Information Patterns of Chemists in a University Environment*

LOUIS VAGIANOS
Dalhousie University, Halifax, Nova Scotia, Canada

Received May 27, 1970

A literature survey isolates many characteristics and information use patterns of university chemists. Information volume and usage patterns indicate problems on the physical, operational, and intellectual levels. Long-term solutions are indicated and reviewed. Thirteen short-term solution elements are listed; each requires active participation by university chemists.

Over the past 50 years, it has become increasingly clear that the communication of information in social relationships and in all institutions has become one of the single most important tasks of civilized man. This is particularly true of science where all practitioners view the transfer of information as an inseparable part of research and development. No one would quarrel with the declaration of President John F. Kennedy that "one of the major opportunities for enhancing the effectiveness of our national scientific and technical effort and the efficiency of government management of research and development lies in the improvement of our ability to communicate information about current research efforts and the results of past efforts."[1]

During the past decade, in view of the information explosion and the so-called communications revolution and in view of the importance of information to decision making, entirely new service industries developed. Much effort was made, in government and nongovernment circles and on all levels, to discuss and analyze the nature and possible ramifications of the problem. New solutions and new systems were negligible, but popularization of the subject was so far-reaching that many words coined to characterize and explain it have long since been elevated to the status of cliches. Cliches can be very dangerous for two reasons:

> First, they represent labels—simplified, stereotyped representations, which have long since lost the originality, ingenuity, and impact of the ideas they summarize.
>
> Second, they can, and often do, hinder constructive progress by obfuscating and oversimplifying the exact nature of the problems to be solved. False claims and quack remedies become the order of the day.

WHO IS THE CHEMIST?

The chemist is a scientist who pursues the study of chemistry—i.e., he dedicates himself to the understanding of the physical universe through systematic investigation and measurement of the properties of matter. He is an individual who accepts the premise that chemistry can be organized in no other manner than by granting mature chemists complete independence in the organization and pursuit of their research. He is confident that given the appropriate support and environment, he and his colleagues will distribute themselves productively over the whole field of possible discoveries, applying themselves, according to their own evaluations, to tasks they believe most profitable for them. He believes that the function of public authorities is not to plan research but to provide opportunities and facilities for its pursuit. Moreover, evaluation of existing research programs, including decisions involving performance and need for continuation, must be left in the hands of mature, experienced chemists.

WHO IS THE UNIVERSITY CHEMIST?

All the foregoing comments apply to any chemist, but the university chemist is different from other types in at least four significant ways:

1. He does not operate through a chain of command—i.e., he develops his own investigation program which is often not subject to review or criticism from within the university community.
2. He is required to teach.
3. He is normally involved in pure research, although applied and development research are sometimes pursued.
4. He and most of his colleagues have achieved their PhD.

As a result of the above, one can surmise that a university is the most heterogeneous melting pot within the scientific community, requiring information services which will satisfy requests which range from the simplest to the most sophisticated inquiry, or from the mind of an amateur to the ultimate professional in both applied and theoretical contexts.

In 1966, the Gross National Product of Canada was 60 billion dollars. Scientific R&D represented 1% of this total, rendering it equal to forestry in industrial importance; university scientific R&D amounted to nearly 1/5 of the national total spent. More important, in the last occupational census by the Dominion Bureau of Statistics which shows 10,471 people in Canada calling themselves physical scientists, slightly more than half (5702) listed

* Presented before the Division of Chemical Literature, Joint Meeting of the ACS and the CIC, Toronto, Ontario, Canada, May 24-29, 1970.

themselves as chemists. This means that chemistry, as a profession, involves the largest proportion of working scientists in Canada. It is also necessary to note that only 8% of this group claimed university employment; yet this small number of chemists appears to command a disproportionate amount of influence within its own group and upon society. An analysis indicates several reasons for this. Since most PhD's are involved in basic research and an overwhelming majority are at universities, an influential establishment known to some as the "invisible college," both within and without the university, appears to maintain a dialogue with its own group without any real interests in the problems of colleagues in the development and applied research areas. This creates a communication problem between groups and has a detrimental effect on the information flow.

The situation is made possible because of the freedom the group enjoys through the extraordinary recognition and financial support it receives from government and the influence it exerts within government circles. The potential for perpetuation by controlling the training requirements of the new generation of chemists is also worth underscoring.

The diverse information needs of university chemists have forced university libraries to collect in all aspects of chemistry. This has led to much duplication of effort in services and in material purchased and has created a situation in Canada where, with the exception of the National Science Library and several other government and special libraries, most good science collections are part of university libraries.

This, in turn, caused serious problems within university libraries as they were, unexpectedly, called upon to play a role for which they were not prepared. As a result, by most criteria used by scientists, libraries have failed in their new role. However, much of their failure can be traced to science and the scientists.

GENERAL OBSERVATIONS OF INFORMATION NEEDS AND SOURCES OF CHEMISTS

Over the years, studies have been made on information patterns of needs and uses among applied and pure chemists. These reveal sharp differences in their approach to information. For example:

> Pure scientists tend to do their own literature searches; the applied scientist has them done for him, with references evaluated, extracted, and summarized.
>
> The pure scientist uses advanced textbooks and research periodicals; the applied scientist uses security-classified research reports heavily.
>
> Pure scientists get 70% of their data from domestic sources and 30% from foreign sources; the applied scientist obtains 90% from domestic sources and 10% from foreign sources.[2]
>
> Pure scientists are heavier literature users than applied scientists.
>
> Pure scientists make greater use of libraries than applied scientists.
>
> Pure scientists keep personal indices which are regarded as important.

The sources of scientific information, however, remain constant. Since information sources help to determine information patterns, studies conducted in this field yield interesting results:

> Information directly related to the researcher is generally not immediately available in published form. He needs current ancilliary information about such things as apparatus and measuring methods, and the stimulation of general scientific reading.[3]
>
> It is generally agreed that the three primary printed sources of scientific information are current literature, reference materials (personal and institutional), and informal conversation.[4]
>
> The five important literature sources for pure scientists (in order of importance) are advanced texts and monographs, research journals, handbooks, review publications, and mathematical and physical tables.
>
> The use of indirect sources of information by pure scientists (again in order of importance) are consulting cited references, personal recommendations, regular literature perusal, indices and abstracts, and bibliographies.
>
> As an information source, the university library is the sole library source consulted external to the pure scientists' department. The majority use departmental or divisional libraries.
>
> 64% of the scientists studied in one report depend on libraries for published materials, 7% depend on personal sources, while 29% depend on both sources equally.[5]

However, as one prominent chemist put it, librarians are an impediment. The chemist is the expert in his own field. He must know the library better than the librarian.

Finally, whether from print or from other sources, a chemist needs information in three identifiable areas:[4]

> **Current information.** The need to know what others have recently done or are doing. Personal contacts are important but chemists regard literature as the primary source.
>
> **Specific information.** This need is directly connected with a research or operational problem at hand, the need for a bit of datum, a method, the construction of a piece of apparatus, an equation, or an explanation of a phenomenon. This requirement is of greater importance to the chemist than any other scientific field. It, too, is closely tied to the literature. Conferences, consultation with colleagues, personal indices, and reference works prevail on this level. Rarely does the chemist work independently.
>
> **Exhaustive information.** The need to check through all relevant information existing on a subject. This need arises at the start of work on new research or when the results of investigations are ready to be reported. *Chemical Abstracts* is the favorite tool on this level.

In his study on "Information Gathering and Use Habits of Chemists," Jahoda concluded that the average university chemist spends a 168-hour work week thus:

117.1	hours sleeping and personal and social endeavors
16.5	hours in scientific communication
10.4	hours in equipment and set-up and use
6.7	hours in business communication
3.0	hours data treatment
2.5	hours thinking and planning

He further concluded that:

> Journal reading takes 5.3 to 2.7 hours a week.
>
> Some personal index is used by over half the group studied.
>
> Library use for research and study varies from 3 to 14% of the average chemist's time.[6]

Herner[5] indicates that while the average number of journal subscriptions per pure scientist is two and the average number of books purchased per year is six, university chemists receive the greatest number of technical news, house, and trade publications.

Cole[7] in attempting to determine a pattern of information needs among scientists in chemistry and related technologies also cited chemical literature as being of prime importance to the field.

A 1966 research project at Columbia University,[8] attempting to demonstrate the importance of various sources of information to particular patterns of use by chemists, uncovered similar findings. The study concluded that:

> Deliberate searching is most frequently satisfied through primary literature, especially journals and reprints.
>
> Primary literature, especially books, was even more important for "brushing up" activities.
>
> Pertinent "accidental" information was obtained most often at meetings and, as such, represented one of the real values of conferences.
>
> The contact function of personal communication was significant on the selection of journal articles that would be otherwise missed.

Equally interesting are the reports that show quite distinct differences in requirements according to discipline and type of work.[9] In chemistry these differences in information requirements are emphasized by the diversity and exclusiveness of specialty fields and the variety of information approaches called for. The chemist working in the area of inorganic chemistry will have little cause to consult Beilstein.

Kessler[10-13] of MIT has noted certain properties of citations within the literature of physics which are useful for comparison with citations within the literature of chemistry. For example:

> There exists a definitive journal in physics—*Physical Review*.
>
> This journal plus a small number of additional constant titles account for 95% of all references.
>
> The remaining 5% of references is to a large and ever-growing list of rarely used journals.

For the field of chemistry, Barrett and Barrett[14] indicate that the same is true of chemical literature:

> The majority of literature citations is to recent papers. 40% of the citations in the *Journal of the American Chemical Society* refer to previous publications in the same journal.
>
> More than 45% of the citations in the JACS are to ACS publications.

Bonn[15] corroborates this finding by indicating that the ACS Journal is the most extensively used publication within the area of pure chemistry. Garfield's assertion that for the scientific field as a whole, about 300 journals account for half the articles of importance published is even more revealing. In organic chemistry, less than 50 journals account for more than 90% of the new chemical compounds reported.[16]

IS THERE AN INFORMATION PROBLEM?

One can tentatively conclude from the foregoing commentary that all investigators seem to agree on the importance of the primary literature to the university chemist. They believe the JACS is the definitive journal in chemistry; that the average university chemist can supply most of his needs from a restricted number of sources; that the most important aspect of informal communication takes place at scientific meetings; that libraries are essential but afford very little personal service help because the level and range of specialization within chemistry is so great they place too heavy a burden on most available facilities and staff.

Nevertheless, few university chemists are happy with the existing state of affairs. Most chemists feel there is too much information. It, in fact, hinders the effectiveness of their research. Herner[5] supplies three important complaints:

> 1. Too much publication, the average paper is too short, and too many are worthless.
> 2. Publication is too slow, leads to duplication.
> 3. Helpful information, particularly recent information, is too difficult to find.

To this group the most pressing problem remains to be solved—i.e., directing the right man to the right material at the right time.

But how? In the age of Big Science the volume of publications is incredible. This year 3,000,000 articles in some 35,000 journals are being published in more than 60 languages.[17] There is no sign that the annual literature growth rate will slacken in the foreseeable future. Most scientists who have ever lived are active today, and 50% of them have been produced during the last 10 years.[18] An estimated 1,200,000 significant scientific documents will be generated this year. In biology, chemistry, and engineering (each about ¼ of the total), keeping up with all developments will require studying 1000 articles a day.[19]

The causative forces can be identified. There are two:

> A scientific revolution, funded by government agencies that are mission-oriented, has caused the emergence and rapid development of interdisciplinary science. This phenomenon is being exacerbated by the scientists themselves and is causing major stresses on the institutional forms through which science is conducted, particularly institutions of primary and secondary publication.
>
> A communications revolution, stimulated and supported by the proliferation of information, has rapidly outpaced the efforts and capacity of traditional institutions designed to provide storage and access. More important, new technology for the storage and transmission of information has overtaken traditional institutional capacity for the control of information, resulting in what Overage described as a physical, operational, and intellectual crisis in scientific information handling.[20]

The **physical crisis** has resulted from the sheer bulk of the material generated each year. This has led to limited coverage by secondary sources—about half of the useful articles are not abstracted, while many are repeated in as many as three abstracting services; time lags—the average time it takes to get an article published in a refereed journal is between 1 and 2 years. This has resulted in:

> development of other forms of publication, examples of which are preprints, conference proceedings, and private reports, which are introducing an alarming amount of information noise
>
> gaps in publication
>
> difficulty in determining who is doing research and how far the research has progressed
>
> difficulty in finding and identifying the thousands of products invented that have been marketed and produced.

The **operational crisis,** resulting from the exorbitant cost, in both time and money needed to make each item readily available to the growing number of users has led to:

poverty in many local collections
lack of trained personnel to service user needs
difficulty of access to the information in storage
difficulty in handling new forms for transmission of information (including magnetic tapes, microtexts, and film).

The **intellectual crisis,** resulting from our inability to describe an item by words or numbers that will make it accessible for easy retrieval, is exemplified:

by the interdisciplinary complications arising from the interdependence of knowledge and the growth of mission-oriented research

by the publication of 50% of all scientific and technical publications in languages other than English

by our lack of understanding of the processes of distribution, dissemination, and storage of messages in great quantities for easy access; by our inability to ensure the final connection between great stores of messages and the unknown minds to whom they are not addressed but for whom they were intended; and by our uncertainties relating to the functions of our mental processes, particularly those involved in the interpretation of meaning once the message has been transmitted, which remain controversial in theory and are effectively unknown.[21]

Whatever improvements are made must be made within two additional constraints. The first involves the rate of input and output of information to the human brain. The simple fact seems to be that, regardless of the rate at which recorded knowledge is accumulated through the years, the rate at which we read remains constant.[22] The second constraint arises out of characteristic inertia. As Mooer's law would indicate, an information system will tend not to be used whenever it is more painful or troublesome for a user to have information than for him not to have it.

IS THERE A SOLUTION?

I don't know, but most experts are optimistic. They point to two developments:

Recent efforts on national and international levels for cooperation in the sharing of resources, the development of shared acquisition programs, and in attempts to standardize techniques for cataloging.

The potential, and in some cases the proven value, of audiovisual devices and computers for information retrieval.

Both developments are important and worthy enterprises but they will reduce, not eliminate or solve, the problem. AV devices and computers are limited by restrictions inherent in their systems and beyond their immediate control. Until these restrictions are eliminated and until the causative forces described earlier are corrected, no lasting improvements can occur.

An ultimate information system has been described by Licklider.[23] Ideally, it would include all of the following objectives:

Organize and process information so that the user has immediate access to all facts, concepts, ideas, or other items of information relevant to his particular needs.

Handle a wide variety of inputs, including formal publications, informal notes and comments, and items derived from more than one source.

Reduce the difficulties now caused by the diversity of symbols, languages, jargons, and terminologies.

Eliminate publication lag.

Make possible retrieval of all or nearly all existing information relevant to a search.

Respond immediately to a user's request, interact with him, and adjust to his level of sophistication.

Facilitate interaction of groups of coworkers with each other and with the system.

Permit users to deal with abstract concepts and relationships, or with substantive information, or with both at once.

Provide the user, on demand, with either the flexibility, legibility, portability, and convenience of the printed page or the dynamic quality and immediate responsiveness of the modified TV screen or cathode-ray tube and the light pen.

Store information on the users' interests and needs for the purpose of formulating policies about acquisition and retention, taking the initiative to keep each user informed about new information which is in his field of interest.

Develop flexible working relationships with other systems such as research systems, governmental information-acquisition systems, and industrial application systems.

Standardize cataloging, indexing, and abstracting activities to make them more efficient and valuable to users.

Record and process all bookkeeping, billing, receipts, and disbursements.

Provide special facilities (languages, computer programs, computer subroutines, processors, and displays) for use in making and in implementing decisions that affect system policies and rules.

Handle guidelines, strategies, tactics, and rules of thumb intended to expedite solution of information-processing problems.

Provide for continuing efforts to improve the organization of an easy access to the existing body of knowledge.

It is obvious that many of these solutions are years away. It is also fair to assume that some of the problems may prove insoluble—e.g., organize information so that it will immediately be accessible. Nevertheless, the optimum should be our goal.

What remains, though, is the present and immediate future. What can we do now to help the user? There are many suggestions. Let me list some:

Find a way to restore the scientific journal to its former importance. This cannot be done by an outside agency, such as a library. It must be done by the scientists themselves basing their acceptance on the scientific merit of the work presented. This would reduce the quantity of publication.

Offer strong enough incentives to provide the kind of respectability and prestige that would attract competent scientists into the writing of critical reviews, progress reports, bibliographies, monographs, and textbooks. Secondary publication has faltered. It will always be needed so it should be strengthened. At present all publications, whether they be garbage, scraps, or half-baked items, are being accorded the same treatment when acquired for permanent storage.

Standardize the format of proceedings from symposia, colloquia, and congresses. Ensure publication of all papers; improve, through selectiveness, the quality of distribution. A conservative estimate places the number of yearly conferences at 5000. Each publish some form of proceedings. Liebesny estimates that 48.5% of all papers presented at U.S. conferences are not published and that 32% of those published appear in periodicals other than the one in which summaries appear.

Develop tools which will provide better, quicker, and more reliable current awareness of information.

Provide a mechanism which will produce a single, unified calendar which will yield the information needed by scientists to identify the over 5000 conferences which are held yearly.

Establish an agency which will provide the necessary but difficult information to find about instruments, equipment, techniques, and products.

Develop an effective training program for information scientists, particularly those with strong subject backgrounds, which will have sufficient depth to equip practitioners with motivation for problem solutions.

Develop techniques which will reduce the time lag of publication, for items which are useful, and not uncontrolled "noise."

Help develop useful training programs and orientation courses which will teach the user how to use the resources available to him. Many users do not use information sources effectively.

Support and encourage libraries to develop standardized local cooperative schemes for acquiring, processing, and servicing materials. This will save large quantities of money, minimize duplication, and free staff to introduce or improve existing information services.

Develop criteria from which librarians and systems analysts can discern patterns of use, including priorities of need. The truth is that we know very little about the way users use materials. For example, we know that relevance, accuracy, and availability are the concepts most often associated with useful information, but we also know that those are essentially personal judgments.

Help involve scientists in the development of programs which will improve subject indexing and abstracting technique.

Help develop chemical information specialists who will aid in designing software techniques which will allow the introduction of electronic equipment into the information retrieval function of libraries.

All the items I have noted are possible within the means available through today's technology. The only question is whether we desire to concentrate more of our resources in this area.

One cannot help but remember the words of Barnabe Rich in the year 1613 (as quoted by Price[18]) as he ponders this whole information problem: "One of the diseases of this age is the multiplicity of books, they doth so overcharge the world that it is not able to digest the abundance of idle matter that is every day hatched and brought forth into the world."

REFERENCES

(1) "Science, Government and Information," a report of the President's Science Advisory Committee, The White House, Jan. 10, 1963.

(2) Bach, H., "Scientific Literature Use; A Survey," *Special Libraries*, **48**, 466 (1957).

(3) Bernal, J. D., "Scientific Information and its Uses," *ASLIB Proceedings*, **12**, 432 (1960).

(4) Voigt, Melvin, "Scientists Approach to Information," ARCL Monograph, No. **24**, ALA, Chap. V., p. 49, 1961.

(5) Herner, S., "Information Gathering Habits of Workers in Pure and Applied Sciences," *Ind. Eng. Chem.*, **46**, 228 (1954).

(6) Jahoda, G., "Information Gathering and Use Habits of Chemists," *J. Chem. Doc.*, **9**, 153 (1964).

(7) Cole, P. F., *J. Doc.*, **14**, 197 (1958).

(8) Columbia University, Bureau of Applied Social Research (project leader: Herbert Menzel), "Formal and Informal Satisfaction of the Information Requirements of Chemists," interim reprint, *The University*, Aug. 1966 (PB 173 261).

(9) *The Library Association* (G.B.): "Report of the Committee of Scientific Library Service," 1.43:7 (1968).

(10) Kessler, M. M., "Bibliographic Coupling Extended in Time: Ten Case Histories," *Inform. Stor. Rev.*, 1, 169–87 (1963).

(11) Kessler, M. M., "Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing," *Am. Doc.*, **15** (3), 223–33 (1965).

(12) Kessler, M. M., "Some Statistical Properties of Citations in the Literature of Physics," privately circulated preprint.

(13) Kessler, M. M., and Heart, F. E., "Analysis of Bibliographic Sources in the *Physical Review*," privately circulated preprint.

(14) Barrett, Richard I., and Mildred A., "Journals Most Cited by Chemists and Chemical Engineers," *J. Chem. Educ.*, **34**, 35 (1957).

(15) Bonn, George, "Science-Technology Periodicals," *Library J.*, **88** (1), 954 (1963).

(16) Garfield, Eugene: "The Role of Government, Professional Groups, and Private Enterprise in Scientific Communication," a paper presented to the National Academy of Sciences, Scientific and Technological Committee, New York, N. Y., Dec. 16, 1966.

(17) Sandek, Lawrence, "Man's World of Facts," *Data Processor*, **10** (6), 4 (1968).

(18) Price, Derek J. de Solla, "The Foundations of Science Policy," *Nature*, **106** (4981), 233–8 (1941).

(19) Pasternack, Simon, "Is Journal Publication Obsolete?" *Physics Today*, **19** (5), 38 (1966).

(20) Vagianos, L., and Kurmey, W., "Scientific and Technical Information in Canada," Part II, Chap. 6, pp 4–5, The Queen's Printer, Ottawa, 1969.

(21) Swanson, D. R., "On Improving Communication," *Bulletin of the Atomic Sciences*, **22** (2), 9 (1966).

(22) Bernard, J., C. W. Shilling, and J. W. Tyson, "Information Communication among Bio-Scientists," Washington, D. C., George Washington University Biological Sciences Communications Project, 1964.

(23) Licklider, J. C. R., *Libraries of the Future*, Cambridge, Mass., M.I.T. Press, 1965.