Multidisciplinary Information Sources*

MONROE E. FREEMAN**
Smithsonian Science Information Exchange, Inc.,
1730 M St., N.W., Washington, D.C. 20036

Received September 16, 1971.

For half a century scientific progress has developed toward interdisciplinary and broad-scope technologies. At the same time, the information resources necessary for orderly scientific development have become increasingly fragmented and specialized. The multidisciplinary or interdisciplinary scientist has found it more and more difficult to locate precisely the information he needs. More sophisticated hardware cannot solve the problem. What is needed is more attention to the intellectual input and output of information systems, focusing on actual user needs. The most fruitful approach may be a compromise between a keyword system and a hierarchical system.

For more than 50 years scientific progress and the related information sources have pursued divergent courses.

It was early in this century that science progressively opened new frontiers at the interfaces of traditional disciplines and technologies. Biochemistry became established in the second decade, biophysics in the third, bioengineering more recently. This interdisciplinary trend did not distract many research scientists from exploring "more and more about less and less." At the same time, however, the interdisciplinary trend enabled others to formulate new viewpoints, new definitions, and new conceptual approaches to much broader multidisciplinary problems and programs. The total national program on water resources of 5000 to 6000 currently active projects, for instance, is a selected multidisciplinary mix of hydrology, geology, meteorology, biology, engineering, and even law and political science. Food supply is no longer a matter of merely increased production. The broader programmatic approach includes transportation and distribution systems, sociology, economics, and cultural and ethnic influences. In still another area, urban problems involve every human activity carried on in a crowded environment. This conceptual approach to broad multidisciplinary programs invites the traditional disciplinarians and specialists to descend from their ivory towers and rationally integrate their specialized expertise into the total interdisciplinary mix of these broad problems. To do this they need information and awareness of how their own specialties impact, infringe, and contribute along with the other peripheral disciplines and technologies to the over-all programs.

While all this has been going on in the advancing front of scientific research itself, the other side of the coin reveals the progress of the organized information sources pertaining thereto. For several millenia general libraries have been the primary sources of scientific information—ever since early Mesopotamians scratched hen-tracks on clay tablets. Early in this century, however, the general libraries began to outgrow an individual scientist's capability and patience to sift the accumulated mass of information for the detailed bits and pieces he needed at the moment. The principle of "divide and conquer" was applied and special technical libraries began to appear and proliferate. Because they focused their services on a discipline or a mission oriented technology, the over-all result was fragmentation of the total mass into smaller and more manageable segments.

By mid-century the rapid expansion of research and development outran the capacity of established publication channels. New information increasingly appeared as progress reports, technical bulletins, and the like. This so-called report literature was usually printed in limited editions with limited distribution. In general, it was "hard to find and hard to get," especially for those who were not specialists in that particular subject field. Like their "kissin' cousins"—the technical libraries—some of the documentation centers focused on a single discipline or technology. Others zeroed in on their organizational or corporate research, interests, or missions. In any case, they all specialized in the so-called report literature, the kind that was "hard to find and hard to get." Consequently, the total mass of scientific information was again split into smaller and more specialized fragments.

By about 1965, it became apparent that the doctrine of "divide and conquer" would not suffice when applied indiscriminately to the gross masses of information. There simply was too much junk, and hence the Technical Analysis Centers were devised to eliminate the trivia, the inconsequential—and the junk. Authoritative experts presumed qualified for the discriminating task of deciding what was fit to read and what was not, were assembled in small circumscribed groups. Their efforts, indeed, produced manageable masses of highly refined information, but, at the same time, they were fragmented into even smaller segments of subject specialization.

This progressive fragmentation of information sources did nothing to help those scientists with broad inter- and multidisciplinary interests, problems, and programs. To collect and mobilize the multidisciplinary information specific to their own individual problems, they had to seek out many fragmented and isolated sources, and then sift through each one to find the detailed information they wanted. For example, much of the total information about algae is irrelevant to water resources research. Algal scum in water reservoirs is relevant.

This progressive fragmentation of science and scientific information is not peculiar to organized information

*Presented at 3rd Northeast Regional Meeting, ACS, Buffalo, N. Y., Oct. 11, 1971.
**Retired Dec. 31, 1971.

sources. In a recent editorial in *Science,* Dale Wolfle[1] points out that the department structure of universities along traditional disciplinary lines may not be the optimum organizational structure to cope with these inter- and multidisciplinary problems now facing the nation and the world.

Obviously, some way must be found to put the information fragments back together again if an effective information base for multidisciplinary problems is to be served efficiently. Over the years, there have been several attempts and it will take only minutes to review their successes, failures, and potentials.

Early on, monolithic information centers, were proposed to do all things for all people, inspired no doubt by optimistic enthusiasm for computers as "giant brains." Computer hardware was adequate. "Giant brains" they were not. However, the concept dies hard. Each session of recent Congresses has been introduced to these proposals in one form or another.

The more realistic information practitioners advanced this concept to that of a network system of information sources linked together by real-time remote interrogation computers. The hardware seems to be adequate and available, although currently too costly. The real hang-up appears to be in the intellectual inputs and outputs. When the latter are simple, tangible, and unambiguous, automated network systems are feasible, practicable, and successful. Airline seat reservations, police networks, and inventory controls are familiar examples.

When the inputs and outputs are intellectually complex, intangible, and/or ambiguous, we still seem to be a long, long way from successful network systems. Scientific information is a familiar example. In the context of this symposium the intellectual hang-up is essentially the wordage problem.

The wordage problem of a multidisciplinary information base is further complicated by the fragmentation of information sources. It has several critical aspects.

First, each scientific specialty developed its own technical vocabulary (jargon if you will) often to a point where it is unintelligible to other scientists even in closely related fields. This communication difficulty is moderated by several factors. Within a discipline, technology, or narrow specialty field, the common training and experienee of the members so minimizes vocabulary problems, that the popular panaceas of prescribed word lists, authorized keywords, and "official" standard thesauri have a reasonable chance to succeed. The more limited the subject field, the more successful they are.

When disciplinary frontiers are crossed, however, the vocabulary problem becomes increasingly acute. Plasma means different things to biologists, to physicists, and to mineralogists. Hysteresis means one thing to an electrical engineer and something else to a colloid chemist, especially if they successfully avoided the study of Greek in their early education.

A second related factor is the level of sophistication of the users. Nuclear physicists exchanging ideas about their current research need all the special technical vocabulary they can muster. On the other hand, successful communication with my 10-year old granddaughter demands a severely limited vocabulary. Furthermore, in the more sophisticated realms of scientific research, commonly understood words and technical terms acquire different meanings and nuances when used in the context of novel hypotheses, new ideas, and scientific speculation.

Some interesting experiments have been made—mostly in government circles—to prescribe, arbitrarily, "standard" vocabularies of technical terms. Enthusiasm has been waning for a number of reasons: First, the information specialists promoting these schemes overlooked the important fact that scientists are a rather fractious lot if they are forced to think, speak, or write in a prescribed manner not of their own choosing. From their point of view, this violates every liberal prejudice from Academic Freedom to the Constitutional guarantee of free speech.

Secondly, a prescribed standardized vocabulary graven in stone or printed in thesauri essentially ignores the universal problem of communication—namely, semantics. Many years ago this lesson was succinctly stated by Humpty-Dumpty, "a word means exactly what I want it to mean—no more, no less." In the modern Wonderland of Science, how can this irony by a mathematical logician be so consistently and so expensively ignored?

Thirdly, the proponents of standard vocabularies have focused rather exclusively on the subject content of the discipline rather than on the varied viewpoints of the user-audience. A rose is a rose is a rose to horticulturists, but to an entomologist a rose may be only the preferred domicile of aphids and beetles. A plant pathologist may be interested in roses only as the host of virus diseases, while an enamoured fiance deems a bundle of roses as a declaration of everlasting love and devotion.

More to the point is a research report on the synthesis of steroids for patient therapy. If this were indexed and stored from the viewpoint of a medical clinician, a synthetic organic chemist might have difficulty retrieving the chemical details he wanted and vice versa. Two identical reports on the chemical and physical properties of water in harbors and estuaries were categorized by one sponsor as water resources research and by the other as oceanography. These simple illustrations, could be resolved by judicious cross-reference, but when extended to broad multidisciplinary systems, such resolutions are far from simple even for human brains and well nigh impossible for "giant" ones. The main point, however, is that acceptance by the user is the name of the game. Consequently, the cardinal criteria of systems design necessarily depend on each user's viewpoint and each user's needs. Acceptance depends neither on the exquisite engineering of the system itself nor on the prideful trumpetings of the information engineer who prescribed it. Unfortunately, a broad multidisciplinary information base attracts an equally broad multiplicity of users, with diverse viewpoints and wide diversity of needs. This leads us right back to where we started, a system (network or monolithic) intended to do all things for all people.

A different approach to a system that might possibly do all things for all people was conceived 7 or 8 years ago and has popped up periodically ever since. In essence, this type of system enters full text information (from whatever working documents are being used) into a computer which has been programmed to build a complete vocabulary of all significant words and terms in the text. The user is then on his own, time-wise and money-wise, to retrieve whatever information he can find.

If such a system were successful and accepted by users, it would present very attractive advantages.

First, major costs are transferred from the operational input to the output side. The indexing labor at the input end is eliminated, and this substantially reduces the input cost in many sophisticated systems. The user, via direct access to computer storage, can use any retrieval words, terms, or clues to get the information he wants. He is on his own and absorbs the cost of retrieval. This makes the system manager very happy. It shifts major cost items from his shoulders to the user's.

Second, the users are happy, in the beginning. They have video terminals on their desks directly connected to computer files that may be hundreds of miles away. A video terminal on your desk is quite a status symbol and advertises that the proud owner commands instant answers at any time.

Thirdly, the system can be easily linked to many different information sources in an idealized network system.

Unfortunately some disadvantages have been encountered: First, hardware and software are rather expensive for the average user that has only occasional inquiries. What price glory—or status? Undoubtedly these costs will decrease substantially if the number of users and volume of usage increase. At present the costs are rarely justified except where many inquiries can be concentrated at one terminal.

Second, the wordage problem is still the primary hang-up. Each paper or report in storage necessarily reflects the author's viewpoints, oriented towards his own interests and mission-oriented objectives. Consequently, his vocabulary and semantic nuances may or may not be entirely consistent or compatible with every user's choice of retrieval words and terms. For instance, if the aforementioned report on synthetic steroids had been written by the chemist, a medic searching for clinical details might have to exert considerable time, effort, imagination, and some ingenuity to find the right words or terms to retrieve this report. If he chose all-inclusive retrieval terms like steroids or hormones, he would surely get the report, but only after the time-consuming effort of reviewing and screening out hundreds of irrelevant reports. My previous remarks regarding the complexity of the intellectual inputs and outputs apply equally here. Within the confines of a well-defined discipline or technology where commonality of thought, speech, and writing is the rule, these wordage and retrieval vocabulary problems are minimized. In multidisciplinary information systems, they become acute. According to many practical field tests, only about half the related material actually present in storage can be anticipated under these circumstances. Indeed, a practical field test of a recent version of such a system retrieved only about 65% of the related information. A recall of 60% is considered quite adequate by many information practitioners, but I hasten to remind you that it is the user, not the information specialist, who is the final judge of adequacy, satisfaction, and acceptance. Furthermore, these users have different levels of scientific sophistication and, consequently, different criteria of adequacy.

To summarize briefly, the development of inter- and multidisciplinary problems and programs have been described and their importance emphasized. The countercurrent fragmentation of information sources has been pursued until they can neither effectively nor adequately serve as an information base for such multidisciplinary problems.

Electronic data processing devices have proved to be helpful, but only partial solutions to information problems. Critical analysis leads us to conclude that hardware, software, and systems concepts are available and adequate, and that the real difficulties are centered in the intellectual inputs and outputs. If these are simple, tangible, and unambiguous, the wordage problem is minimized and automated information systems (networks included) are feasible, practicable, and successful. As wordage, technical vocabularies and concomitant semantics become more involved and complex, as in multidisciplinary information bases, for instance, we have not yet approached any satisfactory solution, for even moderately sophisticated scientists.

My contribution is not a smooth, slick panacea but only a diagnosis pointing to the critical point of difficulty. "Giant brains" and elegantly engineered systems have not and, in my opinion, will not achieve solution. The critical point is the intellectual input and output. Real progress in this salience will be made only when systems and designs are refocused on the viewpoints, interests, and needs of the potential user-audience.

After 10 years on the horns of this dilemma, my crystal ball is as cloudy as anyone else's, but I will conjure up the temerity to grasp the bull or the dilemma by the horns and suggest that the most gainful pursuit may be some combination or compromise between a keyword system and a conceptual hierarchical classification system. Only the latter faithfully reflects the innate logic of a scientific discipline or technology. Inasmuch as the potential users are scientists, technologists, and engineers, they learned their trade, they speak, think, work, and pose their queries within the framework of the hierarchical classifications of their sciences.

Having climbed out on this limb so far, I might as well venture farther and add a postscript to the long-range significance of these multidisciplinary, multiagency information resources. As previously noted, they are the information bases necessary to the important scientific and technical problems facing the nation today—the urban problems, environmental, ecological, and food supply. Beyond the scientific and technical problems is an ever-widening gulf between the technologies and the humanities. Nancy Hanks,[2] Chairman of the National Endowment for the Arts, in a recent editorial in *Science*, noted that this gulf has been rapidly widening since C. P. Snow, the world famous scientist and man of letters, developed the theme of two cultures 12 years ago at Cambridge University. Attempted disruptions of scientific meetings and the declining support of research and development are dramatic symptoms which lend even more emphasis to Sir Charles's discernment. If information specialists can bridge the information gaps between the scientific disciplines and technologies, some clear-eyed genius of the future might see a way to bridge the widening cultural gulf with meaningful communication channels that might reconcile the two cultures of science and nonscience.

LITERATURE CITED

(1) Wolfle, Dale, *Science*, **173**, No. 3992, 109 (1971).
(2) Hanks, Nancy, *Ibid.*, **173**, No. 3996, 479 (1971).