The Status of Chemical Information

This report, a summary of the status of chemical information processing, is the latest in the series prepared by the National Academy of Sciences–National Research Council Committee on Chemical Information. The period covered is 1969 to the present. The user, technological developments, publications, services, the Federal government, academia, and industry are the topics reviewed as related to chemical information. In addition to present status, trends are evaluated, problems are stated, and recommendations for action by appropriate bodies are included.

BART E. HOLM,* MARY GERTRUDE HOWELL, H. EDWARD KENNEDY, JOSEPH H. KUNEY, and JAMES E. RUSH
For the Committee on Chemical Information National Academy of Sciences–National Research Council Washington, D. C. 20418

Received August 14, 1973

The National Academy of Sciences–National Research Council, Division of Chemistry and Chemical Technology has had for many years a Committee on Chemical Information. The Committee is concerned with chemical information needs, sources, usage, and methods. It acts as a clearinghouse for assembling, analyzing, and disseminating knowledge about new developments. The Committee conceives, organizes, and coordinates special meetings on important topics requiring presentation and discussion, and proposes, analyzes, prepares, and publishes comprehensive studies in areas needing detailed knowledge as a basis for recommending courses of action. The Committee provides advice on request on problems involving the special technical competence of its members.

During the past 10 years, three notable summaries on the status of chemical information processing and services were issued.[1,2,3] The present report has the objective of continuing that precedent.

In this report, covering the status of chemical information processing, trends are evaluated and present problems are stated. Recommendations for action by appropriate bodies are included throughout. The objective has been to identify the critical issues and to encourage action by responsible individuals in organizations.

Input to this report has come from a number of sources. The primary resource has been the background and experience of Committee members. In addition, appropriate literature sources have been checked. However, this report is not a literature review nor a critical review, but rather an analysis of the key issues as judged by the authors.

Another important input resulted from the contact made with nearly 60 information managers and specialists who were asked for their views of the status and future of chemical information processing. Their thoughts are included, often as a consensus of the majority, and also as singular cogent observations by key individuals. The thanks of the authors go to them but interpretation of the material remains our responsibility.

Our purpose will have been met if this paper affords stimulation for discussion and appropriate action.

THE WORLD OF THE USER

The basic information needs of chemists have changed little in recent years. Chemists need ways to keep up to date in fields of perceived interest and ways to assist in finding answers to questions. Their approaches to meeting these requirements have not changed drastically but rather have evolved as new services are offered. Chemists have not clamored for the new capabilities. Some have not changed their information gathering habits and may always use their established patterns. An increasing number appear to be taking advantage of the new products offered by the suppliers.

While the chemist's basic information needs remain essentially unchanged, his approaches to meeting those needs are slowly evolving. This evolution is the result both of advances and availability of computer technology for information storage and retrieval, and is particularly the result of increasingly dynamic information processing. Other important factors contributing to changes in approach are: increasing demands on a user's time, growing quantities of literature to assimilate, and costs. These factors all contribute to the evolution and adoption of new approaches.

Ideally, new information services are designed to fulfill the needs of the user, but this design must be a delicate blend of understanding of the user's knowledge of his need and of information science technology. In recent years, technological developments in information processing have exceeded the chemist's ability to use them to advantage. Technical capability on the part of information processors is not necessarily ahead of the needs of the user but rather is ahead of his awareness and knowledge of those needs. Thus information services are often provided based on concepts which the user may not perceive as being of value to him.

Technological advances have not met all user needs. For example, automated information systems in operation today are primarily reference systems. We are getting the user to the documents but not to the information or relevant data contained in those documents.

Many organizations throughout the world are engaged in permanent recording and dissemination of data of all kinds for subsequent use by an audience whose specific needs are largely unknown. The user who acquires such data must then seek a means of organizing and using the retrieved material. The user can easily become inundated with undigested, unevaluated data, in larger quantities, much more rapidly than was ever the case before. As Edward David, former Science Advisor to the President puts it, "It's very easy to collect data and put it into a form in which it can be read or recalled, but that is much easier than saying what the data means or how the data should be used in the decision process."[4] The problem is not the ability to find data. The problem is to find the right kind of data, the right amount of data, at the right time.

* Secretary's Department, E. I. du Pont de Nemours & Co., Wilmington, Del. 19898, to whom correspondence concerning this publication should be addressed. The study was supported in part by a grant to the National Academy of Sciences–National Research Council from the American Chemical Society.

In the 1960's, the National Standard Reference Data System was established within the National Bureau of Standards to develop programs for compilation and evaluation of quantitative data. One of the consequences of this program was the introduction of the *Journal of Physical and Chemical Reference Data.*[4a]

The need for data has a parallel in the processing of chemical structures. During the 50's and 60's, a great deal of attention was paid to design of representations which were suitable for computer manipulation of chemical structures and work continues in this area.[5,6] However, it is often not the structure which is of interest to the chemist but the properties of the substance which the structure represents. For instance, one may be interested in the synthesis of the compound, in some of its physical properties, in its behavior in living systems, or in the marketing of the compound. The structure itself is only of interest because it provides the common thread among all the attributes and applications of the compound it represents. Certainly it is necessary to provide for structure representations which are suitable for computer manipulation, but one wonders whether an inordinate amount of attention has been paid to this detail without paying enough attention to the collection, evaluation, and correlation of data about the compound, data which may be more important than the structure itself. More attention must be directed to the acquisition and verification of data, and to the design of systems which disseminate such data and provide the user with the means for data correlation.

At all levels of decision making, in the production, dissemination, and use of information, costs and prices have become a major influencing factor. While economics is an important factor in problems of information transfer, critical information needs should not be totally subservient to economic conditions. A key to improved transfer is the best possible understanding of user needs and attitudes toward information products and services. But, one of the results of the current economic climate has been the almost total elimination of user studies except those that seek to measure willingness to pay for a particular product or service.

In addition, controllers of budgets force users to determine needs in terms of dollars available and not in terms of requirements. Potential users evaluate new products and services based on the cost savings produced. Those who undertake the risks of providing information services seek a high probability of adequate cost recovery before offering new services or continuing existing ones. Any planning for needed research and development finds funding scarce particularly at the Federal level.

Despite the impact of reduced funding for information services and for the development of new systems, key elements in the over-all system have remained unchanged. While there has been a drop from the growth rates of the late 60's, papers submitted to primary journals in chemistry are showing over-all increases of 1 to 2% a year. And growth rates of 8 to 9% a year still occur for the secondary services serving the chemical and related fields. Thus the factors affecting the generation of information seem to be working to increase the flow.

Within this flow, the willingness to pay the cost of information services is an important measure of information need. But there is some danger in leaving support decisions dependent on the economic resources of the ultimate user. Sanford Berg[7] has confirmed mathematically what every publisher has learned through experience: the individual is much more sensitive to price increases than is an institution. Thus the need for price increases has been a factor in pricing individuals out of the market for services and in moving producers of information services toward serving the needs of institutions rather than individuals. Whether part of the cause or the effect, the corollary action is the trend of individuals to become more dependent on various "middleman" services to get the information they want and need.

In turn, the shortage of funds has affected the acquisition policies of institutions. Budgets for information services have tended to hold level. Thus, institutions are not holding their spending for information services at a rate to match the pressures of inflation. We see the results of these acquisition policies in the cutbacks in purchases of books, journals, and the whole variety of specialized primary and secondary services that have been offered in recent years. A further consequence of these policies is the reluctance to experiment with newly offered services, particularly computer based services, to determine their utility and practicality.

As part of the effort to develop and make available new information services and products, we have not developed, on the part of users and those who must pay the bills, a sense of the value of these services and products. One important step in this direction is the education of users. The education process must go beyond the demonstration phase, and also must involve itself in the changing of attitudes and show awareness of the patterns of human behavior. Adequate testing under operating conditions should precede any large-scale educational effort. Only out of such experience will we develop the basics for educating the users of new information services and products.

DEVELOPMENTS IN THE TECHNOLOGY OF CHEMICAL INFORMATION PROCESSING

Background. The computer has been of major assistance in the progress which chemical information processing has achieved during the past two decades. During the 1960's, the discipline of chemistry carried on a major share of the research and development work in information processing. Significant developments in the representation and manipulation of chemical structure were reported during this time.[1,2,3,8] Methods for representing chemical information and the hardware and software for manipulating the information are reported in this section. The work cited represents some of the significant developments in technology since the previous report, "Chemical Structure Information—A Review of the Literature 1962–1968."[3]

During recent years, progress has been modest, with concentration on application rather than development of completely new methods of representation or manipulation.

Methods for representing structural information have included nomenclature, fragmentation schemes, linear notation systems, and tabular and graphic representations. No major developments have occurred recently in fragmentation techniques.

Nomenclature. Significant changes have been made in the past three or four years at the Chemical Abstracts Service in modernizing chemical nomenclature.[9] While the impact of these changes has not been assessed, they represent a determined effort to improve the indexing role of chemical nomenclature. Some work in computer generation of nomenclature has been reported.[10]

Linear Notations. Linear notation systems continue to enjoy a significant place in structural information representation. The most widely used is the Wiswesser Line Notation (WLN). Emphasis has been on extending the areas handled by WLN. The basic reference[11] has been extended by a series of publications offered through the Chemical Notation Association. This association, with representation from 79 organizations, is the official body for standardizing and controlling the rules of the notation and for furthering its development. There is extensive literature on the application of the WLN and its integration into chemical information systems. (See, for example references *12–15.*)

The work by Skolnik on the notation system at Hercules has been extended.[16–18] No major additions have been announced for the other notation systems described in the previous reference.[3]

Tabular and Graphic Representations. The Chemical Abstracts Service continues to include connection tables as the record for its Chemical Registry System. The next version, Registry III, is being designed. Perhaps a dozen organizations including Du Pont, Eastman Kodak, Food and Drug Administration, National Cancer Institute, and the Walter Reed Army Institute of Research use a similar approach.

Numerous references have appeared in the European Literature on the DARC System (*D*escription and *A*utomated *R*esearch of *C*orrelation) developed by Dubois. The system provides for the description and automation of relationships between the topology of a compound and its properties.[19]

An interesting method for the representation of polymer information is TOSAR (Topological Method for the Representation of Synthetic and Analytical Relations of Concepts) developed by Fugmann. The TOSAR System uses a flow diagram of the processes involved in the preparation of polymers with lines representing relations between nodes that represent materials and processes. The TOSAR System is being used for the documentation of polymer patent literature by IDC (International Documentation for Chemistry) in Frankfurt, West Germany for its participating companies.[20–22] TOSAR has also been proposed for a graphical representation of chemical processes for storage and search.

Interconversion of Structure Representations. Work was reported in the previous reference[3] on conversion of WLN to connection tables. Computer programs have been developed at the National Institutes of Health (NIH) to produce canonical WLN for monocyclic, benzene-containing, or acyclic compounds. The user can draw the structure on a Rand Tablet or other writing surface. From the input, connection tables are produced. From rearranged connection tables, the WLN is derived. The programs can also develop WLN from tapes containing connection tables from other sources. The system is designed for terminal input. The user draws the structure and the generated WLN's are typed on the user's terminal.[23] The work at NIH has been extended to include test files for experimentation with graphic input of structural search requests on-line with graphic presentation of structures as responses. The system can also accommodate WLN at input and at output.[24–26]

Indexing. One of the purported advantages of the WLN has been its readability in output listings. Grouping and readability have been improved by further manipulation of KWIC indexes to provide double-KWIC indexes for line notations.[27]

The Multiterm System has been proposed as an indexing system for communicating the information content of documents by coordination of terms in defined directional order. Titles and abstracts are reduced to a concise form which, in the classic case, contains combinations of subjects such as a chemical, prepared from a reactant, by a process, using a catalyst. The terms are laid out in precise order using special symbols for delimiting terms. More complex concepts reportedly can be described following the stated rules. The multiterms can be permuted with the information and order maintained. The method has been proposed for use as a printed index and for a multilingual index.[28,29]

The work on the Mechanical Chemical Code (MCC) was continued at the University of Pennsylvania. Earlier references describe the development of the MCC as a screen system, generated by computer, useful for substructure search. New work includes the use of the MCC to prepare two printed indexes to perform structure search. The first is a KWIC index of the machine generated fragments. The second is an inverted index for search based on logical combinations of the fragments.[30]

The MCC has been applied to a file of pesticides at Esso Research and Engineering Company for manual search of printed indexes.[31] The largest application of the MCC consisted in taking 1.25 million compounds from the Chemical Abstracts Registry System and generating screens using the MCC symbols and preparing indexes for machine or selected manual search. The printed indexes were prepared using computer output microfilm.[32]

It is now possible for an organization to select from a variety of input methods and produce any of a number of forms of output. WLN or other notations can be generated by chemists, stored, and searched. These can be expanded to connection tables if desired. Alternatively, structures can be typed or drawn on graphic input devices for batch or on-line storage. The representations can be converted to connection tables or linear notations. Output can be in the form of registry numbers, WLN, or graphic representation. Thus, the system operator is provided with a complete set of capabilities from which to select the methods meeting his requirements.

With flexibility of interconversion of methods and with increasing amounts of data available for processing from multiple sources, the difficulties of compatibility and convertibility become significant. Maximum use of this information will be possible only if it can be exchanged among a large community of users to reduce the unit cost and to optimize the use of the sources. There are differences in conventions for representation of chemical features, character sets, keyboard input, and connection tables. A thorough survey of the problems of compatibility in chemical information systems with suggested solutions has been published.[33]

Applications. In 1968, the Institute for Scientific Information announced plans for the *Index Chemicus Registry System* (ICRS) using the Wiswesser Line Notation. This commercial service is now operational. Services based on the ICRS are available, including programs for selective dissemination, files for retrospective manipulation, and printed indexes.[34] Bit screens have been developed to improve the economics of retrospective substructure searching of the ICRS tapes.[35] Another by-product is the preparation of a permuted monthly index with WLN called the *Chemical Substructure Index.* The monthly issues are cumulated annually.[36]

The justification for chemical information systems is assistance in the research process. Files of data, presenting to the inquirer tabulations of previous work, are a start toward that objective. Closer to the desired objective of research participation is work done on the computer-assisted design of organic syntheses carried out by Corey and Wipke on the application of machine computation to the generation of chemical pathways for the synthesis of complex organic molecules. Their system stores connection tables for the structures with the major addition being the naming of bonds for explicit representation. In operation, the user describes the target molecule on a Rand Tablet with a CRT display as part of an interactive computer system. The interactive analysis leads to a logically restricted set of structures which may be converted in a single synthetic operation to the synthetic target. The process is carried out repeatedly for each intermediate until a tree of such intermediate structures results which leads down from the synthetic target to structures corresponding to readily available starting materials. The key to success in the operation of such a concept is the design of programs for intelligent analysis. Because of the vast number of possible routes, a brute force approach would not be practical or economical.[37,38]

Technology has been developed to the point where in most areas we have techniques more sophisticated than we have ability to use those techniques. The challenge for the coming period is to use the technological developments in cost effective ways to benefit and not swamp the users.

PRIMARY AND SECONDARY PUBLICATIONS

Primary and secondary publications have remained a vital part of the communication process despite predictions of the demise of this long-basic means of communication. However, several factors, including new technology, new use patterns, and economics have had significant impact on the publication process.

A good deal of talk, planning, and experimentation has been conducted in the computer composition of primary and secondary publications. While publishers are only beginning to utilize computer technology for composition purposes, they do recognize that such a step is fundamental to the total system of publishing and disseminating primary and secondary material. A lack of funding has hampered the research effort in this area, and it is only the promise of cost saving that is leading publishers to seek to adapt computer technology to composition requirements. The realization of economic advantage has been slow, and the movement to computer technology is somewhat slower than had been anticipated. This movement is hampered by the difficulties in handling the complex composition problems of chemistry and other scientific fields.

At the same time, the use to which these publications are put shows increased selection and discretion on the part of the user, and these attitudes have created additional problems for publishers seeking to meet the requirements of users. The technology of photocopying has enabled the user to acquire readily just those items of information he deems important, without dealing with the large residue of unwanted material. Such use suggests changes in packaging that are very much the concern of publishers. But changes in packaging are complicated by unrealistic notions of the cost of photocopying. There is a strong tendency to equate the few cents per page per photocopy as the true cost. The subscriber of a journal may pay for much material of no relevance to him, thus increasing his cost per relevant page. Therefore, if the true cost of photocopying were calculated, it would change considerably the false notion of low cost availability, put the whole problem into better perspective, and lead to more rational solutions on the part of both publishers and users.

The copying problem has intensified as a result of improved technology, and is being reflected in the current interest in copyright legislation. That interest has tended to polarize as a struggle between librarians and publishers. It is clear that all concerned have a stake in working out an equitable solution to the problem. While the situation is polarized for the moment, publishers do recognize that copying technology is an important element in the dissemination of scientific information, and have not sought to curtail this operation except for the purpose of demonstrating the need to work out equitable solutions.

The current economic situation has created problems for both publishers and users, resulting in a decrease in subscription revenue compounded by steadily increasing costs. This in turn, leads the user to be more selective as a way of reducing cost and improving information coverage. Both of these factors are leading to some major reevaluation of the present methods of publishing. We can hope that one benefit of the current critical analysis will be the development of long-range solutions. Financial pressures may result in the failure of some publications and possible reductions in the number of pages published by others. Furthermore, potentially valuable new products and services are less likely to be developed in an unfavorable economic climate. But the demise of a journal is not likely to result in a reduction in the amount of material published. Authors will simply seek new avenues for publication and create added problems for the remaining publishers. The readers of a failed journal will seek other means of acquiring the information they need.

Such changes are not necessarily bad. As the volume of papers to be published increases, journal editors can be more selective in what they publish and we may expect a rise in the quality of such publications. Also, the money the user can afford to spend is spread over a smaller population of information products and services and so one may expect that this will lead to an improvement in the fiscal picture for producers. These arguments simply reflect the free enterprise system in operation. However, there is some justification in the view that evaluation of information products and services in the market place is bad because the system has too much inertia and reacts to changes in the environment too slowly. It is argued, as a consequence, that some level of subsidy is called for to reduce the often violent effects felt when one depends solely on the market place for a determination of value. There are certainly a good many information products and services that are subsidized today, either through public or private funds, and it is likely that many of them would not survive the test of the market place unaided. At the same time, charge of unfair competition may be leveled against these subsidized information services by those organizations whose products and services are not subsidized, yet meet the subsidized products and services as equals in the market place. Whatever one's views may be, there is no doubt that arguments over the matter of subsidy will occupy the attention of many people in the months and years ahead.

Another factor which will play an increasing role in the interaction between primary and secondary publications is that involving copyright and wholesaling. Publishers of primary information sources have recently become sensitive to the potential damage to their markets caused by the incorporation of their products into the data bases of secondary services. Microforms and computer-readable stores are the principal concerns of primary publishers. Although it is doubtful whether any serious effects have so far been felt, primary publishers and secondary services must unquestionably work toward some mutually agreeable solutions, since modern secondary services are dependent on the output of primary publishers. Nevertheless, the question of copying, in any form, is a real one that is being hotly debated and tested in the courts.[39] A timely and comprehensive review of the copyright question will be found in a recent publication.[40]

One last view of the plight of publishers concerns, essentially, the role of the author in the communication process. So far, attention has been focused on the publisher and user, and the author has gotten off lightly. Yet, from all accounts, the author of any work benefits most clearly and most rapidly from its publication. Whether the "publish-or-perish" syndrome is any longer operative to any significant degree is probably inconsequential. The author's salary, prestige, fame, and place in history are often influenced by his publications. It seems surprising therefore that authors bear so little fiscal responsibility for the published versions of their work. The page-charge, which is strongly opposed by many authors and which is employed nonuniformly by the publishing community, is the only financial contribution to the publishing effort made by the author. Most of the cost of publishing his work is borne by the subscriber who would prefer to acquire the publications of a selected handful of authors. Whether the page charge proves to be a viable alternative to spiralling subscription prices or to terminating publication, remains to be seen. It is doubtful, in any case, whether the author can ever be made to bear the full cost of the material he wishes published, but quite clearly, the author's share of the publishing cost should be increased.

A major benefit to the user which has stemmed from economic strictures has been the reemergence of the "total system" concept. While this concept was paid much lip service in early planning and design efforts, most development has taken place in a highly parochial manner. Now, programs of interdisciplinary groups are being launched to reduce duplicative efforts and coverage.[41-43] Such efforts will help, but they are only a part of a larger program that is needed to yield a strong, dynamic publishing industry, an industry that may be expected to be the heart of the communication of information for a long time to come. This program must certainly take advantage of the speed, power, and flexibility of the computer. Several obvious applications include typesetting (metal or photographic), page makeup, line justification, syllabification, graphics, table-of-contents production, indexing, and abstracting. From a generalized data base, many specialized packages could be produced which would more nearly meet the needs of the user. And, multiple uses of the same data base would offer economic advantages. The technology is not wanting, but its intelligent application in a well-conceived system is.

But, in the last analysis, the computer will not solve the really critical problem in publishing: too much material of little or no value. Without adequate evaluation procedures, all else will be in vain. Here, we return to the market place.

CHEMICAL INFORMATION SERVICES

Background. An information service may be defined as "information products or information packages" which can be delivered to the user directly or via intermediaries (librarians, computer centers, etc.). This definition is offered because the term information service has many connotations. Organizations that produce primary or secondary publications are called information services; the primary or secondary journals themselves are referred to as information services; library services are known by some as information services; information dissemination centers, computer centers, and the more recent data analysis centers are all referred to as information services.

Information service concepts are interwoven with tangible products. The tangibles include the services delivered to the user, whether they are secondary journals, abstracts, references retrieved from a bibliography by a librarian, printouts of references from a computer, data displayed on a CRT terminal, and hardcopy generated from a magnification of microfilm or microfiche. Large reference-data banks yield tangible products by making it possible to package subsets of the total content of that bank for the specific needs of specialist user groups. Selected portions of different data bases can be combined to fill the mission-oriented needs of user groups. As mentioned earlier, the user's acceptance of information services in new forms requires development, so special publications, addressed to particular user groups (special packages from more than one base), are often introduced to test new information service concepts.[44]

Impact of Commercial Information Services. Organizations are making increased use of purchased data bases to replace material previously abstracted and indexed internally. At least 50 science and technology-related database services are available to provide organizations collections of material which can be processed for selective dissemination and certain types of retrospective search.[45] In many cases, the route to most efficient use of the acquired data base has been to reformat it to a common form acceptable to in-house software. This is the approach used by many industrial organizations and by the major service centers. For example, the Computer Search Center of the Information Sciences Section of the Illinois Institute of Technology Research Institute has done extensive work and has published the results of the design and operation of software for the reformatting and processing of data bases. Work has been reported on format conversion, profile preparation, word truncation, weighting, search techniques, and extensive analyses of data files.[46-48]

Economic Factors. User needs, information service concepts based on those needs, and the actual delivery of information services cannot be divorced from economic considerations. The speed with which a greater variety of better information services will become available depends on the impact that cost of such services has on the user's ability and willingness to pay. Information services must be economically independent unless the subsidy concept gains favor. Long-term information service subsidies are rare outside the Federal Government, and thus private information services must be self-supporting. Even within the Federal Government, a trend away from "free" services is evolving. The philosophy that information is a commodity for which someone must pay, and that it is worth only what the user is willing to pay, is often discussed and is becoming more frequently applied. Libraries which have traditionally been entirely "free" are beginning to adopt the attitude that the user of library services should be charged directly for its services. In some cases, libraries (notably regional medical libraries) have actually started to charge for certain of their services.[49] An increase in this practice in the future is predicted particularly as machine-based services become more prevalent in libraries.

The Collective Voice of Information Service. The status of information services, as described above, has evolved and grown, especially within the last two decades as a result of interactions between those using services and those who supply them. And, mutual benefit seems to have been gained from interaction among suppliers of information, because many efforts to formalize these relationships have taken place at both the national and international level. The National Federation of Abstracting and Indexing Services (NFAIS) was established in 1958 for the exchange of ideas and experiences among organizations which provide information services to users in the field of science and technology in the United States. Recently, the advantages of such an association having been demonstrated (to the participants if not to the ultimate users of the services) the Federation has yielded to pressures for membership by organizations that provide information services in the social sciences, fine arts, journalism, and other fields. Thus, the Federation will now be made up of organizations in all fields which have the common concern to provide better, more efficient, and economical information services.

The International Council of Scientific Unions Abstracting Board (ICSUAB) is an organization in which groups dealing in information services from all over the world may participate and which has the objective of international cooperation and the mutual benefits to be derived from the better communication of ideas.

The information dissemination centers have organized for better communication and a stronger voice in the information community. The Association for Scientific Information Dissemination Centers (ASIDIC) and its European counterpart, the European Scientific Information Dissemination Centers (EUSIDIC), are organizations whose members provide information services on a regional basis to their constituents. That regional basis may be a state university system, a university network, a state governmental network, or a geographic region, depending on individual agreements between the data-base supplier and the information center. In either case, the user has the benefit of information services designed for and in response to his individual needs and offered on either a current awareness or retrospective basis.

A number of the information dissemination centers were started with Federal Government support on a declining basis. Several of them are no longer subsidized by the Government; if they are to continue, they must become self-supporting or seek other subsidies. It is doubtful that all of them recover costs from sales of services. This concern of the centers has been shared through the relationships established in ASIDIC and EUSIDIC. The unified voice of these groups is being heard; it is suggesting networking and third party agreements between suppliers, primary leasors of data bases and network centers.

Within the United States, the free enterprise system gives rise to innovative information service concepts and marketing. The not-for-profit, for-profit, and "free" services from the Government all compete for users from the same community of scientists. This allows the individual suppliers to respond to the market—each according to its interpretation of that market. This system distinctly enhances the product possibilities. It spurs the creativity of information suppliers with respect to package concepts, the development of systems for better, more innovative products, and a stronger effort to discover user needs and to respond to them in a positive way. The free enterprise system also has a marked effect on the economics of information services. The competitive atmosphere of Government, for-profit, and not-for-profit information services demands that those who provide information services must keep costs to a competitive minimum; they must create and develop services in the most economical way possible. Several Government-based services or segments thereof are provided to the user by commercial information suppliers. Disagreement with this practice of Government subsidy to for-profit organizations has been expressed; the economic impact of this will undoubtedly be known in due time.

The increased concern for standardization in recent years can be attributed to the move toward large mechanized information systems. Information services and ultimately the users of information will benefit from cooperative efforts of information suppliers who provide those services. Umbrella organizations referred to earlier (ASIDIC, NFAIS, as well as the American National Standards Institute's Committee Z39 and X3 and the International Standards Organization, which are directly concerned with the standardization of information transfer) cooperate for better information transfer on a national and worldwide basis. Each of the organizations includes representatives from libraries, universities, and companies that supply information secondarily. For example, an internationally developed list of periodical-title-word abbreviations has been approved by ANSI and ISO, and is currently being implemented in the major countries of the world where informtion services are generated. Several standards which have been developed in the United States are being considered by ISO with the hope that they can be modified to accommodate the individual needs of many of the other countries. Also, ISO and ANSI's Committee Z39 are cooperating on the development of a unique identifier for each individual document within a journal (bibliographic strip).

The accomplishments of many of these organizations are notable; either they have developed standards or they have influenced various committees and organizations concerned with establishing standards in the field of information transfer. These standards directly affect the quality of information services and indirectly serve to fill the needs of the user.

But efforts toward standardization can be more encouraging than actual accomplishments. In assessing the rate at which it is realistic to expect tangible progress, one must keep in mind that international cooperation and efforts toward a world information system require not only mutual understanding of technical matters, but the wherewithal to apply those technologies, the political atmosphere in which such understanding must arise, and the protection of proprietary rights and interests on the part of various national groups. These considerations are further confounded by language barriers, so it is unrealistic to expect rapid progress toward an international information system, but the evidences of cooperation are many, and there is hope.

Cooperation among Information Suppliers. A cooperative study under the joint sponsorship of the Chemical Abstracts Service (CAS), BioSciences Information Service (BIOSIS), and Engineering Index (Ei) has been under way for appproximately three years to determine overlap, to seek ways for reducing redundant effort, and to move toward cooperative programs that will benefit both the users of the services provided by these organizations and the organizations themselves. To date, these organizations have reported the overlap in journals monitored by these three organizations, the amount of document overlap between them,[41,42] and have planned further studies to assess the differences and similarities in machine format, indexing, and editorial considerations.

Recently, BIOSIS and CAS, the two largest scientific abstracting and indexing services in the U. S., announced their decision to undertake an extensive program of operational cooperation and coordination.[43] They will attempt to coordinate coverage, eliminate duplication of intellectual or other processing effort and reach a degree of compatibility that will permit their services to be used effectively in combination. This concept of cooperation, coupled with the fact that the CAS system (one of the first large mechanized systems) is based on extensive planning, research and development, could suggest the "neonatal" stage of a loose U. S. system. The effectiveness of the cooperative program will determine the feasibility of that inference.

INFORMATION AND THE FEDERAL GOVERNMENT

The role of government in information processing is twofold. On the one hand, these organizations have major sources of funds for research and development activities and for the support of on-going information systems work. On the other hand, they have themselves carried out research and development activities and have been major users and operators of systems developed either internally or by outside agencies.

Information policies of national governments affect in a very real way the information services available in the respective countries. For the most part the United States has lacked a coherent national policy on information; numerous U. S. government information programs have existed for years and continue to exist. A national science and information policy has been discussed for some time but tangible evidence of progress is scant. The scientific and technical information program within the U. S. government is largely on autonomous activity based on the operations of the individual Federal agencies and departments.

Chemical information processing within the Federal Government can be characterized as a preponderance of empirically-based, conceptually overlapping, but physically incompatible systems. Thus agency or mission information systems have been developed many of which parallel one another, resulting in considerable duplication of effort.[50-52] The National Technical Information Service (NTIS), however, is an outstanding example of a governmental cooperative effort.

Government agencies are both data gathering and data producing organizations. The characteristics of the data gathered and produced differ with each agency and often this is cited as the principle reason for the development of specialized information systems within respective agencies. It is possible that the apparent differences are, in reality, only superficial differences, and that information processing systems can be developed which are independent of the specific input which they are intended to process. Recently, systems design techniques have emerged which make it possible to develop generalized processing systems which can be tailored to specific data bases by supplying the system with appropriate descriptions of the data to be processed.[53-56] A question of efficiency arises but when the cost of operating generalized systems is contrasted with the cost of developing, implementing, and operating a multitude of specialized systems, it is expected that the generalized system will prove to be much more economical.

Programs on information within the U. S. government have been coordinated by the Committee on Scientific and Technical Information (COSATI) and the former Office of Science and Technology. The much publicized SATCOM report[57] was a source of hope but little impact has been observed. Limitations of the national academies (NAS and NAE) are observed as the private sector looks to them for help which too often does not materialize. To date progress and activities of the U. S. National Commission on Libraries and Information Science,[58,59] which was charged to study information transfer and lack of coordination within the federal sector, have not been reported. Hopefully, the commission will develop a plan that will ultimately lead to elimination of much of the inefficiency and redundancy of the present system which has evolved in the absence of national policy.

A most important influence in the information science community is that of the Office of Science Information Services of the National Science Foundation.[60] This office has established a policy of support for U. S. non-federal scientific and technical information services and programs. It has a strong influence on the discipline oriented secondary services because of its ability to allocate support for research and development in information handling. The most outstanding example of this influence has been observed in the field of chemistry through NSF grants to the ACS for development of a chemical documentation system.[61] Recently, under the influence of new personalities, the philosophy of that office has changed. Since monies for research and development are limited, programs having especially far reaching potential (or for broad-scope systems) will reportedly be supported, but operations support will not be granted.

The prototype in large operational mechanized chemical and biological information retrieval systems is that developed for monitoring drug programs sponsored by the Walter Reed Army Institute of Research (WRAIR). This pioneering work has been reported previously[1,3] and is updated here. The system features input with a chemical typewriter, the storage of connection tables generated from those input structures, stored bit screens, providing either screen searches or atom-by-atom matching. Output of structure diagrams is provided using a drum printer with test data printed along with structures. The collection contains over 200,000 structures with over 2 million data records in the associated biological files.[62,63] It is encouraging to see studies performed and reported on efficient methods for searching large files, work sorely needed if we are to make intelligent use of our growing resources. Understandably at WRAIR, work is under way to develop techniques for efficient querying of large files. Biologically-oriented subfiles are created for further analysis by chemical structure. Their work is indicative of the approaches required to make searches of large files both practical and economical.[64]

Information and Policy Making. Information has been described as the first need for ensuring that changing societies make wise decisions. The Organization for Economic Cooperation and Development (OECD) has published a list of conclusions and recommendations as a result of a study to determine how this need might be fulfilled.[65] Both national and international policy can be influenced by the OECD report (Information for a Changing Society). Several conclusions selected from the 13 listed in the report appear to be noteworthy:

> ". . . information systems designed for research workers are inadequate for users in other disciplines and technology
>
> ". . . traditional discipline oriented systems are not effective in tackling multi-discipline social problems
>
> ". . . emphasis on quality control is needed to ensure that information is not misleading
>
> ". . . existing information systems are at an experimental level of development and will remain so for many years
>
> ". . . information systems of the future must be dynamic . . ."

These are followed by recommendations among which are:

> ". . . government should accord priority attention not only to policies for the generation of scientific and technological information, but also to policies for use in policy making, governmental decisions and R&D management; . . . government should give greater support to mechanisms for ensuring effective interchange of information among scientists . . ."

The effort of OECD in the transfer of information on an international scale hopefully will be complementary to that of UNISIST, rather than compete for limited resources on the part of the participating countries. The stated objective of OECD, to provide information in changing societies, appears consistent with that of UNISIST. UNISIST, which may be described as the framework for information activities of some 90 nations, strongly emphasizes the need to improve information resources for developing countries.

In contrast with the lack of a national information policy within the United States, a highly structured national policy exists in many Western European and Eastern European countries. In those countries where strong national policy exists, all scientific and technical information services are coordinated through one government sponsored and operated branch. This structure can provide for information flow according to discipline as well as systematically provide for interdisciplinary subject coverage in a logical way.[66] VINITI of the U. S. S. R. has the world's largest single information service for science and technology. It publishes the *Referativnyi Zhurnal,* and, although its component on chemistry is smaller than that of *Chemical Abstracts,* the total makeup including chemistry, physics, biology, and other disciplines exceeds that of any single discipline oriented base in the United States or elsewhere.[67]

Several other countries have strong centralized programs—e.g., Canada through its centralized CAN/SDI,[68] Germany through the Institut fur Dokumentation, and the French government through Centre National de la Recherches Scientifique (CNRS)[69] and support to the ISDS (International Serial Data System).[70] Discussions of many East European programs as well as summaries of developments elsewhere in the world may be found in JPRA-53523.[71]

In all respects, governments have a large role to play in the timely and effective development of computer and information science in all disciplines. A great deal of good judgment backed by adequate technical advice and financial support must be exercised in the years ahead.

SOME ACTIVITIES IN ACADEMIA

Background. In contrast with government agencies, which have long played an active role in development of systems, colleges and universities represent latter day entries into the information field in a formal sense. The fact that information science or computer and information science has achieved the status of a discipline within academic circles is encouraging. Such an achievement indicates that information science is a vital activity and that it is considered to be something of value to society in general.

In one sense, colleges and universities have played perhaps the only role in information processing for several millenia. The library in ancient days was the university and from the library grew an institution for educating selected members of the populus.

In the 1930's and 1940's, the rate at which the production of data was growing was considerably ahead of the abilities of libraries to deal effectively with it. In particular, the volume of data produced during and following the second World War made problems of traditional librarianship acutely obvious to all.[72]

While libraries had reached this state of affairs, the art of computing by machine was just coming into its own. Initially, computers were designed for numerical processing. It was only after the first computers had been built that the prospects for processing other kinds of data began to emerge. One of the unfortunate aspects of this stage of development of computing was that those who visualized the use of computers in information processing attacked rather than allied themselves with libraries and librarianship. As a consequence, this fostered an antipathy in librarians for the computer and computing. This unfortunate state of affairs resulted in an artificial schism between the disciplines of library science and computer science which has yet to be fully mended.

Within recent years, academic departments of computer science, information science, or computer and information science, as they are variously called, have evolved on academic campuses.[73] Consequently, much of the research which in other disciplines is traditionally carried on within academic institutions, had, in the information sciences, been carried on by government agencies, industrial firms, and other institutions. Only recently have academic institutions begun to make themselves felt as real forces in this field. Information science departments have now assumed the traditional role of academic departments in the training of individuals for employment and for carrying on research with the goal of discovering the underlying theoretical bases of information processing.

One of the principal roles of academic departments in information science is the attempt to define and more clearly delineate the boundaries of the discipline. It is becoming increasingly clear that the discipline called information science cuts across the boundaries of most other disciplines because, in any discipline, the activities of its practioners are ones of data gathering, data analysis, and decision making. These are the basic concepts to which information scientists address themselves. Thus, the role of academic institutions is becoming quite clear, and the advances expected from these academic institutions are fundamental understandings of the various problems within the discipline. These advances should be substantial.

There is still a considerable conceptual gap between information engineers, the typical practioners within the discipline, and the more theoretically oriented individuals within academic institutions. This gap exists partly because of the terminological differences between the two groups and perhaps in part because of ideological differences. Nevertheless the gap is narrowing. In the next few years, the relationship between information engineers and information scientists will bear similarity to the relationship between chemical engineers and chemists. But such a relationship will come about only when those people who call themselves librarians, library scientists, computer scientists, information scientists, information engineers, or information specialists, all reach the realization that the problems in which they are interested, the problems they are attempting to solve, and the persons they are attempting to serve are the same.

Present Status. Both basic theoretical work and practical development effort are being carried out within academic institutions. On the one hand, fundamental developments in such areas as linguistics, machine theory, pattern recognition, artificial intelligence, applied combinatorial mathematics, have improved our knowledge of what these concepts are and how they are interrelated. On the other hand, application of some of this knowledge is being actively pursued in the design and implementation of automated library systems, in the development of computer-assisted instructional systems, and in the design and implementation of automatic monitoring and analysis systems. One of the areas in which significant achievement has been made, both at the theoretical level and at the developmental level, is in the design and implementation of languages for communication with computers.

The past few years have seen a considerable growth in the number of academic institutions offering degree programs in computer and information science. A major effort has been with the development of curricula which are both flexible and sufficiently uniform that a discipline can be said to exist.[74] A survey has presented considerable data for comparison of academic programs in 45 schools.[75] There are now more than 100 schools with degree programs in computer and information science. Curriculum development continues as the field becomes better defined and as greater knowledge of the subject matter is obtained.

Instruction in computer and information science has not only given rise to new academic departments, but has led to a revamping or expansion of the curricula of the existing departments in other disciplines. In particular, many library science programs have been given a broadened outlook. And other departments such as education, languages and linquistics, the sciences, and engineering, have expanded their programs, either in cooperation with computer and information science departments, or through the introduction of their own courses. Courses in use of the research literature and about information transfer systems have been introduced as part of the chemistry curriculum in some colleges. Educational programs for those who must conduct the courses are offered by a number of information processors.

A related problem surrounds the question of education and information services. Many educators are unaware of the existence of the many information services available, the role the services can play in the transfer of information, and of the importance of conveying knowledge of these tools to the student and developing scientists. Also, educators as well as students are too often ill informed in the efficient use of the information services of which they are already aware. The latter problem falls within the domain of the suppliers and the intermediaries who provide services from various data stores, and the former is both a challenge to those organizations as well as the academic institutions.

The Future. The growth of the field of computer and information science will continue to be rapid and uncoordinated for the next several years. The number of academic departments will stabilize, and the incorporation of facets of computer and information science into other academic departments will continue. Activities in computer and information science in academia will continue along three compatible lines; research, teaching of computer and information science, and application of computer and information science to teaching in other subjects.

Research in computer and information science has often tended to be highly theoretical, abstract, and of little practical value. Substantial emphasis on the solution of practical problems will be required if computer and information science is to avoid becoming sterile. Any academic discipline has a major responsibility to produce graduates who are employable, and this means that these people must be encouraged to acquire skills that are useful to industry and government without extensive additional on-the-job training. Research horizons will continue to expand as we become aware of new ways in which computing can benefit mankind. At the same time, reasonable levels of funding will be necessary from private, industrial, and governmental sources to nourish this research.

Massive funding alone does not guarantee results. Funding agencies must display considerable knowledge and sound judgment in deciding how their funds will be allocated. Likewise, review by peer groups must take on a greater degree of real value judgment. Difficult decisions, calling for courage and imagination, must be made in the allocation of limited funds for research projects. Past practice has been neither satisfying nor rewarding.

There are now some 50 to 60 academic institutions that grant the Ph.D. in computer and information science and the number will grow. But it behooves practioners in the field to accept the evidence provided by other disciplines and to control this growth so that their graduates will be and remain employable.

In the teaching of computer and information science, as in other disciplines, it is important to teach problem solving. The problems to be solved are not, however, just those neat clear-cut problems usually dealt with in academia. The most important problems are those which have no clear solutions, and the student must learn to recognize from among the wealth of useless data those few items which seem to exert a controlling influence in the given situation, and to get control of those items. Life is the choosing, almost continuously, from sets of invariably unsatisfactory alternatives that one which seems to offer the fewest undesirable consequences.

Certainly the discipline of computer and information science is one of the most exciting and challenging areas of endeavor one could hope to be in, but while the potential for obtaining useful results from one's effort is great, so is the potential for producing nonuseful or detrimental results. Computer and information science is concerned with automation, but it must also be concerned that automation for the sake of automation does not become the rule.

SOME ACTIVITIES IN INDUSTRY

Background. Earlier reports issued by this Committee[1,2,3] described major long-range research and development work by industrial firms, and significant chemical information programs executed by them. The recent years have been notable in the decrease or elimination of the long-range, fundamental chemical information research carried out in industry. The significant shift in emphasis is toward internal development work, experimentation with new services, and the operation of those new services.

Industrial information activities grew in proportion to the increasing public and private research and development expenditures in the United States throughout the 60's. As research expenditures leveled off, or decreased in some areas, there has been a parallel slowing of the growth, and in many cases a reduction, in chemical information operations. Information activities in industry are now in an operational phase. Development work is concentrated on pragmatic experimentation, with maximum effort being applied to offering practical, cost recoverable information services, for operation within a challenging economic environment.

Meeting User Needs. Most industrial firms are increasingly oriented toward commercial work with short-term payoff. The product development team is becoming common, and this has caused an increase in the need for more interdisciplinary, commercially-oriented information. Thus, while the chemist needs ways to keep up to date in his field and ways to enable him to find answers to questions, his approach to information acquisition has changed.

Such change has followed the development and availability of new services. Because the chemist is not adequately trained in information processes, he does not clamor for the new capabilities which are available, but reacts to new services and is willing to use them if they meet a need. It is the internal information center that must be aware of the new external services available, must experiment with these to determine their utility to the organization, and must extract from them products appropriate for use by the chemist. The chemist is not interested in searching through long lists to extract items of interest, but prefers to receive relatively well sorted and tailored products which are specific to his requirements.

Users find value in the new alerting services. Selective dissemination of information is becoming a working tool in many organizations. Retrospective search is still a difficult process and involves a combination of effort by the user and the information service that is appropriate to the working patterns of both. An important problem that remains unsolved is the need for access to chemical data, substructure correlations, and reaction information.

The user expects that chemical information, whether internal or external, will be delivered completely and in reasonable time. Two limiting factors are the fiscal position taken by the organization and the services available to it. There are many overlapping services, and yet there are gaps in particular areas of interest. The challenge for the information service is to retrieve information from single services to provide solutions to interdisciplinary problems. The information service must know the sources available, how to use them, and the most economical ways of meeting user requirements. This is going to be a challenge for some time to come.

New Services and Products. Industry is providing access to its internally generated information through a variety of systems. A major shift in access to external information has occurred by acquiring data bases and manipulating these for internal use. The products may take the form of printed materials distributed internally, subscriptions to external computer services providing selective dissemination of information, and acquiring data bases on magnetic tape for internal manipulation. Industry is using a variety of services such as those offered by, American Petroleum Institute, BioSciences Information Service, Chemical Abstracts Service, Derwent Publications, Ltd., Engineering Index, Excerpta Medica, IFI/Plenum Data Corporation, Institute for Scientific Information, National Library of Medicine, Paul DeHaen, United States Patent Office, and regional centers.

The above list is representative but not exhaustive. Many of the regional centers are offering essentially the same service and industry is now engaged in shopping the centers for the best combination of price and service. In this competitive situation those services performing the best work will survive. All will soon be on a more stringent financial basis as many of the Federal support programs are terminated.

The SDI services will be operated by a number of regional centers equipped to provide cost-effective services. The size of the data bases for retrospective search are such that usage and cost will dictate that no more than one or two centers in the United States will supply services from the large retrospective data bases.

Services Curtailed. There is a clear trend toward curtailment of the internal indexing and abstracting of external material. Industry is buying access to external material for current awareness and SDI programs. A similar philosophy is also evident in common areas of interest, such as patents, with increasing numbers of companies subscribing to patent services from IFI/Plenum Data Corporation and Derwent Publications Limited for U. S. and foreign chemical patent information. Industry will use these external services as long as the services are responsive to the changing requirements and as long as a cost-acceptable product is available.

Another trend is clearly evident. Traditionally, many industrial laboratories had reading rooms complete with major references such as *Chemical Abstracts.* As the major references became more expensive, multiple subscriptions were reduced in number and many local reading rooms were either closed, reduced in scope to a resting area for a few less expensive journals, or replaced by centers with microform collections. Multiple copies of major references have been cancelled and centralization of library facilities has taken place.

Hardware and Software Developments in Industry. Specific developments in hardware and software are briefly noted under *Developments in the Technology of Chemical Information Processing.* However, a few generalizations are in order on the status in industry. In the 50's and 60's, industry played a major role along with Federal Government and the academic community in developing and using specialized hardware and software for chemical information processing. At present, industry is concentrating on using the developments already made and is doing only modest work on software extension. There is no known work on hardware development. Industry is more and more oriented toward the use of existing, commercially available, software packages which can be adapted to internal needs. Industry representatives state repeatedly that their limited funds must be spent on operating programs and that needed fundamental research must be left to governmental and academic organizations. Major basic research has been carried out in many industrial organizations, but no extensive basic research in information processing is known to be under way in any industrial firm at present. This is unfortunate because industrial users represent a large community whose special requirements are only incompletely served at present. Some research should be continued to provide these users with needed improvements in chemical information processing.

A few organizations have reported on the work they have done in selecting from the techniques available in their implementation of information services for their personnel. At the Dow Chemical Co., internal information is indexed with the chemical structures represented by WLN with both batch and interactive searching available on selected files.[76,77] Dow has operated for many years selective dissemination services at cost for personnel using a variety of external data bases.[78,79]

Developments on other large systems have taken place. The Agriculture Research and Development Department at Monsanto has revised its information storage and retrieval to take advantage of new concepts. All of the chemical compounds in an existing service were coded in WLN and stored in an on-line system. Chemists have been trained to input new compounds using the WLN. Associated with the structures are biological test data. The objective is to have the research chemist input new structures and associated test data and make the total file available for search or for extraction of test results for reports to laboratory personnel.[80]

Another large system is reported at Du Pont where former internal report and patent files were converted to a system using threaded lists and inverted files. The system features user queries in a free format language which is then optimized by the system to make most efficient use of the file structure.[81] The system has since been extended to include the processing of chemical structure information using keyboard entry of structures, storage of connection tables, and an extensive screen system.

The work of the Pharmaceuticals Division of Imperial Chemical Industries is typical of the flexibility available to operators of chemical information systems. A search system is operational within the synthetic pharmaceutical research organization covering nearly 100,000 compounds represented by over 70,000 samples and incorporating nearly 800,000 test results. Input of structural information is by WLN and three levels of search are provided. A fragmentation search is employed for questions of occurrence. For implicit connectivity, a notation search is made. Atom-by-atom searches on connection tables derived from the notation are made when a detailed structure search is required. The output on cards includes a structure diagram and the availability of biological sample data.[82]

Organizing for Information Processing. Information organizations are staff or service functions. Their strength and effectiveness are often related to their position in the parent organization and are clearly affected by the understanding and acceptance by the corporate technical administrators. Some chemical information centers are part of the corporate library. Other chemical information centers have libraries as a subdivision. Frequently, the functions are divided into services covering internal material and services covering external material. A major trend is for the centers to repackage for internal use the information products offered by outside wholesalers.

There are variations in understanding by management of the center operation. Information operations in the pharmaceutical companies seem to have stronger support and are relied on more heavily than those in chemical companies. Industrial management often has a better understanding of product research, manufacture, and marketing, and a lesser understanding of the possible and proper role for information services.

There is a clear trend in the larger companies in industry toward combining or coordinating separate information activities into centralized services for reasons of economy and designated responsibility. Also, with many information services and products becoming more expensive, centralization has brought about cost reduction. The concern with economy in industry has caused the chemist to think about whether he wishes to do without an information service, do the literature searching and copying himself, or go to the internal service for assistance and pay for that help.

Trends. Industry wants to buy services rather than index material. Therefore, there will be increased use of outside data bases and the use of regional information centers for processing those data bases. These centers have great potential and in many cases are already fulfilling sizable segments of user needs. Economics continue to favor keeping files separate and searching those most likely to produce answers to the questions submitted.

For the use of retrospective data bases, industry will concentrate on going to the one or two large centers offering such services on a reasonably economical basis.

A new trend is emerging in the use of on-line services, the tapping of remote files with a local terminal and a time-sharing network. The National Library of Medicine, Lehigh University, Systems Develoment Corporation, and

Battelle Memorial Institute among others, are offering searches of various data bases. This activity is expected to grow.

There is as yet no significant trend toward the use of terminals by the individual scientist for querying data bases since these are sufficiently complex and the search systems sufficiently involved that information scientists will continue to be used as intermediaries to translate the question of the user into an effective and economical query for the data base.

There is increased interest and activity by industrial trade organizations to work together toward the encouragement of new services and the improvement of existing ones for their members. The Americal Petroleum Institute has worked in this area for a long time. The Synthetic Organic Chemical Manufacturers Association, the Manufacturing Chemists Association, and the Pharmaceutical Manufacturers Association are all becoming active in encouraging new and improved services for their members.

More information services will be charged back to the user with more accurate cost analyses of the full cost of providing information products.

Information services seem to be taking a more active part in the research and technical functions of their organizations. Information is a necessary part of all technical work. The information services can do a better job of keeping decision makers informed of where technology is ready for commercial development. The use of creative, perceptive, highly competent scientists operating in the information field can provide important input to product ventures. This requires a proper balance of personally involved information scientists working directly with their peers in the organization and a proper balance of centralized operations for economies of scale.

Problems. There are a number of problems ready for solution. At the technical level, there are many formats in which the various data bases are offered. The suppliers decry the low usage of their products, and would be well advised to do more toward consistency and standardization among products. Although users do reformat the data bases to run on internal programs, usage might increase substantially if the various data bases were more consistent in the information contained and if the data base could be more readily used by an in-house computer program. The suppliers talk a great deal about standardization, but little leadership and progress are evident.

There is poor access to large files of chemical information at the substructure level. Here is an area where Chemical Abstracts Service leadership or activity would be welcome. The task is large and difficult and requires an organization with the strengths of the American Chemical Society to mount an effort to develop economical ways to effect substructure search using large files; in order to provide a means for associating chemical structures and activity data. ACS and CAS should mount a program toward a viable solution to this problem.

There exist many data bases with significant overlap in coverage and yet gaps in coverage also exist which result in an inability to retrieve complete information on an interdisciplinary basis. The current studies on data base overlap are designed to bring us closer to solutions to this pressing problem.

Information services in industry need to earn management support and to demonstrate that information services are pertinent to user requirements. Cost benefit relationships are slowly coming into view, but a great deal more work needs to be done in industry to demonstrate the value of information and to justify the expenditures for information services.

Information services must also be careful not to delegate scientific information processing solely to systems people who may have little appreciation of the intellectual content of the material they are processing. This often results in the degrading of the quality of information manipulated to meet the constraints of software and hardware.

Action Needed. Present information processing capabilities are still under experimentation and some are being effectively exploited. The never-ending need for attention to the user and his changing requirements continues to challenge the profession. The quality of software and data bases needs to be improved to replace mediocrity with excellence and uncertainty with reliability. Improvements such as these will help develop a payoff that management can appreciate so that information services will earn their fair share of the available resources.

Research expenditures, in information science as elsewhere are being held at constant levels. Once management has been shown the benefits of information usage, then research funds will become available for experimentation and for pragmatic development work needed to move the field forward for the future.

Many challenges exist. Complete retrospective search in chemistry is time-consuming, expensive, and in some areas well-nigh impossible. Improved techniques are needed to handle large files efficiently so that better retrospective searches can be done at reasonable cost. Additional research is needed on methods for correlation of structure-related activity. The cost of input remains a large part of the total information-processing dollar. More low-cost input devices and techniques are needed, such as more material published in standardized fonts and greater latitude from optical character recognition devices. Storing the full representation of a chemical structure provides the maximum information, but requires expensive substructure search. Linear notations are a useful compromise. Improvements are needed across the range of capabilities, including fragmentation, notation, and full structure to allow a choice of methods to fit specific needs.

The national leadership recommended by SATCOM to direct scientific and technical communication is still lacking and does not appear to be forthcoming. Federal information programs are largely uncoordinated and would benefit from stronger central leadership. The professional societies are effectively discharging their responsibilities in their own disciplines. More society cooperation would be beneficial. More industrial groups and trade associations are needed to provide mission-oriented information services.

Industry has gone through a consolidation phase. Research and operating funds have stabilized and center managers are concentrating on maximizing the services they can offer by buying products from information wholesalers and repackaging these in new ways for enhanced usage and more creative exploitation by their own staffs. Greater attention is being placed on the user in industry in what he needs to aid him in his daily work. Hopefully a better appreciation of his information habits and attitudes will result in the center manager's providing services which produce a demonstrable improvement in the conduct of technical work. The goal is to prove to all technical administrators that the cost of information services is a necessary contributing part of the technical program. Information work in industry will continue and will expand as long as tangible benefits can be shown in the research, manufacturing, and marketing operations.

LITERATURE CITED

(1) "Survey of Chemical Notation Systems," National Academy of Sciences, National Research Council, Pub. **1150**, Washington, D. C., 1964.

(2) "Survey of European Non-Conventional Chemical Notation

Systems," National Academy of Sciences, National Research Council Pub. **1278**, Washington, D. C., 1965.

(3) "Chemical Structure Information Handling—A Review of The Literature 1962–1968," National Academy of Sciences, National Research Council, Publ. **1733**, Washington, D. C., 1969.

(4) David, E. E., Jr., Remarks presented at the Conference Board Meeting, New York, New York, February 15, 1972.

(4a) Rossmassler, S. A., "Public/Private Cooperation in Planning and Developing Reference Data Programs," *J. Chem. Doc.*, **13**, 65–8 (1973).

(5) Hosoya, H., "Topological Index as a Sorting Device for Coding Chemical Structures," *Ibid*, **12**, 181–2 (1972).

(6) Deffareit, H., Carie, A., Combe, H., Leveque, S., Malka, A., and Valls, J., "CORA—A Semiautomatic Coding System Application to the Coding of Markush Formulas," *Ibid.*, **12**, 230–3 (1972).

(7) Berg, Sanford, "An Economic Analysis of the Demand for Scientific Journals," *J. Amer. Soc. Inform. Sci.* **23**, 23–9 (January-February 1972).

(8) Davis, C. A., and Rush, J. E., "Information Retrieval and Documentation in Chemistry," Greenwood Press, Inc., Westport, Connecticut, 1973 (in press), Chapters VIII and IX.

(9) Loening, K. L., ACS Committee on Nomenclature Annual Report for 1972, *J. Chem. Doc.* **13**, 51–2 (1973).

(10) Binnendyk, D. van, and Mackay, A. C., "Computer-assisted Generation of IUPAC Names of Polycyclic Bridged Ring Systems," *Can. J. Chem.* **51**(5), 718–23 (1973).

(11) Smith, E. G., "The Wiswesser Line-Formula Chemical Notation," McGraw-Hill, New York, 1968.

(12) Bowman, C. M., Landee, F. A., Lee, N. W., Reslock, M. H., and Smith, B. P., "A Chemically Oriented Information and Retrieval System III Searching a Wiswesser Line Notation File," *J. Chem. Doc.* **10**, 50–4 (1970).

(13) Bond, V. B., Bowman, C. M., Lee, N. L., Petersen, D. R., and Reslock, M. H., "Interactive Searching of a Structure and Biological Activity File," *Ibid.*, **11**, 168–70 (1971).

(14) Granito, C. E., Becker, G. T., Roberts, S., Wiswesser, W. J. and Windlinx, K. J., "Computer-Generated Substructure Codes (Bit Screens)," *Ibid.*, **11**, 106–10 (1971).

(15) Garfield, E., Revesz, G. S., Granito, C. E., Dorr, H. A., Calderon, M. M., and Warner, A., "Index Chemicus Registry System: Pragmatic Approach to Substructure Chemical Retrieval," *Ibid.*, **10**, 54–8 (1970).

(16) Skolnik, H., "A Correlative Notation System for NMR Data," *Ibid.*, **10**, 216–20 (1970).

(17) Skolnik, H., and Hays, J. T., "A New Notation System for Indexing Polymers," *Ibid.*, **10**, 243–47 (1970).

(18) Skolnik, H., "A Notation Symbol Index for Chemical Compounds," *Ibid.*, **11**, 120–4 (1971).

(19) Dubois, J. E., "Principles of the DARC Topological System," *Entropie* **25**, 1–14 (January-February 1969), "French National Policy for Chemical Information and the DARC System as a Potential Tool of this Policy," *J. Chem. Doc.* **13**, 8–13 (1973).

(20) Fugmann, R., Nickelsen, H., Nickelsen, I., and Winter, J. H., "TOSAR—A Topological Method for the Representation of Synthetic and Analytical Relations of Concepts," *Angew. Chem.* **9**, 589–95, English Ed. (August 1970).

(21) Lobeck, M. A., "Use of the IDC System," *Ibid.*, **9**, 576–83, English Ed. (August 1970).

(22) Nickelsen, I., and Nickelsen, H., "Mathematische Analyse Des TOSAR-Verfahrens," *Inform. Stor. Retrieval* **19**, 95–119 (February 1973).

(23) Farrell, C. D., Chauvenet, A. R., and Koniver, D. A., "Computer Generation of Wiswesser Line Notation," *J. Chem. Doc.* **11**, 52–9 (1971).

(24) Feldmann, R. J., and Koniver, D. A., "Interactive Searching of Chemical Files and Structural Diagram Generation From Wiswesser Line Notation," *Ibid.*, **11**, 154–9 (1971).

(25) Feldmann, R. J., Heller, S. R., Shapiro, K. P., and Heller, R. S., "An Application of Interactive Computing—A Chemical Information System," *Ibid.*, **12**, 41–7 (1972).

(26) Feldmann, R. J., Heller, S. R., "An Application of Interactive Graphics—The Nested Retrieval of Chemical Structures," *Ibid.*, **12**, 48–54 (1972).

(27) Petrarca, A. E., Laitinen, S. V. , and Lay, W. M., "Use of the Double-KWIC Coordinate Indexing Technique for Chemical Line Notations," *Ibid.*, **11**, 148–53 (1971).

(28) Skolnik, H., "The Multiterm Index: A New Concept in Information Storage and Retrieval," *Ibid.*, **10**, 81–4 (1970).

(29) Skolnik, H., "A Multilingual Index Via the Multiterm System," *Ibid.*, **12**, 128–32 (1972).

(30) Lefkovitz, D., and Gennaro, A. R., "A Utility Analysis for the MCC Topological Screen System," *Ibid.*, **10**, 86–94 (1970).

(31) Morneweck, S. T. and Hawthorne, D. G., "Application of the MCC Topological Screen System to a Small File of Pesticides," *Ibid.*, **11**, 116–19 (1971).

(32) Milne, M., Lefkovitz, D., Hill, H., and Powers, R., "Search of the CA Registry (1.25 million compounds) with the Topological Screens System," *Ibid.*, **12**, 183–9 (1972).

(33) Jacobus, D. P., Zabriskie, K. H., and Gordon, M., "Compatibility in Chemical Information Systems," *Ibid.*, **9**, 118–25 (1969).

(34) Garfield, E., *et al.*, "Index Chemicus Registry System: Pragmatic Approach to Substructure Chemical Retrieval," *Ibid.*, **10**, 54–8 (1970).

(35) Granito, C. E., Becker, G. T., And Roberts, S., "Computer-Generated Substructure Codes (bit screens)," *Ibid.*, **11**, 106–10, (1971).

(36) Granito, C. E., and Rosenberg, M. B., "Chemical Substructure Index (CSI)—A New Research Tool," *Ibid.*, **11**, 251–6 (1971).

(37) Corey, E. J., and Wipke, W. T., "Computer-Assisted Design of Complex Organic Syntheses," *Science* **166**, 178–192 (October 1969).

(38) Wipke, W. T., "A New Approach to Computer-Assisted Design of Organic Syntheses," Proceedings of the Conference on Computers in Chemical Education and Research, Northern Illinois University, July 1971.

(39) In the United States Court of Claims No. 7368 (Filed February 16, 1972) The Williams & Wilkins Co. versus The United States. Report of the Commissioner to the Court, Consisting of Opinion, Findings of Fact, and the Recommended Conclusion of Law Submitted Under Order of Reference and Rule 134(H).

(40) *Omnibus Copyright Revision—Comparative Analysis of the Issues,* prepared by the Cambridge Research Institute, Cambridge, Mass., and published by the American Society for Information Science, Washington, D. C., 1973.

(41) Wood, James L., Flanagan, Carolyn, Kennedy, H. E., "The BIOSIS-CAS-Ei Overlap Study: Phase I: Monitored Journals," Presented at NFAIS Conference, Washington, D. C., February, 1971.

(42) Wood, James L., Flanagan, Carolyn, Kennedy, H. E., "Overlap in the Lists of Journals Monitored by BIOSIS, CAS and Ei," *J. Amer. Soc. Inform. Sci,* **23**, (January-February 1972).

(43) "BIOSIS and CAS Announce Program of Operational Cooperation and Coordination," *NFAIS Newsletter* **14**(5), 3 (October 1972).

(44) "BIOSIS Announces New Service." *Information-Part I* **4**(4), 187 (July-August 1972).

(45) "Survey of Commercially Available Computer-Readable Bibliographic Data Bases," Eds., Schneider, J. H., Gechman, M., And Furth, S. E., American Society for Information Science, Washington, D. C., 1973.

(46) Williams, M. E., Schipma, P. B., "Design and Operation of a Computer Search Center for Chemical Information," *J. Chem. Doc.* **10**, 158–62 (1970).

(47) Schipma, P. B., Williams, M. E., and Shafton, A. L., *J. Amer. Soc. Inform. Sci.* **22**, 326–32 (September-October 1971).

(48) Williams, M. E., *et al.*, "Education and Commercial Utilization of a Chemical Information Center," IITRI Report No.

C6156-18, July 1972.
(49) "Allocation of MEDLINE Terminals in Canada," *Information-Part I* 4(6), 315, (November-December 1972).
(50) Herner, S., and Vellucci, M. J. (Eds.), "Selected Federal Computer-Based Information Systems," Information Resources Press, Washington, D. C., 1972.
(51) Kruzas, A. T., "Encyclopedia of Information Systems and Services," Edwards Brothers, Ann Arbor, Mich., 1971.
(52) "Information Retrieval Systems," General Services Administration, National Archives and Records Service, FPMR 101-11.3 (Federal Stock No. **7610-181-7577**), Washington, D. C., 1970.
(53) CODASYL Data Base Task Group, October 69 Report (available from the Association for Computing Machinery, 1133 Avenue of the Americas, New York, New York, 10036).
(54) *Ibid.*, April 71 Report.
(55) *Ibid.*, June 73 Report.
(56) *Feature Analysis of Generalized Data Base Management Systems,* Report of the CODASYL System Committee, 1971 (available from the Association for Computing Machinery).
(57) "Scientific and Technical Communication: A Pressing National Problem and Recommendations for its Solution," A synopsis of the report of the Committee on Scientific and Technical Communication of the National Academy of Sciences-National Academy of Engineering, NAS, Washington, D. C., 1969.
(58) "National Commission on Libraries and Information Science," *NFAIS Newsletter* **13**(3), 14 (June 1972).
(59) "National Commission on Libraries and Information Science Act," U. S. Congress, House Rept. **10666,** 91st Congress, First Session, U. S. Government Printing Office, Washington, D. C., 6, May, 1969.
(60) Cima, L., "Changes at NSF Concerning the Abstracting and Indexing Community," (unpublished paper presented at NFAIS Annual Meeting, March 6-8, 1972, New York).
(61) Tate, F. A., "Handling Chemical Compounds in Information Systems," Annual Review of Information Science and Technology, *Encyclopedia Britannica, Inc.*, **2,** 285-309 (1970).
(62) Jacobus, D. P., *et al.,* "Experience With the Mechanized Chemical and Biological Information Retrieval System," *J. Chem. Doc.* **10,** 135-140 (1970).
(63) Eckerman, E. H., *et al.,* "Processing Data From a Large Drug Development Program," *Ibid.*, **12,** 38-40 (1972).
(64) Pick, R. O., Eckermann, E. H., Schafer, J. A., and Waters, J. F., "Strategy of Data Retrieval and Analysis From Large Biological and Chemical Files," *Ibid.*, **12,** 35-7 (1972).
(65) Brady, Edward, and Branscomb, Lewis M., "Information for a Changing Society," *Science* **175,** 961-6 (March 1972).
(66) Arutiunov, N. B., "State System of Scientific and Technical Information," *Nauchno-Tekhnicheskaya Informatsiya,* Ser. 1,(1), 1-6 (January 1973).
(67) Chernyi, A. I., "Integrated Information Systems," in, "Problems of Information Science," FID 478, Moscow, VINITI, 1972.
(68) Brown, Jack E., "National Scientific and Technological System for Canada," presented at NFAIS Annual Meeting, March 6-8, 1972, New York, to be published in NFAIS Report Series.
(69) "Centre National de la Recherche Scientifique (CNRS) P.A.S.C.A.L. System," *NFAIS Newsletter* **14**(2), 12 (April 1972).
(70) "UNISIST International Serials Data System," *Ibid.*, **14,**(5), 16-17 (October 1972).
(71) "Information Retrieval in the Field of Chemistry," JPRA-53523, National Technical Information Service, July 1971.
(72) Shera, J. H., Foreward to "Documentation and Information Retrieval," J. W. Perry and Allen Kent, The Press of Western Reserve University and Interscience Publishers, Inc., Cleveland, Ohio, 1957.
(73) "U. S. Institutions Offering Degree Programs in Computer Science Data Processing and Related Programs," *Commun. Assoc. Computing Machinery* **10**(10), 675-6 (1967).
(74) "Curriculum 68" *Ibid.*, **11**(3), 151-97 (1968).
(75) Belzer, J., Isaac, A., Finkelstein, E., and Williams, J., "Curricula in Information Science: Analysis and Development," *J. Amer. Soc. Info. Sci.* **22**(3), 193-223 (1971).
(76) Bowman, C. M., *et al.*, "A Chemically Oriented Information Storage and Retrieval System. III. Searching a Wiswesser Line Notation File," *J. Chem. Doc.* **10,** 50-4 (1970).
(77) Bond, V. B., *et al.*, "Interactive Searching of a Structure and Biological Activity File," *Ibid.*, **11,** 168-70 (1971).
(78) Bond, L. *et al.*, "A Computerized Current Awareness Service Using Chemical-Biological Activities (CBAC)," *Ibid.*, **9,** 158-61 (1969).
(79) Petersen, D. R., "Computer-Assisted Information Services in The Dow Chemical Company," Paper presented at the American Institute of Chemical Engineers Meeting in Houston, Texas, March 1971.
(80) Schrodt, V. N., *et al.,* "Information Communication System," *J. Chem. Doc.* **11,** 174-8 (1971).
(81) Hoffman, W. S., "Du Pont Information Flow System," *Ibid.*, **12,** 116-2 (1972).
(82) Hyde, E., Lambourne, D. R., and McArdle, L. A., "Use of a Multi-Level Substructure Search System—Survey of User Queries," Private communication from ICI, March 1972.