FEATURE ARTICLES

Electronic Publishing: Potential Benefits and Problems for Authors, Publishers, and Libraries

LORRIN R. GARSON*

Books and Journals Division, American Chemical Society, Washington, DC 20036

JEANNE G. HOWARD

University Libraries, The University of Oklahoma, Norman, Oklahoma 73019

Received February 14, 1984

The appearance of scholarly publications in electronic form represents a profound change in communication. This new technology offers exciting possibilities as well as potential dangers for authors, publishers, and librarians. A number of issues are discussed that potentially have a variety of effects on these three groups—ownership and archiving of information, quality control of publications, access to, payment for, and retrieval of information, and downloading of machine-readable files. Education of scientists and librarians is viewed as crucial to the acceptance and utility of these electronic systems. As a result of this new technology, it is expected that libraries will probably undergo the greatest change and scientists are likely to be the principal beneficiaries.

In June 1983, the American Chemical Society (ACS) made the full text of its 18 primary journals, in electronic form, available to the public through Bibliographic Retrieval Services (BRS) in Latham, NY, for on-line searching. Although in the past few years, the full-text files of other publications have been available on-line (Mead Data Central's *Lexis* and *Nexis* and John Wiley & Sons' *Encyclopedia of Chemical Technology*), the ACS is the first scholarly association to provide its publications in this manner.

Undoubtedly, other scientific associations and publishers of technical information also will make their publications available electronically for on-line searching and retrieval. This new method of communication represents a profound change for scientific research and scholarly activity. Just as scientific information and data are the resources that fuel the engine of science and technology, like fossil fuels, dissemination of information on paper will eventually be substituted by something else; neither are likely to be replaced immediately or completely. The issue is not *whether* substitution will take place, but when and by what.

Technological changes in information processing, communication, and use will occur as have other institutions and processes of our society. Consider transportation. Late in the 19th century and throughout the 20th, the internal-combustion engine brought unprecedented mobility to our society, first through the railroads and shipping on the high seas, then most profoundly through the automobile, and finally by aircraft. The stately, full-rigged clippers were replaced by coal-burning steamships, which were superseded by oil-fired liners with steam turbines. The beautiful liners no longer make the North Atlantic run; they have been replaced by jumbo jets. These transformations in transportation also created serious disruptions to many institutions and industries. Those individuals involved in the manufacture of naval stores and horse-drawn buggies or horse breeding were very much adversely affected. Society also accepts the disadvantages of technology with astonishing equanimity. Few are distressed by the deaths of 200–300 individuals every weekend on U.S. highways, except those directly involved. Computer technology promises to be to the information industry what the internal-combustion engine has been to the transportation industry, a profound disruption to institutions and industries. However, as with other new technologies, electronic publishing systems offer exciting new possibilities as well as potential dangers. It is the purpose of this feature article to suggest what these possibilities may be—who the potential beneficiaries are and what institutions and individuals are likely to be disturbed and why.

In the science establishment, authors or "author–scientists", publishers, and libraries[1] comprise a triad. Like the circulatory, respiratory, and nervous systems, one cannot function without the other. Publishers rely on the submission of manuscripts by author–scientists to produce manuscripts for books and journals. For revenues, publishers depend upon individual scientists and libraries to purchase their products. As for libraries, their customary role is to serve the information needs of scientists and the public in general by purchasing materials and lending them to users. Thus, libraries house and dispense that which comes from author–scientists via publishers. Scientists depend upon libraries for many of their information requirements, for the information itself, and for assistance in efficiently locating that information. Scientists also depend upon publishers to communicate their thoughts and findings in a format that is accessible by others. In addition, publication is the mechanism by which a scientist's priority and reputation are established.

In this triad of information transfer, all is not harmonious. In spite of the interdependency of publishers, libraries, and author–scientists, there are conflicting concerns among these groups. Who owns information? How much is it worth and to whom? Who should pay for information? Who can most efficiently search for it? Who is responsible for maintaining its quality and accuracy? Who should establish standards for

0095-2338/84/1624-0119$01.50/0

its format and presentation? Who should have access to it and who should control that access? Who is responsible for maintaining it for future generations—for how long and at what cost? What formats are most suitable for reading, storing, searching? Who should profit the most, and take the concomitant financial risks, from intellectual creativity in a free-enterprise system? Whether you view a Boeing 747 jumbo jet with benevolence, malice, or disinterest depends on whether you are a pilot, seaman, or sailmaker—and in what age you live. So it is with the electronic journal and other electronic publications.

ELECTRONIC PUBLISHING

It is common wisdom in the publishing industry that electronic publication,[2] in some form, is the "wave of the future". There certainly is not unanimous agreement on this point; it is a matter of presupposition and not fact. In part because of the success many publishers have experienced in cost savings with computer-controlled photocomposition, they are willing to experiment with computer-delivery systems such as the electronic journal.[3] However, it is only the larger publishers who can afford to do this as it requires hundreds of thousands of dollars and skilled staff knowledgeable about publishing, the subject content of the published material, and computer use and technology to develop these systems. Such experimentation involves a long-term commitment. For example, the American Chemical Society has been working toward the concept of the electronic journal for over 10 years with intensive activity during the past 3. Thus, for publishers involved in such experimentation, there is considerable financial risk. Yet because publishing is a competitive business, the major publishers dare not neglect electronic delivery systems for fear that they may lose their competitive position. Also, publishers have been faced with markedly increased postage costs in recent years and paper costs in past years, so there is considerable interest in a delivery system that circumvents the mails and the use of paper.

OWNERSHIP OF INFORMATION

The issue of ownership is important to publishers as well as librarians. Since very few technical works provide significant financial reward to authors, scientists have not been particularly interested in retaining ownership or copyright. Many publishers, like the ACS, require the transfer of copyright from author to publisher as a condition for publication. Once libraries purchase a book or journal, that resource belongs to them, and libraries are then free to use and disseminate the information that they contain within the limits established by copyright law. Thus, ownership has been reasonably well-defined and shared for traditional, hard-copy publications.

However, what would happen to ownership rights in an electronic journals system? At this point another stakeholder enters, the database vendor. These vendors play a role somewhat analogous to libraries in that they store information, but in electronic form, and make it accessible to users, who may be either librarians, as intermediates, or scientists directly.

Publishers are concerned that they may lose their revenue base by database vendors gaining ownership and control of information. Librarians are concerned that they may lose ownership and therefore the right to lend materials freely to users. Scientists are concerned that information may be restricted to only those who can afford the relatively high cost of on-line searching and retrieval.

Most publishers have been careful in their contractual arrangements with database vendors to explicitly retain their copyrights. Database vendors have offered favorable terms to publishers, who will agree to surrender ownership or make long-term and/or exclusive commitments. It is possible that some of the larger publishers may also become database vendors themselves as in the case of The New York Times and Chemical Abstracts Service. At the present time, it is not at all clear how all this will be resolved, but it is an important issue.

For libraries, the concept of ownership of information centers upon the possession of hard-copy documents, principally books and journals. The hard-copy document is the primary tool with which librarians perform their service mission and represent their basic connection to society. The purchase of large numbers of printed materials also requires libraries to store and organize the material for use. Increasingly, librarians have become aware of their responsibility to preserve this comprehensive record of knowledge for future generations. In addition, libraries are interested in building collection strength in a variety of areas or disciplines.

Electronic publishing of magazines, scholarly journals, or even books that have no hard-copy equivalent would present unique problems for librarians. On the one hand, space requirements to house materials would diminish, and the necessity to classify and index materials would disappear (presumably the publisher would provide that information). Card catalogs or even on-line catalogs could disappear. Libraries, instead of being able to serve patrons with the printed record, would act as message-switching centers, accessing research information from remote databases through electronic communication systems. The pay-as-you-use concept would be in full force, which would markedly affect how library budgets are established, and could dramatically increase costs for libraries and users. The mission to preserve the recorded knowledge for future generations would undergo great change as information in an electronic environment is vulnerable to being lost.

These scenarios are extreme, and publishers are unlikely to stop the publication of hard copy in the next 5–10 years. There are other compelling reasons for publishers to keep producing hard copy. First, it is their principal source of revenue, and no one knows for certain if and how quickly electronic delivery of information will replace traditional formats. Second, it will be a number of years before true graphic material (line drawings, half-tones, colored photographs, etc.) can be handled practically in an on-line environment. The reason is not the inability to reduce graphic material to digital form but rather (a) the large storage capacity required, (b) telecommunication of vast quantities of data, and (c) the lack of graphic-quality terminals. Moreover, there are no standards for graphic material as there are for printed character sets (such as ASCII and EBCDIC). Although the cost for hard copy could increase markedly should the electronic counterpart cause serious erosion of hard-copy sales, most publishers would probably be willing to produce hard copy as long as they could gain revenues.

New electronic journals, for which there is no hard copy, are likely to be published. There are now several such publications: *I. B. Magazette*, *Mentor*, *Microzine*, and *Window*, all of which are oriented toward home-computer enthusiasts.[4] It will be interesting to see if these journals prosper.

ARCHIVAL INFORMATION

There is no argument that information must be archived in some manner; the questions are who should do it and in what form the information should be stored.

Traditionally, libraries have played the role of custodian–archivist. Besides the binding of hard-copy materials, libraries have found that microforms are suitable for archival material. However, the issue of what type of media is used for archiving electronic information will be crucial. Archival information

from an electronic system could be stored in three types of media, or combinations thereof: (a) on paper or microform, (b) in computer-readable form, (c) on videodisc. In the future, other possible types of media are likely to become practical, for example, holograms.

Although information on paper will continue to have value and use, it would seem like a step backward to relegate electronic journals to paper or film for archival purposes. The single greatest virtue of the electronic journal format is one's ability to search rapidly and to recall data interactively and practically instantly. These features would be lost if the archival form were paper or film.

Archiving electronic journals to a computer-readable form, however, implies that at some point it will be impractical to maintain files on-line. With current technology, archival files would be on magnetic tape. Searching of back-files would be done off-line in a batch mode. However, who would maintain the back-files, the computer, and the software to do the retrospective searching? With the high cost of current technology, it would have to be the database vendors or publishers, not libraries. However, it is believed that in the not too distant future, the equipment for doing batch-mode, archival searching may be sufficiently inexpensive that many libraries could hold tape collections and perform the processing in-house.

Videodisc technology offers considerable promise for archiving information. Both digital and analog, i.e., binary coding as on magnetic media or traditional pages or pictures, can be stored, even both forms on the same disk. This technology could permit libraries to maintain archival information and allow their patrons access on site, uncontrolled by the publisher, database vendor, or government. At the same time, if the videodisc devices were merged with a small, inexpensive computer, publishers and/or database vendors could be compensated for use of the information with the computer acting "like a water meter". At present, there are at least two companies that are marketing such videodisc systems. Videodiscs may also be a likely candidate as a mass-storage device for on-line searching.

Not too many years ago, it was practical for a library to physically maintain a large portion of the world's materials in a particular discipline. Now, unless the discipline is extremely narrow and little related to other topics, it is impossible for a library to afford to purchase or store such quantities of information. Interlibrary lending is a partial answer to this situation. Electronic publishing offers the potential for greatly expanding the access to both individual scientists and libraries and to speed up access to information that no longer can be held locally or only obtained slowly through interlibrary loans. In addition, libraries could store information in machine-readable form and dispense it electronically on demand. The American Chemical Society has cooperated with one large library network to investigate this possibility. This network is investigating the feasibility of converting archival journal information (on tape) to a format that could be telecommunicated to a remote library upon request—like an interlibrary loan—and would drive a laser printer to produce the typeset hard copy. The costs for this type of service and the conditions under which publishers would be willing to provide their publications in machine-readable form are yet to be established. The costs are certainly not as prohibitive as one would imagine, and the technology to accomplish this scenario is nearly at hand.

DOWNLOADING

In the context of this topic, downloading is the process of copying information from a database vendor into a local computer, whether into a mainframe or small computer in a library or corporation or into a microcomputer owned by an individual user. No technical difficulties are encountered for an on-line user to copy whole issues or volumes or even complete databases, provided the user is willing to pay the fee to transmit large quantities of data. However, for large volumes of data, this process could be very costly.

The situation is analogous to photocopying of hard copy and, like that issue, of concern to publishers as well as database vendors. Clearly, it is beneficial to scientists to be able to download data into local computers for subsequent processing and retrieval. It is not the use of relatively small quantities of information by scientists that is of concern to publishers but fear that large files will be pirated and used without compensation.

Software is currently available for even most microcomputers to easily accomplish downloading. There are sophisticated software packages that provide for downloading and the management of download files, as well as locally generated files, word processing, etc. BRS offers a software package that operates on several computers under the UNIX operating system. Carlos Cuadra & Associates have developed an even more sophisticated software package called "The Star", which runs on an Alpha Microsystems computer (based on the Motorola 68 000-CPU chip). As the demand for such software increases, other packages will become available, at decreasing prices. To protect their interests and to gain additional revenues, publishers and/or database vendors may very well develop such software themselves, as BRS has already done. Libraries also may develop these software products as the investment can be rather small.

The potential for downloading large quantities of data that would seriously affect publishers or database vendors is rather small. First, database vendors know to whom and how much information is being transmitted because it is a simple process for a database vendor to monitor exactly what is being transmitted. Second, downloaded files can be searched only sequentially, that is, from the beginning to the end of a file, unless considerable data processing is done on the download file. Searching sequentially, as performed by word-processing software, is satisfactory for small files but not for files with millions of characters.[5] Third, searching of large files requires an "inverted" index, which in turn requires precise data-element identification. Each field or paragraph, sentence, and word must be identified and "mapped" so that the location of each word can be pinpointed to its location in the field, in the paragraph, or in the sentence within the field or paragraph and the specific word in the sentence.[6] Except for those identifying the field or paragraph, no other data-element identifiers are available by downloading. To generate these data elements and produce an inverted file would require a considerable financial investment. Such an investment is likely to be made for only two purposes: (a) by a large organization for extensive internal use (large firms, like Caesar's wife, must be above suspicion and are unlikely to embark on such a program); (b) an organization with the intent to resell the information or provide competing services. Many such activities are clearly in violation of copyright law.

CONTROL OF ACCESS

It would seem that whoever has ownership of information controls access to it. That is not completely true; even with complete ownership, access to that information may be controlled by others. For example, there is consternation in Europe because the U.S. government denied access to on-line files residing in the U.S. to French firms who had made contractual agreements to build the Soviet natural gas pipeline. The issue in this case is not whether the government acted appropriately but that governments can and do control access to information. Over the years, there have been debates in various European

countries concerning requirements that the databases that are accessed by their nationals must be physically located in that country.

Taxation and regulation is another mechanism by which governments may control access to information. In the U.S., postal rates for books and journals have been set at a relatively low rate to foster the exchange of information; however, there seems to be a reversal of this policy in recent years. The postage costs for mailing printed material have increased at a much greater rate than those of first-class mail. Similarly, tariff rates for telecommunications are set directly by governments or pseudogovermental agencies in most countries. In the U.S., there is concern that with the breakup of AT&T, rates for telecommunications will increase markedly. One telephone company has already imposed a $50 per month surcharge on telephone lines that use modems. Such policies and practices will have a profound impact on electronic delivery of information.

PAYING FOR INFORMATION

Within institutions, whether they be academic, industrial, or governmental, libraries have traditionally been funded as part of general overhead expenses. Libraries, like general administration and financial operations, are a necessary component of doing business. Until very recently, there has been no precedent for charging patrons of libraries on the basis of the extent they use the library's resources. And for good reason, there has not been a practical way to do so! "Now let's see, that book was purchased in 1978 at a price of $68.23, it's being depreciated over ten years, and you used it for two weeks, that brings the charge to $1.23, when you factor in the libraries overhead. Ah, but you're a student? Well, the rate is different for students. Are you a graduate or undergraduate student? It makes a difference you know...." However, with the advent of on-line bibliographic databases, many libraries, including public libraries, now pass on the direct cost of accessing the information in the electronic systems to the end user. One principal reason that direct costs are passed on is that these costs are easily identified—being billed by the computer system that did the searching. Charging library patrons directly to access electronic information has a restraining effect on the use of the resource. Moreover, librarians are concerned that only those who can pay can access the systems, thus creating an information "elite".

The "cost" and "value" of information are terms frequently used interchangeably, but they are different concepts. In fact, value has nothing at all to do with cost. Cost refers to what funds must be exchanged to have access and use, whether in hard-copy or electronic form. Value connotes relative worth, utility, and importance. Thus, the value of information is a very difficult issue to address. Paradoxically, a particular piece of information may be worth millions of dollars to one organization or individual and completely worthless to another. Information that is of no apparent value today may be priceless tomorrow. Like the family basement, when its contents are to be moved, it is all junk, but to fix a broken toy, or as a source of merchandise for a garage sale—ah, what treasures! The fact that data and information are essential and of priceless value to scholarly activity and the fact that modern society is just as dependent upon information as it is on energy are not seriously questioned. But the issues of cost and value will always be debated.

QUALITY OF INFORMATION

Traditionally, the peer review system has been the principal mechanism for providing quality control in scholarly publishing. It has been the scientist who has been largely responsible for this quality control. In almost all scholarly journals, manuscripts are submitted to one or more of the authors' peers, who are asked to pass judgment on the quality of the science and, to a lesser extent, its presentation. For ACS journals, many manuscripts (a majority for some publications) undergo revision after they have been submitted for publication, principally because of reviewers' comments. Many scientists who publish regularly have, at one time or another, been saved embarrassment by a sharp reviewer.

It has been suggested that in the interest of timeliness and because electronic formats can "be easily corrected", manuscripts should be published without peer review. The GIGO principle (Garbage In, Garbage Out) is a commonly accepted maxim in the computer sciences. There is no need for further verification of this principle. Electronic publication also requires quality control, and the peer review system can be applied equally in this situation.

However, the electronic medium offers another mechanism for quality control that is impossible with hard copy. It is technically feasible to provide readers with an "electronic note pad" so that they could make comments on specific articles, which then could be made available to others on-line. However, there are a number of policy issues that then must be addressed: Should the identity of commentators be made known and should all comments—no matter how scurrilous—be included? Who should decide what comments to exclude and on what basis, etc.? This mechanism has already been used in several computer network systems, but not for electronic publishing per se.

Electronic information, in magnetic form, is not permanent; such computer files are indeed mutable. If a file is on a single system, it can be altered and past history totally obliterated. In order to preserve the integrity of the publication system, it is important that at the time of acceptance for publication the "text be frozen" in the same manner as is currently done for hard-copy publication. This mechanism allows for corrections of, and additions to, or withdrawal of a manuscript. But once publication is accomplished, whether printed on paper or mounted on disk, that is it!

Electronic publishing also offers the potential for reducing the amount of time required for authors to submit articles and for the peer review process to take place. In the not too distant future, it may be practical for many authors to submit their manuscripts in machine-readable form rather than on paper. Such manuscripts could be transmitted via telecommunications or on some magnetic medium (tape or diskette). Eventually, it may be practical for an editor to send the manuscript to the reviewer by telecommunications, or permit the reviewer to access it electronically, and thus save considerable time in the evaluation process. However, there are many problems with these scenarios, quality control being just one of them.

INFORMATION RETRIEVAL

There is a question of who should do on-line searching, and at what location. At the present time, it is common practice for library staff to do the searching—having received a request from a scientist to do so. In general, it has not been practical for most scientists to search bibliographic databases in a productive, cost-effective manner because they do not use on-line systems on a sufficiently regular basis to maintain searching skills, nor do they have an adequate understanding of indexing policies for the databases—and each database has a different indexing policy. Also, until quite recently, there were few terminals in laboratories and offices that were not dedicated solely to local computer and/or word-processing equipment. Libraries or information centers were often the only source of terminals and modems that could be used to access external computer systems. All this is changing very rapidly.

First, software systems are becoming more "user friendly", that is, easier for nonspecialists or occasional users to use in an effective manner. Scientists are quickly becoming more knowledgeable about computers; this is particularly true of chemists and other physical scientists. The trend toward greater computer literacy is most apparent among high-school and undergraduate students. Many of these individuals will be the scientists of the future, and they will not be inhibited or intimidated in searching on-line files. Second, with full-text files, there is no need for an indexing policy as the full text itself is the index. Every word in the text, with the exception of common prepositions and conjunctions, is indexed. Third, terminals, modems, stand-alone microcomputers, and telecommunication facilities on word-processors, mainframes, and minicomputers are all becoming more common in laboratories and offices.

Finally, there is the issue of control, or to express it pejoratively, "politics". If a library or information center is charged with the responsibility for conducting on-line searches, other information retrieval activities, maintenance of collections, etc., it is certainly reasonable that library staff should desire the necessary control and authority to accomplish their responsibilities. But will the demand for library services and the expertise of library staff decrease? Of course no one really knows, but probably the opposite will happen. As end users become experienced in using electronic media, they will have more questions and problems, and they will require the assistance of specialists to perform complex retrieval tasks. Students and scientists must be trained to use electronic resources. One does not leap from the womb, fingers poised over a keyboard, counting in octal or hexadecimal, with notions of Boolean operators coursing through the brain. Librarians have a tremendous opportunity to educate end users.

EDUCATION

The education of scientists and librarians, in all aspects of electronic information, is crucial to the acceptance and utility of electronic systems. This includes, but is not limited to, on-line searching of bibliographic, full-text, and numeric files.

During the past 2 years, when several hundred librarians and chemists had the opportunity to access the ACS' experimental, full-text file, it was obvious that most experienced on-line users searched the full-text file in the same manner as they would search a bibliographic file. Although the structure of the full-text file requires the use of the same Boolean logic used in searching a bibliographic file, the manner in which the Boolean operators are used and the handling of retrieved information are quite different with full-text files. Chemists, few of whom had any previous on-line experience, did not fall into this trap and found searching the file quite straightforward. While experience in searching one type of file is helpful in learning to use other types of files, clearly training and education are necessary.

CHANGES TO COME

Clearly, it is the library that will probably undergo the greatest change as the electronic journal and other forms of electronic publication become accepted. The library's traditions of ownership of information, archival responsibility, interlibrary lending, training and educational responsibilities, etc. have already been discussed. Librarians soon will realize that unless they adapt to these new technologies, add value to information, and/or provide worthwhile services, that they and their institutions will be circumvented. Ships that are propelled by the wind are no longer used in commerce; they are used for pleasure. Libraries that do not change may well become the information yachts of tomorrow.

Publishers too must educate themselves and appropriately alter the manner in which they conduct their affairs, not only in how to efficiently produce traditional hard copy and at the same time its electronic counterpart but also be sensitive to users' needs and sufficiently flexible to make changes as circumstances indicate. For example, the electronic display format of information is patterned entirely after the format used in printed material; there is little reason to believe that this is the best format for electronic display. However, it will take more experience before a different form of electronic display can be expected to evolve, but we can be sure that this evolution will take place.

The principal beneficiaries of electronic publishing are likely to be scientists, the end users who need information. But scientists too will have to change the manner in which they seek and use information, changes that are under way now. Scientists, in rapidly increasing numbers, are using computers to generate data, to produce reports and manuscripts, and now to locate information within their own organizations and externally. Both publishers and libraries *must* not forget that it is the transmission of information at an affordable cost to the end user that is the single most important consideration.

REFERENCES AND NOTES

(1) The term scientist is used here in a general sense. "End user" is commonly used in this context, meaning scientist, engineer, physician, etc.
(2) Electronic publishing is the production of published materials (newspapers, magazines, books, journals, etc.) in which computers are employed as a major production tool. Composition, if required, is accomplished with computer-controlled photocomposition. Electronic publishing employs complex word-processing-like systems whereas traditional publishing (hot metal, for example, Linotype) uses typewriter-like, mechanical devices. In electronic publishing, the information that is being processed is stored in digital form and is processed electronically rather than mechanically.
(3) An electronic journal is a publication (journal) that is available in digital form. Currently, most electronic publications are available via on-line database vendors (BRS, DIALOG, SDC, STN International, and others) who provide access to these publications via phone lines (telecommunications) for users employing computer terminals and modems. Electronic publications could also be distributed on magnetic media (magnetic tape, tape cartridges, diskettes, or digital videodiscs in the future) for processing and use with on-site computers. At present, most electronic journals have a traditional, printed counterpart.
(4) For more details, see *InfoWorld* **1983**, *5* (48), 75.
(5) This is not completely true. Sequential, parallel searching can be done by using multiple computers, but such systems are very expensive and not commonly available.
(6) In *Lexis* and *Nexis*, Mead does not require such a precise data-element identification as searching is done on the basis that words must be within "*n*" words of one another. Thus, paragraphs, sentences, and position of words within sentences are not required. However, a degree of precision in searching is lost in this scheme, and there are other disadvantages.