Documentation and the Individual

Individual scientists use small segments of the literature. These segments can be provided by good libraries.

Ralph R. Shaw

Creative science, now as through all time, depends upon trained individual minds, regardless of whether these minds work in isolation or in the concert hall of team research.

A new kind of numbers game has tended to obfuscate our thinking about scientific documentation. The game consists in showing how publications in field x double every y years. The purpose of the game is to scare us into accepting radical solutions to the documentation problem, for the alternative pictured is that of science dead from autointoxication, from immersion in its own effluvia. Clear thinking about documentation, however, begins in a rather different fashion. It begins by noting that creative science, now as through all time, depends upon the trained minds of real persons. The central issue in documentation then becomes whether we have the will to provide each working scientist with the intellectual tools essential for optimum fulfillment of *his* potential. Whether such provision requires this classification system or that one, or even many systems—whether it requires printed books, walking encyclopedias, or computers—is of lesser importance. It does require consideration of the total system of communication rather than of any one of its parts, and the end product of this system of communication must be the information needed by and usable to each scientist, wherever he may be and whatever his needs may be at the moment.

Let us face the fact that our intellectual system is not in balance. Creative research is a mysterious product compounded of mechanical manipulations and creative intellectual processes. We commonly spend some $20,000 to $50,000 a year or more to provide physical support for each bench scientist, yet rarely do we allocate as much as 1 percent of that amount each year to provide him with intellectual support. What is even worse, we have not, apparently, even begun to think deeply about what is truly needed for providing optimum intellectual support. It is easy to prove that a scientist who is working on plants and animals that are too small to yield their mysteries to the normal human eye must have a microscope. And if his materials are still smaller he must have an electron microscope. Inadequate mechanical equipment means no program; it is as easy and as absolute as that.

Proof of the need for intellectual support is a little more difficult. The evidence here is an infinitely variable array of shades of gray. Creative minds can postulate the existence of microbes without having read of them and without being able to see them, while, given a pattern of iron filings to look at, noncreative minds, or creative minds not triggered by that particular stimulus, will not create a theory of magnetic fields. The evidence that we have for the importance of intellectual preconditioning and intellectual support really consists of little more than testimonials.

While we do not know very much about continuing education of adult minds in science as a factor in creativity, or about the value of particular methods in this regard, we consider intellectual preparation to be absolutely indispensable for research. We subject scientists to fairly rigorous intellectual exercises before turning them loose in the laboratory, and it is reasonable to assume that this academic discipline is essential for modern scientific achievement. Since the basic theories and data that science students absorb in college are subject to constant and rapid change, it seems reasonable to assume in addition that, if the original intellectual training is necessary, constant updating and broadening of the intellectual training of each scientist is also indispensable. And we do indeed assume that. Otherwise, why do we publish journals, hold meetings, or even use the telephone or carry on correspondence to keep up with what is happening in science? It can be argued that the nature and degree of this need vary with the field and the particular task to be performed. The nature of the intellectual equipment and of the stimuli that are required may well vary from man to man and from task to task, but we all give lip service, at least, to the concept of intellectual support as a continuing need for all research workers.

The paths that we follow from this point diverge widely. The most common path is one that leads, from considerations of the mass of scientific material, to the argument for machine classification. However, the size of the literature is not the only problem facing the scholar; this is not a new problem, and it is probably not the central problem at all. To be sure, no scholar can read everything that might conceivably be of some value to some facet of his total personality—but then, he never has been able to read everything, and he probably never will be able to do so. He cannot listen to the conversations of all his peers all over the world, but he never has been, and never will be, able to do so; he cannot attend all of the simultaneous sessions at every conference on all of the subjects in which he is or might be interested, but again, he never has been, and never will be, able to do so.

Notwithstanding the fact that no one, in recent centuries at least, has been able to read everything, science has made what some of us consider substantial progress and it should continue to make substantial progress. This would appear to indicate that, despite all the obvious shortcomings of the communication system, scientists have, by one means or another, sometimes wastefully and sometimes not, succeeded in achieving sufficient mastery of the

The author is a member of the staff of the Graduate School of Library Service, Rutgers University, New Brunswick, N.J.

intellectual resources required so that science has *not* come to a standstill. The issue, then, is not whether science is dying of an overabundance of information but whether we can improve the flow of information to the individual scientist so as to make him more effective than he would otherwise be.

Two Types of Use

Before I go any further with this argument it appears essential to point out that there seem to be at least two fundamentally different levels of organization of literature, and two fundamentally different types of use of literature, even though neither of these groupings is mutually exclusive. We documentalists tend to concentrate on the use of systems and machines for retrospective searching of the literature, but this is only one of the two areas of use. The other is the type of use represented by the input to the scientist's mind from his casual day-to-day contacts with the literature of science and with that of other fields which may or may not appear to other minds to be related to his basic professional pursuits.

It would therefore be a mistake to concentrate only on finding ways to provide material that is asked for, or on solving the problems of retrospective searching of subject literature. A very large percentage of the scientific material read by scientists is outside their own fields, and we do not know enough about the processes of research creativity to determine the relative importance of purposeful reading and chance reading. We do not know much about the process of creativity that is triggered in the trained mind by what, to others, appears to be nonpertinent reading. Such studies of reading habits as have been made indicate that we must make more than a cursory obeisance in the general direction of Walpole's Princes of Serendip. A substantial part of communication through recorded information stems from a casual rather than a purposeful search for facts. A substantial portion of a specialist's reading is of material outside his specialty. And there is considerable evidence that creative thought is generated, or at least stimulated, in the trained mind by apparently nonpertinent written materials. Until we know a great deal more than we do now about the nature of the intellectual processes that result in creation of new knowledge, it will be difficult for us to program a computer to provide this stimulus. While we may eventually learn enough about the "genius of creativity" to routinize it and then to mechanize it, at present we know so little of the nature of these processes that we shall have to rely, as we have relied in the past, on encouraged and supported chance encounters between trained minds and a wealth of stimuli in the still relatively unorganized mass of information to be found in Gutenburgian form.

This means that we have to make available general scientific and technical literature, as well as specialized materials, to the extent that the research staff wants them and will use them.

Even in retrospective searches there appears to be a difference between what is needed in the basically static file of a research library or central file room and what is needed in the individual file for a group or individual working on a particular research project or for an individual working in a limited number of fields. The techniques used for standby service in great storage warehouses may, or more probably may not, be the same as those that are most suitable for small files for individual research projects in narrowly specialized fields. The average special file contains from 5000 to 25,000 pieces, and very few of the files of documents reported in the National Science Foundation's list contain as many as 50,000 pieces.

We need great research collections, and it is exceedingly doubtful that any of the machine techniques currently suggested can be applied with reasonable operational or program efficiency to files of millions of books and documents. It appears that here too we need to handle the material in two or more steps of varying lengths, the first step being the basic one of organizing the raw materials so that we can get at them, in a second step, for further analysis for special needs.

Classes of Services

So far as the individual scientist is concerned, it appears that he must have access to at least five classes of services.

1) He needs to have access to as broad a file for browsing as he feels he can use; the size of this file will vary from scientist to scientist, from field to field, and from time to time.

2) He needs to be able to browse in bibliographical and abstracting journals which may give him leads that he may or may not be looking for.

3) He needs "current-awareness" service to the extent that his field calls for it and that he can use it.

4) He needs retrospective approaches to literature, through annual reviews, or reviews of the state of the art, or special bibliographic searches.

5) Above all, he needs access to material that he wants when he wants it, regardless of where it is housed, and he needs some screening service, so that, in searches of older or of current material, at least the obviously redundant and nonpertinent can be deleted.

Discussions in the field of scientific documentation are primarily concerned with the mechanisms and tools by which these general objectives can be attained. The proposed mechanisms should be subjected to rigorous management studies under each of varying sets of conditions, so that the most suitable tools can be devised.

Classification Schemes

The literature of documentation, to date, has been heavily weighted with arguments favoring one or another particular detail of a system of operation. Objective descriptions of operations, giving full operational data for the total cycle of storage and retrieval, are almost nonexistent.

We have many articles describing classification schemes out of the context in which they need to be applied, and without regard for the obvious fact that classification, or indexing or coding, is just one step in the input (or output) routine, and that this, in turn, is simply one small step in the total cycle of storing and retrieving information. It has yet to be established that any given classification or coding scheme can provide anything that cannot be provided by other schemes, or that any mechanism, including the book, provides anything that cannot be provided by other mechanisms. We need to determine which of the tools available is best for the particular task to be performed under the conditions under which it has to be performed, and here science would be served better if we would apply the scientific method to finding the answers.

Problems such as these cannot be solved by specious arguments about classification schemes used for other purposes and not designed for indexing in depth. One of the almost classic arguments in the documentation literature starts by pointing out that the Dewey Decimal System may be all right for small public libraries but breaks down completely when it is applied to indexing scientific literature in depth, and that we must therefore go to . . . (whatever it is that is being proposed). Now, nobody who has the least familiarity with bibliographical method has proposed that we use Dewey for this purpose, and other tools have commonly been used. For example, the Index to the Literature of American Economic Entomology, prepared by the Department of Agriculture Library in cooperation with the American Association of Economic Entomologists, has for more than a generation provided "concept bibliography," in which each insect, insecticide, parasite and host, and so on, in the entomological literature is indexed, consistently and usably and inexpensively. Far wider general dissemination has been achieved with this index than with any of the so-called newer devices. This example is but one of many systems which have resulted from the cooperation of subject specialists and bibliographers (or subject specialists acting as bibliographers and *working* at it).

Furthermore, no system is any better than it proves to be in operation. By way of example, let us consider a study I made of a very large information center, which shall be nameless. It had a sophisticated code book and trained subject specialists to apply it. It had trained librarians on the output side and a very complete and well-managed punched-card shop, with much interesting associated hardware. To test the system, we submitted clean copies of a large number of documents to the analysts who had coded them several months earlier, and in no case did we get the same coding for the document as that given earlier. We then resubmitted inquiries to the reference librarians who had handled these same inquiries several months earlier and asked them to supply the codes for machine search. In no case did we get the same coding for machine search for a given subject as had been given earlier. And then, as a final test in this series, we sent a sizable sample of inquiries to the punched-card shop for searching. We selected searches that had been made in the punched-card files 3 to 6 months before, and we sent the tab shop the codes for these subjects that had been submitted for the previous search. In no case did we get identical literature citations. It did not take much investigating to determine the causes of these wide discrepancies, and corrective measures were applied. But this experience, which is not uncommon, indicates reasonably clearly that not all of our problems are problems of lack of adequate theory or tools.

In concluding this brief appeal for sanity and hard work, may I submit that we need to learn more about the parameters that apply to each species of tools and to each tool under varying conditions, and that we need to consider applications of these tools as integral parts of systems for providing information services to scientists rather than as ends in themselves. This means that we need to work on basic theory in our field, and it also means that someone has to tend the store while we are doing this, unless we expect science to do without information services while we are designing theoretically optimum systems. This means that, important as the improvement of techniques is, we ought to take the plumbing out of the front office and put it back into the workrooms and the associated documentation research laboratories.

The front office should be designed to get each research worker what he needs when he needs it, and in the form in which it is most useful to him, regardless of what we have to do behind the scenes to achieve this, and regardless of how we do it. Only insofar as we achieve this objective currently and continuously can scientific information services contribute to the advancement of science.

Arthur Russell Moore, General Physiologist

The science of physiology has had its historical focal point in medicine, but it has had, as well, a broader tradition in the more general aspects of the subject. This tradition has developed in part through the missionary efforts of a few conspicuous figures, but it owes much to the balance contributed by devoted, if less famous, scholars whose interests centered in the most fundamental aspects of vital phenomena. Arthur Russell Moore was one of these scholars, and one of a dwindling group inducted into physiology by Jacques Loeb.

Moore was born in Beaver, Furnas County, Nebraska, on 10 November 1882 and received the B.A. degree from the University of Nebraska in 1904. After 3 years as a school teacher, he went as a graduate student to the Spreckels laboratory of physiology at the University of California, Berkeley, where he received a Ph.D. in 1911.

Those were great days in Berkeley, Jacques Loeb had been brought out from Chicago to establish a physiology department, and his mission was to clear the last lingering mists of *Naturphilosophie* from biology. Fresh from his triumphant demonstration of artificial parthenogenesis, Loeb was going to solve the problems of cellular biology by studying the properties of proteins, and the problems of behavior by studying reflexes and tropisms. Moore's colleagues and teachers included S. S. Maxwell, J. B. MacCallum, T. Brailsford Robertson, and C. L. A. Schmidt, and his own first scientific paper was a treatise on the biochemical concept of dominance, an early essay in biochemical genetics. Throughout his long and productive career, represented in more than 100 scholarly publications, Moore retained a primary interest in