

A Network of Chemical Information Handling*

GEORGE P. HAGER

College of Pharmacy, University of Minnesota, Minneapolis, Minnesota 55455

Received August 17, 1965

A review is given of studies on national-, discipline-, and mission-oriented information centers. Qualitative aspects of the problem are discussed.

As one contemplates the magnitude of the work involved in the handling of information by modern methods and for modern purposes and as one envisions the investment in manpower and facilities that may be needed to get the job done, there tends to evolve a concept of a large central establishment that will be all things to all men as regards information needs and their satisfaction. This Utopian idea of the documentalist's destiny has been satirized by Simpson (1) in his statements about a "World Scientific Information System." Very serious statements have also been published about the Soviet Union's VINITI, *i.e.*, the "All-Union Institute of Scientific and Technological Information of the Academy of Sciences of the U.S.S.R. of the State Committee of the Council of Ministers of the U.S.S.R. on Coordination of Scientific Research." Among the many sage and entertaining comments by Bingaman (2) about VINITI, the following is especially significant: "VINITI prints a horde of scientific articles and abstracts. Among them the most famous is 'The Journal of Abstracts,' which is not, as the title implies, simply a journal. It is fifteen journals! Each one of them has to do with a different branch of science. The closest the United States can come to this sort of a grand accomplishment is the publication known as 'Chemical Abstracts,' which is published by the American Chemical Society. But, you see, it has to do with only one science, and here the Russians have fifteen . . . , but on the drawing board are plans for an assortment of other editions to take care of the humanities and linguistics."

It is true that "a man's reach should exceed his grasp— or what's a heaven for?" On the other hand, omniscience has never been an earthly attribute nor is it likely to become one even by the intervention of a black box. Perhaps the pecking order engendered by an unequal distribution of information or knowledge and certainly the exciting synergism that operates at the interfaces among various sciences are human virtues that should *not* be jeopardized.

At this time a greater emphasis should be given to the qualitative aspects of the information problem. The over-

all need to be informed can be factored into segments that will include the particular *unmet* needs that are peculiar to various discipline-oriented groups and, within a discipline, the many mission-oriented groups that use information. As the qualitative aspects of the differing *unmet* needs of various groups of information users are brought into focus and priorities are judiciously evaluated, the feasibility of a single large central establishment fades; and numerous specialized centers, user-oriented, and, to whatever degree may be possible, integrated by way of a network begin to take on a realistic form.

Lack of information is not the primary concern of the specialized center. In the recent past, a lack of information was one of the most serious limitations with respect to productive research. It was appropriate at that time to invest the bulk of the available resources for research and development that would generate the information that was lacking. Certainly the generation of new information and data still is an essential part of a productive research effort. However, much of the information void has been filled. Research productivity now requires a greater emphasis on the *transfer of information* and the efficient application of information in modern research endeavors. According to the January 1963 Report of the President's Science Advisory Committee, "Science, Information, and Government" (3), "*Transfer of information* is an inseparable part of research and development. All those concerned with research and development . . . must accept responsibility for the transfer of information in the same degree and spirit that they accept responsibility for research and development itself . . . the individual theoretical scientist will, on the average, maximize his overall productivity if he spends half of his time trying to create new scientific information and half of his time digesting other work and communicating his own." Actually, a much smaller proportion of the investment now made in generating new information would go a long way towards financing the transfer of information even in accordance with the most ambitious schemes.

What once was an information void is now being filled not only by generation of new information at an ever-increasing rate but also by a redundant *regeneration* of

* Presented before the Divisions of Chemical Literature and Medicinal Chemistry, Symposium on Drug Information, 148th National Meeting of the American Chemical Society, Chicago, Ill., Sept. 1, 1964.

existing information. The information bank is approaching a critical mass and we often hear about an information explosion. We now wonder if the researcher is somehow being deprived of the content of the primary publications he used to read so diligently and the other information tools he formerly was able to use with great effect. We fear that he is inclined to settle, albeit unwillingly, for merely the fall-out of the information explosion; for example, the titles or permuted titles of the articles he formerly read, regardless of the notoriously low relevancy of titles.

The particular information needs of chemists and the many nonchemist users of chemical information are, of course, related to the massive volume and the dynamic nature of chemical information. According to recent estimates made by the Chemical Abstracts Service, the annual output of chemical literature is now doubling every 7.8 years. The cumulative bank is doubling every 11–12 years. In 1975, the total knowledge of chemistry and chemical engineering will be twice what it is today. At that time, the chemical information problem will also be at least twice as great unless some radical changes are made in the "information transfer chain" that is presented in the PSAC Report (3):

- (1) Generation of information and data
- (2) Recording and exposition
- (3) Cataloging
- (4) Storage and dissemination
- (5) Retrieval
- (6) Exploitation by the user

The role of the bench scientist is obvious in "generation of information and data" and, to some extent, in "recording and exposition." Editors, publishers, and librarians have major roles in "recording and exposition," "cataloging," and "storage and dissemination." The "Central Depositories," very wisely recommended by PSAC, seem to relate to these particular links of the chain and to the users' need for *current awareness* of the work being done in his own field, for *browsing* among the reports of the work proceeding throughout his discipline and at the interfaces with other disciplines, and for *being alerted* to discipline-wide and interdisciplinary developments that may have an impact on the users' mission.

The PSAC also recommended "specialized information centers." These should be rather strictly oriented with respect to discipline and mission. A specialized information center should be "primarily a technical institute rather than a technical library. It must be led by professional working scientists and engineers who maintain the closest contact with their technical professions and who, by being near the data, can make new syntheses that are denied those who do not have all the data at their fingertips." In many ways, the specialized information centers may be considered indispensable for the last two links in the information transfer chain: "retrieval" and "exploitation by the user." From the operational viewpoint of those who use chemical information, the specialized information centers would provide, through retrospective searching of the information bank, a new information commodity (4): a compilation of information and data that are pertinent to a current, specific, research project—up-to-date, precisely selected, and comprehensive both as regards their chronological and interdisciplinary aspects.

Perhaps a single specialized information center should not be expected to satisfy all of the needs of the users even within a given discipline, especially a discipline as broad as chemistry involving missions as diverse as those of chemists. It was recognized that the chemistry information bank is now very large. It is growing at an alarming rate. It is interdisciplinary and very heterogeneous. On the other side, the users of chemical information are very numerous. Their interests and their missions are distributed over a broad spectrum. Moreover, new publics with their peculiar information needs are joining the ranks of customers for chemical information. Switching *all of the information bank* to *all of the users* would be a useless and wasteful exercise in documentation and communication.

At a conference in March, arranged by the NAS-NRC Committee on Modern Methods of Handling Chemical Information and sponsored by the National Science Foundation, attention was focused on the problem of handling chemical structures and their retrieval in conjunction with a *limited* amount of the information with which the structures would often be correlated. This hard core of information and data with which all users are concerned is a relatively small segment that can be kept up to date. It can be precisely and comprehensively sampled in accordance with the users' specifications. It can be switched with minimum resistance and at minimum cost between the large bank of chemical information and the vast and varied public that uses chemical information.

Although all of the elements of the hard core have not yet been identified, there is little doubt that all will agree that the composition and identity of chemical entities are *sine qua non* in the effective use of chemical information and data. The composition and identity of chemical entities are most conveniently expressed and most widely understood in terms of the two-dimensional structural formula. Most individuals who have sought to establish a specialized information facility involving chemical information have recognized at the outset the need for switching information about molecular structures. Structures are descriptive and, therefore, readily handled by modern methods. They serve very well as addresses or access points for other components of the information bank. They are an important part of most of the chemical information employed by users of all types. Solution of the chemical information problem across-the-board requires the initial solution of the structure-switching problem.

Much of the effort to solve the structure-switching problem has been redundant. Lack of agreement on a universal solution to the problem, or at least parallel but compatible solutions to the problem, have forestalled the establishment of the specialized information center or network that could serve most or all of the users of chemical information. These aspects of the problem are unnecessary and inconsistent with scientific objectivity. No solution of the problem will completely satisfy everyone who is dealing with it. No solution today will withstand the modifications growing out of future research and development (nor should it). A survey of chemical notation systems and their use in retrieval of structures was undertaken by a staff under I. Moyer Hunsberger's direction. The results of the survey (5) were most informative as regards the "state-of-the-art." Considerable progress has also been

made in this field through development of the Army Chemical Typewriter at Walter Reed Army Institute of Research and work is being carried forward in the Army's CIDS program. The development of effective methods and the urgency of the need for expeditious retrieval of hard core information has led to the conclusion that the time has come when a choice must be made with whatever compromises and self-sacrifices that choice may entail.

The Committee on Modern Methods of Handling Chemical Information (CMMHCI), at its March meeting, prepared a summary statement concerning a specialized center for retrieval of structures and hard core information. The statement is in the form of a resolution (6):

"WHEREAS, there is a finite segment of the large and growing volume of chemical information (the 'hard core' consisting of the registry number, structural formula, identifying names and numbers, etc.) that is of vital concern in research and development activities of practically all chemists and other users of chemical information regardless of their affiliation (academic, government, industry, or other) and regardless of their mission, and

WHEREAS, modern methods of handling information of this type have now been developed to the point where they can be implemented (and later improved through further research and development), and

WHEREAS, the President's Science Advisory Committee has emphasized that 'the special sensitivity of non-government, decentralized information services to the needs of the user as well as the variety of approaches offered by these services is precious and must be preserved' (Science, Government and Information, The White House, Jan. 10, 1963, p. 4), and

WHEREAS, the great majority of the users of chemical information are, and traditionally have been, closely affiliated with the American Chemical Society (ACS) and the Chemical Abstracts Service (CAS) and depend upon these agencies for the information tools and services they most urgently require, and

WHEREAS, the present services being rendered by the CAS must be preserved and, now, expanded to meet the new and growing needs of the users of chemical information, and

WHEREAS, the Committee on Modern Methods of Handling Chemical Information notes with great interest the action of the Board of Directors of the ACS taken in September 1963, authorizing that CAS seek additional support for research and development activities, and commends that action, and

WHEREAS, in the opinion of this Committee the CAS is the proper agency for the creation of an integrated mechanized system for handling the 'hard core' of chemical information and the primary source of a service involving the 'hard core' for users of all types,

RESOLVED, that the committee recommends that the ACS and the CAS assume this larger role with the responsibility implicit therein, and

RESOLVED, that the committee urges sympathetic consideration by all federal and private agencies from which the CAS may solicit support for the achievement of the above goals."

Solving the "hard core problem" may seem to be a small step in relation to the over-all problem facing the users of chemical information. Of necessity, however, it is the first step; and it is, of itself, a most useful step. It may be the nucleus from which the snowball can grow. The hypothetical center or network handling the hard core may be a mainstream from which will proliferate the other specialized information centers that will be necessary to handle in depth those large segments of the information bank that are of particular interest to specific groups of users with common missions.

As long as the effort at the potential mission-oriented centers is being dissipated in redundant efforts to solve the discipline-wide hard core problem, major aspects of the information problem peculiar to those centers will remain unsolved. When the common problem has been solved, the problems related to each particular mission can be attacked effectively. We should now proceed with a firm conviction that an integrated system for handling chemical information is within reach. As we work diligently and unselfishly toward our common goals, we shall be able to grasp what, perhaps, we could have reached long ago.

LITERATURE CITED

- (1) Simpson, G. S., "Wiley Bulletin, Science and Industry Ed.," Vol. 5, John Wiley and Sons, Inc., New York, N. Y., 1961, pp. 1, 4, 5.
- (2) Bingaman, R. L., "The World of Wiley," John Wiley and Sons, Inc., New York, N. Y., spring, 1964, pp. 4, 5, 10.
- (3) "Science, Government, and Information," U. S. Government Printing Office, Washington 25, D. C., 1963, pp. 1, 3, 10, 13.
- (4) "Surgeon General's Conference on Health Communications," U. S. Department of Health, Education, and Welfare, Public Health Service, February 1963, p. 22.
- (5) "Survey of Chemical Notation Systems," National Academy of Sciences-National Research Council, Publication 1150, Washington, D. C., 1964.
- (6) "Conference on Mechanical Processing of Chemical Information," National Academy of Sciences-National Research Council, Washington, D. C., 1964.