Chemical Abstracts Service's Secondary Chemical Information Services

DALE B. BAKER

Chemical Abstracts Service, Columbus, Ohio 43210

Received February 22, 1985

A current environment of the information-transfer activities, involving backward and forward integration, is presented. Chemical Abstracts Service's (CAS) activities as a secondary database builder, a systems developer, an information distributor, and an international network node operator are described. Its strategic planning and the continuing interaction with users are highlighted. The future of CAS secondary services is discussed with reference to overcoming barriers to flow of information, taking advantage of technological advances, and striving toward international cooperation.

INTRODUCTION

I have had the great good fortune to be intimately associated for more than 40 years with a remarkable organization made up of remarkable people whose outstanding intelligence, knowledge, creativity, energy, and devotion to a worthy cause have assured their place in history. Not their place, personally, perhaps, but certainly the place of the institution, Chemical Abstracts Service (CAS), that they and I belonged to—literally—is assured.

What follows is a reminiscence, a highly personal retrospective view as well as a projection into the future, centered on CAS. What have we done that was important? Why did we do it? What was the environment? What else might we have done? What else should we do in the years ahead?

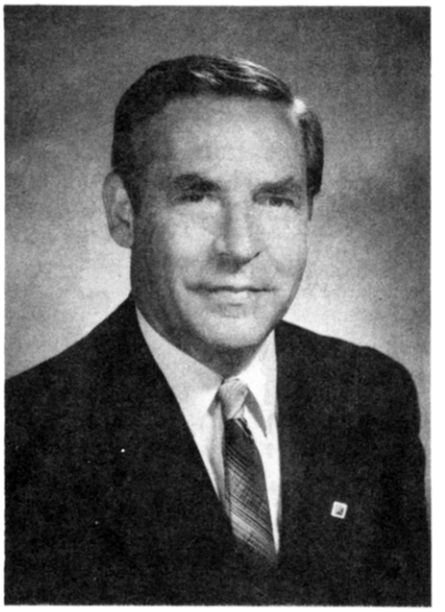
Certain major trends, accented by a few revolutions, are readily visible over the long haul of history of human communication. There was the invention of language, whenever, and the invention of writing some 5000 years ago. The printing press again revolutionized communications some 500 years ago. And, electromagnetic communication via telegraph some 150 years ago permitted delivery of messages at the speed of light. With the inventions of telephones, radio, and television, electronic modulation today transmits voice and pictures as well. Today the two dominant electronic technologies revolutionizing our field are computers as they are used primarily to manipulate information in very general ways and communications technology as the mechanism for electronic transporting of information from place to place. Some scholars consider the computer on a par with the printing press in significance to humanity.

There is virtually no limit to what can be done today in the manipulation and communication of information via the spectacular developments of recent years. The technology today exceeds our capacity to utilize it. The major challenges (or sources of conflict) in information and communication systems today are not technological, but economic, political, and social.

This brief sketch of technological developments describes the platform on which virtually all information services stand today: for the balance of my comments, I will focus more closely on the information systems for chemistry and chemical engineering and still more closely on CAS.

No other scientific or technological disciplines enjoy older, larger, or more widely utilized information systems or services than does chemistry. The many 19th century landmarks of German chemical documentation attest to the dedication and accomplishments of chemists advancing their discipline. For 110 years, the American Chemical Society (ACS) has pioneered in all areas of the information-transfer chain. We have not been the only innovators, but ACS has been the first to provide many useful features to the information flow from authors to users.

The major thrust for the ACS toward what some call the "paperless society" stems from the application of computers to publishing, which started in 1959. Computer-controlled photocomposition or electronic publishing has now been applied



Dale. B. Baker, Deputy Executive Director, American Chemical Society (ACS), and Director of Chemical Abstracts Service (CAS), began work for CAS as a part-time employee in 1938 while attending the Ohio State University (OSU). He obtained from OSU a B.Ch.E. in 1942 and a M.Sc. in 1948. He worked as a chemist and supervisor for Du Pont in 1942–1946. He returned to CAS full-time in 1946 as an Assistant Editor and became Associate Editor in 1951, Associate Director in 1958, Director of CAS in 1958, Chief Operating Officer ACS, Columbus Offices, in 1980, and Deputy Executive Director, ACS, in 1983. He has served on many boards and committees, some of which include International Council for Scientific and Technical Information (presently General Secretary), International Union of Pure and Applied Chemistry, UNESCO, American Association for the Advancement of Science (Vice-president and Chairman, Section T), ACS, American Society for Information Science (ASIS) (President), American Standards Institute Z-39, National Academy of Sciences/National Research Council, National Federation of Abstracting and Information Services (NFAIS) (President), and Institute for International Information Programs. Awards include an ASIS Award of Merit, an ACS Patterson-Crane award, and a NFAIS Miles Conrad award.

0095-2338/85/1625-0186$01.50/0

Chart I

Structure of the Information Industry (1984)

backward integration ←

→ forward integration

information sources → database producers → vendors → communications services → users

to the full array of ACS publications: scholarly journals, abstracting and indexing services, books, monographs, and other publication types. Forecasts indicate that most scientific and technical publications worldwide will be produced under computer control by 1990–1995.

In the past, most people thought of the information-transfer process, if they thought it at all, as a sequence of discrete, value-added steps, each step performed by a different organization or by different units within an organization, each with special competence to do its particular job. These once sharp lines of demarcation between the links in the information chain have broken and have been replaced by a continuum. Steps in the process have begun to blur. The changes in who does what in information handling are analogous to the well-known phenomenon of business integration. Integration occurs when an organization begins to perform business activities that are also performed by either its suppliers or its distributors. When an organization moves closer to its sources of supply, it is integrating backward (Chart I). Forward integration occurs when an organization moves closer to the ultimate consumer or user of its products and services. Integration, in both directions, is a very strong trend in information organizations today.

While technological advances are making it easier for integration to occur, it is basically the information users who are driving these developments and who are the ultimate beneficiaries of these changes because of the following: information services will become more responsive to users' needs. The information-transfer chain will be shortened and will become more efficient. More innovation and rapid improvement in information services are likely to result. Overall costs to users will be less.

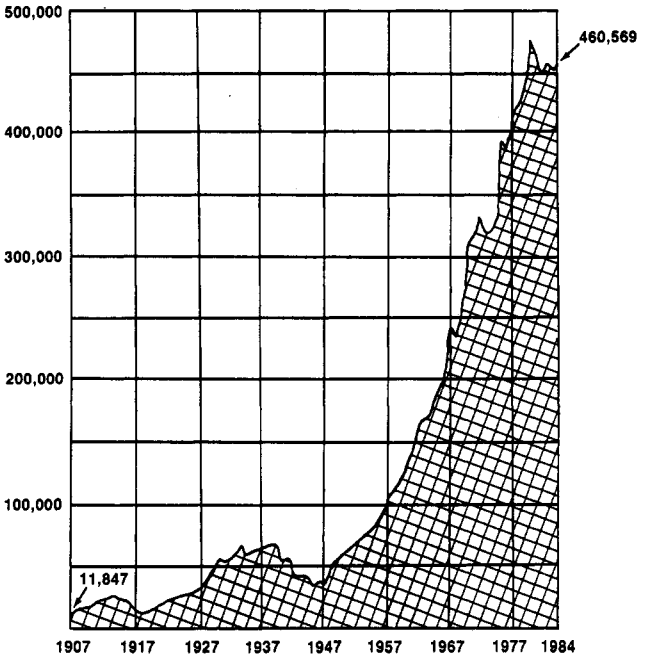
CAS AS A SECONDARY DATABASE BUILDER

Against this terse and simplified background, let us now focus more closely on secondary chemical information services. My assignment in this paper is not to provide a scholarly review of the more than 50 secondary services in the world that cover a large or small part of the literature of chemistry and chemical engineering. I will, as I suggested at the outset, concentrate on the oldest and largest remaining integrated service that covers the entire discipline, namely, Chemical Abstracts Service.

In its first 50 years, CAS employed a simple and direct strategy. We intended to be comprehensive, accurate, and as timely as possible. CAS still maintains these objectives, and we try to keep the prices of our services as low as feasible. In the past 27 years, the world has become much more complex than when we first adopted that basic strategy.[1]

The mission of CAS today is to be the "leader in providing a family of diversified information services related to worldwide activities in chemistry and chemical engineering to meet the needs of the international scientific and technical community." In order to accomplish this mission, CAS policies and programs are guided by several principles. Among these principles is a basic one that states that CAS services will strive for a balance among coverage, quality, timeliness, consistency, continuity of service, and price. The user community will only gain maximum satisfaction from CAS publications and services if they are derived from a database of the highest technical integrity. While CAS is motivated to *maximize*

Chart II. Annual Growth of Number of CA Abstracts Published (1907-1984)

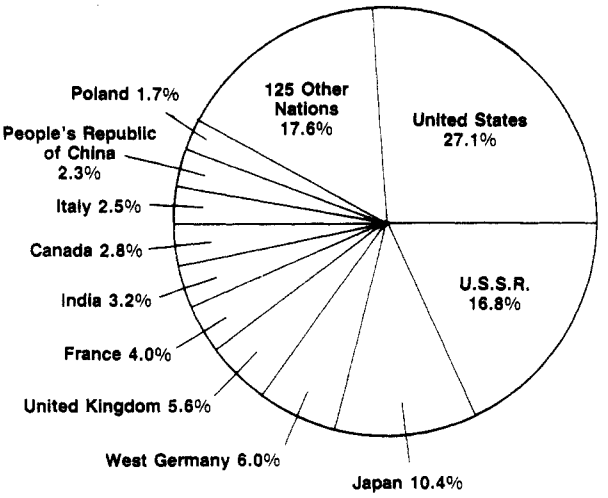


coverage, *maximize* quality, and *maximize* timeliness, achieving all three of those goals is impossible: CAS must achieve an *optimal balance* among those characteristics. CAS will seek to maintain that balance within the limits of prudent economics, pricing its publications and services carefully so that they are realistically accessible to as many potential users as is consistent with the organization remaining financially healthy.

CAS subject scope is chemistry and chemical engineering, definitions of which are not only very broad but also change to keep pace with changes in the disciplines. In addition, CAS will continue to collaborate with organizations in other disciplines for mutual benefit in serving interdisciplinary user groups. CAS is international in scope, with respect to the primary sources we cover and users we serve, whom we reach with our printed publications and computer-based services. CAS document analysis is designed to ai d access to primary information sources rather than to supplement them. The ACS has a strong, active program of projects interrelating the primary journals and secondary services with the objective of improving the efficiency of production of these media and increasing overall services to users. The CAS document delivery service today is operated to support publishers rather than compete with them in helping users find the information they want.

CAS monitors the scientific and technical literature of the world to select documents of chemical and chemical engineering interest. Periodicals, monographic series, technical reports, conference proceedings, books, and patents are monitored. CAS does not deliberately restrict its coverage of documents on the basis of country except for its coverage of patent documents. In 1984, CAS covered documents in 45 different languages reporting on research performed in 141 countries. The CAS patent document coverage extends presently to 26 countries and 2 international patent organizations. Chart II shows the annual growth of the number of *Chemical Abstracts* (CA) abstracts published, increasing from 11 847 in 1907 to 460 569 in 1984. Additionally, there were 111 239 equivalent patents, for a record yearly total of 571 808 documents cited in 1984. Chart III shows the national origin of papers abstracted in CA in 1984. In 1984, there were 1 258 849 general-subject index entries and 2 161 083 chemical-substances index entries generated for the semiannual

Chart III. National Origin of Papers Abstracted in CA in 1984



volume indexes. There were 563 390 new chemical substances added to the CAS Registry, which totaled 6 910 103 unique substances in the 20-year-old file along with 10 570 000 total names.

CAS AS A SYSTEMS DEVELOPER

The database is the traditional basis of CAS's worldwide reputation and remains the organization's most valuable asset, but CAS has become very strong as a systems developer. A very large component of the secondary database building process is labor—highly skilled scientific and technical experts (as well as supporting staff)—required to intellectually analyze the literature, abstract, index and name chemical substances systematically, etc. CAS recognized early that automation was urgently needed to assist these labor-intensive tasks, to increase productivity as far as possible, and to improve the economics of this part of the information chain. But, there was no software in the early 1960s to run the computers to support the CAS editors and to automate the CAS operations. Thus, in the last 25 years, CAS has had to become a major software producer for its production facilities for secondary services (as well as for the primary journals of the ACS), to continue to be cost competitive.

There have been many demands from large industries and academic, publishing, and government groups for the kind of special functions that the CAS software supports. CAS has applied its software and systems skills to undertake special projects for the National Cancer Institute, the Environmental Protection Agency, and the European Economic Community, for example, and most recently for the U.S. Patent and Trademark Office. Major corporations are also using CAS software to process proprietary information. CAS is now developing business strategies to provide customers a range of software products under controlled conditions, in effect creating an important diversification of CAS business, consistent with the CAS mission.

CAS AS AN INFORMATION DISTRIBUTOR

CAS has always supplied information directly to users (individuals now often referred to as "end users") through printed publications and microforms of those publications. With the new computer-based services, CAS had moved toward being an *indirect* supplier as it was a major source of tape files in the early years (since 1962) of computerized scientific and technical information. These files were used largely by information specialists who, in turn, served end users. After a few years in that mode, CAS realized that if it was to continue to be truly user oriented, CAS must strengthen its direct relations with its users. Excellence in database building *alone* was not sufficient to assure a future that is economically vigorous and sufficient to satisfy the CAS stated mission. Discussions with the CAS Advisory Board over several years indicated that *delivery* and *use* was where the economic growth would occur and that segment of the activity would have the leverage to "call the shots" in the rest of the information chain for policies of *coverage*, *content*, and *delivery form* as well as for the *acceptable costs* of database building. Delivery and use are clearly the parts of the information chain nearest to the user, and as use-oriented payment for information becomes a more dominant factor in information industry economics, its leverage increases. Thus, the CAS database building and service delivery strategies became closely interrelated, each supporting the other.

Direct *on-line* delivery of CAS Registry services from Columbus, OH, started in 1980 as a major move toward returning to closeness with users. The development of CAS's Messenger search software and advances in on-line technology helped open the way for greater variety in the kinds of information and data that can be made available to users, by making possible greater breadth and depth of CAS offerings. CAS considered the investment in the development of the CAS ONLINE delivery system as prudent as the earlier investment in the production systems. CAS's unparalleled expertise in manipulating chemical information gives us the flexibility to optimize the service in ways not practical for other vendors, who must strive for general-purpose systems and services.

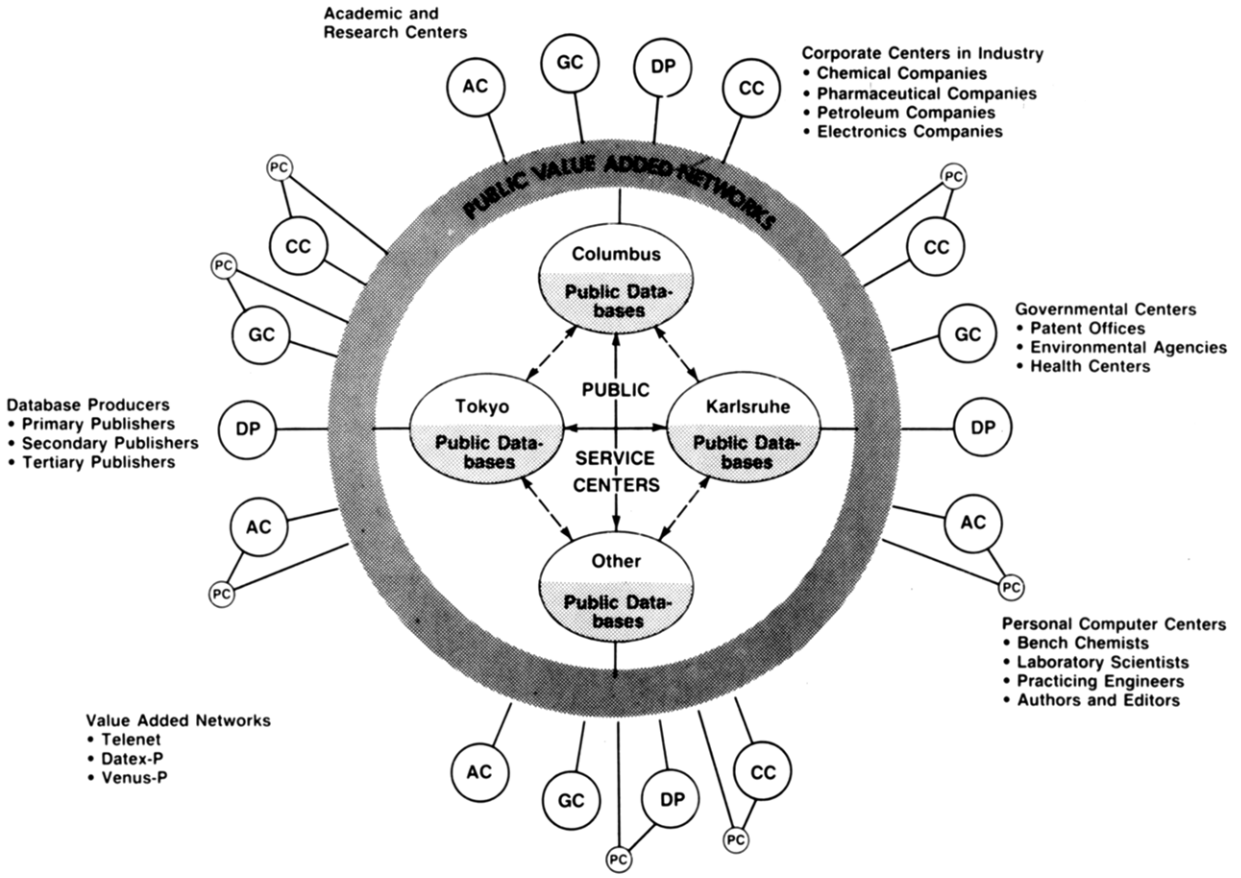
For years, users have expressed their wishes and needs for types of information and data that are in the primary sources analyzed by CAS but that are not presently covered in the CAS databases. To meet these needs, the present databases are being enhanced, and new databases are being built. Improving the patent-related content of CAS files and building new databases to support new search services for reactions and for Markush structures in patents are several examples of new and improved services that will appear over the next several years, beginning in 1986.

CAS AS AN INTERNATIONAL NETWORK NODE OPERATOR

Advances in telecommunication technology provided both the motivation and the means for new approaches to international cooperation. The opportunities for development of a network of international scientific and technical information centers, located in various parts of the world, are more favorable today than ever before. Networking among systems is inevitable because not all knowledge will be accumulated into a single system or center. This is true because of the pluralistic nature of our society and business environment and of the science disciplines themselves and not because of any fundamental technical limits. The major, current motivating forces for networking are as follows: Digital data stored at distant locations can be made available rapidly and efficiently through mutually compatible systems. All nations and organizations are eager to improve their systems and services. Financial pressures are forcing organizations to consider ways of sharing rather than duplicating materials, resources, and expenses.

In London, in December 1983, CAS was one of the parties announcing creation of STN International, a homogeneous network that will function as a nonprofit, on-line service bureau for database producers who wish to offer their own files on-line, under their own control. Joining CAS in the announcement was FIZ-4, the Fachinformationszentrum Energie, Physik, Mathematik GmbH. CAS in Columbus, OH, and FIZ-4 in Karlsruhe, West Germany, have formed the first STN link. Services are now being provided through the link, and work

Chart IV. STN International: A Worldwide Integrated Information Network



is actively going on to upgrade and expand network functions (see Chart IV).

In early 1984, a dedicated communication link was established connecting Tokyo and Osaka with the Columbus STN Service Center, and the Japan Association for International Chemical Information has become an interim service center for the network. Negotiations are currently under way toward establishing a fully operational service center in Tokyo.

Each of the public service centers will load databases of interest to the scientific and technical community. The STN user through personal computers or corporate centers (industry, academia, or government) will in time be able to search the full range of databases in all major areas of science and technology, including those physically loaded at other participating public centers. The user will conduct such multiple database searches in a single session by using a single command language (available in natural languages other than English at some of the centers).

CAS STRATEGIC PLANNING AND ITS IMPLEMENTATION

CAS's first 5-year planning goes back to 1964, and objectives of that plan to achieve a modern man–machine information system was essentially realized in 10 years (1965–1975).

CAS's annual strategic planning since 1975 has become more structured and rigorous. The CAS mission and "living" goals are reviewed and refined as needed, at least annually. CAS's understanding of the environment, including customers, competitors, and the state of information handling technology, is almost continuously scrutinized as are our internal strengths and weaknesses. CAS's selection of strategic alternatives and the development of business plans are critical procedures before any significant new program is implemented. As our production systems and service offerings rapidly became more complex and the economics of CAS became radically different than during the first 75 years, the task of obtaining deep and detailed understanding of the financial implications of all policies and activities has also become much more complex as it has become more critical. It was essential that we devote considerable effort to achieve a more intimate strategy–finance relation and understanding of that relation. That has been a 3-year task. The guidance, counsel, and opinion of the ACS governance are also essential to CAS policy formulation and decision making on all new product development.

In general, CAS's long-range future of present and new services will be based on the approach to planning identified above, which has confirmed CAS's current major strengths, as follows: (1) Experienced document acquisition staff with worldwide contacts. (2) Strong foreign language skills. (3) Large number of subject-trained, literature analysis staff. (4) Large computer systems research and development staff oriented toward information handling, concepts, and techniques. (5) Extensive decentralized computer hardware facility and skilled management thereof. (6) Sophisticated, diversified marketing staff specialized in scientific and technical information (STI) activities, including training of users. (7) Excellent reputation worldwide for quality and dependability. (8) Established national and international associations with scientific societies, industry, academia, and government organizations.

Coupled with the principal themes in users' lists of most wanted information services, CAS explores almost daily the ways in which its corporate strengths can be turned to advantage in providing new services and/or in the generation of new revenue.

Users, of course, are ultimately CAS's best critics. They look very carefully at how we have described their own reported work and that of their colleagues. They call our attention to omissions, inaccuracies, inconsistencies, and problem

areas. We especially invite and welcome input from our users on the quality and utility of our services. It is only in response to such input that we can seek to assure that our services do truly continue to meet the chemical information needs of our users.

Users continue to identify additional detail or information that they would like to see incorporated into the CAS database. CAS, in fact, actively and sincerely seeks such input from its users. CAS has three established user councils—North American, European, and Japanese, as well as an International Advisory Committee. Extensions and enhancements to the CAS database content, priorities, and strategies are among the principal discussion topics at those meetings.

It is not possible or practical for CAS to implement all of the suggestions that have grown and will continue to grow out of those discussions, or will arise from other internal and external sources. Nevertheless, those discussions and suggestions do help to give us valuable insight into our users' problems and needs. This input will help us to set our planning, research, and development priorities for the months and years ahead.

Most suggestions for database enhancements and extensions carry with them the requirement for additional intellectual effort. This means the need for more professional staff. It also means increased costs for CAS services unless those enhancements result in increased or new uses of CAS services, and thus additional revenue. The determination of which database enhancements CAS will pursue is based on a very careful consideration of the most pressing and general user needs and the organization's capabilities.

FUTURE OF CAS SECONDARY SERVICES

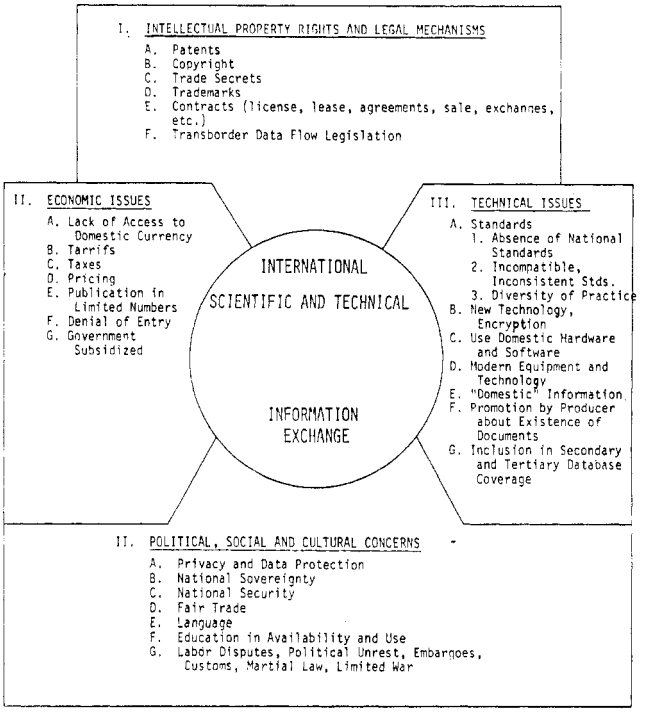
At the beginning of this paper, I pointed out that the major challenges today are not technological but economic, political, and social. Recognition of the role and of worth of information and of its potential economic and political power in our societies has already prompted some nations to develop policies to regulate, manage, and in some cases, especially, to control the electronic flow.

Some countries have called for a world information order, and others have suggested or established protectionist barriers that would filter information into and out of the country according to the sovereign interests of that country. The range of concerns are legal, economic, political, social, and cultural (see Chart V).

I call attention to these barriers as, in my opinion, they can and will, more than anything else, affect the future nature of ACS/CAS information activities. It seems imperative that the international scientific and technical information (STI) communities try to ascertain the potential adverse effect of these complex and critical business issues on the "free" and uninhibited flow of STI. The database producers as centralized collectors of information and as deliverers of information have a great stake in these developments and should guard against the imposition of restrictive constraints by governments. In my opinion, we must work together continuously to act as necessary to meet and overcome these challenges and assure the uninhibited flow of scientific and technical information and the development of open, global telecommunications systems.

This brings us to what we see in the crystal ball as to the future of secondary services. If you accept the adage "What is past is prologue" and the visible trends cited above, then marvelous things are in store for the secondary database builders and for the on-line delivery systems. The future of CAS will be shaped by the agility and imagination of management in coping with emerging technology and in responding to its impact on the information commodity markets.

Chart V. Barriers to International Flow of Scientific and Technical Information



All of us as stakeholders must continuously strive toward an idealized international chemical information system and services that share resources, increase efficiencies in collecting, processing, and disseminating information, and reduce overall costs. World politics and the nature of the existing STI complex suggest that there is not an internationally shared holistic view of the ideal-world STI system. And, we are apparently not motivated to change our habits easily in order to build the desired systems. But we can, and all nations should, concentrate on creating significant contributions to a better world information order. Each of us can offer something to exchange or barter, rather than trying to create individual islands of independence, at great duplication of effort and expense to the community of scientific users. We should strive continuously for information exchange and information peace in the years ahead.

As indicated earlier, rapid and profound changes will be made in the next decade in the information chain in areas of (a) input, (b) storage and processing, and (c) delivery–retrieval.

The use of abstracts and indexes that has now been around for centuries has provided scholars with quicker and easier methods of keeping up with and retrieving information from a growing literature in their subject fields. The growth of the use of abstracts paralleled the growth of the subject literature. As the latter became very large, indexing of both the abstracts and the corresponding original document was required. We see no fully suitable alternatives today to the use of abstracts and indexes. The automatic on-line search of indexes and full text of abstracts today is aiding users to determine which full-text documents are to be searched. Even with very low-cost computer storage, we should keep in mind the proportional sizes of the files that will be called for. Files of index entries and abstracts are already huge but will seem small compared to full-text primary paper files. The advantages of the screening or filtering processes and the value-added efforts of the secondary services will long be continued for alerting and retrieval purposes. Improved cooperation between secondary services and primary publishers, however, is expected to open new vistas not dreamed about today.

Many scientists and engineers today, as well as secretaries and typists, are using word processors and computers to put

STI into electronic form for storage, formating, editing, and transmission. Although the number submitting machine-readable manuscripts now directly to publishers is small, some important experiments are under way to help determine the economics and feasibility of such systems. These processes have compelling advantages to both the author and publisher, such as reduced proof-reading, decreased costs, improved quality of production, and faster production time. In my opinion, the major obstacles we see today will be overcome rapidly in the next 5–10 years so that this electronic input will be available and used for input to primary, as well as secondary and tertiary, services as an integrated set of activities.

Capturing electronic input will allow for all kinds of numerical, factual, and other specialized, in-depth information databases to be processed and stored. It is clear that the users mainly want to retrieve factual information directly from the information systems of the future. For this purpose, the organization of the database becomes a more important consideration, because efficient retrieval from the conventional linear text format may not be possible. In other words, not only abstracts or primary articles need to be organized within a database, but all the facts in the abstracts or primary papers need to be organized on the basis of content rather than merely of structure, which is a more complex task.

The retrieval process itself will become sophisticated enough to take advantage of the relationships built into the organization of the data and other relationships that may only be implicit in the organization and that may be specific to the chemistry discipline involved. Instead of arbitrarily restricted query languages, new systems will accept questions posed in natural languages, or in a reasonably "natural" artificial language. I have confidence that the insights gained from artificial intelligence research and the expert systems being developed will help greatly in this respect in the years ahead.

It seems obvious that the functions and roles of the centralized, comprehensive database builder, processor, and disseminator of chemical information, such as CAS, will continue to be essential for an information-demanding world for long into the future. A pivotal issue is how to distribute information from a central store to a worldwide community of users and do so in such a way that each receives what is needed in a timely and convenient fashion but pays only for that received and, conversely, receives only what is paid for. Of great, possibly paramount, importance is the flow of revenue enabling the maintenance of the collecting, analysis, and processing functions essential to building the databases. The information business side of ACS/CAS is going to become more like that of the telephone company, which sells and distributes huge quantities of relatively inexpensive items rather than small numbers of costly items. A consequence of this change is that revenue will become harder to predict, at least until the new-order information commodity market is established and its economics understood.

Because CAS is the only single comprehensive chemical information service in the English language, its continued viability is important to all nations who wish to build their future information activities on this 78-year-old record and store of chemistry and chemical engineering knowledge. We have renewed and extended our longstanding agreements of cooperation with The Royal Society of Chemistry in the United Kingdom, Centre National de l'Information Chimique in France, VCH Verlagsgesellschaft in the Federal Republic of Germany, and the Japan Association for International Chemical Information, and we have entered into an additional agreement with Fachinformationszentrum Chemie in the Federal Republic of Germany.

An overriding purpose of ACS/CAS today, as it has been for more than 75 years, is to promote advancement of science and technology through a conscientious effort to serve the chemical information needs of the worldwide scientific community with services of the highest quality. The American Chemical Society's Board of Directors reaffirmed the Society's dedication to that purpose at its December 1984 meeting when it adopted unanimously the following important statement of policy and intent:

> "The Board of Directors of the American Chemical Society recognizes that the Chemical Abstracts Service database is an irreplaceable international resource that could not have been built without international support and the unimpeded flow of information among many nations. The Board is deeply aware of the American Chemical Society's obligations to ensure that the content of the CAS database remains accessible to information users in all nations and that it is not exploited for purely commercial or for nationalistic purposes."
>
> "In keeping with the Society's Charter, it has been a major, longstanding goal of the Society to provide the best possible chemistry-based information services. We reaffirm the Society's commitment to that goal. We also reaffirm the Society's longstanding commitment to the philosophy and principle of cooperating and sharing resources in practical ways with scientific and technical societies and other organizations in the U.S. and other nations, with the aim of developing comprehensive international information systems, services, and networking arrangements."
>
> "The Board reaffirms its intent that the Chemical Abstracts Service will remain an integral part of the American Chemical Society and will continue to be operated as a non-profit, self-supporting, service-oriented Division of the Society."

This statement recognizes the mutual interdependence that exists among scientific and technical organizations in many nations and reaffirms the Society's commitment to cooperate and share resources with those organizations to ensure effective collection and dissemination of chemical information and convenient access to the CAS database in all nations.

REFERENCES AND NOTES

(1) Baker, D. B.; Horiszny, J. W.; Metanomski, W. V. "History of Abstracting at Chemical Abstracts Service". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 193–201.