A Look at the Division of Chemical Information†

CARLOS M. BOWMAN

Central Research Computation Laboratory, The Dow Chemical Company, Midland, Michigan 48640

Some of the past developments in chemical information are reviewed and some expected developments are outlined. The role of the information chemist in this process is examined.

Thank you very much, and I hope that the time that we spend this afternoon will be worthwhile to all of us. It is my hope that as a result you will be moved to do some things differently than what you are doing today.

First of all, let us review some past history. My first acquaintance with the then Division of Chemical Literature showed that the papers presented before the Division had titles such as the "Literature of Fluorine Chemistry" and the "Literature of Pesticides". Computers were just starting to make their appearance felt in the area. One of the most novel developments at the time were KWIC indexes. IBM was doing a fantastic job of selling computers and computer time using the selective dissemination of information concept. The searching of chemical structures was done primarily by the use of fragmentation codes and card sorters. At this time several linear notations had been invented, and there were many arguments over which one was best. In fact, I have seen grown men before this Division lose their tempers arguing over which notation would save the most spaces on a punched card.

Coordinate indexing and concept coordination were sold to us as the answer to all subject indexing problems. Of course, going along with the need for coordinate indexing was the matter of a controlled vocabulary. I wonder how many millions of dollars were spent on the development of authority lists of indexing terms, thesauri, and other such compilations. TERMATREX CARDS were considered, also, an answer to many of our information problems.

Recognizing the need for the dissemination of information, the National Science Foundation put a great deal of money into the development of information centers which would serve as an interface between the major information producers and the individuals. The National Science Foundation also elected to support the development of automated processing for the Chemical Abstracts Service. They supported the development of the Registry System and all of the automation of *Chemical Abstracts* can be traced back to that early support.

Without support and fighting an uphill battle all the way, Gene Garfield brought out the Science Citation Index and has established the Institute for Scientific Information as a reliable entity which performs a very useful service to the community.

With the advent of automation in some of the major abstracting services, the concept of on-line bases and on-line searching became a reality. We now have a large number of options available where we can dial-up and inquire.

Our history has been a colorful one and also very interesting and challenging. Let me stop at this point and try to draw some conclusions I see that we ought to be able to make from examining our history.

Some of the very early systems devised to handle chemical information were attempts made to devise a system that a chemist could use to find information easily. Think of Beilstein

where once you actually found the information rather than finding pointers or references to that information. Even the Chemical Abstracts Decennial Indexes were devised so that it would be easier to locate the information. Attempts were made in *Chemical Abstracts* to actually include the pertinent information so that a chemist would not always have to go back to the original article to get the information that he wanted. The use of chemical information was an important skill. Courses were taught in schools on how to use the literature. We spent quite a number of hours in the library learning how to use the various tools. It was an integral part of the needs of a chemist. It was recognized that information was needed in order to innovate.

As the amount of information increased, finding information became more and more difficult. Indexing and classifying information are nothing more than guessing how information is going to be asked for. If you guess right, you do a good job of indexing. If you guess wrong, nobody will ever know because the information will never be found. As we tried to cope with large amounts of information, we turned more and more attention to indexing systems and we started to lose track of the primary goal. We developed a corollary to Moore's law which states that the use of information is adversely proportional to the difficulty of the work required in obtaining that information. As information chemists we lost track, to a great degree, of our primary mission which is to provide information to the user, and we spent most of our time devising systems and talking to each other.

Let us examine the state of the art at this point. Most of the information systems today provide references. They do not provide information. We produce long lists of numbers, and in the more sophisticated systems you may get the title, author, and reference. From these you are asked to make a decision as to whether the reference contains appropriate information or not. The systems that we have devised are very complex. Chemical nomenclature, for example, has become something that only an expert can tackle. In an effort to accommodate the computer system, *Chemical Abstracts* has gone, what I consider as a giant step, toward the standardization of nomenclature. While this change has provoked much heat and is a change in many respects, it will eventually have the effect of systematizing things and making them easier to find. Chemical notations can only be used by experts. The searching of on-line data bases requires a person to be trained in order to use them. We send people away to two-, three-, and four-day schools in order to become acquainted with a data base, how to use it, and how to interact with it. A chemist cannot just walk up to a terminal, inquire, and get the information that he wants.

He more than likely will be frustrated in trying to use the system and if he is successful in getting some answers, they will only point to other places, perhaps to an obscure journal that he cannot obtain and he has absolutely no assurance that he has got all the data that he wants. We advertise retro-

spective data bases that go all the way back to 1972. At this time when we are celebrating the ACS Centennial, going all the way back to 1972 and calling it retrospective, I think is a little naive and presumptuous.

Let us return to *Chemical Abstracts*. With all the automation and all the advances which have been made, is it easier to use? No! It is more complex by the sheer magnitude of the amount of information. The automation effort was directed at making publication more efficient. The user was not considered in the initial design.

In summary, we have strayed a long way from providing information, although we have learned how to cope with huge volumes of material.

Let us turn to the future in light of some of the present developments. What is one of the most exciting developments in chemical information today? In my mind, it is the development of techniques in pattern recognition and artificial intelligence. By this I mean the techniques by which we are able to take data about compounds or materials, about their behavior and properties, and try to infer or predict the behavior of similar compounds or predict the types of compounds that would be required in order to obtain certain behavior. It is very interesting to see this work progress. One of the areas that causes me a great deal of concern is when I examine who is doing this work. Are information chemists doing it? No. It is the chemists who have recognized the need for this kind of technique who are delving into the data and developing systems, techniques, and approaches for solving their problems. I am afraid it is a repetition of old patterns. Look back into history and ask yourself the question, "Who developed some of the more advanced techniques that we are using today?" Was it the information chemist of the time who developed the chemical notations, coordinate indexing, or some of the computerized systems that we have? No! In most cases it was a chemist, the user, the person who became frustrated with the systems available to him at the time and struck out on his own to develop a system or to devise the tools which would help him to solve his problems. We are doing that all over again today.

Reflect for a minute on the papers presented on the first day of this ACS meeting in the Division of Chemical Information. I would venture to say that very few of those of you here today at this luncheon understood some of the techniques and mathematics that were being talked about. In fact, there were very few of you even present at the meeting. Yet, I believe it will be the application of these mathematical techniques that is going to present the new developments in the handling of chemical information. They are perhaps difficult for us to grasp because of our lack of training or lack of familiarity with these concepts. It makes us uncomfortable and so we would rather avoid it in the same way that librarians and other information chemists avoided Boolean logic, concept coordination, and chemical notations only to have the rug pulled out from underneath their feet by the user who needed tools and new ways of doing things.

It would be presumptuous for me to try, in detail, to predict the future, but I think the cycles are repeating themselves. As information chemists we have developed systems that are very sophisticated, interesting, and challenging to us. Unfortunately, to a great extent, we have lost track of the primary purpose of information systems. That is, to provide information to the user, to the chemist. We are setting ourselves up as filters and the chemist is not comfortable with filters of his information. He wants to do it himself. We must recognize this or we will become obsolete like some of our information folks became obsolete 10 or 15 years ago.

The challenge today is to serve the chemist. It is no different than it ever was before. We need to provide him with the information that he needs as well as the techniques that he needs for manipulating it. He does not want sophisticated systems; he wants information. He wants ways of evaluating information. He wants to be able to display his information. We need to provide this for him.

I hope my remarks today have not been too negative. They were intended to stimulate your imagination, to cause you to do some soul searching, and to try to evaluate just what it is you are trying to accomplish, what are you trying to do, and does it really coincide with the people you have been asked to serve. After all, chemical information is not in itself a science or a branch of chemistry, it is a service part of chemistry. An information chemist is somebody that is engaged in doing something for others. We must remember that and if we do so, we will be very successful. If, as a result of this very fine dinner and this talk, you find yourself better serving the needs of the chemists in your organization, I will feel that the talk has been a success. Thank you.