HUGH FOLK
Professor of Economics
Institute of Labor and Industrial Relations
University of Illinois
Urbana-Champaign

The Impact of Computers on Book and Journal Publication

Rising wages have made the cost of composition and printing editions of a few thousand copies exorbitant. Rising book and journal prices have contributed to the rising operating expenses of libraries. At the same time, the volume of scientific and technical literature is increasing rapidly and the publication and library system is increasingly incapable of satisfying user needs.

Electronic publication of scientific and technical literature is technically feasible. One machine-readable copy of a document may be stored in a computer and accessed at any remote location by a user with a computer terminal. Current costs of preparing a machine-readable text, storing it on a computer, communicating with the remote computer, and computer time for the user to read or print the document are low enough that, for many applications in scientific publishing, a computer-based system may be less expensive than the existing paper-and-ink system. Rapidly decreasing computer and communications costs indicate that electronic publication will be increasingly cost-effective compared to alternative systems.

The benefits of a comprehensive electronic system for scientific and technical publishing are substantial. Not only would such a system be more complete and less expensive than conventional publication and distribution, but it also would be more accessible. Many users who do not now have access to the scientific literature could use the system. An electronic system would increase scientific productivity; not only would the scientist spend less time learning about what has been done, but he/she would obtain much more current information than is now available. Moreover, a computer-based scientific information system could provide a medium for two-way communications between users that would be more convenient and effective than the telephone or mail service.

PUBLICATION OF SCIENTIFIC BOOKS AND JOURNALS

Most scientific literature is published by professional societies, universities, and government, with few scientific books and journals selling more than a few thousand copies. Scientific literature is complex in terms of composition; it is characterized by equations, tables, and diagrams, the composition cost of which may exceed printing and distribution costs. The amount of scientific literature is growing rapidly. The stock of scientific publication has been estimated to increase at 7 percent annually;[1] the total number of journals is increasing at 4 percent per year.[2]

The growth of scientific literature threatens to overwhelm readers. Libraries, cramped for space and short of money, cannot hope to provide complete coverage of the growing literature. Indexing and abstracting by computer may identify what is in the literature, but this makes the tasks of libraries more complex.[3] Interlibrary loan and photoduplicates are increasingly used to provide missing documents, but seekers of current literature frequently pursue the author by mail and request a reprint. Because of journals such as *Current Contents*, a parallel system of author-to-reader distribution has forced upon authors the burden and cost of distributing their work.[4]

A huge literature of technical reports has also emerged;[5] unreviewed, unread, and unsung (at least by promotion committees), technical reports are cited with increasing frequency. Most are not indexed and consequently are difficult to find, but federal government-supported research reports are increasingly included in the National Technical Information Service (NTIS), are indexed, and are readily obtainable as microfiche, photocopy, or magnetic tape.

Attempts to stifle the growth of the literature create more problems than they solve. Proposals to tighten reviewing and to exclude literature from journal publication are idealistic but unrealistic. Authors must publish even if they have nothing to say and, if necessary, they will create new journals.

Attempts to lower the costs of publication by using microfiche, separates, or abstract or condensed journals simply fragment the literature further and increase the gadgets and oddly shaped cabinets in the library.[6] These new media also increase the pressures on authors for direct distribution.

A scientific information system that is complete, unified, rapid, and inexpensive is the goal. Many people claim that by the 1990s such a computerized scientific information system will exist. Computers are considered to be too dumb, too small, and too costly to do the job today. They are thought to be too dumb to process the multilevel math, tables, and diagrams that fill technical writing, too small to store and access the trillions of characters required, and too expensive to replace the existing paper/ink and film system.

Many who have worked with computers are acutely aware of their limitations, but those who are involved in the development of new computer systems realize that large complex systems are being created at an accelerating pace. Lockheed Information Systems offers in DIALOG a remarkable set of bibliographical indexes usable interactively by telephone or Telenet. The LEXIS system of Mead Data Central, Inc., provides on-line retrieval of legal text for several states. Large data-base applications in business are widespread; some firms have hundreds of terminals accessing a central computer and processing millions of transactions daily. Moreover, computer costs are decreasing rapidly. It is these recent developments that make an economical, computer-based scientific information system possible.

COMPOSITION OF SCIENTIFIC BOOKS AND JOURNALS

Between Gutenberg in 1450 and Mergenthaler in 1886, hand composition experienced no significant technological change. Slug casting, or hot-metal composition, reduced production costs sharply. In 1890, Alfred Marshall's *Principles of Economics* sold for $4, a worker's weekly wage, while in 1936, Keynes's *General Theory of Employment, Interest, and Money* sold for $2, a worker's daily wage.

Book and journal prices remained remarkably stable for the first half of the twentieth century, material and labor price increases apparently offset by increases in productivity. From 1967 to 1974, the price of hardcover books increased by 67 percent,[7] and from 1967 to 1975, U.S. periodicals increased by 130 percent, with even greater increases for chemistry and physics (214 percent) and for engineering (166 percent). The large increases in the cost of scientific and technical periodicals result from their typographical complexity.

Cost increases during the past two decades would have been even greater had not publishing technology improved. Offset printing has made it easy to reproduce simple typescript, but typescript is ugly. Lines are unjustified, letters are usually the same width, and footnotes and tables are unattractive. Strike-on composition with even margins was made possible with justifying typewriters, and more sophisticated strike-on compostion is possible with new devices such as magnetic card typewriters, computer terminals, or computer line-printers.

Strike-on composition is inferior in speed and elegance to photocomposition, which can solve all of the typographical problems of the compositor. Subscripts, superscripts, non-Roman alphabets, and several type sizes are readily composed. Phototypesetters are normally driven from magnetic tape produced by a computer.

Strike-on and phototypesetter composition, or "cold-type" composition, have been widely adopted. At its best, the quality equals that of hot type. For pure text composition the cost can be as little as $3 or $4 a page, or comparable to typewritten text composition. A typist working with a computer text editor can produce copy with speed and accuracy. With the economy and speed of offset printing, press runs of a few hundred copies can be quite inexpensive.

Offset printing has contributed to the growth of scientific publication. In the 1960s, Praeger, using typewriter composition, produced a number of books from doctoral dissertations and technical reports. In the late 1960s, Heath Lexington adopted cold copy to produce justified-text books. The first of these were not very attractive—footnotes dangled and white space was uneven—but libraries bought them and the publishers apparently made money.

During the 1960s cold copy was widely adopted by technical publishers. Today, text is set on a phototypesetter, but tables and equations are usually monotyped, proofed and stripped with the text photocopy to produce the page copy.[8] Composition combining photocopy, monotype, and typescript requires several suppliers and is slow and costly.[9] Moreover, the complexity of the process drives production managers wild. (Parisi gives a fascinating history of composition at the American Institute of Civil Engineers.[10])

Publishers agree that a completely computer-based composition system is needed and will take over the market as soon as it becomes available. It need not be inexpensive: publishers want speed and simplicity. Economy is important and would accelerate the adoption of computer photocomposition, but high-quality appearance is necessary.

COMPUTER-BASED COMPOSITION

Computer-based composition produces a single computer file that contains all text, tables, equations, footnotes, references, and page makeup instructions to run an on- or off-line phototypesetter. Several such systems are used commercially to compose scientific journals.[11] The TROFF system developed by Bell Laboratories[12] is used to compose in-house technical reports, documentation, and at least one technical newsletter.[13] This system has several major advantages and represents the current state of the art: (1) it runs on inexpensive computer hardware, (2) it is usable by persons with little training, and (3) it uses an inexpensive phototypesetter. TROFF runs on the UNIX time sharing system developed by Bell Laboratories as a general operating system for the Digital Equipment Corporation PDP 11/45 and 11/70 computers.[14] UNIX includes a powerful context editor that allows a typist to input text at either an on-line terminal or an off-line cassette terminal.[15] The typist inputs equations and table using special programs that can be learned in a few hours.[16] Once in the computer, the draft can be corrected by using special commands, such as *spell*, which looks up every word in an on-line dictionary and produces a list of words not in the dictionary, and *typo*, which uses the rules of English spelling (such as they are) to find possible typographical errors. Other programs have been written by various users to help authors improve their style. A word-frequency command provides a list of words used in a document and a count of how many times each was used. Another program prints the lines in which some troublesome homophones (such as *there/their*) or pet words (such as *obviously* and *clearly*) occur so that a writer can avoid error.

The UNIX typist then inserts commands to provide for centering, underlining, footnoting, equations, tables, indentation and other typographical specifications. The TROFF program then processes the file with text and interspersed command language and produces a decent-looking, justified typescript.

Three kinds of output files can be produced by the TROFF program. Text without footnotes, superscripts, subscripts, or multilevel equations can be printed correctly on a line-printer. Text including these complications can be printed on a special printer, such as a Diablo. Text with phototypesetter commands interspersed can be printed on the phototypesetter.

The emergence of commercial, computer-based phototypesetting means that the day of hot-metal or mixed-mode composition for technical publishers is ending. The availability of machine-readable copy from phototypesetter files suggests that concurrent electronic and printed publication is now possible.

A COMPUTER-BASED SCIENTIFIC INFORMATION SYSTEM

No serious technical problems prevent development of a computer-based scientific information system (SIS). Computer-based photocomposition demonstrates that computers can store and print typographically complex documents.[17] On-line disc systems can store billions of characters, any of which can be accessed in milliseconds. Computers can be accessed remotely by telephone, or interconnected through data-communications networks. Computer time-sharing allows hundreds of users to use the same computer interactively and simultaneously. SIS is technically feasible, but most people think the computer costs are too high to build one now.

SIS, like any storage and retrieval system for scientific documents, must accommodate an enormous volume of material and low usage. Any such system is dominated by the cost of data input and storage, with computer system costs a very small fraction of the total.

Computer-based photocomposition can reduce the cost of data entry substantially. Several publishers, such as the American Institute of Civil Engineers and the American Chemical Society, are now using computer-based photocomposition to print their journals; others, such as the American Institute of Physics, use computer composition for parts of each article, such as title, author, and abstracts. Multiple use of part of the material is increasingly common.[18] Machine-readable abstracts are routinely used in the production of secondary journals and indexes.

It is said that each article in a scientific journal is read ten to twenty times in its life. This statistic depresses authors and computer system designers alike. It hardly seems possible that it would be economical to retain scientific literature of such limited popularity throughout eternity on a computer disc with an access time of a few milliseconds, yet it is.[19]

The telephone is inexpensive for local calls and costly for long distance. Economical electronic publishing requires minimization of the combined storage, communication, and computer cost. Storage costs are minimized with one universally accessible copy on disc, but this requires the use of long-distance telephone lines. To print a 10-page, 50,000-byte article might require five minutes using a common thirty-character-per-second terminal. This could cost from $1 to $5, depending on the locations of the reader and the computer, and the time of day. This is inexpensive compared to costs of $7-$8 for interlibrary loan.

The telephone line can transmit information faster than the terminal can type it, so it is necessary to have either a faster terminal or a local computer to buffer the transmission. Computer networks can reduce the cost of data communications. Each node in the network has a computer which receives messages (or documents) and stores them until the user wants them. Because many users' messages are passing through a communications circuit simultaneously, each message (or "packet") must be labeled with identifying information. Research networks (such as ARPANET) and commercial networks (such as Telenet) using packet-switching technology are operating and obtaining substantial economies in data communications.

Packet-switching networks such as ARPANET and Telenet allow interconnection of dissimilar computers and permit users to send and receive files from remote computers easily. Minicomputers can be interfaced to a network without using any significant part of their memory or sacrificing much of their local processing power.[20] Packet-switching technology makes the cost of data transmission essentially independent of distance or intensity of usage. The user pays for the data actually transmitted plus the cost of occupying a permanent port on the network and users may pay an hourly connect charge to cover the costs of direct terminal access.[21]

Computer time-sharing is a well-established technology, and almost everyone has used such a system or has stood by helplessly while one was being used. Most time-sharing systems operate on medium-sized computers costing several million dollars to support a few dozen simultaneous users. Some systems, such as PLATO IV, support several hundred simultaneous users in sophisticated (but computationally limited) applications. The cost of an hour of computer time is usually based on how much hardware is used, but commercial time on full-sized systems usually averages from $10 to $25 per connect hour. Limited service systems, such as PLATO IV, may be able to provide services at a cost of from $2 to $3 per hour. The development of time-sharing operating systems for low-cost minicomputers promises to reduce general-purpose time-sharing costs to $1-$2 per hour. Special limited-purpose usage, such as that required from printing documents and inputting data, should be somewhat less expensive.[22]

Data input costs are very low for journals using computer-based photocomposition. New disc systems are inexpensive, computer networking provides low-cost communications, and minicomputers can provide low-cost time-sharing services.

A distributed SIS is emerging without any plan or central direction. Not only is DIALOG available on Telenet, but several universities (such as MIT) have computers on the network. Thus, the indexing system and user hosts already are interconnected. All that is necessary for SIS to exist is that one or more publishers place their machine-readable texts on a network computer. An organization such as the American Chemical Society or the Institute of Electrical and Electronic Engineers could make part or all of their future publications available in this manner. Indexes could indicate the articles that were available and the host address. As journal hosts joined the network, a computerized scientific information system would emerge.

IMPACT OF A COMPUTERIZED SCIENTIFIC INFORMATION SYSTEM

A computerized scientific information system is emerging. It promises eventually to cost less and be far more convenient for the user than the existing publishing and library system. Computer cost trends suggest that we should not be niggardly in designing the system. A computerized system with the capability of two-way communications must not merely imitate the paper-and-ink system of today. SIS need not be a limited document storage and retrieval system; for example, a network "mail" system would permit readers to comment on a document and the author to reply.[23] The comments and replies could be linked to the document file so that subsequent readers could be brought up to date on the state of discussion. Authors might also use the system to prepare and submit papers for publication, and editors and referees would use the system to speed publication by on-line reviewing, using the network mail system. The mail facility would permit scientific publication to be quite rapid; for example, a few days would suffice for refereeing, author's corrections, and copy editing. The system would become complete, and a user could access every document in the system from anywhere in the world, and no journal would ever be in the bindery.

Such a system would have at least as much garbage in it as libraries do today—probably more—but the garbage would not clog the system. Computer-based indexing would guide a new breed of scholars to the literature, and users could retrieve everything bearing on a subject they wish to investigate deeply, or skim the cream by requesting only widely cited and reviewed articles. Inaccurate, slovenly, and plagiarized articles would be panned; at last it would be possible for authors to publish and perish at the same time.

The impact of SIS on authors and readers would be revolutionary: no document will be condemned to obscurity, or hidden from a reader who wants it. The impact of SIS on publishers would also be revolutionary. At first they would attempt to collect a copyright fee, but ultimately SIS would become the exclusive system. Nonprofit publishers who are only attempting to cover expenses would find that modest page charges can cover the cost of publication. Journals in hardcover would wither away, until only the table of contents is left as evidence that the editors have approved publication. Finally, libraries would also wither away, their historic duty done. Perhaps we will call the local user host, through which the user accesses the network to read and write, a "library."

Developments in computer-based photocomposition now hold the promise of producing inexpensive photocopy from a single sequential file. In the process of producing this file, electronic publication of scientific literature becomes available as a low-cost alternative to conventional distribution. Recent decreases in the costs of disc storage, time-sharing computers, and the development of computer networking permit cheap storage, access and transmission of text files. As a result, computer composition and electronic publication now appear to be less expensive than conventional publication. To gain these advantages, libraries should be prepared to participate as user sites and to install user network hosts to provide access to the network as electronically published journals become available. Publishers should participate in developing the system by providing machine-readable copies of their publications to experimental and demonstration systems.

The widespread adoption of electronic publishing will herald an important new day in science. The act of publication will become the first step in scientific communication, rather than the last step, as it is too often today. The scientific literature will become unified, reversing the recent trend toward diverse forms of publication. Scientists everywhere will have equal access to the scientific literature, so that the advantages of being in a famous center of research will be substantially lessened. Scientists in obscure universities or poor countries will be able to participate in scientific discourse more readily. When that day finally dawns, scientists will look back on the problems of authors, publishers, and librarians of today with sympathy. Let us hope that they will be grateful for the work that was done to make electronic publishing possible.

REFERENCES

1. Price, Derek J. de Solla. "Citation Measures of Hard Science, Soft Science, Technology, and Nonscience." *In* Carnot E. Nelson and Donald K. Pollock, eds. *Communication Among Scientists and Engineers.* Lexington, Heath Lexington Books, 1970, pp. 3-23.

2. Barr, Diana R. *Trends in Book Production and Prices.* London, National Central Library, 1972, p. 24.

3. Many people believe that computer indexing and abstracting is not very reliable, efficient or cost-effective. Approaches such as Salton's SMART, written in appropriate languages and implemented on appropriate hardware, can be extremely useful. It is important to distinguish between interactive systems which the user can direct and redirect to desired documents, and batch systems in which a once-and-for-all request must be specified. An interactive system can be highly efficient in retrieving all desired documents, and is limited only by the user's imagination in specifying objectives.

4. An author has, perhaps, the right to hope that having paid submission, review or page charges for publishing his article, he has done his part to finance distribution. The heavy cost of distribution by reprints in some fields is burdensome for the author, but it is much more inexpensive for the requester than is interlibrary loan or subscribing to the journal. Requesters seem to expect free, postpaid copies, and sometimes grow abusive when a charge is imposed. As it is, a reader with a stack of prepaid and preprinted postcards can build an excellent library at the expense of the authors.

5. Brearley, Neil. "The Role of Technical Reports in Scientific and Technical Communication," *IEEE Transactions on Professional Communication* PC-16:117-19, Sept. 1973.

6. Bovee, Warren G. "Scientific and Technical Journals on Microfiche," *IEEE Transactions on Professional Communication*, PC-16:113-16, Sept. 1973; and Staiger, David L. "Separate Article Distribution as an Alternative to Journal Publication," *IEEE Transactions on Professional Communication* PC-16:107-12+, Sept. 1973.

7. Brown, Norman B. "Price Indexes for 1975," *Library Journal* 100:1291-95, July 1975.

8. The trouble or expense in setting mathematical equations has led to the exportation of much of the mathematical typesetting to Europe. This is much less expensive for books, but it is obviously impractical for journals that attempt to be current. G. Wroughton of Superior Printing, Champaign, Ill., has developed a highly effective and high-quality phototypesetter (using Alphatype as a foundation) that allows on-line setting of up to eight-level math that is used particularly for mathematical texts.

9. Metzner, A.W. Kenneth. "Multiple Use and Other Benefits of Computerized Publishing," *IEEE Transactions on Professional Communication* PC-18: 274-78, Sept. 1975.

10. Parisi, Paul A. "Composition Innovations at the American Society of Civil Engineers," *IEEE Transactions on Professional Communication* PC-18: 244-73, Sept. 1975.

11. Mack, Paul F. "Lower Composition Costs Through Optical Scanning and Photocomposition," *IEEE Transactions on Professional Communication* PC-18:279-80, Sept. 1975; Metaner, *op. cit.*; and Parisi, *op. cit.*

12. Ossanna, J.F. "TROFF User's Manual." Bell Laboratories internal memorandum, 1974.

13. Lesk, M. "Cheap Typesetters," *SIGLASH Newsletter* 6:14-15, Oct. 1973.

14. Ritchie, Dennis M., and Thompson, Ken L. "The UNIX Time-sharing System," *Communications of the ACM* 17:365-75, July 1974.

15. For a description of several on-line editors and their use, *see* Roistacher, Richard C. "On-line Computer Text Processing: A Tutorial," *Behavior Research Methods & Instrumentation* 6:159-66, 1974.

16. Kernighan, Brian W., and Cherry, Lorinda L. "A System for Typesetting Mathematics," *Communications of the ACM* 18:151, March 1975. Equation typing on other systems is usually difficult or tedious; *see* Korbuly, Dorothy K. "A New Approach to Coding Displayed Mathematics for Photocomposition," *IEEE Transactions on Professional Communication* PC-18:283-87, Sept. 1975.

17. Diagrams and line-drawings can be digitized, stored and printed on printers such as the Diablo or graphics scope terminals. Even half-tones or television frames can be stored and displayed. In terms of storage, a picture is worth considerably more than a thousand words. Doing these things in a commercial system is simply a matter of cost.

18. Metzner, *op. cit.*

19. Suppose one printed volume of a journal costs $25 and is purchased and shelved by 100 libraries at a total cost of $2,500. Storing 100 volumes requires about ten square feet of library floor space or, at current construction costs, about $400. Thus, the capital cost of the volume to the library system is $2,900. A volume is about 5 megabytes (MB), or 1,000 pages of 5,000 bytes (or characters) each. An AED controller and disc system for 536 MB (formatted) costs about $75,000. This system would hold 100 different journals for a cost of $750 per journal. Mass storage devices with much cheaper costs (less than $1/MB) are commercially available, but for a distributed system using minicomputers, a single system will not exceed a few hundred megabytes.

20. For a clear, but technical, discussion of this, *see* Chesson, G. "The Network Unix System." In *Proceedings of the 5th Symposium on Operating Systems Principles.* Austin, Tex., 1975.

21. The current charge for Telenet is about $6/MB and $2/hr. for connect time. A user accessing the network through a user host would pay the connect charge to that host and not the Telenet connect charge. The average cost of host-to-host data transmission is about $7/MB. This cost will be reduced substantially in the future.

22. A Digital Equipment Corporation PDP 11/70, supporting up to thirty-two simultaneous users, with sixty-seven megabytes of AED fast swap disc, 64k words of memory, and hardware and telephone interfaces can be purchased for about $100,000. With the system running unattended, operating, space, and maintenance costs should not be more than $20,000 a year. If 25 percent utilization were achieved (about 60,000 connect hours) the system could break even on a five-year depreciation schedule for about $.67/hr. for connect time. A system with hardware similar to the one described is operating at the University of California, Berkeley, and is informally reported to achieve the operating characteristics specified.

23. This facility exists on many time-sharing systems as a message system. I typed this paper on the Center for Advanced Computation's computer system and sent a message to a colleague to comment on it. He copied the paper from my directory and sent his comments to me overnight. The network mail facility is in constant use over the ARPANET. Every registered user has a mail address, and a message is sent from a user on his own system by typing a command such as "netmain high at 111-nts" followed by the message or file that is to be sent.

ADDITIONAL REFERENCES

Burchinal, Lee G. "Microforms and Electronic Publication: Emerging Bases for Scientific Communications," *IEEE Transactions on Professional Communication* PC-18:174-76, Sept. 1975.

Licklider, J.C.R. *Libraries of the Future.* Cambridge, Mass., M.I.T. Press, 1965.

Morris, Robert, and Cherry, Lorinda L. "Computer Detection of Typographical Errors," *IEEE Transactions on Professional Communication* PC-18:54-55, March 1975.

Senders, John W. "Information Storage Requirements for the Contents of the World's Libraries," *Science* 141:1067-68, 1963.

Shirrell, Robert. "New Production Options for the Journal Publisher," *IEEE Transactions on Professional Communication* PC-18:150-51, Sept. 1975.

Soma, John T. "The U.S. Computer Industry: An Economic-Legal Analysis of its Major Growth Factors." Ph.D. dissertation prepared for the University of Illinois, 1975.