Information and Research—Blood Relatives or In-laws?

Dissemination of the results of experimentation is an integral part of the total research process.

Dwight E. Gray

That the control and effective dissemination of today's flood of scientific and technical information presents serious problems, no one will deny. On just how grave these problems are and how best to solve them, there are many opinions. Speeches, articles, books, symposia, panel discussions, studies, and surveys that view this situation with alarm and propose remedies are themselves assuming torrent-like proportions. Corrective measures that have been, and are being, recommended are legion and involve a complex variety of combinations of major and minor aspects of the total problem. Many appear to be well thought out and promising. Regarding some others, one occasionally gets the impression that the vigor with which they are promoted varies inversely with the impressiveness of the evidence adduced to support them, and that their potential for creating new diseases may equal or exceed the probability of their curing old ills.

But my purpose in this article is not to recite and attempt to evaluate this multitude of remedial efforts. My objective rather is to suggest that while we have been worrying like mad about how to arrange the second-floor furniture and where to locate the bathrooms in our information split-level of the future, we may have neglected to insure for it a firm and permanent foundation. I believe that prerequisite to the establishment of such a foundation is realistic recognition of the fundamental relationship that scientific information bears to scientific research. It is the basic thesis of this article that, whereas this kinship actually is of a blood-relation kind, information has been treated by the overall research and development community as a slightly suspect in-law or a cousin several times removed. I shall try to justify this proposition and then explore some of its implications.

An early ancestor who wished to plan and construct a better thatched hut than the one next door undoubtedly consulted with and obtained information from other hut builders of his day. (And, who knows, perhaps even earlier forebears held symposia and exchanged data on coconut throwing and swinging by the tail!) In those days of thatched-hut suburbia, "specs" did not need to be highly precise. A few words and gestures, with maybe a diagram or two drawn in the sand with a stick, probably sufficed as media of information exchange. Nevertheless, even then research and development began with the acquisition of technical information.

Today our scientific information system is immensely complex, and engineers and scientists flounder in an informational sea of journals, reports, monographs, conference proceedings, handbooks, theses, abstracts, indexes, data sheets, and personal correspondence. As the difficulty of locating particular items or aggregates of items of scientific information has increased, so have both the need for prompt access to the results of previous research and the importance of having data that are precise and reliable. In short, every research project uses information as an essential raw material. This is my first conclusion.

But certainly all research and development also produces new information—sometimes positive, sometimes negative. Both kinds are important, since knowing how *not* to proceed or what *not* to try often is as valuable as having positive data that can be directly incorporated into a research project. At what one might call the ultrabasic end of the R&D spectrum, information is the principal—indeed, frequently the only—product. At the infra-applied extreme, where research and development also turns out "things," new knowledge is at least an important coproduct. My second conclusion, then, is that since every research project generates knowledge, information is an important *product* of research as well as a valuable raw material.

In the Middle West, farmers sometimes are said to raise more corn to feed more hogs to sell more meat to buy more land to raise more corn to feed more hogs, and so forth. I believe the relation of information input and output to research can be validly illustrated by an analogous expression characterizing research and development as consisting, to a considerable extent, of a process of obtaining more information to trigger more new ideas to lead to more experimentation to produce more information to trigger more new ideas, and so forth. In short, combining the two conclusions already mentioned leads one inescapably, it seems to me, to a third, and I believe very fundamental, conclusion: that the processing and dissemination of the results of research—that is, of scientific information—is as integral a part of the total research sequence as experimentation is.

This principle has significant implications for the support of the processing and dissemination of scientific information. Typically, in our economy, three fundamental steps are required to bring any product into being and make it available to its consumer public: (i) obtaining the necessary raw materials; (ii) employing people and equipment to convert these raw materials into the finished product; and (iii) making this product available to users—both initially and continuously. An automobile manufacturer, for example, who doubled or tripled the capacity of his factory to produce motor cars would, as a matter of course, take whatever steps were necessary to augment the other two phases of the overall process. If a total of $\$Z$ million were to be devoted to the expansion, appropriate fractions of these funds would be used to insure increased supplies of raw materials and to augment distribution facilities to take care of the increased output. To do otherwise in a manufacturing situation would be

The author is science liaison officer of the Office of Science Information Service, National Science Foundation, Washington, D.C.

recognized by everyone as unintelligent, short sighted, and a forerunner of disaster. Now, let us see what kind of situation has evolved in research and development with respect to the use, production, and distribution of the two general kinds of scientific information —the results of fundamental research and the data from applied or developmental investigation.

Traditional Communication of Research Results

Traditionally, prior to World War II, fundamental scientific research in the United States was almost entirely privately sponsored, most of it being conducted by or in universities. The principal motivation for any basic research project was simply the personal interest of the particular scientist or scientists concerned. Pressures to rush the findings of fundamental experimentation into "practical" applications were relatively small and often essentially nonexistent.

Consequently, communication among scientists regarding the progress and results of their work was prompted largely by the spontaneous, unpressured desire to share ideas with, and learn from, their colleagues.

Except for personal conversation and correspondence, the natural and accepted media for such communication were those provided by the scientists' own professional organizations—namely, society meetings and society-sponsored journals. In fact, improvement of communication by these means was a major objective in the formation of most scientific societies; in some cases it was almost the only one. Also, the number of papers published in scientific journals became an important criterion for judging a scientist's professional stature—an aspect of scientific communication that has tended both to increase the quantity of the literature more rapidly than its quality and to complicate the problem of improving the overall communication system. As more unknown areas of more sciences were explored and charted, scientific societies, the meetings they held, and the journals they published all grew in number, size, and cost. And, as it became increasingly difficult for a scientist to attend all of the meetings and read all of the journals he should to keep abreast of his field of research, abstracting and indexing services—those important Baedekers to the primary literature—assumed greater and greater importance. Funds to support the meetings of scientific societies, and the journals and abstracting services they sponsored, came almost entirely from society dues and registration fees and from subscriptions of individuals and libraries. With a few modifications that are mentioned later, this still is largely the situation.

In so-called applied research the pattern of information dissemination has been to a considerable degree similar to that just described for fundamental research. As I mentioned earlier, usually closely associated with applied research is the production of "things," as well as of information on which to base further developmental research that will lead to the production of still other things. The principal media for making the results of applied research available, however, have been those noted above—meetings and publications. Again, most meetings for the exchange of technical data and know-how have been sponsored by professional societies—mostly those in the engineering fields. These groups long have published journals carrying papers on applied research and development, and here again the publications have received a significant part of their support from the parent societies and from subscriptions. The close identification of the content of these journals with applications has made them attractive as advertising media, thereby providing a major source of income that has been largely unavailable to the basic-research publications. Consequently, there are many technical or applied-research journals that are published strictly as commercial ventures.

In neither case—fundamental or applied research—has the dissemination of the results of experimentation really been treated as an integral element in the research process. Thus, the system has had the basic defect that variations in the magnitude of the effort in the experimental phases of research are not accompanied automatically by corresponding changes in the information handling and dissemination capabilities.

The Growing Problem

Prior to World War II this illogical scheme for dealing with the information phase of research worked quite well. The total amount of research and development was less by some orders of magnitude than it is today; the rate of expansion was relatively low; and existing dissemination media were for the most part able to keep up with the output of new scientific knowledge. Then came the deluge—of federal funds to support R&D, of the R&D these funds spawned, and of the information this R&D generated. Most floods abate after a while; this threefold one hasn't. On the contrary these "waters," far from receding, have continued to rise. Derek Price (*1*) has pointed out that most nonscientific aspects of our culture double in magnitude every 30 to 50 years but that science in this country is doubling every 10 years. Whereas, just prior to World War II, U.S. expenditures for pure and applied science totaled less than \$300 million (*2*), in fiscal year 1961 the amount was some \$14 billion. It is estimated that the nation's R&D bill in fiscal 1963 will be in the neighborhood of \$18 billion, of which roughly two-thirds will come directly or indirectly from federal funds. The output of scientific information has increased accordingly.

Two factors in addition to growth in quantity have served to magnify and complicate the problem of maintaining the results of research readily available, and therefore to increase the cost still more. One stems from the fact that this enormous expansion in scientific research and development has been world-wide. Only a few years ago the bulk of significant scientific publication appeared in just a few languages which most scientists could read with at least some facility. Today, it has been estimated, perhaps a third of the world's scientific literature can be read by fewer than 5 percent of U.S. scientists. In chemistry, for example, almost 30 percent of the material covered by *Chemical Abstracts* appears in Russian, Japanese, Italian, Polish, and Chinese —five languages read by fewer than 2 percent of our scientists (*3*).

The other complicating factor is the progressive breakdown of the boundaries between the traditional subject fields. No longer can the average research scientist limit his professional interest to one or two of the conventionally defined disciplines; increasingly he is finding that he must be knowledgeable in areas he formerly considered foreign to his field of specialization. Under the conditions brought about by these three factors—rapid growth in research and development, increase in the significance of foreign scientific literature, and overlap of traditional disciplines—the system's inherent defect, which I mentioned earlier, very quickly became of much more than academic significance.

I hasten to add at this point that it would be both incorrect and unfair to say, or even to imply, that those responsible for planning and administering research and development have given no thought to the growing seriousness of the scientific information problem. The picture I have sketched, like all generalized descriptions of complex situations, suffers from oversimplification. Industrial subscriptions to journals at rates higher than those for individuals, organizational memberships in scientific societies, direct industrial and federal subsidies under certain conditions, and page-charge levies all have served to provide journals and abstracting and indexing services with substantial additional income.

Support from Research Funds

However, except for page charges, which I shall discuss later, none of these mechanisms for information support is tied appreciably more closely to the funding of experimentation than are the society dues and subscriptions of individuals, which I mentioned earlier. But if dissemination of the results of experimentation really is an integral phase of the research process—if, say, $95 or $96 worth of experimentation plus $4 or $5 spent to make the results available actually is preferable to $100 worth of experimentation that no one ever hears about—then, support for information control and dissemination should vary more or less directly with total R&D funds. One assumes as a matter of course that multiplying an R&D budget by a factor of *X* will, on the average, provide about *X* times as much money for employing personnel and purchasing experimental equipment; it should be equally natural and valid to assume that such an increase in R&D funds would multiply roughly by *X* the financial support available for activities related to disseminating the results of the expanded research program.

Problems of Implementation

So much for my thesis. Now, I should like to comment briefly on certain of the problems associated with implementing it. Two general kinds of problem must be faced whenever one attempts to ticket some fraction of a budget for a particular purpose—first, determination of the exact amount or fraction to be so labeled, and second, selection of the mechanisms to be used in spending it. In both of these respects the ambitious automobile manufacturer I mentioned earlier is in a much better position than is the R&D planner or administrator. The motor car magnate can calculate fairly precisely the additional raw materials and the augmented sales and distribution facilities he will need, and the procedures for satisfying these needs are well standardized. Not so for an expanded R&D effort.

First comes the question of deciding what fraction or amount of R&D funds should be marked for information control and dissemination. Obviously, no one can predict either the quantity or the significance of the new knowledge that will come out of any given project or program. Consequently, the costs of its initial dissemination, by publication or other means, and of abstracting and indexing coverage also are unpredictable. However, to make an order-of-magnitude estimate of the average, overall cost of the information phase of research in relation to total R&D funds is possible.

A recent National Science Foundation report (*4*) shows that for the 3-year period fiscal 1960 through fiscal 1962, federal obligations clearly identifiable as being for scientific and technical information average a little less than 1 percent of the government's total R&D expenditures. (The data show, incidentally, that less than half of these monies identifiable as for information actually comes from R&D funds.) Information-associated activities too closely interrelated with other aspects of research and development to be readily separable probably cost at least another 1 percent, making total federal expenditure for scientific information of the order of 2 percent, plus, of the R&D budget.

A survey made several years ago of a group of companies with R&D programs showed their identifiable expenditures for scientific information activities varying from less than 1 percent to 10 percent of their research budgets, the median being around 2 percent (*5*). Allowing for the present inadequacies of both public and private scientific information systems, one might estimate 4 to 5 percent as a minimum order-of-magnitude portion of R&D funds that could justifiably and effectively be devoted to the control and dissemination of the results of research.

However, the important point is not whether 4, 5, 8, or some other percentage is optimal for the information phase of research and development; it is, rather, that planners and administrators of research and development fully accept the principle involved. The use of appropriated, contracted, and granted R&D funds to support information activities associated with scientific experimentation generally has been permitted. The very fact that such funding has been only permissive, however, has had two bad effects: (i) intellectually, it has helped maintain the in-law or distant-cousin relationship between information dissemination and experimentation, and (ii), practically, it often has meant that a scientist preferred to spend all of his R&D funds in other ways, with dissemination of results falling into an afterthought category. Complete, realistic acceptance of the thesis discussed in the first part of this article would mean that some portion of research funds always would be devoted to dissemination of the results of the experimentation.

The second general problem mentioned earlier—that of the mechanisms for channeling an appropriate fraction of R&D funds into information control and dissemination—has both intra-agency and extra-agency aspects. The former concern an organization's own library and searching services, its internal reporting system, and the like. Although establishing these activities on an effective, well-balanced basis may pose a host of difficult internal problems, the *mechanics* of supporting the activities from a portion of R&D funds labeled for the information phase of research presumably would present no serious difficulties.

It is in the dissemination of research information to the scientific community as a whole that the problems of implementing the thesis of this article principally would arise. When some definite portion of an organization's R&D budget is designated for dissemination of the results of its research, it must face such questions as the following.

1) How should such funds be spread among primary publication, abstracting and indexing, dissemination through other media, and research aimed at developing new and improved techniques for making scientific information available?

2) With respect to primary journal

publication, which periodicals should receive support, how much should they receive, and how can this support be given within the present framework of scientific journal management?

3) How should the questions just raised for primary publications be answered for abstracting and indexing services?

4) With respect to study and research on techniques for information control and dissemination, should the organization conduct such investigations itself, should it join with other such projects, or should it ignore this area?

There are no easy answers to these and many other questions; this fact, however, is not sufficient reason for failing to seek solutions if the principle discussed in this article is sound.

Possible Approaches

The ramifications of possible approaches to solving this implementation problem are far too complex and interdependent to permit comprehensive discussion here. I should like, however, to comment very briefly on two of the possibilities. At present some two-thirds of all scientific research and development being conducted in the United States is supported directly or indirectly by the government. On first thought, therefore, an obvious solution to the information dissemination problem might appear to be the creation of a mammoth federal agency that simply would take over all scientific and technical publishing, abstracting, indexing, and associated activities. Certainly such a move would simplify the mechanism of channeling a portion of R&D funds (at least those of federal origin) into the control and dissemination of research data. Second and third thoughts, however, raise a number of questions of practicability and desirability. Let me mention just two or three.

For example, the existing U.S. scientific information system is extensive, long-established, scientifically accepted, and in large part privately sponsored and operated by scientific societies, commercial publishers, and universities. The practical difficulties in replacing this large and well-established complex with a single, federally controlled operation might be almost insurmountable. But beyond the question of practicability is that of whether a comprehensive centralizing step of this kind would be desirable. Our continued belief in the importance of the role of private enterprise in our system certainly is one strong reason for thinking it would not be. Further, such a move, in addition to possessing the conventional disadvantages characteristic of any major centralization, would be bound to introduce increasing government control and would remove responsibility for quality control and other important aspects of scientific information handling from those best equipped to provide such regulation—the scientists themselves, through their professional societies and other organizations.

A time-tested method that partially solves one phase of this "mechanism" problem involves the so-called page charge—pioneered some three decades ago by the American Institute of Physics and adopted to an increasing extent in recent years by other nonprofit publishers of research periodicals. The American Chemical Society has announced that in 1963 a page-charge policy will be instituted for eight of its primary journals; in discussing this move, the society has said (*6*): "Philosophical justification of the page charge is based upon the contention that the cost of publication is properly a cost of research. Traditionally, research costs cover the author's time and that of his secretary to prepare his report. The ACS and others have adopted the view that research is not complete until its results have been made available to others. Thus it follows that some portion of the cost of preparing the paper for public distribution is a fair charge against a research budget."

Typically, the page charge is set at an amount equal to or less than the cost of setting an article in type and preparing it for the presses—that is, the so-called fixed cost of publication. Payment of the page charge is never a condition for acceptance of a manuscript; the editorial decision to publish is made before the page charge is assessed. The charge is levied, not against the author of a paper, but against the organization supporting the research whose results the paper reports; therefore, to a degree, it accomplishes the objective of channeling a portion of research funds into dissemination of the results of the experimentation. As now used, the page-charge plan obviously achieves this goal only for initial publication in a primary journal—that is, it provides one answer to part of the first of the four questions posed above. It has been proposed that an additional charge of so much per article be made to assure adequate abstracting and indexing of the paper, and some study and experimentation along this line is now in progress. Whether page and article charges prove in the long run to be the best solution to the problem remains to be seen. If other approaches look promising, certainly they should be considered also. As already noted, however, the important point is that if the principle is valid, some effective mechanism or mechanisms for implementing it must be found.

Conclusion

In closing, I might summarize the basic argument of this paper by two mathematical expressions:

$$R \neq E \quad (1)$$
$$R = E + D \quad (2)$$

where R is research, E is experimentation, and D is dissemination of results. It may very well be that if all planners and administrators of R&D programs in the United States were asked whether they think Eq. 2 is sound, almost all of them would reply in the affirmative. But to accept Eq. 2 intellectually while continuing to operate on the basis of Eq. 1 tends to complicate rather than solve the nation's serious scientific information problem. It is the contention of this article that prerequisite to a solution of this problem is realistic recognition of information dissemination as a blood brother of experimentation in the R&D household. Acceptance and implementation of this principle will provide a logical, solid, and relatively permanent foundation on which to base the multitude of specific activities and studies that are essential to the eventual achievement of a sound, effective, overall U.S. scientific information system.

References

1. D. J. S. Price, *Science Since Babylon* (Yale Univ. Press, New Haven, Conn., 1961), p. 108.
2. V. Bush, "Science, the Endless Frontier," *National Science Foundation Publ. No. NSF 60-40* (1957), p. 87.
3. K. Heumann and P. M. Bernays, *J. Chem. Educ.* **36**, 478 (1958); "Lost—One-third of the World's Scientific Literature," *National Science Foundation Publ. No. NSF 60-20* (1960).
4. "Federal Funds for Research X," *National Science Foundation Publ. No. NSF 61-82* (1962).
5. *Ind. Eng. Chem.* **51**, 48A (Mar. 1959).
6. *Chem. Eng. News* **40**, 92 (19 Mar. 1962).