Growth patterns in the National Library of Medicine's serials collection and in *Index Medicus*® journals, 1966–1985

By Betsy L. Humphreys, M.L.S.
Deputy Associate Director, Library Operations

Dianne E. McCutcheon, M.L.S.
Assistant Head, Serial Records Section

National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland 20894

Data from the National Library of Medicine (NLM) automated Master Serials System and its MEDLINE® database were used to chart the growth of NLM's serials collection and of the journals indexed in *Index Medicus*® from 1966 to 1985. The number of live serial titles in the subset of NLM's collection examined increased 30% in the twenty years. The average number of articles per *Index Medicus* journal increased 56%. The average number of articles in U.S. *Index Medicus* journals grew more rapidly than the average number in journals published elsewhere. The NLM data provide clear evidence that the years from 1966 to 1985 saw a substantial increase in the percentage of the biomedical serial literature published in English. The period from 1966 to 1985 saw substantial but uneven growth in the number of serial titles in the NLM collection and in the average number of articles in *Index Medicus* journals. Although data on the number of articles published in *Index Medicus* journals is unlikely to reflect the number of articles in other journals, the pattern of growth in the number of serials held by NLM probably reflects trends in the universe of all biomedical serials.

INTRODUCTION

In 1964, Orr and Leeds observed that "the holdings of very few libraries can be considered as essentially complete for any broad area of science; it is only such 'saturated' collections (e.g., NLM) that could be expected to grow at the same rate as the literature" [1]. In addition to having the only "saturated" collection of biomedical literature, the National Library of Medicine (NLM) indexes many substantive journals to produce MEDLINE®, *Index Medicus*®, and other online bibliographic databases and publications. Perhaps inevitably, data concerning the size and composition of the NLM collection and the journals NLM indexes have therefore been used in many estimates of the size, growth rate, geographic origin, and subject and language distribution of the biomedical journal or serial literature [2–6].

Despite their common reliance on data about NLM's collection and indexing activities, the various studies and their results are rarely comparable. This divergence is due to variations in the definition of a biomedical journal or serial, in the inclusion or exclusion of "living" and "dead" journals, and in comprehensiveness [7]. Such variations reflect differences in availability of data at different time periods and to different investigators, the necessary reliance on cumbersome manual sampling or counting techniques, and the inevitable place of subjective judgment in categorization of journals or serials.

The purpose of this study was to determine how NLM's serials collection and the journals indexed in *Index Medicus* grew and changed between 1966 and 1985. A secondary objective was to assemble machine-readable data on the size of the NLM serials collection and of *Index Medicus* journals that could be updated consistently in the future and made available to other investigators. The study was prompted in part by an observation made by Corning and Cummings in their bicentennial essay on biomedical communication. In looking at the SERLINE® data for the net number of wholly "new" serial publications appearing each year (i.e., excluding title changes and subtracting the number of titles that ceased), they commented that the data "show a decline in new titles for 1973. Whether this will in fact be a sustained decline, a trend toward

a plateau, or an oscillation remains to be determined" [8]. The data reported here show what actually happened after 1973.

METHODOLOGY

The sources of data for this study are the bibliographic module of NLM's Master Serials System and the MEDLINE database and its backfiles [9]. The Master Serials System is a set of files used to control the ordering, receipt, and processing of serial issues for the NLM collection; to produce SERLINE, a publicly searchable file of information on biomedical serials; to build SERHOLD®, a file of serial holdings data for more than 3,000 U.S. health sciences libraries [10]; and to generate a variety of serials publications. The system, which uses the INQUIRE® database management software, contains bibliographic data for virtually every serial title in the NLM collection, including those that have ceased publication.

When new serial titles are cataloged for the NLM collection, they are assigned shelving location numbers, which begin with broad category designations. For example, serial government documents are assigned shelving locations beginning with "W2," and abstracting and indexing tools are assigned locations beginning with "Z." More than half of the serials NLM acquires are assigned shelving locations beginning with "W1." This category includes journals and monographic series (e.g., *Annual Review of Physiology, Monographs in Population Biology*) with substantive, indexable articles as well as a relatively small number of newsletters and annual reports of private organizations. The "W1" category excludes government documents, abstracting and indexing tools, directories, most proceedings of international congresses, and purely administrative publications of organizations. The serial title statistics in this study included only the serial titles from the "W1" category. A small number of titles indexed for *Index Medicus* are not classified in the "W1" category and, therefore, are not reflected in the serial title counts. However, articles from these titles are reflected in the *Index Medicus* article counts.

With the "W1" classification as a basic parameter, search strategies using the INQUIRE user command language were developed to identify the number of new titles and new publications for each decade from the 1660s to the 1970s and the number of new titles, wholly new publications, and ceased titles for each year from 1966 to 1985. In this paper, "new titles" includes both the publications that changed titles in a particular year and those that appeared for the first time during that year. "New publications" excludes the title changes. "Ceased publications" includes only those that were not continued in a new form or superseded by another title. To identify the set of wholly new publications, the search strategies excluded records with certain text strings (e.g., "continues," "supersedes") in the bibliographic note field. In cases where an existing title split into two or more titles, none of the new titles was counted as a new publication. To identify the set of ceased publications, search strategies excluded records with text strings such as "continued by" or "superseded by" in the bibliographic note field. Titles absorbed by other titles were counted as ceased. Information in the fields designated for beginning year, ending year, country of publication, and language also was used to prepare the counts summarized in this paper.

Journals indexed for *Index Medicus* are a subset of the journals indexed in MEDLINE, NLM's online database of indexed citations to articles in selected biomedical journals. The article data used in this study were restricted to *Index Medicus* journals because that is the subset for which data are available for the entire 1966–1985 period. Counts of the number of articles in *Index Medicus* journals were obtained by searching MEDLINE and its backfiles. Counts of the total number of serial titles indexed in *Index Medicus* were taken from the introductions to the *List of Journals Indexed in Index Medicus* (*LJI*) for 1967 to 1985 and from an actual count of the number of titles listed in the *LJI* for 1966.

Counts of the number of U.S. titles indexed in *Index Medicus* in each year from 1966 to 1985 were obtained by reviewing the title listing by country that appeared in each annual *LJI*. The average number of articles per year was computed by dividing the total number of articles indexed by the number of journals indexed in that year.

Numbers derived from these various methods then were entered into the Lotus 1-2-3® microcomputer software package, from which the charts illustrating this paper were generated.

LIMITATIONS OF THE DATA

Title counts

Delays in NLM's acquisition of new serial titles could affect the results. This possible confounding factor was minimized by ending the study with 1985 data and by obtaining counts from NLM's Master Serials System in January 1992. The assumption was that the more than six years between 1992 and the end of 1985 were sufficient for NLM to have identified and acquired most of the titles begun in 1985 that would ever be added to the collection and for NLM to have verified and recorded titles that ceased publication or changed through 1985.

The validity of the new title counts also is dependent on the consistency with which the "W1" classification was assigned to publications initiated during the period under study and on the lack of any

significant NLM backlog of unclassified titles begun in the relevant years. In fact, the NLM policy for classifying these materials did not change in any pertinent way between 1966 and 1985. Unclassified titles for the twenty-year period were reviewed, and data for those belonging to the "W1" category were included in the results.

The success of the method used to distinguish changes in title from wholly new publications is dependent on the accuracy of NLM's identification of relationships between new titles and previously published ones and on the consistency of the terminology used to describe these relationships in NLM bibliographic records. There is no way to determine what percentage of actual title changes was detected by NLM staff and then described in the library's records for serial publications. Anecdotal evidence suggests that NLM may not have been very successful in identifying title changes that took place prior to 1960, but there are no firm data on this point. As revealed in an inspection of records for samples of titles assigned to each category, title continuation information in NLM's bibliographic records was expressed consistently, and the search strategies used in this study interpreted it with a high degree of accuracy.

Finally, the validity of the results depends on the completeness, accuracy, and specificity of the information in NLM's files regarding the beginning and ending dates of serial publications. An examination of these data in NLM's automated records revealed a small number of titles with no beginning dates or with dates specified only to the relevant century (e.g., "19??") and a significant number of pre-1960 titles for which start dates were specified only to the relevant decade (e.g., "194?"). Data on the specific years and even decades in which titles ceased were very incomplete for titles known to have ceased prior to the mid-1960s. For these reasons, counts of new titles and new publications are aggregated at the decade level prior to 1966, as background to the more specific data presented for 1966 to 1985. No attempt has been made to count the number of ceased publications prior to 1966, and, therefore, no data are presented for the net gain in titles per year or decade for the pre-1966 period. The few titles that began or ceased after 1965 but have beginning or ending dates specified only to the nearest decade were excluded from the year counts.

Article counts

The figures for the number of journals indexed and the average number of articles per *Index Medicus* journal in each year are close approximations rather than exact numbers. Because they were taken from the introduction to the *LJI*, the title counts do not reflect titles added or subtracted during the year and may include some titles that yielded no indexable articles in a particular year. The counts include a small number of selectively indexed publications, from which NLM indexed only some articles published in any given year. NLM gradually became more inclusive in its indexing of letters, editorials, and articles from selectively indexed journals during the time period examined, but these categories account for only a small percentage of the total articles in all of the years studied.

The figures for total articles and average number of articles per year reflect the year of publication of the articles, not the year NLM indexed the articles. Because 1985 is the last year for which data are presented, the effect of any indexing delays should be negligible.

RESULTS

Growth in the number of serial titles

Figure 1 shows the distribution by initial decade of serials in the NLM collection that began publication before 1850. Although the oldest serial title at NLM began publication in 1665, the 1750s was the first decade that yielded more than a few serials (i.e., twenty-one publications). From the 1780s on, the number of new titles per decade began to rise quickly, exhibiting a sharp spike in the first decade of the nineteenth century. It was also in the 1780s that the phenomenon of title changes became more noticeable. (The difference between the "new titles" and the "new publications" bars on the graph is the number of serials per decade that changed titles.) The 1840s, the last decade shown in Figure 1, produced 417 wholly new serial publications. Thus, the number of new publications that appeared in the 1840s was nearly twenty times greater than the number published in the 1750s. The increase occurred very unevenly. The percentage increase from decade to decade varied from 0% to 89%, excluding the spike at the start of the nineteenth century.

Figure 2 illustrates the trend from the 1850s to the 1970s, using a substantially different scale than in Figure 1 to accommodate the increasing numbers of new titles per decade. With the exception of the decade of World War I, the data show a significant overall increase in titles per decade from the 1850s to the 1940s. The number of wholly new publications per decade was nearly ten times higher in the 1940s (4,495) than in the 1850s (465). The rate of overall increase for this 100 years was therefore about half of the rate for 1750 to 1849. Again, the increase occurred very unevenly.

The 1950s followed the pattern of the previous several decades. NLM acquired 5,489 new publications initiated during the 1950s, a 22% increase over the 1940s. The pattern changed in the 1960s, with the

Figure 1
Serials in the NLM collection, by decade initiated (1660s–1840s)

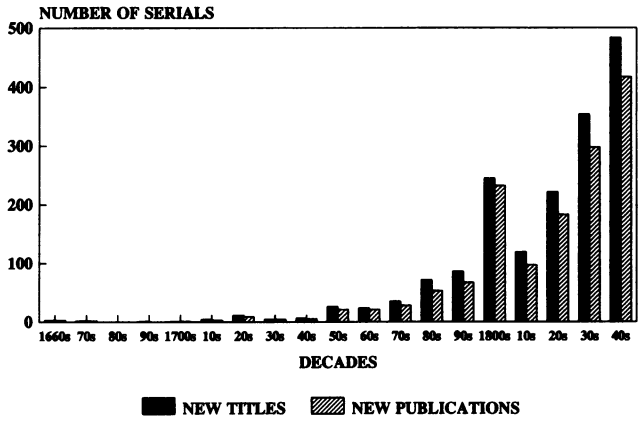
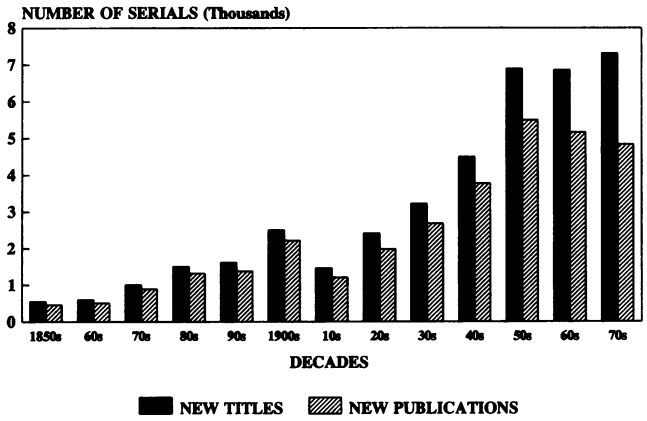


Figure 2
Serials in the NLM collection, by decade initiated (1850s–1970s)



number of new publications acquired by NLM declining 6% from the 1950s level. In the 1970s, the number of wholly new publications declined another 6%—the first two-decade decline since 1750. As the number of new publications declined, the number and percentage of title changes increased. The 1980s are not included in Figure 2, because complete data for the decade are not yet available. Preliminary 1980s data indicate that the change from the 1970s will be an increase of at least 20%.

Figures 1 and 2 provide general background for the more specific data gathered for 1966 to 1985. Of the "W1" serial titles in the NLM collection, 10,591 appear to have been in publication in 1966. NLM's records indicate that 13,780 titles were being published in 1985, 30% more than in 1966. In 1966, 2,655 (25%) of the "W1" serial titles in the NLM collection were published in the United States or its territories. In that same year, 5,203 (49%) of the titles were published entirely or partially in English. Twenty years later, in 1985, U.S. publications accounted for 29% of the current "W1" serial titles, and 60% of the publications were entirely or partially in English.

Figure 3 illustrates how the 30% increase in the "W1" serial titles acquired by NLM occurred between 1966 and 1985. The number of new titles appearing each year remained relatively constant throughout the period, although peaking noticeably in the first year of each decade and again in 1985. There is only a 1.5% difference, for example, between the numbers of new titles appearing in 1967 (719) and in 1984 (730). The number of wholly new publications appearing each year generally followed the trend of the new titles, although the percentage of new titles that were title changes was greater in the 1970s than in the late 1960s or early 1980s. The number of ceased publications per year increased gradually from 320 in 1966 to a high of 419 in 1973 and then generally declined for the next ten years. Excluding the peaks in 1970 and 1980, the net annual increase in "W1" serials acquired by NLM declined overall from 174 in 1966 to 41 in 1973 and then gradually increased to a new high of 379 in 1985.

Growth in the number of titles and articles in *Index Medicus*

In 1966, NLM was indexing 2,282 titles for *Index Medicus*. After a drop from 1966 to 1967 caused by the decision to stop indexing about 200 foreign-language titles, the number of *Index Medicus* titles remained relatively steady from 1967 to 1973, reflecting a con-

Figure 3
New and ceased serials in the NLM collection, by year

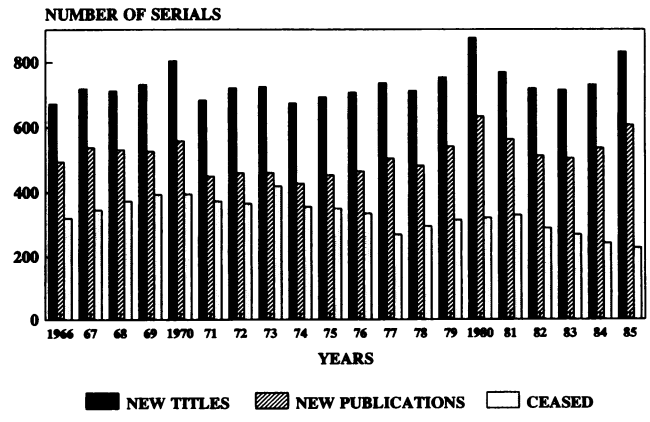
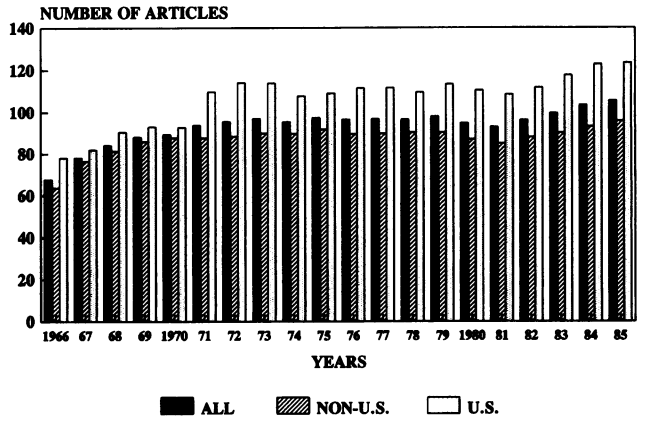


Figure 4
Average number of articles per *Index Medicus* title



scious policy on the part of NLM. After 1973, this policy changed and the total number of journals indexed in *Index Medicus* was allowed to grow very gradually. By 1985, the number of titles indexed in *Index Medicus* had increased by 434 to a total of 2,716, 19% higher than in 1966. The percentage of titles indexed that were published in the United States also increased from 1966 (28%) to 1985 (35%).

The number of articles indexed was 73% higher in 1985 (286,469) compared to 1966 (165,255). In 1966, 32% of the articles came from U.S. titles, and 54% were published in the English language. In 1985, U.S. titles accounted for 41% of the articles, and 78% of the articles were in English.

Because the number of articles in *Index Medicus* increased more rapidly from 1966 to 1985 than did the number of titles being indexed, the average number of articles per indexed title obviously was growing. Figure 4 illustrates this uneven growth. The overall increase during the twenty-year period was 56%, from an average of 67.5 articles per title in 1966 to an average of 105.5 articles per title in 1985. Although the general trend is similar for U.S. and non-U.S. titles, the average number of articles per U.S. title is higher. The disparity between the overall number of articles in U.S. and non-U.S. titles widened in 1971 and remained substantial through 1985. In 1985, the average number of articles in U.S. indexed titles was 123.4, which was 27.7 articles or 29% more than the average number for non-U.S. titles.

DISCUSSION

The data described in this paper indicate that 1966 to 1985 was a period of substantial but not explosive or exponential increase in the volume of serial literature acquired by NLM. The 56% increase from 1966 to 1985 in articles per *Index Medicus* title is unlikely to have been reflected in the nonindexed titles in NLM's collection. Even if this rate of increase did occur throughout the collection, the resulting rise in the total number of articles in the NLM collection still would not qualify as exponential. By comparison, data on the number of new serial publications published from 1750 to 1959 and acquired by NLM appear to reflect an exponential growth rate for that period, although there are insufficient data on ceased publications and on increases in the number of articles to draw a firm conclusion. In general, the NLM data collected for this study appear to support De Solla Price's seminal analysis of the growth of the scientific literature, described as an initial period of exponential growth, followed by saturation and slowdown to a steady rate of increase [11].

Data concerning NLM serial titles and indexed articles offer conclusive proof that the percentage of the world's biomedical literature being published in the English language increased overall in the 1966 to 1985 period. Even if the growing numbers of U.S. publications in the NLM collection and in *Index Medicus* are due to variations in selection effectiveness or policy, these factors do not account for the much larger increase in the percentage of titles and articles published in English from 1966 to 1985.

As first identified by Corning and Cummings, the 1970s were a period of particularly low net growth in the number of serial titles in the NLM collection [12], as well as a period of relative stability in the number of articles published in *Index Medicus* journals. The 1970s also saw an unusually high number of cessation and title changes in serials owned by NLM. The title change phenomenon probably was related to the increasing emphasis on publication in English. In the 1970s, the titles of many serials acquired by NLM changed from another language to English.

The NLM serial title data indicate that biomedical serial publishing activity is atypically high in the first year of a decade and in the first decade of a new century. Library managers should take this into account when projecting serials cataloging workload and estimating shelf space required for serial volumes. If past patterns hold true, then the arrival of the new century will mean a hectic period for staff engaged in serials processing and control. Of course, it is possible that increasing reliance on alternative publication formats (e.g., electronic publishing) will make the near future very different from the recent past.

An obvious question is whether the overall growth in the NLM serials collection from 1966 to 1985 reflects an expansion of the entire universe of biomedical serials during that period. Although there is no

way of determining what percentage of all biomedical serial publications initiated from 1966 to 1985 was acquired by NLM, some pertinent information is available. NLM has the largest collection of biomedical serials in the world. The library collects materials in all languages and in a broad range of fields related to medicine and health, including dentistry, veterinary medicine, basic biomedical sciences, and alternative therapies. In the period under study, NLM's acquisition of new serial publications was not affected by its budget; that is, the library was not forced to exclude titles due to lack of funds. Between 1966 and 1985, NLM modified its collection development policy several times to reflect the expanding and changing boundaries of biomedical and health-related information. The figures for serials in the NLM collection therefore include publications in newer fields of study such as biotechnology, as well as new serial publications in longstanding medical fields such as surgery. As basic scientific disciplines produced publications in subdisciplines specifically related to medicine and health, NLM cut back on acquisition of the general publications that previously had contained some biomedical material. During the period under study, the policy on acquisition of newsletters became more restrictive, but the number of such publications acquired was probably not significant in any of the years examined.

The results of past studies of NLM coverage of key journals—in disciplines as varied as medical informatics, medical behavioral science, immunology, paleopathology, and gene sequencing—indicate that NLM coverage of significant journals in biomedicine and other health-related fields is close to comprehensive for the entire 1966–1985 period. The figures presented in this paper, however, include other types of publications, such as peripheral journals and some newsletters. Inadvertent and undetected changes in acquisition patterns for these materials could have an effect on the numbers derived from NLM files. These caveats aside, it is unlikely that the patterns of growth in the universe of biomedical serial titles differed substantially from the patterns of growth evident in NLM's collection from 1966 to 1985.

In contrast, the article counts for *Index Medicus* titles do not reflect the number of articles in the nonindexed "W1" serial titles in the NLM collection. In 1957, Brodman and Taine found that the average number of articles in titles indexed by NLM was 71.2, and the average number of articles in nonindexed but indexable titles received by NLM was only 46.9 [13]. The magnitude of this disparity is likely to have persisted and even increased since 1957. Anecdotal evidence suggests that from 1966 to 1985 the number of articles per nonindexed title grew at a slower rate than did the number of articles in indexed titles. Therefore, the data presented in this paper should not be used to estimate the percentage of published papers in the NLM collection that NLM indexed during the years studied. This limitation is particularly important given the hundreds of non-*Index Medicus* titles covered in various NLM databases that are not reflected in the indexed article statistics reported here.

Potential research uses of the data

The data gathered for this study may be useful in the investigation of a number of research questions related to the growth of the biomedical literature and the scientific literature at large. Some examples follow.

- Are the growth patterns in journals of other scientific disciplines different from the patterns observed in the biomedical literature acquired by NLM?
- If NLM's collection reflects the biomedical serial literature as a whole, why were the 1970s a period of particularly slow growth?
- Are growth patterns in specific biomedical fields (e.g., molecular biology) significantly different from those observed in the entire NLM "W1" serials collection and among all titles indexed in *Index Medicus*?
- How do growth patterns in NLM's collection and in the articles indexed relate to increases in research funding, in research activity, and in the number of researchers in biomedical disciplines?

NLM intends to update the data reported in this study from time to time and to make them freely available.

Implications for users of the biomedical literature

The data presented here are of little practical significance to the individual researcher or practitioner who is interested, not in literature in the mass, but in information on particular topics. Irrespective of the relative rates of growth in the biomedical serial literature or in the amount of that literature published in English, there is likely to be a daunting amount of literature relevant to any single research or practice question—much of it duplicative and some of questionable value. This fact has not changed since 1966. Keeping current in any field, no matter how specialized, remains difficult and time consuming. Devising new and better ways to help users to identify, to obtain quickly, and to synthesize information in the literature pertinent to a particular problem must continue to be a priority for medical librarians.

ACKNOWLEDGMENT

The authors wish to thank Annette Nahin of NLM's MEDLARS Management Section for obtaining the *Index Medicus* article counts used in this study.

REFERENCES

1. ORR RH, LEEDS AA. Biomedical literature: volume, growth, and other characteristics. Fed Proc 1964 Nov/Dec;23(6):1329.
2. ORR, op. cit., 1310–31.
3. BRODMAN E, TAINE SI. Current medical literature: a quantitative survey of articles and journals. Proceedings of the International Conference on Scientific Information. Washington, DC: National Academy of Sciences, 1959:435–47.
4. CORNING ME, CUMMINGS MM. Biomedical communications. In: Bowers JZ, Purcell EF, eds. Advances in American medicine: essays at the Bicentennial, volume 2. New York: Josiah Macy Jr. Foundation, 1976:725–39.
5. KRONICK DA. Characteristics of the literature. In: Kronick DA. The literature of the life sciences. Philadelphia: ISI Press, 1985:57–62.
6. RODERER NK. U.S. expenditures for biomedical communication 1960–1985. Rockville, MD: King Research, 1979: 40–5.
7. CORNING, op. cit., 727.
8. IBID.
9. FISHEL MR, HUMPHREYS BL. Descriptions of automated serials control systems: the National Library of Medicine. In: Potter WG, Sirkin AF, eds. Serials automation for acquisition and inventory control. Chicago: American Library Association, 1981:134–9.
10. WILLMERING WJ, FISHEL MR, MCCUTCHEON DE. SERHOLD: evolution of the National Biomedical Serials Holding Database. Ser Rev 1988;14(102):7–13.
11. DESOLLA PRICE D. Science since Babylon. Enlarged edition. New Haven: Yale University Press, 1975:165–86.
12. CORNING, op. cit.
13. BRODMAN, op. cit., 437.

Received February 1993; accepted May 1993