

International Sharing of the Production and Distribution of Chemical Information Services[†]

RUSSELL J. ROWLETT, JR.

Chemical Abstracts Service, P.O. Box 3012, Columbus, Ohio 43210

Received May 14, 1979

The international aspects of chemical information services by Chemical Abstracts Service (CAS) are described.

Although originated in 1907 by the American Chemical Society (ACS) as essentially a national service, *Chemical Abstracts* (CA) has always been international in its content. Since the demise of *British Abstracts* (in December 1953) and *Chemisches Zentralblatt* (in December 1969), Chemical Abstracts Service (CAS) has become the principal abstracting and indexing service in chemistry and chemical engineering for most of the world. Today, it is truly an international service. More than 73% of the 498 559 papers, patents, and other documents cited in *Chemical Abstracts* in 1978 came from countries other than the United States. About 65% of the CAS subscribers reside outside of the United States.

The CAS data base has long been produced by an international, cooperative effort. Until recently, most CA abstracts were prepared by volunteer abstractors who lived around the world. Without this cadre of international scientific reporters CA could not have existed. Gradually, in recent years, abstracting has been combined with indexing, about 85% of it within the Columbus offices, into a single intellectual document analysis step. A number of computer systems provide assistance and automatic edits for this unified analysis operation, freeing the scientific staff to concentrate on questions pertaining to technical content. Thus, the number of volunteer abstractors has decreased from a peak of 3600 to less than 1200. Seventy-three percent of the present volunteers live outside of the U.S., chiefly in Japan, Czechoslovakia, India, and Poland. There are obviously language capabilities and subject expertise which continue to be required in this much smaller, but still important, group of international workers.

Our Columbus staff has often been referred to as a "little United Nations". We have native-born scientists from many of the some 135 countries from which we receive chemical documents, and these scientists are capable of handling most of the 55 languages in which these documents appear. There is also direct input of technical content from the United Kingdom which will be described later. A few years ago, in reviewing the handling of chemical information, the National Science Foundation published a paper entitled "It All Comes Through Columbus". Truly, if the information is chemical from anywhere on earth, it eventually comes through the Columbus offices of CAS.

Much of the current international sharing of the production and distribution of chemical information services stems from discussions which began 12 to 15 years ago between the American Chemical Society and the Organization for Economic Cooperation and Development (OECD). In December 1967, an OECD Working Panel on Chemical Information Systems, after three years of study, issued a final report on

"Chemical Information: National and International Responsibilities". This report recommended the establishment of an international chemical information system based on the ACS system developed at CAS. As a consequence of this report, early in 1968, the ACS made a proposal to the OECD Working Panel "for cooperative international action aimed at an early and widespread experimentation on the new mechanized information services being developed by CAS".

Following these actions, over a period of years, 12 scientists and information specialists from 6 OECD nations plus Japan spent approximately one year each as full-time CAS employees in Columbus. These visits constituted an "intern training" program. In almost every instance, the scientists, upon returning home, helped establish a national information service utilizing CAS computer-readable files. Many of these services have grown and include today data bases on subjects other than chemistry. They got their start from the international initiative generated from the joint ACS-OECD discussions and agreement. CAS has continued the practice of "interns" from other nations interested in developing chemical information systems.

The OECD discussions in 1968 and later enunciated the ACS policy that CAS computer-readable services would be made available to all nations on a license basis for use in their own country or countries. Prior to this time and continuing, CAS/ACS had working relationships in R&D with organizations in at least five nations. The Chemical Society (London) established one of the first centers which concentrated on the use of CAS computer-readable services and made many basic contributions to the development of text search capability. BASF, Bayer, and Hoechst, in the Federal Republic of Germany, developed a substructure search and information retrieval system which today accepts input directly from the CAS chemical substance identification files. The loose affiliation of these three German companies was succeeded by the formation of the Internationale Dokumentationsgesellschaft fuer Chemie (IDC) which increased the number of participating companies to eleven.

In the late 1960s and early 1970s, the French developed a sophisticated structure-handling system which provided substructure search capabilities. Their Association for Research and Development in Chemical Information (ARDIC) worked for many years with the CAS Research and Development Division to interconnect the chemical structure records contained in the two organizations' computer files. More recently, ARDIC has implemented a text search system and has integrated this with their structure search capability for a large segment of the CAS abstract-index files.

The three cooperative research efforts are indicative of the important work underway around the world. One of the most used substructure search services which employs CAS files as primary input is provided by the Basel Information Center for

[†] Presented in Division of Chemical Information, American Chemical Society/Chemical Society of Japan Chemical Congress, Honolulu, Hawaii, April 2, 1979.

Chemistry (BASIC), an information center owned by Ciba-Geigy, Hoffmann-La Roche, and Sandoz Limited of Switzerland. Japanese groups, principally the Japan Information Center of Science and Technology (JICST), have also been active in information dissemination research.

With this growing international interest in the use of the CAS files of chemical and chemical engineering information coupled with the OECD agreement, it was only natural that several nations looked to ACS/CAS for more formal ties and for some participation in the decisions associated with the development of the file contents. Understandable also was the ACS/CAS interest in sharing some of the awesome operational, developmental, and financial responsibilities for maintaining the ever-growing data base of chemistry. In addition, the availability of on-line information services worldwide demanded local promotion and local question-answering capabilities. While all chemical information would necessarily continue to be centrally processed in Columbus, it was becoming increasingly difficult to provide all the personal interaction so necessary for effective use of the new developing services.

Four bilateral national agreements have resulted from this philosophy of international sharing of the production and distribution of chemical information services. The Chemical Society entered into a joint agreement with ACS, effective Jan 1, 1970, under which the United Kingdom Chemical Information Service (UKCIS) provided CAS abstracts and specified index entries for papers and patents published in the U.K. and marketed publications and information services derived from the CAS data base in the U.K. and Ireland. The UKCIS was formed as a directorate of The Chemical Society to weld together the interests of ten scientific societies in the U.K. under The Chemical Society's leadership. It has been a very active group in the development of computer-readable services using CAS and other data bases. During 1975, a new ACS-Chemical Society agreement was negotiated and became effective Jan 1, 1976. Like its predecessor, the new agreement continues British input of abstracts and index entries. While all four of the bilateral agreements allow input, only the British today find input economically advantageous.

Also during 1969, the ACS and the Gesellschaft Deutscher Chemiker (GDCh) of The Federal Republic of Germany negotiated an agreement similar to that with The Chemical Society. It too became effective on Jan 1, 1970, and initially included input of abstracts and index entries prepared from West German papers and patent documents. A new agreement was negotiated during 1974 in which GDCh was replaced as the West German representative by IDC. This agreement became effective on Jan 1, 1975, and while it greatly increased the West German participation in CAS activities, IDC chose not to provide input of abstracts and index entries. The increasing inflation rate and the continuing devaluation of the American dollar combined to render input a much less attractive financial venture for IDC.

After some years in the making, a third bilateral agreement was signed in Sept 1977 between ACS and the Japan Association for International Chemical Information (JAICI). The latter group was organized specifically to provide a suitable vehicle for Japanese participation in CAS operations. Its membership includes 27 chemically oriented scientific and technical organizations and more than 200 Japanese corporations.

Marketing responsibility for CAS publications and files was transferred to JAICI in two steps. JAICI immediately took over marketing and distribution of CAS computer-readable files, programs, and documentation in Japan, and CAS worked with the Japanese organization in developing services for the Japanese chemical community from the CAS information

base. Having established the ability to develop services of its own and gained sufficient user support for these services, JAICI assumed full and exclusive responsibility for marketing and distributing all CAS publications, services, computer-readable files, and other materials for use in Japan on September 1, 1979.

Japan contains the largest concentration of subscribers to CAS publications and services outside of the United States. It also at present is the third largest contributor to the world's chemical and chemical engineering literature.

Under the terms of the fourth agreement signed in Paris early in 1978, France's Centre National de l'Information Chimique (CNIC) has the right to use all CAS data and computer programs and the exclusive right to distribute CAS-produced publications, microfilm services, and computer-readable information files in France. CNIC is a nonprofit association established by organizations representing academic, governmental, industrial, and learned society interests in France to coordinate French information activities in chemistry. Its membership includes the Centre National de la Recherche Scientifique (CNRS), which manages the principal public research laboratories in France, operates a bibliographic information retrieval service covering some 9 million scientific and technical documents, and publishes the *Bulletin Signalétique* series of abstracting journals.

Under the British, West German, and French agreements, the participating national organization pays a share of the annual cost of producing the CAS data base. In addition to the marketing rights in their geographical areas for CAS publications and services, the agreement countries also receive the right to participate in discussions on CAS operations and plans in meetings of the two CAS Advisory Boards, the ACS Committee on CAS, and regular ACS Board and Committee meetings concerning CAS activities. The ACS has some reciprocal opportunities to participate in similar Board and Committee meetings in the agreement countries.

In their efforts in the more widespread use of the CAS data base, each of the agreement countries has adopted somewhat different strategies. The UKCIS concentrated initially on providing access to subsets of the total abstracts and to the development of the information center concept. Their early studies of techniques for development of off-line and then on-line search profiles were a major help to all who have followed. Their *Macropfiles* printed service, although it did not include abstract text, was a forerunner of CAS' current successful *CA Selects*.

The West Germans have given their chief attention to use of the CAS chemical substance files. IDC has developed an extensive data base and search system which includes their GREMAS system for coding and searching information on the structures, reactions, and properties of organic compounds. Computer programs have been developed for deriving the structure codes used in their GREMAS system from the structure records in the CAS Chemical Registry System. IDC plans to combine Registry structure data with other information extracted from the CAS data base to permit correlations of structure with activity and computer searches for specific chemical reactions.

CAS and CNIC in France have been cooperating since 1972 on the development of chemical information retrieval methods. The ARDIC research group at the University of Paris under the direction of Professor Jacques-Emile Dubois has converted the CAS computer-readable structure notations for more than 500 000 substances into the topological code used to represent structures in the French DARC computer-based system. The group also has developed computer programs for conducting searches of the structure notations in combination with searches for associated information on properties contained

in the abstracts in CAS' *Chemical-Biological Activities* computer-readable file. CNIC expects to make this integrated text-structure search system available for general use in France.

Because CAS is the only single comprehensive chemical information system, its continued financial viability is important to all nations who wish to build their future information activities on this 72-year-old record of chemistry and chemical engineering. The financial support from the organizations in the agreement countries assists greatly in preserving this viability. The responsibility to assure the continuation of the service is thus shared among organizations in five nations rather than resting only on the shoulders of the ACS and the United States. There is one difference in the basic financing of the agreement country operations and those of the ACS. The difference may become more significant in years to come. The organizations in each of the agreement countries have received financial assistance from their national governments. There is no such support, at present, for the ACS information operations from the U.S. government, although in the 1965-75 period CAS did receive some \$25 million for R&D.

Along with the obvious financial advantages and the joint R&D efforts which have been described, this five-nation international cooperation provides the support of respected leaders and organizations in the four countries and helps to build understanding of mutual interests, philosophies, principles, and policies. The organizations bring local expertise into marketing efforts in their nation's culture and environment which would not be possible otherwise. They also provide a base for user education and training for all types of local scientists and their supporting staffs.

Undoubtedly, one of the most valuable benefits from these international agreements is the opportunity to interact with large organized groups of users of CAS publications and services. This is something which has been very difficult for CAS in the past. CAS subscription files contain addresses of many intermediaries, librarians, book dealers, etc., but few actual bench scientist users. The agreement organizations are supplying very necessary interactions with these end users. IDC has been outstanding in these efforts. They represent a large group of industrial scientists and operate an efficient and growing patent information service. Their constructive suggestions have included a wide range of interests: all the way from the simple addition of volume numbers to abstract numbers in each printed CA abstract, a very popular improvement, to complex improvements in the CAS patent coverage and indexing policies. Similarly, the UKCIS has assisted materially in the improvement of new *CA Selects* profiles and in the documentation, instructions, and plans for improved on-line services. The French and Japanese agreements are relatively new, but equally important user assistance is anticipated from these two major groups of international scientists.

CAS began by sharing the responsibility for abstract production with a large group of international volunteer abstractors. While their activity has decreased in importance, the international sharing has grown via four existing bilateral national agreements which provide marketing, user interactions, R&D assistance, consulting advice, and financial support. This international sharing is expected to grow in the years to come and to be of even more importance in the total chemical information system.

Present and Future Prospects for Structural Searching of the Journal and Patent Literature

JOHN A. SILK

Imperial Chemical Industries Limited, Plant Protection Division, Jealott's Hill Research Station, Bracknell, Berkshire, England

Received May 14, 1979

Available systems for structural searching of organic compounds in the journal and patent literature are briefly reviewed. From a survey of the various methods of substructure search it is concluded that an algorithmic notation able to deal with both specific and generic descriptions of structure could be an extremely valuable development, since it would occupy a key position intermediate between canonical connection tables for individual compounds and degenerate fragment coding for groups of compounds.

The question, "How can we best attain comprehensive substructure searching of the journal and patent literature for low molecular (nonpolymeric) compounds?", is an important practical one for industrial information services in the pharmaceutical, agrochemical, and general organic business areas. The development of on-line searching has had a major impact on the retrieval of text-based information, and this includes chemicals which can be identified adequately by names or registry numbers. It has, however, only served to emphasize the limitations of available systems for searching generically for classes of compounds or those containing specified structural units.

To set the problem in perspective we must examine its component parts, namely, the available data bases and the systems for searching them, and, moreover, we need to make

a distinction between past and future. Whatever new systems and data bases might be developed over the next few years, there is little likelihood of improvement in the keys by which the older literature can be searched. Consequently, for the past we need to identify the best of what is available, and to consider how to utilize it most effectively. For the future we need to examine the current state of affairs and try to decide what developments would be most desirable.

PRESENT-DAY SOURCES

Table I lists the major compilations which cover the journal and patent literature in a manner providing some degree of retrospective search. Foremost is *Chemical Abstracts*, which covers both journals and patents and provides an excellent