject Heading list (MESH) has taken the form it has. In medical libraries where the NLM list is used in cataloging, major revisions can be an expensive nuisance.

In connection with the subject heading list, it should be pointed out that MEDLARS provides an automatic mechanism for testing provisional subject tags, for recording how useful they are, for incorporating selected ones into the main body of MESH, and then for publishing regular, up-dated cumulated lists. Perhaps Mr. Taine would like to explain how this works.
It is significant that MEDLARS will permit a great increase in the number of subject tags that can and will be tied to each unit citation. From the present two subject headings, it is expected that the average citation will, under MEDLARS, get eight to ten tags, or descriptors. However, only three (on the average) will be coded for out-put for the printed Index Medicus. Why not all ten? The monthly issues would be just too big to handle—to say nothing of the annual cumulation. The other subject tags, that is, those that do not appear as headings in Index Medicus, are in the computer store as coded tags for retrieval in the case of the demand bibliographies.

Because the imagination is captured by the MEDLARS' capabilities for compiling Index Medicus, one tends to overlook the importance of the other half of the program: the demand bibliographies. It is this aspect that opens new horizons and has great potentialities. Incidentally, it is worth pointing out that the plan for starting the indexing input as of January 1963—one year ahead of the first MEDLARS-produced Index Medicus—serves not only to provide a year for testing the system, it also serves to build a one year’s store of citations (i.e., 1963) for its obvious usefulness in the out-put of demand bibliographies.

The National Library of Medicine now responds to mail requests for literature citations on many subjects. Its Reference Department does some manual compiling of bibliographies, as staff resources permit. With MEDLARS, these special bibliography requests can be answered by in-putting properly coded paper tape search requests, whereupon the computer will grind out the desired citation list, using fast print-out techniques. This kind of automation holds great promise. But also, it is likely to be the first place where the MEDLARS load capacity will be reached and exceeded.

Out of anticipation of this problem, the National Library of Medicine has announced, somewhat tentatively, a plan for decentralization, specifically a plan for placing in selected, geographically-distributed medical libraries duplicate magnetic tapes, that is, the computer store of citations. Assuming for the moment that such regional libraries could make full use of these tapes for producing demand bibliographies locally, the scheme would have a number of other obvious advantages:

1. Reduce the load at the national center, that is at the NLM.
2. Provide for faster processing of requests.
3. Improve the quality of the search by permitting a face-to-face exchange between the individual making the request and the search-specialist.
4. Increase experience in the use of MEDLARS and increase knowledge of its capabilities.
5. Lead to the discovery of new applications; the more operators, the greater the probability of new ideas being generated.
It is apparent that the possibilities here are enormous. Literature-searching is a time-consuming activity in medical libraries. Often it is carried on by top-flight research men and women, whose time ought to be saved by the automated screening process that decentralized MEDLARS could theoretically provide. But there are more problems to a decentralized MEDLARS than meet the eye. In the first place, MEDLARS in Bethesda is centered in a Honeywell 800 computer. The computers locally available to my library happen not to be Honeywell machines and cannot read Honeywell’s tape. And even if Honeywell’s tape were to be converted, there is still non-compatibility at the systems or operational level. I should like to hope that Mr. Taine would tell us more about this very serious problem.

But let us assume that this problem of non-compatibility can be resolved, and there is evidence that it will be. In this event, the impact of MEDLARS will be enormous. It will be, for librarianship, one of our most pleasant dreams come true, since it will embody a principle which we have for a long time embraced, namely: Let the intellectual labor (of indexing, cataloging, analyzing) be done just once—centrally; and then let our out-post libraries exploit the product in the form of the endless variations which individual requests and needs require. Let the NLM staff, with its subject and language specialists, index the 250,000 journal articles and catalog the current American and foreign medical imprints, and let the rest of us, relieved of this one-time task, concentrate our attention on making use of their product on behalf of our readers.

The availability of published periodical indexes has made homemade periodical indexing in individual libraries unnecessary for a long time. MEDLARS may now eliminate the need for cataloging many books in our individual medical libraries. Is this perhaps the beginning of the end of the card catalog as we have known it up to now?

My remarks, so far, have tended to sing the virtues of MEDLARS. I have admitted that I am enthusiastic about this project. But I am here presumably as a critic, and a critic ought to criticize.

Let me discharge my responsibility now by raising a few questions. These are questions that some medical librarians are asking about MEDLARS, and I address them to Mr. Taine with the expectation that he will clarify these points for all of us.

**Question 1.** The reels of magnetic tape, which are the memory store, have a finite capacity. We are told that, in the case of the “Compressed Tape” (for demand searches) six reels are required to hold one year’s citations. Thirty reels are needed for five years. Since it takes the computer at least five minutes to search one reel, it follows that a complete search of five years of citations would take two and one-half hours. How then can MEDLARS undertake to compile more than two or three demand-search bibliographies in a working day? And when does the over-worked computer do its other tasks?
Answer by Taine. The demand search capability of MEDLARS was planned to permit up to about 100 searches daily. These searches are not run in the computer individually; rather, depending on the number of parameters contained in the individual requests, they are "batched" and many, perhaps all of the daily work load will be run simultaneously. Of the eight hours of computer time available in a single shift operation, it is anticipated that the demand search requirements will be met in about 85 minutes when this peak is reached in 1969; initially, the time required on the computer will be close to 45 minutes.

Question 2. If the distribution of duplicate tapes to regional libraries is a part of the MEDLARS plan, why was the Honeywell 800 computer selected for MEDLARS, when other types are so much more prevalent and available in the academic communities where medical libraries are located, and when the Honeywell tapes, being non-compatible cannot be used on these other computers?

Answer by Taine. At the inception of the MEDLARS Project, a firm decision was made to divide the aims of the system into two parts. First priority was given to the successful attainment of the primary objectives, essentially the development of an in-house computer search and publication capability. It was recognized that useful, even essential, secondary objectives would also be generated but that it would be desirable to wait until the complex requirements of the primary objectives first be accomplished. This, incidentally, has proven to be a wise course of action.

The selection of the computer to be used by MEDLARS, therefore, had to be primarily based on these requirements and secondarily on those of the vaguer and more distant goals. Since no computer exists that is ideal for all purposes and, an "average" need is not possible to establish, the computer selection had to be based on the most urgent needs of the system. Following a very careful study of available configurations, the Honeywell 800 was finally chosen.

It should also be pointed out that Honeywell is not alone among the computer manufacturers in being incompatible with other computers. In fact, there is even a considerable degree of incompatibility between different computers from the same company. However, this is not a serious obstacle to the effective distribution of duplicate MEDLARS tapes to outside computer centers. The conversion of M-H 800 magnetic tapes to other tape formats is economically and technically feasible at the present time.

Question 3. Why is MEDLARS, and the resulting Index Medicus, greatly extending its coverage to include many more medical journals (14,000 journal issues now; 25,000 by 1969), when it is strongly suspected that these additional journals are of an inconsequential nature? Does not this effort to "increase coverage" really amount to a cluttering up of the columns of Index Medicus with thousands of citations to low-quality articles?
Answer by Taine. If objective techniques actually did exist to permit one to distinguish the "junk" from the "worthwhile," a major contribution to bibliographic control would have been achieved. These desirable characteristics do not, however, exist and until such a universally acceptable yardstick comes into being, there appears to be no alternative to admitting virtually the entire literature corpus into the system. Although a sizeable quantity of inferior material will thereby be indexed, it is also true that there are still very large numbers of truly valuable items, yet unindexed, that will be included as a result. Subsequently, other refining techniques can be used to screen out the undesirable material if the specific search warrants such an approach.

Question 4. The subject scope of MEDLARS and Index Medicus is much narrower than that of most medical school libraries, which need to provide materials in chemistry, biology, physics, psychology, etc.—subjects which are out of scope for MEDLARS. Why doesn't MEDLARS extend its subject scope instead of trying to increase coverage to embrace further strictly medical journals?

Answer by Taine. This is a dilemma that most, if not all, scientific information services, must face today. As a result of the prevalent trend towards increased interdisciplinary relationships, the traditional boundaries between subject fields have become rather tenuous. An exhaustive literature search is therefore not likely to be successful if it is based on the use of a single bibliographic aid or on the multiple tools in the same discipline. The size of the total literature does not permit each disciplinary index or abstracting service the luxury of duplicating the peripheral material. As the index stretches to encompass these fringe items, it quickly finds itself in the untenable position of attempting to be a universal bibliography. MEDLARS, therefore, strives to first cover the medical and para-medical literature. We look to the future for the real solution to this problem, perhaps in the form of an integrated scientific information system in which the efforts and contributions of various individual disciplinary services are combined.