Information—A Problem of the Research Scientist

By Burton W. Adkinson*

"What you don't know won't hurt you" is a cliché whose lack of validity probably is exceeded only by its age. Think what an unlighted, open manhole at night can do to you! A paraphrase of this old saw that I think *can* be defended, at least in the context of my remarks today, is that "What you don't need, needn't worry you." The really serious problems are posed by the things and services you need badly and are unable to obtain. It is only because prompt, reliable access to accurate scientific information is so necessary to good research that the title of this talk makes any sense.

But, let me preface my discussion of scientific information as a problem for the research scientist, with one or two remarks relative to the over-all university research complex. There are many reasons why a university can provide an ideal environment in which to conduct research. An extremely important one, is the integrated combination of "teaching" and "trying" that is inherent in the university concept. All good education involves experimentation, and all good experimentation involves education. The three fundamental requirements for any effective research program, and their translation into the specific terms of a university situation may, I believe, be stated as follows:

1. *Competent personnel:* A well-trained, imaginative, dedicated faculty; alert, intelligent, enthusiastic students;
2. *Good physical facilities*: Adequate laboratories, apparatus, and associated equipment; and
3. *Adequate supporting services*: A research-minded, far-seeing university administration; effective library and other information facilities.

Having stated this arbitrary breakdown of what seem to me to be the principal elements of any research program, I hasten to add that I recognize that here, as in most classification schemes, the subgroups are not independent of each other, and changes occurring in

* Head, Office of Science Information Service, National Science Foundation, Washington, D.C. Paper delivered April 5, 1962, before the Association of Texas Graduate Schools.

any one will have implications for the others. Man loves to categorize, and Nature seems to take even greater delight in preventing his carefully designed categories from ever being mutually exclusive. I certainly need not emphasize to this audience the wheels-within-wheels relationships that exist among the basic elements into which I have divided university research capability. You have had much greater, and far more direct, experience than I with the extent to which, for example, the good faculty, good equipment, and good library problems are functions of each other.

My three-category breakdown of the fundamental ingredients of a good research program hinted at the place of scientific information in this picture. Let me discuss this particular aspect in greater detail. Here I shall try to give you brief answers to five questions:

1. What is the place of scientific information in the research process?
2. What are "good" information services in the research context?
3. Why does scientific information dissemination present a particularly serious problem today?
4. What principal kinds of remedial measures are being taken?
5. What can universities do toward coping with the scientific information problem?

Scientific Information's Role in Research.—Every research project starts from a base of some body of knowledge or information; also every research project produces information useful for future experimentation. The latter part of the preceding statement is true even when results are negative; to know what *not* to do can be just as important as knowing what *to* do. Thus, information is both a raw material and a finished product of research, with the informational output of one project or program becoming informational input for another one. In very fundamental research, information invariably is the principal and often is the sole product. In developmental or applied research, it is at least an important product along with the concrete, three-dimensional objects that also may be turned out. In short, the conclusion seems inescapable that prompt and dependable information services are as essential to maximum effectiveness of research as are good equipment and competent scientists.

Characteristics of Good Information Service.—The perfect scientific information system for research scientists presumably would automatically place in front of every scientist, at just the time he needed them, exactly (no more, no less) the results and analyses of previous research that would be pertinent to his work at that instant. We probably need not consider this alternative further.

Perhaps a minimum, or bare subsistence, level of scientific information service can be defined in terms of a statement Lord Rayleigh once made. He said, "By a fiction as remarkable as any to be found in the law, what has once been published (no matter the language) is usually spoken of as known, and it is often forgotten that rediscovery in the library may be a more difficult and uncertain process than was the first discovery in the laboratory." In a related vein, Santayana said, "Those who cannot remember the past are doomed to repeat it." So, perhaps we can say that, at the very least, a scientific information system should allow a scientist to locate readily the results of previous research pertinent to his work. Somewhere between these limits lies the maximum level of scientific information service that realistically can be provided in any given case.

This level varies, of course, both with the particular research situation and with time, with "realistically" obviously being a very significant word in the preceding sentence. Among the elements that must be considered are: (1) access to primary journals, monographs, abstracting-indexing sources, conference proceedings, and the like, pertinent to the research programs concerned; (2) availability of competent information searching assistance; (3) opportunity to attend and participate in professional conferences and symposia; and (4), perhaps most important of all, administrative officers who realize the importance of this aspect of the research process and who do their best to provide effective information services.

Why the Scientific Information Problem is Serious.—Traditionally in the U. S., as you are aware, dissemination of the results of scientific research has been accomplished principally through the complex system of scientific publications and meetings mentioned above. Personal conversations and correspondence also have been important media of scientific communication. This pattern has the extremely important advantage that it is largely controlled by the scientific community it serves, with this control being exercised principally through the scientific societies. This scheme always has had the inherent disadvantage, however, that the information phase of research and development is rather completely divorced from the experimentation component, so that changes in the latter do not automatically affect the support available for the former. Even so, up to World War II, our total capability for reporting and keeping track of the results of research was essentially in equilibrium with the research and development production of new information, and the system worked quite satisfactorily.

Three major developments during and since the war have destroyed this balance and made increasingly serious the problem of providing the research scientist with the scientific and technical information he needs. First, and basic to the others, has been simply the vast, and still growing, expansion in the research and development programs. Just prior to World War II, total U. S. expenditures for pure and applied science totalled between 250 and 300 million dollars.[1] In recent years the nation's expenditures for research and development have been doubling about every seven years and in fiscal 1961 reached some 14 billion dollars. (In comparison, most nonscientific aspects of our culture have been doubling every 30 to 50 years.[2]) Something over 60% of all U.S. scientific research and development is supported directly or indirectly by Government funds; in the fiscal 1963 Federal budget, one dollar in every eight is to be spent for research and development. Even allowing for rapidly mounting costs of personnel and equipment, this growth represents a staggering increase in both the relative and the absolute magnitude of this phase of our national life.

Just as tripling the size of a nuts-and bolts factory can logically be expected to provide roughly three times as many nuts and bolts for sale, so the many-fold expansion of research and development has resulted in a deluge of new scientific information to which research scientists should have access. Another way to emphasize how natural and expectable this growth in the scientific literature has been, is in terms of scientific manpower. Derek Price has shown that while the average production of scientific papers per man per year has remained about constant for decades, the number of scientists has grown in

---

[1] Bush, Vannevar, *Science, the Endless Frontier,* National Science Foundation (NSF 60-40), 1960, page 87.

[2] Price, Derek J., *Science Since Babylon, Yale University Press,* New Haven, Conn., 1961, p. 107.

the last 300 years from a few to hundreds of thousands. In fact, the rate of increase has been such that some 80 to 90 per cent of all scientists who have ever lived, are alive today.[3] Extensive statistics to demonstrate the enormous quantity of scientific literature that exists today and the rapid rate at which it has been increasing could be cited from almost every field of science and technology. I shall mention here only a few representative figures.

Chemical Abstracts has for years covered the world's scientific literature of interest to chemists. From 1907 to 1958, it carried some 20 million abstracts; of these, 40% appeared in the last eight years of this 51-year span. The more recent rate of increase of the literature in this field is indicated by the fact that this abstracting journal published 42,000 abstracts in 1948 and 145,000 in 1961. *Engineering Index*, which selectively indexes and annotates articles of engineering significance from over 1,500 publications published in more than 40 countries in over 20 languages, carried approximately 38,000 items in the annual volume published July 1961. *IRE Transactions* for 1960 consists of six volumes which contain about 1,200 articles as compared with approximately 500 articles in three volumes in 1955. Similar multifold increases in the literature have occurred in substantially every area of science and technology.

As I am sure you know, this phenomenon of explosive research and development expansion—and, therefore, of the growth of scientific literature—has not been limited to the U. S. or even to the English speaking countries; it has been world-wide. This leads me to the second major reason for the increasing seriousness of the scientific information problem. Not too many years ago, English, with occasional excursions into German and French (in which, of course, by definition all U. S. scientists are proficient) enabled most research scientists to keep reasonably well informed in their respective fields of interest. This is far from true today. Estimates show that nearly 30% of the world's scientific literature appears today in languages read by fewer than five per cent of U. S. scientists. In chemistry, almost 30% of the material covered by *Chemical Abstracts* is in the following five languages that fewer than two percent of our scientists read: Russian (16.8%), Japanese (6.1%), Italian (3.6%), Polish

[3] *Ibid.*

(1.5%) and Chinese (0.7%).[4] Data for other scientific fields are not greatly different. The old idea that everything good eventually comes out in English probably always was more a salve for conscience than statement of fact; any validity it may have had in the past certainly has vanished today.

The third major reason why the scientific information problem has become so critical involves the rapid breakdown in recent years of our tidy "line fences" between the subject areas into which traditionally we have so neatly divided science. I have already mentioned Nature's contemptuous attitude toward man's proneness to establish pigeon holes. This fact is nowhere better exemplified than in the increasing overlap of the subject fields about which we used to feel so comfortable. Are biology and chemistry separate disciplines? And, if so, what about biochemistry—and what are its boundaries? Similarly, we have geology, physics, and geophysics; astronomy, radio, and radio astronomy; chemistry, physics, physical chemistry, and chemical physics; navigation, astronomy, and astronautics which, married to biology, has produced bioastronautics; and so forth.

These new fields, sub-fields, interdisciplinary areas—call them what you will—did not spring simply from scientists' carelessness and mental indolence. They came into being and have grown in importance because, as science has advanced, individual scientists have found their research more and more involving phenomena and making use of results from two or more of the conventionally defined disciplines.

These three major developments—growth in U. S. research and development, in the importance of foreign research, and in the interdisciplinary implications of research—have combined to produce numerous difficult problems. You, as officials in universities that have budgets and departments organized along conventional disciplinary lines, are painfully familiar with many of these. I am concerned here with their effect on the availability to the average research scientist or engineer of the scientific information he needs. From this standpoint, the consequences have been little short of catastrophic. No longer can he keep up with research related to his own work by reading a couple of professional journals, periodically

[4] Heumann, Karl & Peter M. Bernays, "Fifty Foreign Languages at *Chemical Abstracts*": *J. Chem. Educ.*, 36, p. 478, October 1958; also "Lost—One-Third of the World's Scientific Literature," National Science Foundation publication, NSF 60-20.

glancing through issues of a small or medium-sized abstracting publication, and occasionally attending a scientific meeting. In addition to being swamped by the literature in English, he finds that he also needs to be aware of what is being published in a number of other languages (many unfamiliar to him), and that increasing amounts of information pertinent to his work are turning up in subject areas that formerly did not need to concern him.

Equally troublesome problems have been posed to the information specialists and librarians who have to store, control, and service these constantly expanding and increasingly complex bodies of scientific knowledge, and to the administrative organizations—private and public—that have broad responsibility for the adequacy of the system as a whole.

Remedial Measures Under Way and Contemplated.—I shall take time here only to give a few examples of some of the principal components of the general attack being made on these problems. Any realistic, long-range remedial program must both plan imaginatively for the future and "keep the store open" while the new designs and procedures are being developed and tested. Consequently, the measures being taken toward solving the over-all problem are of two basic kinds. One group of activities is directed primarily toward the development of new and better methods of recording, processing, disseminating, storing, and retrieving scientific information; the other has for its main objective the improvement of existing facilities and techniques to make them more effective.

In the category of planning for the future, study and research are, of course, the predominant activities; four major areas of such investigation are:

1. Studies to identify the information needs of scientists and engineers with some degree of precision. Such knowledge obviously is prerequisite to an intelligent approach to developing improved procedures. Here exploratory work has been done by Case Institute of Technology and others looking toward the development of measures of the value of recorded information. Subject-oriented studies in this area are being conducted by the American Institute of Biological Sciences, the American Institute of Physics, and the American Psychological Association.

2. Research on information organization and searching. In this

area, the application of mechanized techniques unquestionably offers the principal hope for the future for rapid and reliable storage and retrieval of scientific information. Current research in this field is extensive and requires the combined talents of computer experts, mathematicians, linguists, logicians, librarians, and information specialists. That work in this area is being pursued on a fairly broad front is indicated by the fact that a partial list of organizations involved includes: The University of Pennsylvania, the Itek Corporation, the Electrada Corporation, Western Reserve University, the British Association of Special Libraries and Information Bureaux, Arthur Anderson and Company, and the Stanford Research Institute.

3. Research on machine translation of one natural language into another. In addition to the obvious and practical application that success in this area would have, the research has important long-range implications for linguistic analysis and for other aspects of mechanizing information handling. Among the organizations where research in this area is under way are the University of Texas, the University of California, Georgetown University, Harvard University, and Massachusetts Institute of Technology.

4. Research on the use of microforms. Expansion of the scientific literature at the rates described above obviously poses an increasingly serious storage problem for libraries and other document repositories. Research on the use of microforms is directed toward improving the acceptability and effectiveness of some of the newer types (Microcards, for example), toward minimizing the disadvantages of microfilm, and toward developing better and cheaper reading machines and other associated equipment. The Council on Library Resources, Inc., is particularly active in supporting investigations of these kinds.

Now, let me turn to my second general category—that of improving existing media of dissemination and the services they provide. A great many different kinds of such activity are under way and I shall have to limit my discussion to a few representative examples:

1. Improvement in primary journal coverage. Here, existing journals are being enabled to expand their coverage and eliminate backlogs of accepted manuscripts. The *Journal of Aerospace Sciences* and the *American Rocket Society Journal* are two recent examples. Also, new journals are helped to get under way in fields where authoritative

scientists agree that a clear need exists. For example, assistance to initiate *Applied Optics* was given in 1961. Also, scientific societies are urged to cooperate in defining more sharply the areas of coverage of their respective publications and in eliminating unnecessary overlap.

2. Improvement in abstracting-indexing coverage. Abstracts and indexes are the principal tools that scientists use to search for and retrieve previously published information. Remedial emphasis is principally on strengthening the major services through appropriate expansion of coverage in their respective fields and speedier publication of indexes. *Biological Abstracts, Sociological Abstracts* and *Meteorological and Geoastrophysical Abstracts* are services that have been helped in this way. Also, assistance has been provided for initiation of two needed new services—*International Aero/Space Abstracts* and *International Abstracts of Operations Research.* The National Federation of Science Abstracting and Indexing Services, organized in 1959, is working toward better coordination of all of the major secondary publications.

3. Provision of prompter announcement of published papers. Two new publications directed toward this end are *Chemical Titles* and *Current Contents.* The former, published by the Chemical Abstracts Service, is a permuted title index in which the title of every paper in a selected list of 600 journals covered by *Chemical Abstracts* appears separately with each significant word in the title, these words being arranged alphabetically. Similar title indexes are being started in other fields. *Current Contents,* issued by the Institute for Scientific Information in Philadelphia, consists of photoreproductions of the actual tables of contents of a large number of primary journals. Both of these journals are successful as judged by the interrelated criteria of acceptance by scientists and subscription income.

4. Improved access to Government technical reports. This peculiarly formidable body of literature, running to upwards of 75,000 unclassified documents annually, poses special problems of availability because in general it has not been covered by the conventional bibliographic control media. Two existing programs that are improving this situation are (1) maintenance of a report collection and reference service in the Library of Congress, and (2) publica-

tion by the Office of Technical Services, Department of Commerce, of a comprehensive abstracting service on the technical reports issued by AEC, NASA, the Department of Defense, and their respective contractors. Soon to be initiated are a comprehensive permuted title index of all reports and a network of regional depositories of Government report literature.

5. Improved access to foreign language literature. Pending the successful development of machine translation, the obvious bull-by-the-horns approach to this problem is simply to translate the material into English. Some 85 key Soviet scientific journals in 10 broad scientific disciplines are now completely available in the U. S. in English translation. Selected translations also are being made from Russian, Japanese and Chinese publications.

This brief summary of five kinds of remedial activity gives some indication of the scope of the total program concerned with improving existing scientific information services. Other areas in which work is under way include provision of pre-publication information on research in progress, development and coordination of specialized data centers, and improvement in the education and training, both of scientists in the use of scientific information and of information specialists to assist scientists in acquiring information. I shall refer to this last activity again.

What Universities Can Do.—The *general* approaches that can be made to assisting the scientist with his information problem mostly are implicit in what I have been saying about the problem and the remedial measures now under way. On this very broad basis, what a university can do is substantially no different from what any organization that conducts scientific research can do. I should like, however, to discuss several specific areas where, it seems to me, universities have particular responsibilities. That they have these special obligations (I would rather call them opportunities) results largely, I believe, from certain factors inherent in the University pattern of conducting research.

One of these factors is that university-sponsored research and development is, for the most part, directed more toward the search for new knowledge than it is toward the production of "things." And it is at the basic end of the research and development spectrum that information's dual role of both "raw material for" and "product of" the research process is of maximum significance. In my opinion, this fact places upon a university a particularly strong responsibility (1) to provide its research scientists with adequate access to the published and other information they need, and (2) to make certain that the results of the research it sponsors become promptly available to the scientific community as a whole.

Fulfilling the first of these obligations obviously will have a major impact upon the university library system. I heard recently of a college administration that, faced with the problem of expanding the school's library to meet accrediting standards, adopted a policy of requiring a book as part of the admission price to all football games. This probably is not the optimum approach to building a good research library. Instead there must be active, intelligent participation in the process of all concerned, and especially by the scientists and engineers the library is to serve. The biggest library, of course, is not necessarily the best one. Quality is more important than quantity, with excellence of the collections and services depending upon how fully they meet the needs of the particular people who will use them.

To discharge properly the second responsibility mentioned above, a university must recognize realistically—not just intellectually—a principle I stated earlier. This is that dissemination of the results of experimentation is an integral phase of the research process. Expressed in its simplest terms, this says that spending $1,000 to support certain experimentation *and* make its results generally available, is preferable to spending $1,000 for a slightly larger amount of experimentation, the results of which no one ever hears about.

To ticket a suitable fraction of the research dollar for information dissemination obviously is difficult at the appropriation stage because no one can predict accurately the quantity or quality of the results that will come from any given experiment. This fact, however, reduces neither the validity of the principle nor the urgency, in view of the seriousness of the situation I have described, of achieving some reasonable approximation to its application. The page charge, which has been employed successfully for many years by U. S. physics journals and is being adopted increasingly by other scientific periodicals, is one mechanism whereby a small fraction of research funds can be channeled into communicating research results. Since these charges are levied against the organization that supports the experimentation, this cost (actually only part) of initial dissemination of data is linked directly with cost of personnel, equipment, heat and light, and all other phases of the research that produces the data. It remains to be seen whether this particular kind of approach will in the long run prove to be the best means of incorporating the cost of dissemination of information into total research expenditures. The important point is that present conditions make fairly clear the necessity of developing some means of achieving this objective.

But before research results can be published they must be prepared for publication. Herein lies another aspect of the responsibility, both of scientists and of university officials, to recognize that no research project is complete until its significant findings have been made available. Although, as I have emphasized, the total existing quantity of scientific information is enormous, much research still is inadequately reported or not reported at all. Universities that sponsor and conduct research have an obligation to make certain that the scientists in these programs report their research findings fully and effectively.

However, every information situation has elements of "too much" as well as of "too little." The former, in the context of scientific research, is exemplified by the extensive and unwarranted duplication and multiple publication that plagues the scientific information field. The "publish-or-perish" philosophy, by which a scientist's professional standing depends more on how many papers he publishes than on how good they are, has been to a considerable extent responsible for this problem. Much of the blame for this situation lies, I think, with those university officials and others who emphasize quantity of publication more than quality in hiring and promoting scientists. I suggest that with these same officials lies the principal responsibility for ending this vicious practice.

The heavy emphasis on fundamental research in university research and development programs places universities in a particularly favorable position to participate in and take advantage of cooperative scientific information activities. Time scales are somewhat less demanding than in the highly developmental aspects of research and development, and there are likely to be relatively few problems of a proprietary interest or of a "company confidential" nature to impede cooperation. Universities that have overlapping or complementary research interests, and are not too widely separated geographically, can interrelate their scientific information programs for the mutual benefit of all concerned. A program currently being developed by Columbia, Harvard, and Yale illustrates one form this can take. The respective medical library holdings of these three schools are being combined in a single computer system. Thus each school will have rapid access at all times—through automatic print-out of catalogs, bibliographies, and the like—to the medical collections of all three libraries.

A university that has a scientific research program and is located in an area where there also is extensive industrial research and development, has a particularly good opportunity to develop research library and information facilities that will provide improved service to both the academic and the industrial scientists and engineers. There are a number of regions in which this kind of cooperative information program is being studied and considered. Under a recent grant from the National Science Foundation, the University of West Virginia is conducting a careful investigation of the potential uses that can be made of a large academic research library by the libraries of nearby industrial organizations and by smaller academic libraries. We believe that this study will provide valuable guide lines for this kind of cooperative information effort.

I referred earlier to the interrelationship between education and experimentation inherent in our basic concept of a university's function. This characteristic, in my opinion, gives the university that conducts research still another important opportunity (and responsibility) in the scientific communication field. Three factors that obviously affect the influence of scientific information upon the advancement of research are: (1) the skill with which the information is presented, (2) the appreciation that scientists and engineers have of its potential value, and (3) the competence of the librarians and information specialists involved in making it available to research people. Greatly increased educational emphasis is needed in all of these areas.

Prominently lettered across the front wall of a large chemistry lecture room at the Ohio State University is this sentence, "The

English language is the most important scientific instrument at your disposal. Learn to use it with precision." It is unnecessary for me to emphasize to this audience either the soundness of this idea in theory or the appalling extent to which it is disregarded in practice in much scientific and technical communication. Undergraduate and graduate training in many sciences far too largely ignores instruction in the use of scientific literature and of bibliographic tools in general. Finally, no institution is now equipped to provide librarians and information specialists with adequate training in modern methods of processing, storing, and retrieving scientific data, and in the new approaches that are being made to cope with the explosive growth of the technical literature. As the ideal agents to initiate and carry on remedial programs in these areas, the twin fingers of opportunity and responsibility seem again to point naturally at the universities engaged in scientific research.

Several weeks ago in the Miss Peach comic strip, two little girls seized control of their fourth grade civics club by a method that they described as "the fair and square way by which any group takes over any club—capture of the mimeograph machine." I would not contend that the scientific information "machine" all by itself should play quite this crucial a role in scientific research. I am convinced, however, that since scientific information is the beginning and the end of all scientific experimentation, its prompt dissemination is complementary to sound investigation. I have tried to identify and emphasize the very important part I believe the university can and should play in integrating these phases. To the extent that information services are not adequate to a research scientist's needs, the progress of his investigations is slowed down, his own stature in the scientific world and that of his university suffer, and the nation's scientific and technical advancement is impeded.