is one example out of many in which Information Researchers worked as an integral part of the research team.

As the need for Information Researchers continues to grow, the questions arise, "Where does one find an Information Researcher? How are Information Researchers trained?"

Information groups at Esso and elsewhere have found that new Information Researchers are often hard to find. This frequently reflects a general lack of knowledge about the operations and rewards of what is still a very young and not well known field. At present, no academic program for the training of Information Researchers is available. The Information Research personnel of TID were either selected from within the company or hired direct from school. None have had previous training in the information field. There is a good deal that can be done to stimulate topnotch young scientists to turn to Information Research as a career.

Academic institutions can go a long way toward instilling in the student an appreciation for the importance of the literature. They must emphasize the value of making full use of the literature—including abstract journals and patents—and must further teach how to use this literature most effectively. They must promote the precise communication of information. They must encourage the student to develop and make use of a reading knowledge of foreign languages, particularly Russian. One excellent way of accomplishing these objectives would be the widespread establishment of Chemical Literature courses such as that described by Dermer in the ACS monograph, "Training of Literature Chemists."[4]

The industrial organization itself can help greatly by emphasizing to educators and students the importance which it attaches to the work of Information Research Chemists. It can explain its needs for such men to campus contacts, placement officers, and potential new employees, in promotional literature and in job interviews. The Esso Speakers' Bureau has prepared a talk on Information Research Careers, for presentation to interested student groups.[5]

Today's deluge of information can be of enormous value if adequate techniques are developed to use it. Experience at Esso has shown that Information Research can contribute greatly to the effective use of the literature, thereby increasing the productivity of current research efforts and stimulating the creativity of the entire research organization.

Acknowledgement.—The author wishes to express his gratitude to Mr. W. T. Knox and Drs. H. J. Hall, I. Kirshenbaum, and R. W. Scott, whose ideas on Information Research have been used freely in preparing this paper.

REFERENCES

(1) D. J. de S. Price, "Science Since Babylon," Yale Univ. Press, 1961 p. 107.
(2) Ibid., p. 97.
(3) *Chem. and Eng. News.* Feb. 27, 1961, p. 68.
(4) O. C. Dermer in "Training of Literature Chemists," ACS Advances in Chemistry Series, No. 17, p. 27 (1956).
(5) "Information Research—A New Research Career," by H. J. Hall and R. W. Scott; Technical Talks, P.O. Box 175, Linden, N. J.

The Documentation Research Chemist[1]

By. P. N. CRAIG
Science Information Department, Smith Kline & French Laboratories, Philadelphia, Pa.
Received April 3, 1962

The dictionary defines a document as a means of storing information. Documentation chemists prefer to be much more positive and, in a broader sense, we interpret chemical documentation work to include both the storage and retrieval of chemical information; also we like to include in the definition any necessary interpretation of such information when retrieved. The handling of large numbers of documents usually requires special equipment such as punched card machines or computers. The term "Chemical Documentation" should not convey merely the special techniques, but rather the whole process of storage, retrieval, and interpretation of chemical information.

What do we include in the term "chemical information"? The three major types of chemical information to be considered in this paper are:

(A) structures of compounds;
(B) properties of compounds or mixtures thereof;
(C) chemical processes, reaction conditions, *etc.*

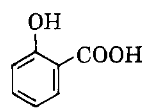
Structures indicate how atoms are joined to form molecules. The structure is the basic information which is conveyed best from chemist to chemist by a structural formula; it may be most easily printed in form of a name (although several names may represent the same compound) or may be represented unambiguously by means of a linear chemical notation system, often called a cipher.[2,3] A cipher would place every compound in one definite position in an index. A translation step is required to regenerate the structural formula from a name or a linear notation. This is critical, because the average chemist, especially one dealing with organic compounds, wants to use the structural formula rather than a name, number, or notation.

The problems of indexing are important, since the hundreds of thousands of known compounds require organization in indexes for retrieval. The subject and molecular formula indexes of *Chemical Abstracts* enable retrieval of information pertaining to a single compound or closely related compounds, but it is almost impossible to run a truly generic search with *Chemical Abstracts.*

An example which illustrates this is typical. By a search which was both exhaustive and exhausting, a literature chemist found that about 30% of the compounds containing a phenothiazine ring were subject-indexed under entries other than "phenothiazine," and these were scattered from A to Z throughout an entire annual subject index.

Although it usually is not necessary to run a completely generic search such as the one just mentioned, most large research organizations are finding that they definitely need partially generic retrieval of information especially for internal research project data. This need arises as an organization grows in size and gathers a backlog of thousands of compounds, prepared or collected for some type of experimentation. It becomes particularly important for young chemists beginning their career with a new organization to be able to draw upon the background of experience obtained by their predecessors in any particular subject area, since the largest part of the internal data is confidential and unpublished, and therefore cannot be found in *Chemical Abstracts*.

The best solution to date for the problem of generic retrieval is the use of a chemical structure code, by which the various structural features of a molecule are assigned descriptors such as those shown in Figure 1.



HC11 = 1-Aromatic carboxyl group
H751 = 1-Aromatic hydroxyl group
NY81 = 1-Unfused benzene ring
OOA1 = 1-Carbon in a functional group
Z201 = Ortho arrangement of hydroxyl and carboxyl groups

Fig. 1.

The use of such a code enables generic searches to be complete, but the assignment of several descriptors per compound dictates the use of machine methods of retrieval when large numbers of compounds are concerned. A structural code[4] such as that developed by the Chemical Biological Coordination Center, National Research Council, usually requires > 8 descriptor terms per compound. Thus, the 63,000 compounds encoded by the CBCC would require > 504,000 individual entries in an alphabetical index. But by means of punched cards, only 63,000 are needed. Furthermore, the use of mechanical punched card equipment enables any desired combination of descriptors to be searched. The great advantage of punched cards is that every entry is equivalent to an index entry. Those who try to keep up with the current literature by means of personal card files will appreciate the desirability of, and problems caused by, multiple indexing; these are met adequately by punched cards.

The documentation chemist must decide between the many known types of structure codes, and usually must adapt or combine certain features of several codes in order to develop one best suited to handle the particular type of data needing to be stored and retrieved.

It is evident from the number of published codes that no one code satisfies all needs. Some codes require no mechanical equipment; others use edge-notched cards.[5] Some codes have been developed for use with computers[6,7]; these are topologically designed so that each compound can be retrieved individually, if desired. Searching of such records can be done either specifically or generically, but a large capacity computer is required for such codes. Experiments are under way at *C.A.* on the feasibility of searching ciphers generically on computers.[8]

The properties of compounds or mixtures also require special handling for ease of retrieval. Such data are much more varied in kind and type of treatment from field to field. For example, pharmaceutical research organizations and heavy chemicals manufacturers both need to store and retrieve chemical structures, and may use similar codes for that purpose. But the best ways to store biological test data may vary from the best ways to store data on physical and chemical properties, such as tensile strength, thermodynamic properties, weathering data on formulation, *etc.* Again, the documentation chemist must compare existing code systems, and usually must design his own special "use" code. (A typical example is described in reference 9.)

The ultimate in efficient use of chemical research data is a retrieval system which allows questions to be asked on either a chemical structure basis, a use-property or test-results basis, or any desired combination of these approaches. Such systems are gradually evolving, and the experiences gained with their use are most promising. Our experience has shown that the documentation chemist is in an excellent position to design and operate such a combined system.

The third category of chemical information, that pertaining to reaction conditions, processes, *etc.*, is a subject which is poorly retrievable in the chemical literature. Internal records of this type sometimes have been ignored or have been coded in connection with chemical engineering data or with patent records. The usual background for documentation chemists working with this type of data is chemical engineering or chemical patent work.

It is difficult to realize, but some research organizations are quite unaware of their inability to use efficiently the background of knowledge gained in the course of years of their own research. Worse, some are even convinced that this is not a serious problem! However, more and more companies, research institutes and government labortories are becoming aware of their need to make their research more efficient, and more people will be needed to enter the field of chemical documentation. A prime example of an organization which is faced with an overwhelming research information problem is the Cancer Chemotherapy National Service Center, which has accumulated 80,000 compounds and hundreds of thousands of test results over a period of only a few years.

What, then, is the proper background for documentation chemists? Historically, most of the workers in this field have made the transition from the laboratory research bench to documentation research within the same organization. This is not merely because formal educational training is lacking for documentation chemists, but because a working knowledge of the organization's research techniques and goals, and a personal aquaintance with laboratory scientists is an excellent background for the documentation chemist. This is not meant to imply that no courses should be slanted toward chemical documentation in college or graduate school. It would be a valuable step for all chemists to take a course which would include a survey of how to use the existing chemical literature, and which would alert budding young scientists to the problems involved in storing their own and other's scientific contributions so that they may not go unused or unconsidered.

Finally, emphasis should be placed on the creative aspects of documentation chemistry. The storage and retrieval of chemical research data is a means to the end of ensuring that as much pertinent knowledge and experience as is possible be considered by both those planning new research and those trying to capitalize upon the knowledge presently available. The documentation chemist often can aid the research chemist materially in this planning phase by pointing out unexpected extensions, or new research parameters, based upon consideration of the peripheral data usually accompanying specific pertinent data. This advantage of the generic approach to information retrieval gives the documentation chemist the opportunity to make suggestions not considered prior to his use of the storage-retrieval system. This enables the documentation chemist to have an outlet for his creative abilities.

REFERENCES

(1) Presented at the 4th Delaware Valley Regional Meeting, American Chemical Society, Philadelphia, Pa., Jan. 25, 1962.

(2) W. J. Wiswesser "A Line-Formula Chemical Notation," Thomas Y. Crowell Co., New York, N. Y., 1954.

(3) G. M. Dyson, "A New Notation and Enumeration System for Organic Compounds," Longmans Green & Co., New York, N. Y., 1949 and 1961.

(4) "A Method of Coding Chemicals for Correlation and Classification," National Research Council, Washington, D. C., 1950.

(5) R. S. Casey & J. W. Perry, Editors, "Punched Cards; Their Applications to Science and Industry," Reinhold Publishing Corp., New York, N. Y., 1951.

(6) A. Opler, *Chem. and Eng. News*, **35**(33), 92, (Aug. 19, 1957).

(7) W. H. Waldo, R. S. Gordon, and J. D. Porter, *American Documentation*. **9**, 28–31 (1958).

(8) G. M. Dyson & E. F. Riley, *Chem. and Eng. News*, **39**(47) 74 (Nov. 20, 1961).

(9) E. B. Edge, N. G. Fisher & L. A. Bannister, *American Documentation*. **8**, 275 (1957).

Panel Discussion on Education for the Literature Chemist*

By HERMAN SKOLNIK
Research Center, Hercules Powder Company, Wilmington, Delaware
Received January 24, 1962

The purpose of this panel discussion is to explore the role that colleges and universities might play in the further growth of chemical documentation as a discipline of chemistry.

As of now, there is no bridge between chemical documentation and universities. Both the discipline and the professional literature chemist have arisen in response to the needs of the chemical industry. Although chemical documentation is flourishing in the industrial environment, the literature chemist is not a direct product of the universities, and research in chemical documentation is practically non-existent in universities. Even more serious is the fact that many chemistry teachers are not aware of chemical documentation as a discipline of chemistry.

In the light of the picture I have just drawn, why should we want to involve colleges and universities with chemical documentation? I think that the primary reason is that no discipline of science can truly exist without roots in our educational structure: We need these roots to attract able young scientists to chemical documentation. We need these roots to develop procedures, knowledge, and principles of documentation through academic research.

On the other hand, a chemistry curriculum devoid of chemical documentation cannot prepare students for a science undergoing constant change. Science and technology are not tied to the knowledge and mode of thought acquired in four years of college and three years in graduate school. For in these seven years, at the present rate of growth of the chemical literature, the total chemical knowledge will have doubled, new disciplines of chemistry will have arisen, and some older ones will have died.

I do not advocate that we build a large bridge between chemical documentation and the universities. I do not think that there is a need for an undergraduate curriculum in chemical documentation. A student entering college and for most of his collegiate life has only a vague knowledge of the relationship between his chosen profession and the world at large. Specialization too early handicaps a student in his professional life. The greatest disservice to chemical documentation, in my opinion, would be the training of literature chemists before they become the best possible chemists they are capable of being. No curriculum in chemistry should have as its objective the teaching of crafts and end-use disciplines.

The chemistry curriculum should not distinguish the potential literature chemist from other potential chemists. Chemical documentation should be a part of the curriculum for all chemistry majors, not necessarily as a special course, but as part of every course in chemistry. Chemical documentation is primarily a process of asking questions and answering them. This, in essence, is chemistry.

The potential literature chemist can be distinguished from other potential chemists by certain characteristics. He usually has superior talents and strong interests in skills such as writing, language, law, mathematics, philosophy, history, or even literature. He likes to rearrange data and organize them in new arrays. He likes to handle more data than he could possibly obtain himself. He likes being a perennial student of chemistry. These characteristics do not make him a lesser chemist. They merely make him a different kind of chemist.

There are many opportunities for literature chemists in the chemical industry. The symposium on the literature chemist in the chemical industry has delineated some of these opportunities. Professors of chemistry, if aware of chemical documentation as a discipline of chemistry, can lead potential literature chemists over the bridge to an exciting and important professional career.

* Presented at the Fourth Delaware Valley Regional Meeting, American Chemical Society, January 25, 1962, Philadelphia, Penna.