Communication or Chaos?

Effective transfer of scientific and technical information continues to be a pressing national problem.

Dale B. Baker

In writing or speaking at the beginning of a new decade, it is fashionable to speculate about what the next 10 years may hold in store in one's field of endeavor. In peering into the future of scientific and technical communication, it might be possible to conjure visions, if not of sugarplums themselves, then of televiewers, megabit storage devices, and instantaneous recall systems that constitute the sugarplums of this field. Or, it might be just as easy to foresee a real potential for chaos in scientific and technical information transfer and predict, as some already have done, the partial or total breakdown of our formal information transfer channels.

But the field of scientific and technical communication has perhaps too long and too often been characterized by glowing visions of future possibilities on the one hand and dark foretellings of doom on the other, with little concrete action toward solving the immediate and urgent problems at hand. So, rather than speculate about the future, I intend to review some of the progress made in the past decade toward solving what the recent report of the National Academy of Sciences has called "a pressing national problem" (*1*).

My comments have two theses. First is that information processors in many scientific disciplines made significant technological advances in the decade of the 1960's, though perhaps too few scientists have recognized these advances and their potential, and that this technological progress gives us a solid base from which to move rapidly ahead in the 1970's. My second thesis is that despite the progress we are making in modernizing the individual discipline-oriented and mission-oriented information systems in science and technology, we perhaps have not yet recognized that a truly effective solution to our scientific and technical communiation problems can come only through a coordinated, multinational, multidisciplinary effort.

The author is director of Chemical Abstracts Service, Ohio State University, Columbus 43210. This article is adapted from an address to the AAAS Section on Information and Communication on 29 December 1969 in Boston, Massachusetts.

A Mythical Explosion

At the beginning of the 1960's we were being bombarded by outcries from many directions about an information explosion in science. In fact, occasional echoes are still being heard today. But as we should all be aware by now, the information explosion was a myth. Scientific information has been growing at pretty much the same, predictable, exponential rate for several centuries.

Nevertheless, the sheer volume of scientific information that confronts us today—some 2 million new reports of research and development per year—and the rate at which our store of useful scientific information is now growing are placing a severe strain on our traditional means of communication in science and technology. And so, we stand on the threshold of the eighth decade of the 20th century wondering if, somehow, computer technology can do now for scientific and technical communication what the advent of abstracting and indexing services did more than a century ago.

One hears a wide range of views expressed on this question. At one extreme are those who assure us that we will not achieve Nirvana until we are all on-line with a reactive, zero-defect, negative-cycle-time system that calls heuristically on its many banks of trillion-bit memories to tell us the answers to the questions we should have asked instead of those we did ask. At the other extreme are some who resist any change in the traditional system and curse the machines that are increasingly encroaching upon every aspect of their lives.

In the midst of this turmoil, a number of scientific information processors have been moving steadily ahead on a middle course, trying to preserve what is best in the old ways and using what is feasible and practical in the new technology to provide better information services for scientists and engineers. I suppose that, in the catch phrase of the day, we might term ourselves the "silent majority" of the information transfer field.

Perhaps we have been too silent, for while at the end of the 1960's more scientists and engineers were getting more information in a more effective manner than ever before, many still were unaware of, or unwilling to recognize, the array of new information tools now becoming available to them—from printed computer-produced alerting services to current-awareness information selected by computer according to individual interest profiles. Information processors are now a different factor in science and technology than they were a decade ago. The scientist, engineer, or research manager who continues to assign to us only the quiet, scholarly corner of his mind—the lonely trip to the library before the project starts—is as surely hobbling himself and denying his responsibilities as if he threw out all of his modern analytical equipment in favor of a return to the classical laboratory techniques of the turn of the century.

Modernizing the Systems

The growing volume of new scientific and technical information and the changing information needs of scientists and engineers have had the greatest impact on the abstracting and indexing services, and, in the past decade, a number of these services have made significant progress in applying computer technology to information handling on a large scale. At Chemical Abstracts Service, for example, we are now handling nearly 1 billion characters of information per year in a computer processing system.

At CAS we began working with computers in 1960 and, with financial assistance from the National Science

Foundation, began seriously to convert all of our information-handling operations to a computer base in 1964. The CAS computer processing system is built on the premise that sound intellectual input is essential to optimum secondary information services. As in the past, we continue to depend upon scientifically competent analysts to summarize, condense, and index the primary literature. But now the results of this intellectual analysis are recorded in a unified machine-language information store, and essentially all subsequent manipulation, organization, and checking is performed by a computer system that produces output in the form of both computer-composed text ready for offset printing and computer-searchable magnetic tape files.

The new system not only allows us to conserve our limited professional manpower and gives us the capacity to keep pace with the growth of the chemical literature through the next decade, it also enables us to produce more accurate and timely services and to make digested and indexed information available to scientists and engineers in a variety of useful forms—to derive multiple benefits from the investment of intellectual effort. Thus CAS is no longer simply the publisher of an abstracting and indexing journal but a man-machine information-processing system that produces publications as one form of output.

The publication is by no means downgraded in importance; it is merely placed in proper perspective as one of several important media for disseminating information today. Nor have we simply adopted the computer as a high-speed printing press. Rather, we have restudied in complete detail how to provide the most effective access to the primary record of science and technology.

Some of our more visionary colleagues might tend to miss this point and accuse us of responding to computer technology the way the shipbuilders of the last century responded to the steam engine. The shipbuilders put engines on sailing vessels, but kept the sails. We are using the computer to produce traditional printed publications as well as newer forms of information service.

There are those who might suggest that we are perhaps wasting our efforts in developing a system that produces printed output, that modern information technology will soon make printed abstracts and indexes obsolete. We cannot accept that line of reasoning, at least for the next decade or so. Only a minority of the world's scientists at present have access to the necessary equipment or have had the opportunity to develop the necessary experience to use even the more rudimentary computer-based information tools available today. We hope that this situation can be remedied to a considerable extent in the next several years, but even so, the printed page, or a microform that requires many of the same production steps, is likely to remain the principal information tool for many of the world's scientists for some years to come.

If you were to ask me when our conversion to computer processing would be complete, I would have to answer, "probably never," for we and others are constantly evolving new and better ways to do the job. Some of the first subsystems to go into operation have been redesigned several times to achieve better overall efficiency and to take advantage of newer and more powerful equipment. Our rate of progress has also been restrained somewhat by the need to maintain a continuing flow of information through our operations while implementing the new system. We could not shut down for a year, or even a month, install new equipment and procedures, then start up again. Nor was it economically feasible to build the new system completely in parallel with the old, while the old system continued intact and operative. Our problem is somewhat like that of building a new bridge on the site of an old one without diverting or obstructing a constant and heavy load of traffic.

But even though our system is still some years away from operating in an ideal state, much of it is functioning in a very useful manner now. This year, the complete bibliographic citations and subject indexing terms for some 270,000 papers, patents, and reports will be recorded in the machine-language data base, as will be the full text of the abstracts for about 25 percent of these documents and the structure diagrams, names, and molecular formulas for approximately three-quarters of a million chemical substances. Our chemical registry subsystem, which is bringing together in one data bank information on the structures, names, and sources of published information for chemical substances, contained machine-searchable structural records for some 1.3 million substances, along with almost 3 million associated references, at the end of 1969.

Services now produced through the CAS computer system include *CA Condensates*, weekly machine-searchable tapes containing the complete bibliographic citations and keyword index terms for each journal article, report, and patent covered in the corresponding weekly issue of *Chemical Abstracts*; *Chemical Titles*, a biweekly alert to the titles of papers appearing in some 600 journals of pure and applied chemistry; *Chemical-Biological Activities*, a biweekly abstracting and indexing service covering publications in pharmaceutical chemistry and biochemistry; and *Polymer Science & Technology*, a weekly guide to the journal, report, and patent literature on the chemistry and technology of polymers. *Condensates* is produced in machine-searchable form only. The other three are produced simultaneously in printed and machine-searchable magnetic tape editions.

In addition, the author and molecular formula indexes of *Chemical Abstracts* are now organized and composed through the computer system, and the subject indexes will be so produced before the end of 1970. We will convert the weekly abstract issues of *Chemical Abstracts* to computer production over the next several years.

Thus, at the end of the 1960's we were well on the way toward a comprehensive computer-based system for processing and disseminating secondary chemical information. Processors of biological, engineering, and medical information have made similar progress, and others are moving in the same direction. This technological progress gives us a solid base from which to move ahead rapidly in the 1970's.

But technology alone will not bring us communication rather than chaos in the decade of the 70's. Despite the progress we are making toward modernizing and improving the efficiency of individual information systems, our overall system for recording, locating, and disseminating scientific and technical information remains highly redundant, decidedly uneconomic to operate, and, to a growing extent, unresponsive to the needs of scientists and engineers today. A truly effective solution of the overall scientific and technical communication problem can come only through a coordinated effort across national and disciplinary boundaries.

International Cooperation

Obviously, scientific and technical information is an international as well as a national resource. At Chemical Abstracts Service some 73 percent of the information we process originates outside the United States. Secondary chemical information services in the Soviet Union, Germany, Japan, and France are attempting to keep pace with major parts of the same body of literature. All are facing the same economic and manpower problems as CAS, and all are considering the same alternatives. We need to find ways to share the burden to the mutual benefit of all.

We are beginning to make some progress in this direction. In 1969, the United Kingdom Consortium on Chemical Information, an organization representing ten chemically oriented scientific and scholarly societies in the United Kingdom, and West Germany's chemical society, the Gesellschaft Deutscher Chemiker, agreed to share with the American Chemical Society in the development and operation of a common computerized chemical information system. The British and German organizations will assume responsibility for selecting, organizing, and recording information from an agreed-upon portion of the primary chemical literature for input to a central computer-processing system at CAS's Columbus headquarters, and will share in the output from this system.

The agreement with the German group in effect unites the operations of CAS and Chemisches Zentralblatt, the world's oldest scientific abstracting and indexing service. Chemisches Zentralblatt ceased publishing its own comprehensive abstracting and indexing journal at the end of 1969, and input to the common system from Germany will be prepared by members of the West German editorial staff of CZ, now the Information and Documentation Division of Gesellschaft Deutscher Chemiker.

Scientific communities in several other nations have expressed interest in exploring similar forms of cooperation. While no formal agreements are likely to follow immediately the outlook for further cooperation is encouraging.

Over the next few years, the British and German groups will develop the capability of transmitting abstracts and index entries to the Columbus processing center directly in machine-readable form, and the degree of refinement of data that can be performed on such a decentralized basis will be determined in a series of carefully controlled experiments. Ultimately, as much of the intellectual processing effort as possible will be performed by these groups.

Another important function of the British and German organizations under these agreements will be the exploitation of the machine-readable output from the central processing system. They will manipulate this output, separately or in combination with data from other sources, to develop specialized publications and services of their own and to provide computer searches of the data for individuals and organizations in their respective nations.

This approach to providing service to the ultimate user from the computer data store seems to offer the most promising means of bringing the benefits of computer-based information to greater numbers of scientists and engineers. While the use of computers for computational and numerical data-processing applications is almost commonplace today, only a limited number of organizations have yet developed the necessary expertise to manipulate complex machine-readable files of nonnumerical data to extract useful services from them. The major scientific information processors have neither the manpower nor the financial resources to develop by themselves more than a fraction of the potentially useful services that could be derived from these growing stores of machine-readable data, nor do they operate in a context of sufficiently close contact with the ultimate users of the information to provide the most responsive form of service to these users. Decentralized information centers staffed by personnel versed in both the subject content of the data base and the computer hardware and software involved, each working closely with a relatively small group of users, can provide the most economical and responsive service to these users. Through their close contact with users, these centers can also provide feedback to the information processors that can help make these processors' systems more responsive to user needs.

Since late 1967, the Organization for Economic Cooperation and Development has been encouraging cooperation between CAS and groups in its member nations in the development of such information distribution centers. So far, centers have been established in Belgium, Canada, Denmark, the Netherlands, and Sweden under this program. Six organizations in the United States, including one commercial organization, also are now offering public searches of CAS machine-readable data files.

These cooperative ventures with sister societies abroad and the experiments with independent information centers in the U.S. and abroad represent the first tentative steps toward the development of a coordinated, international network for chemical information, a network that will provide both decentralized input to a single processing system and decentralized exploitation of output from this system on behalf of the user. Because of the staff, equipment, and management requirements, the number of national centers participating directly in input to the common processing system will necessarily be limited to four or five in the world, including the one operated in Columbus. The number of information distribution centers sharing in output from the system would be limited only by the centers' ability to maintain financial viability.

Interdisciplinary Coordination

It also should be obvious to all by now that research and development in the physical and natural sciences and technology are highly interdisciplinary today and growing ever more so. As research in chemistry, physics, and biology probes to more fundamental levels of understanding, the boundaries between these disciplines are becoming less and less distinct. Similarly, research in other scientific fields and much technological development are more and more drawing upon and contributing to the information store of these fundamental disciplines.

Today, almost every report of progress published in the primary scientific and technical literature contains information of use to the practitioners of several disciplines. To provide reasonably complete coverage of its discipline, in recent years each discipline-oriented secondary information service has been covering more and more reports in which the majority of the information is of concern to another discipline. The result has been increasing overlap of coverage among these services that has caused them to grow at a rate of about 9 percent per year, while the primary literature of science as a whole has been increasing at an average annual rate of only 6 percent.

While the techniques for measuring duplication of coverage do not always yield comparable results, and exact figures are subject to question, a substantial percentage of papers covered by *Physics Abstracts* and *Biological Abstracts* today are also covered by *Chemical Abstracts*. Still greater overlap exists between these discipline-oriented services and the various mission-oriented abstracting and indexing services that have been started in recent decades.

Historically, overlap in coverage among secondary services in printed form has often been necessary to assure that scientists in a particular discipline have access to important information. While the abstracts of a given document used by two different discipline-oriented secondary services generally do not differ substantially, each service tends to index the document from the viewpoint of and in the specialized terminology of its own discipline. For this reason, it has often been difficult for a scientist trained in one discipline to locate information through the indexes of a secondary service oriented toward another discipline, and, in the absence of any strong incentive for coordination of indexing practices, overlap of coverage has been necessary.

Now, computer technology provides an effective and relatively inexpensive means of interlinking the information stores of the various disciplines by creating bridges between the indexing nomenclature used by these disciplines. An example of this type of interdisciplinary networking is the mechanized interface that has been developed between CAS and the National Library of Medicine's *Index Medicus* and MEDLARS data base. A substantial amount of information in the chemical-biomedical borderline area, most of it concerned with chemical substances, is being indexed by both CAS and NLM. Over the past several years, the two organizations have worked together to develop a bridge that links the terms used to index a substance in *Index Medicus* and the *Chemical Abstracts* indexing nomenclature for the substance through a computer-based structure recognition system. Through this link, it is possible to convert automatically from one indexing system to another and to search for references to a given substance in both the CAS and NLM data bases, even though each service uses a different system of nomenclature in indexing substances. While this is only a small step toward interdisciplinary coordination, it is representative of the kind of interface between discipline-oriented services that computer technology makes possible.

With the current volume and continuing growth of publication in each discipline and the increasing integration of the disciplines, no one information processor can reasonably expect to satisfy the full range of information needs of the scientists in a given discipline, nor can the scientific community afford to support highly redundant information systems for each of the scientific and technical disciplines. There is an urgent need for an agreed-upon division of coverage among the various discipline-oriented and mission-oriented information processors, and this division must necessarily be accompanied by a coordination of indexing and data recording practices among these processors, so that a scientist can readily shift from one information store to another.

Furthermore, as the emphasis on cross-disciplinary research grows, as it undoubtedly will, it is likely that the information needs of many scientists will be best satisfied by combining selected machine-readable data from several discipline-oriented or mission-oriented processing systems. Such combination of data can be performed very efficiently if the discipline-oriented systems are compatible, and such compatibility can be achieved at a minimum of expense if coordination and standardization of data recording practices begin while each system is in a relatively early stage of development.

Coordinated application of computer technology throughout the scientific information-handling community would make it possible to interlink the various discipline-oriented and mission-oriented, government and private, computer processing systems now under development into an effective information network, substantially reduce the overall cost of operating our scientific and technical information complex by eliminating much of the duplication of effort that now exists, and result in far more useful information services for scientists and engineers in all disciplines. This is not something that can be accomplished in a few years or, perhaps, even in a decade. But the necessary technology exists. We need only the will and the means to cooperate, and we need to start soon.

My review of the 1960's convinces me that we are making progress, though perhaps not as rapidly as some might like. Scientific information processors who have recognized that traditional methods are no longer adequate for the task to be performed have set out to modernize their systems, and have accomplished a great deal in a very practical way.

Most of my comments have concerned secondary information services, for that is where the most progress has been made in the 1960's. But primary journal publishers also are now modernizing their operations through the installation of computerized typesetting procedures. This opens up the possibility of highly efficient and reliable machine transfer of data between compatible primary and secondary processing systems with attendant elimination of much of the redundant intellectual effort that now exists between the primary and secondary publication operations. We must not forget that secondary information systems exist mainly to lead the seeker to information recorded in the primary literature, and that most users of scientific information depend upon the libraries to supply the primary documents they need. Our modern access systems will be of little value without effective links to the library system.

We also have a tremendous educational task to perform. Scientists, both young and old, are generally unaware of the power of the new information tools now becoming available, nor are they being taught to use them effectively. This is a task the information processors cannot undertake alone. We need the help of the colleges and universities in promoting and teaching the use of modern information methods.

But the greatest challenge facing us in the 1970's is the coordination of the individual information systems into an effective information network for science and technology. The integration of scientific and technical information services is a natural and necessary consequence of the integration of science itself. If we recognize this and begin to work together now, we may have the prospect for more communication and less chaos in the sciences in the decades ahead.

Reference

1. "Scientific and technical communication. A pressing national problem and recommendations for its solution," report of the Committee on Scientific and Technical Communication of the National Academy of Sciences-National Academy of Engineering (Washington, D.C., 1969).