

Chapter 6

MEDLARS and Private Health Information Systems: Discussion of Domestic Policy Issues

MEDLARS and Private Health Information Systems: Discussion of Domestic Policy Issues

Advances in the application of computer and communications technologies to information transfer have focused attention on on-line systems in the public dialog concerning the Government's role in the creation and dissemination of information goods and services. Recently, the National Library of Medicine (NLM or the Library) has figured prominently in the debate. One view is that the achievements of its computerized bibliographic system, MEDLARS, indicate that NLM is fulfilling its mandate to create and disseminate health-related information. Another view is that MEDLARS' success prevents the growth of private sector organizations that create health-related data bases and commercial firms that vend on-line access to health-related data bases.

This chapter examines two sets of issues. One set concerns the range of NLM's computerized products and services. The Library regards MEDLARS as an extension of its library functions, while others say MEDLARS does not fall within traditional library functions and is an inappropriate NLM activity.

The second set of issues pertains to the pricing of NLM's computerized bibliographic products and services, specifically, NLM's fees for leasing NLM data base tapes and NLM's charges for on-line access to its data bases. NLM's position is that its pricing policies aid in the dissemination of health information to all who seek it; others argue that MEDLARS' subsidized prices give NLM a competitive edge over private sector firms and that NLM should recover the "full costs" of its computerized products and services. The debate concerning full costs remains active, although it

appears that NLM is moving toward recovering the full costs of its computerized products and services. Nevertheless, as will be discussed, the definition of full costs is open to various interpretations.

The issues examined in this chapter are considered within a general framework of the Government's role in the allocation of resources to information development and distribution; the effect of the Government's involvement in information activities on certain segments of the private information sector, and the health community; and the historic role of the Government in health information activities (see ch. 5). Specific criteria used to examine the issues are product differentiation, historic precedence, the presence of positive externalities (the social benefits received from a product or service exceed the sum of the benefits received by separate individuals), Governmental costs, present or potential private sector involvement, the effects of private sector participation on the creation and dissemination of computerized bibliographic health-related information, and international implications. (International issues are discussed in app. I.)

Although the analysis in this chapter focuses on current issues, that focus is not at all intended to minimize the importance of new and emerging technologies on biomedical information policy. Current issues are likely to be altered by changes in computing and communications technologies. Thus, the analysis in this chapter should be considered in the context of information concerning future information technologies. (These technologies are discussed in app. H.)

RANGE OF NLM'S COMPUTERIZED PRODUCTS AND SERVICES

Data Bases

MEDLARS, NLM's computerized bibliographic retrieval and technical processing system, includes MEDLINE (NLM's original data base) and many other health-related data bases. The rationale for NLM's creating or maintaining data bases varies from data base to data base, because of their diversity. (See ch. 2 for a detailed description of MEDLINE and, NLM's other health-related data bases.) MEDLINE is the focus of this chapter's discussion. For the most part, the other health-related data bases are used to illustrate specific points: it is not possible to generalize about them and it is beyond the scope of this report to assess individual ones.

The immediate concern with NLM's creating and making available health-related data bases arises from the claim that NLM is competing with the private sector's creation of such bases. Basic to the concern is the relation of NLM's health-related data bases to those created by the private sector. Do these data bases provide identical, similar, or complementary information? Can one base be substituted for another?

It is generally accepted that MEDLINE and "similar" health-related data bases do not duplicate one another, but that there is a degree of overlap. Overlap in coverage can occur at two levels. Journal overlap occurs when two or more data bases contain articles from the same journal. Journal article overlap occurs when two or more data bases contain citations of the same article from the same journal. Journal overlap does not necessarily indicate journal article overlap (112), because different data bases may include citations of different articles from the same journal. Moreover, the contents of the same article are often analyzed differently for different data bases in order to meet user needs. For example, an article on the biologic differences between two stages of a particular cancer can lend itself to a variety of analytic approaches centered on basic biologic processes, or on diagnostic techniques, or on chemical analysis, or on clinical manifestations of the disease.

EXCERPTA MEDICA and BIOSIS PREVIEWS are the data bases that appear to be the most similar to MEDLINE in subject coverage, but the extent of overlap among them has not been accurately determined. An early study comparing the results of searches on biological, medical, and veterinary subjects in EXCERPTA MEDICA and MEDLARS data bases reports that of the 226 relevant references retrieved from EXCERPTA MEDICA data bases and 467 relevant references retrieved from MEDLARS data bases, only 94 references (or 13.6 percent) were found in both secondary sources. The authors' tentative conclusion is that the systems complement each other, since if one system answers a question poorly, the other system answers it well (154).

A similar conclusion is reached in a later study that compared the two on-line search files as they existed in 1978 (19). An estimated 42 percent of the on-line MEDLINE file was covered also by the EXCERPTA MEDICA file and about 31 percent of the EXCERPTA MEDICA file was covered by the MEDLINE file. The methodology employed an author search: a search using indexing terms might have disclosed a different degree of overlap. The author concludes that "given the relatively large numbers of unique records contributed by each file, they are clearly complementary services and any comprehensive search should make use of both files" (19).

MEDLINE and BIOSIS PREVIEWS were included in an analysis of 14 major scientific and technical data bases for both journal and journal article overlap.¹ Among all the 14 data bases, there was approximately 20 percent journal overlap, and among these journals there was 23 percent journal article overlap—i.e., only 4.6 percent of all the articles in all the journals cited in any one data base were cited in more than one data base. The authors concluded that the amount of overlap was much less extensive than they

¹T. C. Bearman and W. A. Kunberger, *A Study of Coverage Overlap Among Fourteen Major Science and Technology Abstracting and Indexing Services* (Philadelphia: National Federation of Abstracting and Indexing Services, February 1977).

hypothesized before the study began. In any event, overlap is often useful, particularly when comprehensive retrieval is important, and MEDLINE and BIOSIS PREVIEWS have been used successfully in such a complementary fashion.

The data bases differ in other ways, including the professional fields of interest and focus of each of their abstracting and indexing services. For example, MEDLINE includes journals in nursing, dentistry, and allied health fields; EXCERPTA MEDICA does not. On the other hand, EXCERPTA MEDICA includes materials related to other health-related disciplines not found in MEDLINE, as do BIOSIS and the other health-related bases. The bases also differ in their coverage of English language publications. Sixty-five percent of the journals indexed for MEDLINE are in English, 60 percent of the journals indexed for BIOSIS are in English, and 50 percent of the journals indexed for EXCERPTA MEDICA are in English. Other differences include the percentage of abstracts in each data base.

There are few rigorous studies of overlap concerning other NLM health-related bases and "similar" data bases created by the private sector. One analysis of toxicology information abstracting and indexing services found that each service contributes somewhat differently in terms of scope of coverage and type of coverage, and that no one service is comprehensive or exhaustive in the field (112). The study points out the advantages of diversity of biomedical data bases and printed products, particularly in the toxicology and chemical fields.

Product differentiation, however, does not completely negate the possibility of substitutability among health-related data bases or printed indexes and abstracts, and at this time, there are no hard data on the extent to which the existence of one influences the use of another. Some claim that libraries with limited budgets, such as hospital libraries, choose MEDLINE instead of "similar" data bases on the basis of price alone. Others claim that price is only one criteria and that hospital libraries choose MEDLINE instead of similar data bases because the literature cited in MEDLINE is more relevant to their needs: the

practitioner who uses the hospital library will choose MEDLINE over other data bases because its content and scope are more related to clinical medicine (155).

The little evidence currently available suggests that the citations in MEDLINE, and other NLM health-related data bases, do not duplicate those available in any one base currently created by the private sector, and that the availability of a diversity of biomedical data bases may be advantageous. It is on this basis that the following arguments for and against NLM's creating and making available MEDLINE, and other health-related bases, are presented.

NLM created MEDLINE as a result of the computerized production of *Index Medicus*, and the medical and information communities throughout the world perceive this data base as an extension of the printed publication. * The printed index has a long and respected history, and is used worldwide: of the 5,888 copies sold in 1981, 2,623 were sold abroad. MEDLINE is also used extensively overseas. Foreign centers obtained access to the data base just shortly after it became available in the mid-1960's through quid pro quo arrangements (see app. I). In this country and abroad, the medical and scientific communities depend heavily on *Index Medicus* and MEDLINE and, with few exceptions, support their continued production.

The historic relationship between *Index Medicus* and MEDLINE in conduction with NLM's legislative mandate to publish and make available "catalogs, indexes and bibliographies . . . to the material it collects" (Public Law 84-941) is an important argument in favor of NLM's continuing to create and make available MEDLINE. The argument is strengthened by the joint production of the two products. *Index Medicus* and MEDLINE are tied together by the production process, as the computer tape which is accessed in MEDLINE serves as a vehicle for the production of *Index Medicus* and derivative products. Indeed, the computerization process that results in MEDLINE is the least expensive way of producing *Index Medicus*.

*There is little, if any, expressed interest in NLM's discontinuing the creation of *Index Medicus*.

The presence of social benefits (externalities—see ch. 5) is another criterion by which to assess NLM's continued creation of MEDLINE and other health-related data bases. Insofar as it is believed that the general public health requires expenditures for information products and services which individuals will not incur for their own benefit, Government provision of such information may be necessary. More specifically, if MEDLINE and NLM's other bases provide information to physicians, researchers, and other health professionals that results in an improvement in the health of various members of the general public, society may want the information created and disseminated, even if the physicians, researchers, and other health professionals do not perceive that they get enough benefit from the information to pay the price of buying it on the private market.

In considering how far such conditions exist for MEDLINE and NLM's other health-related data bases, it is assumed that some of the primary literature cited in the data bases contains information that results in an improvement in health. But do secondary sources, which refer to the primary literature, contain information that is needed to protect the general public's health? They would appear to, since the primary literature is voluminous, scattered in increasing numbers of journals, and virtually inaccessible without a well-constructed index (see app. D). Thus, secondary information is considered to have social value because it "leads to the use of primary information and can reduce costs of identifying and locating information for the Government and other users" (71).

The more significant questions are whether MEDLINE and NLM's other data bases cite the primary literature that contains information to improve health, and whether the citations are readily available to institutions and individuals that benefit the public's health. There is no simple answer to the questions. It is likely that the public receives the most immediate benefit indirectly from information that is available to practicing physicians and other health practitioners. MEDLINE contains references to literature in many fields that contribute to the science and practice of medicine and public health, with an emphasis on the basic medical sciences and clinical

medicine. While there is evidence that MEDLINE is responsive to the needs of practitioners (24), there are practitioners who suggest its responsiveness could be improved.

Because of subject content, subject headings used in analyzing articles, and the organization of the headings, searchers do not always easily retrieve references to the literature that physicians practicing in patient care disciplines and settings find useful. This is not to say that Index Medicus/MEDLINE is not used by practicing physicians and other health practitioners providing patient care. Indeed, it is the most used secondary source of all bibliographic biomedical sources (135), and there is extensive anecdotal and indirect evidence of its use by patient care practitioners.

On the other hand, information required by biomedical researchers is identified and organized in MEDLINE for easy retrieval. If one believes that biomedical research benefits the general public health, then one may very well believe that NLM is justified in producing MEDLINE. Indeed, Congress established NLM to "aid the dissemination of scientific information important to the *progress* of medicine and to the public health" (emphasis added) (Public Law 84-941).

NLM's other health-related bases require individual evaluation with respect to the presence of social benefits. Each contains citations of literature that varies in form and content: a few of the data bases contain numeric or representational data. The category of user varies from data base to data base as well. For instance, TOXLINE contains references to information that has an immediate, or delayed, beneficial effect on the public health. But 50 percent of its usage, as measured by connect hours, is by commercial firms, including drug companies. It could be argued that commercial enterprises receive enough personal benefits from the information they receive to pay the cost of buying it on the market.

The same reasoning can be applied to CHEMLINE, which receives 50 percent of its usage from commercial firms, but not to HEALTH, POPLINE, BIOETHICSLINE, and other bases. For example, only 1 percent of the approximately 318 hours of HEALTH usage in 1981 was by

commercial firms, while 80 percent was by hospitals, allied health and medical schools, and research institutions. Whether such users receive sufficient personal benefits from the information to pay for it on the market at current levels of usage is in doubt.

Related to the above discussion of social benefits is the Government's allocation of resources to health-related research, since the Government funds research because of the social benefits derived from its findings. The Government supports more than 50 percent of the biomedical research conducted in the United States, and over the years has noted the importance of organizing and disseminating the results of biomedical experimentation. (See the discussion of the history of health information in ch. 5.) It sometimes pays charges to professional journals to publish research results, and it has developed and maintains an organizing tool, *Index Medicus/MEDLINE*, for accessing the literature on Government-sponsored and other research. If the Library were to stop producing these bibliographic sources, it could be argued that meeting the goals and objectives of the Government's biomedical research policy would be made more difficult.

On the premise stated above, as Government decreases its funding for a type of research, its justification in sponsoring the results of that research is weakened. To illustrate, HEALTH is a data base that centers on the health services research and health care delivery literature. The appropriations for two of the major Federal organizations that sponsor health services research, the Office of Research and Demonstrations and the National Center for Health Services Research in the Department of Health and Human Services, decreased from \$69 million in 1980 to \$40 million in 1982. Should the Government, therefore, decrease or discontinue its support of HEALTH?

The issue is extremely complicated and must be measured against other criteria. The data base is used by hospital administrators and health practitioners as well as health services researchers. As Government expenditures for health care continue to increase annually at enormous rates, and to the extent that health service research suggests ways to contain health care costs, the rationale for Gov-

ernment sponsorship of access to this information is strengthened.

The private sector might be inclined to create MEDLINE and NLM's other health-related data bases with the same degree of consistency, quality, and comprehensiveness as NLM has. But the desire or lack of desire of the private sector to do so is purely speculative. At the present time, no established private concern has expressed an inclination to undertake the creation and maintenance of MEDLINE. It would be difficult for a private concern to rationalize the expenditure of money required to create and maintain the data base, since it requires the acquisition of a large collection, library staff to index and process the collection, and other capital-intensive equipment and facilities. Nonetheless, venture capital is available for many information activities. The possibility exists that if NLM were to announce now that it would cease the publication of *Index Medicus/MEDLINE* in 5 years or so, some organizations might be interested in the project (137).

However, there might be potential disadvantages associated with private sector creation of MEDLINE, including the potential risk of not being able to follow through on the proposed substitute system. There would be no assurances of continuity, quality, and comprehensiveness. It is doubtful that another organization could establish a relationship similar to the one the Library has with the medical, research, and library communities. The interaction is of great value in developing and maintaining the quality of *Index Medicus/MEDLINE* in areas such as developing medical subject headings (see app. E).

If *Index Medicus/MEDLINE* were to be produced by the for-profit sector, the data base might not be as comprehensive. MEDLINE contains rarely used citations to articles in journals with low circulation and citations of possible future use. For-profit firms tend to be more selective and utilitarian in their operations, and might concentrate on common information and information of immediate use. The uncertainty of future demand may limit the information to be preserved by the for-profit sector, although the potential societal value of the rarely used information makes its preservation important.

Finally, the price to the user of Index Medicus/MEDLINE might be higher if it were prepared by a private firm. The implications of higher prices are discussed below in the section entitled "Pricing Issues."

It is not possible to generalize when speculating about the private sector's inclination to create and make available NLM's other health-related data bases, because of their number and the diversity of their contents. However, the number of records in NLM's other bases and usage are low compared with MEDLINE (see ch. 2) and they may or may not be profitable for a private firm to create.

In September 1981, users interacted with MEDLINE for more than 5,500 hours, but they used TOXLINE, the next most utilized data base, for only 826 hours during the same period and CHEMLINE for only 285 hours. Nonetheless, in fiscal year 1981, with the NLM user charges then in effect, NLM recovered 110 percent of its accessing and tape costs associated with providing on-line access to TOXLINE and 75 percent of its accessing and tape costs associated with providing on-line access to CHEMLINE. If NLM's 1982 user charges had been in effect, NLM would have recovered 126 percent of its accessing and tape costs associated with providing on-line access to TOXLINE and 126 percent of its accessing and tape costs associated with providing such access to CHEMLINE (56).

In September 1981, BIOETHICSLINE and CANCERPROJ received only 31 hours of usage each and would not recover costs with such usage. But some data bases may reach a wider audience if produced by the private sector. For-profit enterprises generally have much more sophisticated marketing technique than the Government. Data base size and current usage are only two of many factors to consider in determining if a data base would bring in revenues. A few others include the availability of source material, the costs of production, and the possible publication of a print product from the data tapes. The last is essential, because data bases are not profitable currently unless produced along with a hard copy publication. Each of the factors varies with the individual MEDLARS data base.

If the commercial sector were to find it profitable to produce some of NLM's data bases, NLM might be responsible for creating and distributing only those bases that do not meet the market test. In that case, there is a chance that a perception might develop that NLM is creating and distributing data bases of little value.

On-Line Services

In addition to creating MEDLINE and other bases, NLM provides on-line access to them. Although the service is relatively new with respect to other NLM operations, it is a very visible activity and is associated with NLM's history and leadership role in bibliographic retrieval systems both here and abroad. It was established before commercial services provided on-line access to NLM's or to most other health-related data bases.

Currently, two commercial information services (vendors) —DIALOG Information Services, Inc. (DIALOG) and Bibliographic Retrieval Services (BRS)—lease the tapes of MEDLINE and a few other NLM health-related bases and vend on-line access to them. Are NLM, DIALOG, and BRS providing services in the same market? If NLM's on-line services are sufficiently different from those offered by the private sector, they may be operating in different markets. It is generally recognized that the private sector search services, as a result of better services and advanced software, are more efficient than those currently available from NLM (see ch. 4). It is not clear that NLM and private information services are different enough to conclude that the services are operating in different and discrete markets.

Another way of defining discrete markets is by the types of users being served by similar products or services. The markets are discrete if the users are sufficiently dissimilar. On this premise, there is some, but inconclusive, evidence that NLM and DIALOG, one of the private information services that sells access to NLM data bases, are providing services in different markets. As measured by connect hours, the majority of users that have on-line access to MEDLINE through NLM are in hospitals (40 percent) and academic institutions (19

percent). Although the exact figures are proprietary, there are indications that a large percentage of users that access MEDLINE through DIALOG are in commercial organizations. There is some overlap, since 10 percent of the users that access MEDLINE through NLM are commercial organizations, and 28 percent of DIALOG users are in academic and other nonprofit institutions. Nonetheless, DIALOG appears to be providing services to a market different from that of the Library. However, NLM and BRS appear to be serving the same market, as the majority of BRS users (55 percent) are in medical schools and academic institutions. But the boundaries of the markets are not necessarily firm and might shift if NLM were to stop providing on-line services.

Only very indirect evidence is available with which to assess the effect of NLM's provision of on-line services on commercial information services (vendors). Computer-based information services, a very much larger category of business firms than the three information services discussed in this report, realized almost \$8 billion in revenues in 1979 and anticipate increasing their revenues by 29 percent in 1980 (42). Another indication is that BRS was recently acquired by a multinational firm, Thyssen-Bornemisza, for a sum higher than BRS' original capital investment. However, in May 1982, BRS said NLM's new pricing structures did not allow BRS to vend MEDLINE profitably (46). (This issue is discussed further in the next section.)

NLM and others are concerned that serious damage to the integrity of NLM's data bases might result if NLM were to discontinue providing on-line services. Because of on-line access, user training, and user services, NLM continuously interacts with MEDLARS users. The users provide information about any inadequacies of the data bases, thereby facilitating another quality check of the data bases by NLM. A counter argument is that NLM might find other ways to communicate with users if it did not provide on-line access to NLM's data bases. Among others, BIOSIS creates a quality data base without providing on-line services.

A parallel factor is the collection of quantitative data on users. The Library requires such data, if it is to create bases that reflect user needs. No mat-

ter who provides the search service-NLM or private sector firms-they will be precluded from obtaining comprehensive statistical data on users because of the privacy of the search process. With that limitation in mind, it would be possible for private information services to obtain and provide only general statistical information on users for NLM's use, for a fee.

A most important factor to weigh in assessing NLM's provision of on-line access to its data bases is the Government's cost of providing the service. The additional cost to NLM of providing on-line access is relatively low because of user charges. For fiscal year 1980, according to NLM, approximately \$400,000 in appropriated funds would have been saved if NLM had not provided on-line access to MEDLINE but had continued to create the data base. Even if NLM were not to provide on-line access to MEDLINE, creation and maintenance of the data base would still be required. Service support, data base testing, disk storage, and computer hardware would still be maintained in order to provide the data base to vendors (103).

Somewhat higher costs related to NLM's provision of on-line access to its data bases are reported for fiscal year 1981 (56). In fiscal year 1981, NLM incurred costs directly assignable to offering MEDLINE publicly for on-line searching of \$3.241 million. NLM recovered \$2.336 million in user fees (including \$42,000 in services by NLM for MEDLARS carrying out its basic library activities). Thus, it cost the Government \$905,000 to provide on-line access to MEDLINE in fiscal year 1981, or 28 percent of the accessing and tape costs associated with providing on-line access. The accessing and tape costs associated with providing on-line access to all the NLM data bases were \$4.757 million. of this, NLM recovered \$1.271 million, or 28 percent, in user charges. If NLM's current (higher) user charges had been in effect during 1981, the provision of on-line access would have cost the Government only \$166,000 or 6 percent of the costs associated with providing access to MEDLINE and only \$235,000 and 5 percent of the costs associated with providing access to all NLM data bases (56).

PRICING ISSUES

Pricing issues related to leasing NLM's data base and pricing issues related to NLM's on-line services receive separate consideration in the following pages. But these pricing issues are interdependent, because the leasing fee that private information services pay for NLM's data base tapes is a factor in information services' pricing of on-line access to the data bases. The relationship between leasing fees for NLM data base tapes and NLM's on-line access charges illustrates the heterogeneity and diverse interests of the private information sector. If NLM's leasing fees for its data base tapes are high, private sector firms that produce health-related data bases presumably will be in a better position to market their product. This will especially be true if the leasing fees for NLM's data base tapes are higher than the leasing fees for the tapes of privately created data bases. At the same time, however, private information services that lease the NLM tapes and subsequently vend them will have higher costs in providing on-line services.

The different effects on different members of the private sector are reflected in the current dialog concerning tape leasing fees for NLM's data bases. NLM's current leasing fees for MEDLINE tapes are new, having become operational as of January 1982. Thus, it cannot be determined whether the full costs of creating the data base are recovered by the leasing fee, although the fee is much higher than the costs of reproducing the data tapes and much higher than the leasing fees for almost all other Government-sponsored data base tapes.

Some data base producers contend that the NLM fee for leasing data base tapes should recover the full costs of creating the data base. On the other hand, information services (vendors) in the private sector feel that NLM's leasing fees are currently too high, and some propose that NLM follow the model used by the National Agricultural Library for AGRICOLA and by the National Institute of Education for ERIC (see app. G) (137). The tapes of ERIC are leased by the General Accounting Office (GAO) at the cost of reproduction: the tapes of AGRICOLA are leased by the National Technical Information Service

(NTIS) at \$1,220 for domestic use plus a fee. (See table 12 for NTIS leasing fees for Government data bases.) Both computerized data bases have a wide offering, from in-house systems of computers at universities, companies, and other organizations, and are offered by commercial vendors at relatively low rates (see table 8 in ch. 4).

Fees for Leasing Data Base Tapes

The principal pricing issues concerning the leasing of NLM's data base tapes are: 1) whether leasing fees are to recover the costs of reproducing the tapes only, or to recover both the costs of reproducing the tapes and the costs of creating the data bases; and 2) whether foreign lessees are to pay a different fee from domestic lessees.

One view is that leasing the data base tapes at the relatively low cost of reproducing the tapes would allow for private sector participation and at the same time widen the distribution of health information. Another view is that private producers of health-related data bases could be adversely affected by this practice. The American Psychological Association, for instance, claims its data base cannot effectively compete with a base prepared by the National Institute of Mental Health, which leases its data base tapes to commercial information services at the cost of duplicating the data tapes (9). Another argument is that all U.S. taxpayers pay for the data bases, but only a few commercial information services would benefit monetarily from leasing the data base tapes based on reproduction costs only.

The private sector, in general, does not lease its machine-readable data base tapes to include the costs of creating the data base. If creation costs were included, the on-line access cost would be so high as to discourage usage. For the most part, the print products associated with the bases subsidize their production. The use of machine-readable data bases, although extensive, is new; the user communities in the health fields are not quite prepared to pay high costs for information (69). In addition, data base production is in a period of great transition, with a declining demand for print materials and new technological develop-

Table 12.—Selected Data Bases Distributed by NTIS, December 1981

| Department | Agency | Data base | Acronym | Annual lease | Use fee |
|---------------------------|---------------------------------------------|--------------------------------------------|----------|------------------------------------------|--------------|
| Agriculture | National Agriculture Library | The Agriculture On-Line Data Base | AGRICOLA | \$1,220 domestic 1,820 foreign | Yes |
| Commerce | National Oceanic and Atmospheric Agency | Aquatic Sciences and Fisheries Abstracts | ASFA | 2,500 | Yes |
| | U.S. Patent and Trademark Office | Patent Full Text File | | 10,920 domestic 16,380 foreign | Foreign only |
| | | Patent Bibliographic File | | 5,460 domestic 8,190 foreign | Foreign only |
| | | Patent Classification File | | 900 domestic 1,150 foreign | None |
| | National Technical Information Service | NTIS Bibliographic Data Base | NTIS | 5,350 domestic 7,150 foreign | Yes |
| Defense | Defense Technical Information Center | Technical Abstracts | TAB | | |
| Energy | Technical Information Center | Energy Research Abstracts | EDB | 2,950 domestic Not available—foreign | Yes |
| Health and Human Services | National Library of Medicine | MEDLARS On-Line | MEDLINE | 20,000 | Yes |
| Interior | Office of Water Research and Technology | Selected Water Resources Abstracts | SWRA | 2,300 4,600 | Yes |
| | U.S. Fish and Wildlife Service | Pacific Island Ecosystem | PIE | 500 750 | Yes |
| NASA | Scientific and Technical Information Office | Scientific and Technical Aerospace Reports | STAR | | |
| | Smithsonian Science Information Exchange | Notice of Research Project Data Base | SSIE | 8,000 16,000 | Yes |

SOURCE: National Technical Information Service, 1982

ments occurring rapidly. Accordingly, it appears unlikely that the private sector will lease its data tapes at a price that includes the costs of creating the data bases in the near future.

The problem is further complicated by the quid pro quo arrangements that NLM has with foreign centers. In exchange for either lease of the data base tapes from NLM or access to them by means of telecommunication to the NLM computer, these centers contributed almost \$500,000 worth of services, mainly indexing services, in 1980 and more than \$600,000 in 1981. If NLM were to provide the data base tapes at reproduction costs to the foreign centers, there would be a large drop in the quantity of indexing performed for the Library. Although no money is received by the Library in this transaction, existing or additional appropriations would be required by NLM in order to maintain the same level of indexing. Whether the foreign centers would agree to be responsible for the indexing of the foreign literature under a non quid pro quo arrangement is strictly conjectural.

NLM's quid pro quo arrangements with foreign centers have additional advantages. A rough examination of the MEDLINE citations in recent years shows that the percentage originating in the United States has remained relatively stable at about one-third of the total biomedical serial literature indexed (133). Furthermore, NLM's quid pro quo exchange agreements are quite important for the future, because the costs of publishing overseas have continued to remain below those of the United States and the probabilities of using electronic methods for dissemination of research results may reduce the U.S. percentage of MEDLINE citations. With a large percentage of medical journals being published overseas, the cooperative international system, mostly among governments or quasi-governmental organizations, benefits U.S. citizens as much as, if not more than, citizens of other countries.

A suggested alternative pricing policy to the current policy is to lease the data bases to private enterprises in the United States at the cost of reproducing the tapes, but to lease them to foreign

countries to recover the full costs of their creation, i.e., differential pricing. On the surface this idea appears attractive. If the data bases are considered to represent a national investment, it seems appropriate that U.S. citizens benefit from the investment, but that foreign nationals not necessarily share in the benefits.

However, the use of differential pricing might have serious international policy ramifications. Foreign nations might view this change in pricing policy as an antagonistic act on the part of the United States. They might conceivably reconsider their arrangements with the Library, thereby interrupting the flow of some medical information from abroad. In addition, technical considerations reduce the possibility of enforcing differential pricing for domestic and foreign users of the on-line computerized system. It is very difficult to distinguish between domestic and foreign users, and foreign sources often find a domestic source of information where price barriers exist. Unauthorized copies of data base tapes are also feasible with new technologies.

On-Line Service Charges

If NLM continues to provide on-line access to its health-related data bases, several specific pricing issues need to be addressed are: 1) whether to establish charges for on-line access to the data bases to recover the full costs of accessing the data base; 2) whether to establish charges for on-line access to the data bases to recover the full costs of accessing the data bases and of creating the data bases; and 3) whether to subsidize MEDLARS or MEDLARS users. The issues are discussed here in the context of structuring charges for on-line access to MEDLARS data bases to best serve the public purpose for which NLM was created without inhibiting the operations and development of private information enterprises. The debate on these issues has been complicated even more by lack of definition of the term "full cost recovery."

There are advocates of guaranteeing all citizens free and equal access to publicly funded library and information services (161). Conversely, there are advocates of full cost recovery of Government information products and services, including the costs of creating the information. In establishing

the prices for on-line access to MEDLARS, NLM is between these two positions. (See ch. 5 for the history of NLM's pricing policies.) It has adhered to the policy of sharing the costs of on-line services with the user community, with the taxpayers assuming the costs of creating the data bases and the user community assuming the costs of accessing the system.

Within the above framework, however, NLM has shifted its emphasis on the specific goals of its pricing structure. An economic analysis of NLM's on-line services in 1974 and 1975 investigated alternative pricing strategies and their relationship to social benefit (81). The study recommended marginal cost pricing for access in order to benefit the health of the Nation's people. It supported pricing access to MEDLINE at \$8 to \$15 per connect hour, which was the estimated price of recovering the full costs of that access, defined at that time by NLM as costs associated with on-line access "outside the walls of the library," such as telecommunications and backup computer costs.

In October 1981, NLM increased its price for on-line access to MEDLINE to \$15 to \$22 per connect hour. However, it did not justify the increase on the basis of social benefit as well as on the basis of full cost recovery of costs, as it did in 1975. Rather, the justification was on the basis of full recovery of access costs and on the basis of bringing on-line costs more in line with accessing other Government data bases and other discipline oriented data bases. The premise of the Library's cost calculations was also different, as the Library redefined access costs to include "all related computer and communication charges, all direct labor for system and network maintenance, all printing or duplicating charges, overhead and indirect costs, and training costs" (103) (emphasis added). NLM excluded from the cost base those costs which it would necessarily incur in maintaining the bibliographic apparatus of a national library, including capital costs.

According to NLM's definition of the "full cost" of accessing MEDLINE (i.e., the costs of accessing the data base but not the costs of creating the data base), and on the basis of NLM's analysis, NLM appears to be approaching full cost recovery

for accessing MEDLINE on-line (103). * A recent analysis (56) confirms NLM's conclusion that it is arriving at full cost recovery for accessing MEDLARS data bases. For fiscal year 1981, NLM recovered 72 percent of the accessing and tape costs associated with providing on-line access to MEDLINE and to all the other NLM data bases. This analysis was based on NLM's charges for assessing MEDLINE which had been in effect before 1981. If NLM's current (higher) charges had been in effect during fiscal year 1981, NLM would have recovered about 94 percent of the accessing and tape costs associated with providing on-line access to MEDLINE and about 95 percent of those associated with providing on-line access to all of NLM's data bases. NLM's method of identifying and allocating costs was in general considered reasonable for achieving order to achieve full cost recovery of its accessing costs (56).

It is too early to evaluate the effect of the increased charges for accessing MEDLINE on the private information sector. The lower pricing structure for on-line services that was in place until October 1981 did not appear demonstrably to have impaired the growth of the industry. It may indeed have fostered its development. As the first to offer computerized services in the medical bibliographic field, NLM with its low prices may have been instrumental in breaking down initial resistance to new products and services, thereby opening up markets for private firms (137, 163).

In any case, there are problems with generalizing the effects of NLM's charges on the private

information services (vendors). Not only are the firms diverse, with sometimes conflicting interests, but the relationship between leasing fees for NLM data base tapes and on-line access charges masks the effects of the latter. For example, BRS was able to successfully vend MEDLINE at the same price as NLM under NLM's pre-October 1981 prices for on-line services and NLM's pre-January leasing fee structure. As of January 1982, NLM changed its leasing fee structure from a flat annual fee to an annual lease fee plus a \$4 per hour usage fee and a \$0.01 per citation print fee all of which BRS passes through to users. NLM, however, does not impose a per hour usage fee or a per citation print fee on users who access MEDLINE on NLM computers. Because BRS customers are high-volume users and BRS provides on-line access to MEDLINE in approximately the same market as NLM, BRS finds it can no longer provide access to MEDLINE at the same price as NLM does, and obtain a reasonable profit. Thus, BRS is concerned that its users will switch from its service to NLM's and considers NLM's current leasing fee structure and on-line service charges unfair (46).

On the other hand, NLM's increased charges for on-line access to MEDLARS data bases are now more in line with DIALOG's charges. DIALOG has more of MEDLINE on-line than either NLM or BRS, and thus incurs higher costs which are reflected in its prices. DIALOG does not seem to be unduly affected by NLM's new leasing fee structure: the majority of DIALOG's users appear to be commercial organizations, and the majority of NLM's users are not (137).

Any assessment of the relation between NLM's increased on-line access charges to its data bases and the commercial viability of similar health-related data bases produced in the private sector would be based on very little evidence. The only information available are the statements of one or two private data base producers that, even at the current higher rates, NLM's charges for accessing MEDLARS in conjunction with NLM's leasing fees are detrimental to the sale of their products. Since financial data on private companies are proprietary, there is no way to substantiate the statements. As noted earlier, MEDLINE and similar bases are complementary and not identical, but there are no data on the extent to which the pres-

*The rationale for NLM's cost allocation which leads to this conclusion is that each cost for fiscal year 1980 was allocated entirely or in part to one of the five categories: 1) essential national library services (ENLS), 2) data creation, 3) publications, 4) on-line services, or 5) tape distribution. ENLS were considered to be those activities, such as selection and acquisitions, technical processing, reference services, etc., that are basic and essential for a national library to perform its mission. These are services that would be performed by a national library whether or not there were products such as publications, on-line services or tape distribution. The cost of indexing journal articles was considered to be partly basic service and was divided equally between ENLS and data base creation. Costs assigned to data base creation were those which were incurred in gathering or preparing the data on which the publications, on-line services, or tape distribution are based but which cannot be specifically assigned to one of the three. Each cost which could be allocated specifically was appropriately assigned. Portions of salaries, benefits, and other costs associated with personnel (travel and training) were assigned according to an estimation of the time spent.

ence of one in a market influences the use of another.

A new issue in calculating the full costs of accessing NLM's data bases is whether or not to include the costs of creating the data base as well as the costs of accessing the system in the calculation. The issue is of unique interest, mainly because it is of such recent origin. Until the unsuccessfully offered cost recovery amendment of the Health Care Research and Research Training Amendments of 1981 (S. 800), the creation of information was not referred to in any Government directive or practice regarding the pricing of information products or services; even the 1982 GAO report (56) on MEDLARS does not consider the costs of creating the data bases in its pricing analysis of NLM's on-line services.

Principle 6 of the report of the Public Sector/Private Sector Task Force of the National Commission on Libraries and Information Science (96) recommends that pricing policies for distributing information by the Federal Government should reflect the true cost of access and reproduction; however, the costs of data base creation are not included in the recommendation (see ch. 5 and app. F). The concept of creation costs is vague because the costs of creation can be considered indexing or abstracting costs only or can include acquisition costs and overhead costs as well. As mentioned above, private data base producers do not include creation costs in pricing the leasing fees of their data base tapes, and accordingly creation costs are not included in the on-line access charges set by commercial information services.

Most important is that a mandated function of NLM is to acquire, process, and index health information and funds are appropriated for that purpose. It is questionable whether MEDLARS users should pay for a process that is a general library function which benefits all users of the Library.

The major advantages of including data base creation costs in NLM's charges for on-line access to MEDLARS data bases is the increased revenues to the Government and the possible beneficial effects to the private sector producers of similar data bases if charges were increased.

As of May 1982, the effects of the new pricing structure of on-line services for the *users* that access MEDLARS were unknown. There are a variety of user institutions with a large variety of charging procedures for on-line access. Most pass the costs through to the users in varying degrees, but others absorb the costs. It is reasonable to suppose that commercial firms would not be unduly affected by higher costs of on-line access even if creation costs were to be included in the full cost formula that includes creation costs. Even some of the major academic health center libraries might financially accommodate to the approximate additional cost of \$1 or \$2 per search which would result from the added costs.

But there is valid concern that students, nurses, and allied health personnel and researchers who are not working on grants or contracts might find it difficult to pay much more than NLM currently charges. The libraries in small hospitals might be adversely affected even by small increases in charges. Indeed, there is anecdotal evidence that there has been a 30-percent decrease in searches performed in a Northeast consortium of small-sized hospitals since the increase to \$15 to \$22 per connect hour by NLM (50). Smaller institutions are extremely limited in their ability to control costs, and the information center or the library is one of the few areas subject to cost control.

Differential Pricing

The price to access NLM data bases might be prohibitive to some users either if the full costs of accessing the system, or if the full costs of accessing the system and creating the data bases were recovered in on-line charges. The Government could decide to continue to provide access to all users by subsidizing the system. Other Government alternatives would be to set different charges for accessing MEDLARS data bases on-line for different types of users (i.e., differential pricing) or offer a subsidy (through grants for example) to particular institutions or individuals to allow them to select the service and data base they wish to use.

If the Government subsidized the system, theoretically, then as now, MEDLARS users would

not have a choice of information services and would be dependent on NLM, which might or might not be responsive to their needs. This assumption about users' influence holds if NLM is providing on-line access to MEDLARS at a lower price than commercial information services and if price is the only, or even the main, factor in the choice of services. If the Government subsidizes the system another way—i.e., by paying private information services to provide the information at a low rate—users also would have little influence on the system. They would find it difficult to withhold payment or to switch services if they were dissatisfied.

Differential pricing is the alternative to governmental subsidization of MEDLARS that is currently receiving the greatest attention. Many proponents of differential pricing feel that the Government's subsidization of MEDLARS results in NLM's having an unfair competitive advantage over private information organizations. Some proponents also think it unfair that for-profit organizations pay the same on-line access charges as not-for-profit organizations to access the NLM data bases on MEDLARS. Differential pricing might eliminate or modify NLM's supposed unfair advantage and promote the information activities of the private information sector. The Government might also receive more revenues from increasing NLM's on-line access charges to commercial firms and other for-profit organizations.

But there are a number of arguments against differential pricing. NLM has been opposed to it in the belief that all users should have equal access to NLM services and all sectors of the user community should be charged the same amount for NLM's products and services. "Domestic commercial enterprises presumably pay their legislated share for supporting Federal Government activities and should enjoy the fruits of those activities on an equal footing with the non-profit sector" (110). In addition, establishing different charges for different institutions may require the imposition of a means test, which is not only demeaning to those who have to prove their inability to pay, but is usually cumbersome and costly to administer.

A problem in differential pricing is determining the category or categories of users eligible for

lower prices. One could argue that not-for-profit organizations fall into this category. The Federal Government has historically recognized a distinction between not-for-profit and for-profit organizations. This distinction recognizes the important role played by nonprofit organizations in America. It recognizes the contribution which the private nonprofit sector has made towards achievement of social goals.

Precedent for the distinction between treatment by the Government of profit and nonprofit organizations can be found in many laws and policies. For example, the Internal Revenue Code recognizes, for tax purposes, a fundamental distinction between the private, nonprofit corporation organized under 501(C)(3) of the code, and the for-profit corporation. One pays taxes, and one does not. The code also recognizes that a deduction can be made for contributions and bequests by individuals to private, nonprofit organizations. Such contributions are not taxable. But one could also argue that not all not-for-profit organizations benefit the public and that some for-profit firms provide greater social benefits. In addition, the distinction between not-for-profit organizations is increasingly blurred. For example, the National Institutes of Health, which once limited its grants to not-for-profit organizations, now provides grants to for-profit firms as well.

The other alternative to governmental subsidization of MEDLARS is to subsidize MEDLARS users by grants. This may also require adherence to some type of qualifying criteria ("means test"). The key disadvantages of subsidizing users in this manner are the complexity and the costs both of making payments to hundreds of users and of enforcing the regulation. Furthermore, the potential for Government regulation of private information firms exists if the Government provides funds that could be used to purchase on-line access from private information services. Indeed, a Government subsidy using a mechanism similar to grants could be structured even if NLM were not providing on-line access to MEDLINE and its other data bases.

However, there are advantages in subsidizing some users of MEDLARS rather than subsidizing the system. Subsidizing users might encourage competition and stimulate private sector informa-

tion activities, and might increase the efficiency of on-line information systems. Also, if the administrative costs do not exceed the charges recovered, the Government would recover more from user charges for on-line services if only some, as opposed to all, users did not pay the full costs for the services.

One observer has suggested that an office within the Department of Health and Human Services could offer credits for bibliographic search services of up to a given amount, such as \$300 to applicants without other sources of funds (43). If the applicants stated that they were not working on sponsored research and that they had incomes and net worths below specified minimums, they would be granted the credits. A somewhat looser procedure could be followed by schools without asking for income and net worth, much as computing funds are now allocated in many schools, depending on the total number of students in the categories of interests. The above techniques should be capable of being administered at a cost of perhaps 10 to 20 percent of the subsidies granted. Naturally, more assurance of the legitimacy of the requests could be obtained, but only at higher cost (43).

Research and Development

NLM conducts, supports, and promotes basic and applied research in information science and its technologies. Research conducted at the Library's Lister Hill National Center for Biomedical Communications and under research grants and contracts, along with the efforts of many other governmental bodies, played a major

role in the formation and development of the computerized data base and on-line information industry. Subsequently, private enterprise developed more advanced and innovative technologies (see chs. 2 and 4).

Research conducted and sponsored by NLM continues to benefit the private information sector. Among its other contributions, this research has laid the groundwork for the private production of master tapes for the subsequent production of video disks and the private development of video disk files of graphical data from patents, and has been responsible for the establishment of many private information firms (165).

NLM has also been among the first to recognize the need of practicing health professionals for more direct access to biomedical information than bibliographic sources afford. Thus, need is particularly acute in areas where primary information is limited or unavailable, as in developing countries. NLM developed the Hepatitis Knowledge Base as a "prototype information system" to enable users to quickly and efficiently find the proper information for their needs. Because of the intellectual input (the contents of the data base are evaluated by a consensus of experts), it is very expensive to produce. The Hepatitis Knowledge Base may serve only as a model of a refereed data base. Indeed, the American Medical Association and General Telephone & Electronics Corp. are launching a knowledge base containing drug-related information this year, but it will be of lesser magnitude and sophistication than the Hepatitis Knowledge Base and is expected to be used in concert with bibliographic data bases (128).

FINAL COMMENTS

This chapter has presented arguments concerning NLM's creation of computerized health-related data bases, its provision of on-line access to data bases, and different strategies for pricing the data base tapes and on-line access to the base. The discussion has focused on MEDLINE, the original and major data base in the Library's computerized retrieval and technical processing system,

MEDLARS. The findings with respect to the issues follow.

OTA finds that over the years NLM has established strong and intimate ties with health and information communities worldwide who rely on the Library's information sources, and, in turn, contribute to the high quality and comprehensiveness of NLM's bibliographic sources.

A major finding is that the creation of MEDLINE by the Library seems to be warranted by NLM's extensive collection of biomedical materials, by its legislative mandate, and on economic grounds. NLM has the world's largest collection of biomedical literature and is mandated by law to organize and make its acquisitions available. For more than a century, *Index Medicus* has been the guide to the Library's collection of biomedical journals, and today the data base tape that is used in printing *Index Medicus* is also used in preparing MEDLINE. MEDLINE is also part of the Government's investment in biomedical research and assures ongoing access by researchers and practitioners to information needed to maintain and improve the public's health.

In addition, OTA finds that the possibility that the private sector would be inclined to create MEDLINE if NLM were to cease doing the activity is a matter of speculation. The information industry is young: it has been functioning for approximately 15 years. It is uncertain that if NLM were to cease creating MEDLINE a new or established private firm would have the desire to produce a similar product. As noted previously in this chapter, the other major health-related data bases complement rather than duplicate MEDLINE, and the availability of diverse bases is advantageous for the user.

Another finding is that there is no convincing argument that clearly supports any specific method of setting leasing fees for the MEDLINE data base tape. There is a wide range of interests between the public and private sector and within the private sector resulting in equally good reasons for leasing the data tapes to domestic firms at the cost of reproduction or at the costs of reproduction and creation. The economic arguments for instituting differential leasing fees are opposed on technical, scientific, and international grounds. On balance, arguments for changing the present policy are not convincing.

OTA finds no compelling reasons at present for NLM either to continue or to discontinue providing on-line access to MEDLINE. NLM has nearly achieved recovering the full costs of access-

ing the system, thus making its on-line charges more in line with, although still lower than, the charges of commercial information services. NLM's current on-line charges appear to be sufficiently low to adversely affect the ability of one of the two commercial vendors of MEDLINE to realize a profit in providing the same service. At the same time, preliminary evidence indicates that the charges are sufficiently high to prevent some small hospitals from continuing the same level of searching they previously performed on MEDLINE using NLM's system.

Another finding is that there is inconclusive evidence with which to weigh the advantages of charging the users of MEDLINE who cannot pay NLM's current on-line rates (or any future increase in costs) a lower rate than those who can pay such charges against the advantages of keeping the charges at a level all users can pay.

In summary, OTA finds that many of the arguments presented by proponents or opponents on the issues pertaining to the creation, provision, and pricing of NLM products and services seem more reflective of philosophical perspectives than objective analysis, and there are few empirical data to support them. Thus, any changes in the range of NLM's computerized products or services or in their pricing structure require caution.

OTA also finds that the rapidly changing nature of the computer and communications fields *gives* additional credence to the need for care in current information policies and practices. The information field is in a period of flux: the shape of current on-line information systems is expected to change within a few years, particularly in the areas of remote processing and software. Along with technological changes, the economic issues pertaining to information systems, including MEDLARS, may change. For this reason, *OTA concludes that decisions made today in reaction to current problems should be, to the maximum extent feasible, informed by the ways that technological advances might change those very issues. OTA urges that this report be considered in the context of the material presented in appendix H on future information technologies.*