Appendix H.—Future Information Technologies: Implications for Biomedical Retrieval Systems*

Introduction

Electronic computation used to retrieve information from large data bases is an evolving technology, still only in its early stages. Its course can only be dimly perceived at this time. In the relatively near future, very little of today's techniques may have any more than historical interest.

Today's biomedical data base retrieval systems are primarily based on remote time-sharing computers, which means that the processing time of the computer is shared among several completely independent activities. Each user is unaware that there are other simultaneous users of the system, as the processor spends only fractions of seconds with each activity in turn before proceeding to the next. The computers compare indexed citations against a list of search terms. These terms are typed at remote terminals by researchers or trained computer search analysts. Terminals are connected to time-sharing computers via conventional telephone systems, specialized data networks, or both.

Projections for the next few decades indicate that changes are going to be even more radical than that of the past few decades. Machines are getting faster and more powerful. Telecommunications links are getting faster and capable of more sophisticated processing. Both are getting cheaper, but, relative to telecommunications, computation tends to become more cost effective at each technological jump.

Tomorrow's data base technologies may range over a wide variety of systems more powerful than the current methods. Full-text retrieval of articles, computer-aided searching of data, and other possibilities are being demonstrated today in experiments; some of these are expected to have an impact in the near-term future. The physical location where information is processed may alter within the next few years. High-speed data networks are expected to connect microcomputers sitting on researchers' desks; and it may prove more cost effective to move massive amounts of data processing locally than on distant time-sharing computers.

The present information policy issues are the result of usage patterns that have evolved over the past decade and may change greatly in the next few years as new technology based on distributed data processing, particularly the personal microcomputer, is applied to data base access. This observation is not meant to minimize the current difficulties. The present problems are important for the information industries today, and they are also important because what is done about them by policymakers now will affect the information systems of the future.

Past Application of Information Technology to Biomedical Bibliographic Retrieval Systems

Computers were first applied to information storage and retrieval systems in the late 1950's when a number of systems were developed in the United States by the Armed Services Technical Information Agency (now the Defense Technical Information Center), National Ordnance Laboratory, National Aeronautics and Space Administration, and the National Library of Medicine (NLM or the Library). The system developed by NLM, MEDLARS, was the largest of these both in terms of the size of its files and in terms of its user population. When MEDLARS was implemented, it was the first large computerized retrieval system to be made widely available without security and other restrictions.

The characteristics of MEDLARS are typical of the information systems that became available during the 1960's. The systems were operated in an off-line batch processing mode. This technique of batch processing required data to be brought to a data center where they were punched onto cards, and then were subsequently read into the computer. The desired program, usually stored on magnetic tape, was also selected and entered from the tape into the computer memory. The program acted on the data and the results of the processing were output via a printer, or sometimes stored on a second deck of cards for further processing. The user of the data then came to the data center to collect the results. This form of processing worked quite well for applications that occurred at periodic intervals.

MEDLARS was originally designed not as a retrospective search system for bibliographic material, but as a publication system to produce the printed index, Index *Medicus.* The computer was used to manipulate bibliographic records in machine-readable form. It was used to check for errors, perform sorting and formatting, and interface directly with photocomposition equipment. For this, it was necessary to put the index records into machine-readable form and then update it, say monthly. Once this was done and the index produced, for example monthly, the machine-readable data base was available for further exploitation.

●l%is appendix is based on papers prepared for OTA by Richard Solomon, of the Research Program on Communications Policy of the Massachusetts Institute of Technology, Cambridge, Mass., and by Jose Marie Griffiths, of King Research Associates, Rockville, Md.

Such data bases could be used in a number of different ways. First, they could be used to conduct retrospective searches. These are searches through a body of recorded literature to find items on a specifically defined topic. Another application was for selective dissemination of information (SDI). In SDI, the current interests of users or groups of users are defined in interest profiles which are stored in machine-readable form. After the data base has been updated the additions are checked against the interest profiles. Matching records are printed out and sent to the appropriate users, thereby enabling users to keep up-to-date with the literature of their field of interest on a regular basis.

Computerized information systems developed in the 1960's offered significant advantages over their printed counterparts. These advantages are outlined by Lancaster as (79):

1. The possibility through batch processing of conducting many searches at the same time.
2. The ability to provide several access points to a document extremely economically.
3. The ability to handle complex searches involving large numbers of terms in complex relationships.
4. The ability to generate output in the form of a printed bibliography, and even to produce high-quality publications by interfacing the retrieval system with a photocomposition device. Output can also be made directly to microfilm.
5. The ability to collect, on a regular basis and essentially as a byproduct of normal systems operation, management data on how and how much the system is used.
6. The ability to produce many outputs and services from a single input operation. MEDLARS tapes, for example, although produced as a result of one indexing operation and one procedure for reducing the index records to machine-readable form, can be used to generate a general printed index, specialized bibliographies, retrospective search, and SDI searches.
7. The data base, once captured in machine-readable form, can be duplicated simply and cheaply; it is easily shipped around and thus can be used in the provision of information services by a number of different centers. This is perhaps the most important advantage of all. The growth of machine-readable data bases has had a dramatic impact on the provision of information services in the last 20 years.

Although the computer offered many advantages in information handling activities, the off-line batch processing systems have considerable disadvantages. Most significant is the delay in obtaining results, and if the search was unsuccessful on the first run, then it may have to be modified and reprocessed. In the same way, there is no facility for browsing through the literature. There was clearly room for further development of such systems.

In the early **1950's,** it occurred to a number of people that the computer could be useful as a tool for researchers whose problems were too small to justify the initiation of a formal batch operation. Teletype terminals were sited in laboratories at the Massachusetts Institute of Technology, and these enabled researchers, each from his or her own teletype, to program the computer remotely. The computer would process each program in turn and return the results to the appropriate user's teletype as printout. This development marked the introduction of the on-line systems we are familiar with today.

The term "on-line" refers to the searcher's being in direct communication with the system he or she wishes to use. The user provides input to the system and it, in turn, reports back to the user. The user then makes decisions based on that report and provides further input. For this reason the system is also referred to as being "interactive" or "conversational. " The development that made such systems feasible was time-sharing.

All these activities were carried out on mainframe computers and, to a large extent, still are. However, the late 1960's saw the introduction of minicomputers. Minicomputers were initially developed as low-cost computers with a minimum of processing power and memory. They were used primarily for control processing. It was soon recognized that minicomputers had a potential use in a number of unrelated areas and that they should be adaptable to the requirements of each. This resulted in the development of a number of general purpose minicomputers and the "unbundling" of software. Previously, machines were sold for specific applications with software provided as part of the overall system package—all "bundled in" together.

The advances in computer technology that have had a significant impact on information handling can be separated into two categories—processing and storage.

Processing

The first computer ENIAC (Electronic Number Integrator and Computer) was completed in 1946 at the University of Pennsylvania for the U.S. Army. It consisted of about 18,000 vacuum tubes and was very large, requiring a room 60 ft by 25 ft to contain it and weighing more than 30 tons. This computer (and others based on vacuum tube technology—later known as "first generation computers' ')–had high power consumption, had failures every few hours or so, and required cooling plants often as large and complex as the computer itself.

Two years later, the first transistor was developed at the Bell Telephone Laboratories. Transistors, while comparable to vacuum tubes in terms of the tasks they could perform, required much less electrical power, generated very little heat, and were more reliable. They became the basic component of "second generation computers."

Early efforts to miniaturize electronic components were not motivated by computer designers and engineers. Various satellite and missile projects, however, called for complex electronic systems to be installed in equipment in which size, weight, and power requirements were severely constrained. Thus, the effort to miniaturize was promoted by military and aerospace agencies.

The latest type of miniaturization is the semiconductor integrated circuit. Several researchers saw that the characteristics of semiconductors such as silicon or germanium that had been exploited to make transistors might be further exploited. The physical composition of semiconductors contained equivalents of individual electronic components (resistors, capacitors, etc.), and by combining them with transistors in the same material complex circuits could be created.

The integrated circuit was the basic component of the "third generation computers." In the early 1970's integrated circuits were being produced with about 1,000 components. The first microprocessor was developed by Intel Corp. in 1971, was about one-fourth inch square, and carried the equivalent of 2,250 transistors. By 1976, large-scale integration (LSI) produced chips carrying over 30,000 components, and by 1980 very large-scale integration (VLSI) saw chips with over 1 million components.

The individual elements on the silicon chips are defined by a photographic process. The smallest component dimensions on the chips are currently 3 to 4 micrometers (1 micrometer = 1 millionth of a meter). It is expected that this will be reduced in the near future to less than 1 micrometer and that by the 1980's it will lie between 0.05 and 0.005 micrometers.

The overall effect of these advances in processor technologies on information handling has been the reduction in size, cost, and operating requirements of digital computing equipment. Prior to the late 1960's, equipment had been far too expensive to justify use for what were then considered peripheral applications of information handling. However, by the mid-1970's many information-handling organizations were able to justify the cost of purchasing computing equipment, either alone or as part of a cooperative. Many large libraries and information services have their own small-scale equipment, but the continued reduction in size and cost of such equipment is beginning to influence the small and special libraries in particular.

Storage

The technology of digital storage is probably the most rapidly changing sector in all of microelectronics. Over the past decade, operating speed and reliability have been increased by at least an order of magnitude at the same time as physical size, power consumption, and cost per bit of storage have been reduced by factors of up to 1,000. Similar improvements are predicted for the next decade before physical limitations are encountered.

The newest electronic memory systems have been made possible by modern semiconductor technology. In the 1950's and 1960's, electronic memories were arrays of cores (rings) of ferrite material a millimeter or less in diameter. Ferrite core memories have now been succeeded, on the whole, by semiconductor memories which provide faster data access, smaller size, and lower power consumption at significantly lower cost. The overall effect of advancing digital memory technologies on information handling is an increase in the capacity and reliability of information systems, faster and more efficient retrieval of information (although because of increased capacity of storage, searches are made through larger volumes of information, and the faster access may not reduce the retrieval times as far as the user is concerned), together with an accompanying decrease in cost.

Current Status of Information Technology

Computer Processors

Although significant advances have been made in computer processor technology, very few of the commercially available retrieval systems have taken advantage of them. To a large extent, this is understandable. During the 1960's and 1970's, considerable investments were made in the acquisition of computer hardware. The effective lifetime of a single computer processor is 7 to 10 years, so organizations are unwilling (and often unable) to replace a processor within its lifetime. This is the so-called "technology trap." Those who hesitate to invest in a rapidly changing technology because new, improved products are imminent never reach the point of acquisition, and those who do make a decision to acquire a product acquire one which is usually no longer state of the art by the time it is in operation within an organization.

A second factor in the lack of change in commercial bibliographic retrieval services is that the functional capabilities of minicomputers and microcomputers are not as great as with the larger mainframe machines. Over the next 5 years or so, this situation is likely to change. The larger end of the minicomputer range (the 32-bit machines)—the superminis or megaminis as they are sometimes called—will be able to compete with the large IBM, Univac, Burroughs, and similar machines, both in terms of system software (operating systems and system utilities) and number of users that can be simultaneously serviced.

Areas where smaller machines have been applied are in individual organizations and libraries. Minicomputers have been used successfully in libraries for many years now, primarily for automation of many of the day-to-day library operations—acquisitions, cataloging, and circulation control. As the costs of computer processors fell dramatically, a number of libraries began to acquire these low-cost machines for maintaining bibliographic files for relatively small collections. The retrieval programs available on such small machines are not as sophisticated as those for larger systems, nor is retrieval as fast, but they are quite adequate for the environment within which they are applied.

Storage Devices

Digital storage devices are probably the fastest moving area of microelectronics. In particular, two new types of digital storage devices are likely to have a significant impact on retrieval systems in the future: 1) video disks, and 2) charge-coupled devices (CCDs) and bubble memories.

There are two approaches to recording information on video disks (or optical disks). One is an analog method which yields a color television (TV) picture, and the other is a digital method which stores data in digital form. It is the latter type, in particular, that could revolutionize information handling in the not too distant future. The current capacity of such disks is of the order of 10^{11} bits, This means that the equivalent of the Library of Congress card catalog could potentially be stored on one side of a digitally encoded optical disk. However, this is unlikely to happen for several years yet.

The disks can be written once only. A laser is used to burn away a thin metallic film from the disk surface to create a hole about 1 micrometer (one ten-thousandth of a centimeter) in size. The presence of a hole signifies a 1, and the absence of a hole signifies a zero. The projected cost for the disk is $10, and copies could be made for almost the same amount. A disk can store about 10,000 books (the full text), making the projected cost per book 0.1 cents. The bulk of the cost (over the $10 disk cost) will be the cost of digitizing the textual material and the disk reader.

The other type of storage devices under development are electronic serial access memories, in which the stored bits of information circulate as if they were in an enclosed pipeline. Each bit that is stored is transferred sequentially through 64 or more storage locations. These memories are smaller and cheaper to produce than other forms of electronic memory, mainly because the circuitry can be simpler. They cannot, however, compete in terms of speed with other electronic memories. Their most attractive potential application is the replacement of tape and disk memories with a capacity of on-board storage of between 1 million and 10 million bits.

COMMUNICATIONS

Most data communications systems currently under development are based on the concept of packet switching. The rationale for packet switching is that the length of a message is very short in relation to the time it takes to establish a conventional communications connection, and that it is more economical to transmit that message stage by stage through the network, storing the message segments at each node in a network until a link is available for transmission to the next node. It is now possible to purchase services from a commercial packet switched communications network.

To some extent, the microprocessor counteracts the rapid growth of networks. The central idea of networking is to share both functions and data. As small computer systems have become more powerful and compact, they have also become more independent in function. People are now able to buy low-cost computers to solve most of their routine problems. The general trend is away from network designs based on functions shared over a network of computers towards data sharing across networks.

The major trend in telecommunications at present is the fundamental shift from analog to digital modes of transmission. The new transmission channels (based on optical fiber technology) will have enormous capacities. The shift involves all types of communications—voice, facsimile, computer transmissions, TV communications, microwave and satellite communications, and radio links. Digital circuits are less prone to interference and noise than analog ones, and their cost-performance ratio is constantly improving.

Communications may be expected to play an increasingly central role in information handling. Satellite-based communications systems offer a different approach to the present ground-based systems in terms of technology, but the services offered by the two systems should be similar. Satellite communications will probably develop as an extension to the ground-based systems, and will provide a more cost-effective solution in sparsely populated areas, for example.

INPUT/OUTPUT DEVICES

New and improved means of communicating with computers are pouring onto the marketplace at an ever-increasing rate. The standard teletype terminal and cathode ray tube terminals for interacting with computers are being replaced with sophisticated graphics devices, color screens, touch-sensitive screens, letter quality printers, laser printers, and so on. Many of these input/output devices themselves contain microcircuits, thereby providing the user with localized intelligence.

The key change in the mode of interaction with biomedical bibliographic retrieval systems has been in the use of graphic terminals (especially for displaying organic compounds in schematic form) and/or intelligent terminals for search development and storage on input, and bibliography editing, sorting, and merging on output. However, until the retrieval systems themselves change, the mode of user-system interaction is unlikely to change much at all.

MASS MEDIA COMMUNICATIONS

One of the growing applications of communications systems to information handling is the provision of electronic mail services. It is possible to convert any type of message into digitized for transmission. Naturally, different messages require varying numbers of bits for their representation, for example:

- high-quality color photograph — 2 million bits
- newspaper-quality photograph — 100,000 bits
- color television frame — 1 million bits
- brief telephone voice message — 1 million bits
- document page in facsimile form — 200,000 bits
- document page in computer code — 10,000 bits
- coded request for library document — 200 bits

Recent additions to information services are those based on the broadcast systems. Two types of systems exist: Videotex (sometimes called Viewdata) and teletext. Although their services and technology differ, both are accessed through the adapted domestic TV set, distinguishing them from other home information utilities that require computer terminals.

Videotex is an interactive system linking computer data bases to the adapted television set through the switched telephone network. Canada, France, and the United Kingdom have developed systems of their own and are selling the technology to other nations. A variety of Videotex systems exist, but they all operate in basically the same way. To access the signals transmitted via the telephone network, users must have special decoders and a modem built-in or attached to their TV sets. To connect to the central data base the user must first dial the appropriate telephone number and place the receiver in the modem. When the connection has been accomplished successfully, an index page appears on the TV screen and users begin to search for the information they require by pressing numbered keys on their hand-held control panel (keypad). Instructions appear on the screen telling the user which keys to depress for particular types of data,

The central Videotex data base may contain an almost unlimited amount of information provided by sources ranging from local newspapers to travel agents, department stores, and libraries. The data are stored in "frames" or screenfuls and can be updated instantly. Several frames of information on the same topic comprise a "page" and may be accessed sequentially. To retrieve information from the data base, users employ a "tree structure" search method, starting with broad subject headings and narrowing down their choices until they arrive at the frame of information they require. Such a system is fairly limited in capability and has no cross-referencing.

Unlike Videotex, teletext is a noninteractive system linking the information provider to the home via regular or cable TV broadcast signals. Once again, only TV sets with special decoders are able to pick up teletext. Pages of information are broadcast one at a time in recurring cycles. To access them, users consult a contents page, then use the keypad to key in the numbers of the pages to be retrieved. The decoder then selects the appropriate pages when they cycle by, and the information is displayed on the TV screen.

Teletext's chief virtue is its ability to be updated continuously for a large viewing audience. It can provide users with the most current information on a range of subjects and is easily accessible. Because it is broadcast rather than telephone based, teletext is also less expensive than Videotex, which requires users to pay for the telephone service and for each individual frame accessed.

Teletext has disadvantages, the greatest being its limited data base size. To access information, users have to wait until the specific page they are seeking cycles by, giving the decoder time to read, decode, and display the data. The wait time becomes excessive when the data base exceeds about 100 frames. Thus, teletext is severely limited in the amount of information it can carry efficiently.

Developments in the areas of processing, memory and storage, and communications interact to produce some very sophisticated information systems. On the whole, these systems are more compact, less expensive, considerably more reliable, store larger quantities of information, and offer newer types of service than the traditional type of information service. Moreover, access to the systems from remote locations is simpler and more efficient than ever before. Can we cope with the volumes of stored information that are available to us?

It is interesting to note that with the emergence of new technologies there has been a fundamental shift in the concerns of the information profession. This shift can be expressed, in simple terms, as a change from the attitude—''How can we give the user more information to solve his problem?" to "How can we make sure that the information we give the user is accurate, reliable, and up-to-date?" The basic questions concerning the organizations, classification, indexing, and retrieval of information are regaining importance, having been overshadowed over the last few years by the concerns of automation of systems and services.

With increasing volumes of information being stored and transmitted in digital form, and access to the stores being effected increasingly from the home environment, attention must return to the traditional areas of information handling mentioned above. It is envisaged, therefore, that a number of investigations will be initiated in the near future to consider potential solution to the problems of selection and quality assurance.

Future Trends

There are several trends under way today that will change the way researchers access data bases within the next few years. They cannot be ignored, for unlike the pioneers in the first generation of information retrieval, today's pioneers are outside of the biomedical field—and the evolution of the technology will go on no matter what resolution is made of current problems with access to MEDLARS. These trends are summarized here, since detailed discussion is beyond the scope of this paper. Many of the problems that private vendors have with NLM's policies reflect tensions being created by these changes.

1) *The widespread use of microcomputers is expected in the future.* We are now in the age of desktop microcomputers that have the power of the large computers of about a decade ago. Within the next 2 years, desktop machines with the ability to make sophisticated searches of the entire MEDLARS data base will be for sale in computer stores. Many, if not most of MEDLARS users in the developed nations might have such machines as general-purpose office research tools just as they now have typewriters, calculators, and lab equipment.

The key to matching indexes to a citation or abstract data base lies more in the size of a computer's core memory and the amount of external data it can access rapidly than in its speed of operation. About l/z million to 1 million characters (1 megabyte) or more would be adequate for a stand-alone desktop machine for data base retrieval. Such machines have been on the market for the past 2 years, costing about ***$30,000*** to **$40,000.** Their price should drop by half or more in the next few years. These are also precisely the type of machines that many nonscientists would use for general number crunching, and they may make central processing on time-sharing systems obsolescent.

2) *Timesharing, as a model for data base access is expected to be radically altered as distributed computing becomes more widespread.* In the 1970's, data bases were placed on large central computers to which many terminals had simultaneous access, That was called "time-sharing. " MEDLARS data bases and most of today's electronic data bases reflect that technology, which is undergoing transformation. The coming crop of 16-bit and 32-bit microcomputers attached to software represents a several orders-of-magnitude advance in computing, while commensurate improvements in long-distance data communications, and thus in time-sharing, during the next few years may be only one or two orders of magnitude beyond that of today.

With a microcomputer, a user would only have to slide in the data base mass storage device—perhaps a video disk of the data base without waiting for a time-sharing connection or paying time charges. The latest information, always a tiny percentage of the total data base, could still come via telecommunications lines, but not necessarily at that instant. Each evening, for instance, a microcomputer can be programed to call in for the latest citations, articles, or specialized data. The amount should be able to easily fit on the disk drives, which hold some 5 million to ***20*** million characters of data. There are now available for desktop machines at costs ranging from ***$2,500*** to $6,000.

Any user will tell how constrained one can feel when trying to browse through an on-line data base and the meter is ticking; in fact, browsing and experimenting with searches is the antithesis of current data base networks, but would be encouraged by distributed procession on a user's own machine, Many other features would be possible—e.g., generating a private profile of interests, reuse of material often needed, retransmission of important information to colleagues, reformatting information to one's own needs or tastes.

3) Computing costs are expected to drop faster than communicating costs. This relationship has been true for the past **30** years and all indicators show that the trend will continue for the near future. While these technologies are linked—the cost of telecommunications drops because of computer-related and electronics advances—the decline of both do not follow uniform parallel lines. While the cost of computing has been dropping regularly and in half every 2 to **3 years,** universal telecommunications advances tend to come in somewhat unpredictable spurts (for reasons that are beyond the scope of this study) and on the average at a slower rate.

As a result, the "time windows" are different for the introduction of computing and communications products. It appears that computing devices for local data base processing will enter the market about **2 to 3 years** before the advanced telecommunications systems. These data base products are now available: the microcomputers already mentioned, and laser video disks capable of storing all of MEDLARS on one or two disks.

Laser disks, which can store digital data, are being manufactured commercially today for about $5 in small lot quantities. The master disk costs about $2,000, not including the cost of the original videotape. Converting computer data to a video pattern is no major obstacle. But problems of data integrity and software for search and access on such disks are still to be resolved.

If MEDLARS had **5,000** customers worldwide, each paying $500 per year for a complete updated data base to date (assuming two disks), then the cost of producing 10,000 disks would come to **$50,000 + $2,000 +** distribution. The revenue of **$2,450,000** annually surely would cover distribution and a great deal of the cost of production of the data base. Another $500 per year could be charged for updating via floppy diskettes or some other system, and the user would still be saving a great deal of money, not to mention the added revenue to NLM or a private vendor providing the service. These figures are only meant as an illustration of the economics of the new medium of video disk plus microcomputers for data base access; the normal scientific marketing algorithms for pricing such a service might come out with more appropriate fees.

4) *The future telecommunications networks are expected to mix data, voice, and other traffic making it simple to connect distributed machines but also confusing national boundaries, property rights based on print, and what is information and what is communications.* Even Euronet is expected to disappear as new technology is introduced that will merge telephone, Telex, and data systems into "integrated services digital networks" (ISDN). With ISDN, virtually all local telephone connections will be able to handle voice as well as high-speed data; and the data links will connect directly to a packet-like network permitting instantaneous connections, the ability to transfer hundreds of pages per minute between computers or desktop workstations, and most likely distance-insensitive pricing for data. Most major cities, however, will offer some sort of ISDN service by the end of the decade both in Europe and the United States, if replacement projections for older telephone equipment continues at the current rate. In the United States, such networks should begin to emerge within the next 3 to 4 years, at first in industrial suburban and some rural areas.

5) *Copyright questions in the computerized information* ***age will continue to be extremely complex.*** International property rights are particularly complicated, since property rights are defined differently in different jurisdictions. Though U.S. copyright law forbids NLM from claiming copyright protection domestically, NLM places a copyright notice in its printed material claiming protection internationally, and claims contractual rights both domestically and internationally preventing computer centers from replicating its tapes.

Furthermore, applying property concepts derived for the printed press to computer-based scientific information necessary for worldwide health and biomedical research creates a need to reconcile some mutually exclusive goals: Do we want to minimize U.S. Government expenditures in regard to medical information at the risk of encouraging the biomedical data bases which may create a negative balance of trade in information? Do we want to restrict access by foreigners to Government originated publicly available information? Is it strategically effective to try to use this resource to trade for similar rights to foreign data bases? Do we want the originators of the literature to share in the proceeds from abstracting or retrieval or do we want to encourage the widest dissemination of information at the risk of discouraging future information providers? (Abstracters or bibliographers in the print media have never had to pay royalties to the authors of the material described.) How do we fix property rights if **a** computer does the abstracting? What rights do the authors of the abstracting program have? Countries are likely to answer these questions differently—once more making costs in different countries very different.

6) *Machine retrieval of biomedical information has only scratched the surface with on-line literature citations and abstracts; the technology is evolving to permit retrieval of full text of articles, raw data, and eventually automated search,* Precedents established for the rather primitive technologies of today can hamper further developments in the future; the converse could also be true: imaginative solutions to the domestic public/private mix of MEDLARS data base vending may encourage new techniques and industries to develop in the United States, rather than elsewhere. Several complex issues come to bear here: 1) the ability to enforce a form of international copyright on U.S. Government-generated data; 2) the property rights of the originator and "publisher" of the cited material; and 3) the comparative economics of data bases development and input outside and inside of the United States.

Solutions to these problems are not going to be simple. Transnational as well as cross-industry agreements will be necessary to protect information property rights. In some cases, there will be conflict between societal goals, particularly where information important to policy and technological decisionmaking is involved. Designing laws that will withstand decades of change is going to take a great deal of skill, both in anticipating technology and in anticipating new applications. The effects of new interventions are often unpredictable.