THE ART OF INFORMATION MANAGEMENT

Frank C. Shephard, Editor
Oceanographic Literature Review

The amount of primary information published or otherwise available can be expected to continue to grow at rates which pose formidable challenges to its management. Many new technologies either currently or soon to be available can facilitate this task, but will require imaginative exploration of their utility. The main problems of information management are no longer technical per se, but organizational, conceptual, even cognitive and linguistic. The traditional conceptual bases of literature organization - cataloging, indexing, classifying and so on - may well yield to more flexible, 'customizable' and multifaceted approaches. Likewise, traditional services and products may yield, or at least give ground, to innovative services and products yet to be thought of, but made possible by the new technologies. Some of Eugene Garfield's products are illustrative of the possibilities. Many of the new applications are likely to emanate from librarians and other information managers, rather than from computer scientists and engineers. It is the former groups who are frustrated on a daily basis in trying to serve their users efficiently. In any event, information science has truly come of age, with exciting multidisciplinary developments on both theoretical and applied fronts. It is a profession with a bright and demanding future, bright because of the technical possibilities; demanding because scholarship, good judgment and experience will always be required to best serve the user.

INTRODUCTION

I appreciate the invitation to speak before this group. Most of us, including me, are so bound up in the day-to-day detail of our jobs that we rarely enjoy the luxury of reflecting on the underpinnings and implications
of what we do. But having to speak converts such reflection from luxury to obligation, and that is why I am grateful - this half-hour talk has forced some 50 or 60 hours of reading on me which should have been done long ago.

I'd like to start where Paul Gross left off yesterday. He issued a challenge to the information community: the scientists, he said, were preoccupied with their own work; even though aware that the information problem is acute, they would not solve it themselves. Could we solve it, Gross wanted to know.

The answer is yes. From initial obscurity, our roles have been pushed into the limelight by historical and contemporary forces. The mechanisms of information management and dissemination are now bottlenecks in society's absorption of knowledge. But we, the information managers, are the only ones looking at the whole picture and thus are ideally positioned to solve these problems, which mainly have to do with the rate at which information can flow socially and the means of targeting it accurately. Ten or fifteen years ago we were enthralled by recent developments in on-line capabilities largely having to do with the advent of random access storage media. Today it is possible that the very people at this meeting have access to more database terminals than existed nationwide when IAMSLIC was first established. But on-line databases, for all their utility, were simply not the panacea they were said to be. Once the technological problems were solved (and, to the extent they have not been, soon will be) a new set of difficult problems emerged which transcend technology, but which we must now address. The general dilemma may be articulated this way:

The modern age has a false sense of superiority because of the great mass
of data at its disposal, but the valid criterion of distinction is rather the extent to which man knows how to form and master the material at his command.

While this could have been written yesterday or tomorrow it is an 1810 quote from Johaan von Goethe who as I'm sure you know was no slouch in the information business and a scientist as well as a man of letters. So, with that as introduction, we can talk a little bit about information, its management, and ourselves, its managers.

INFORMATION

First, what is information? Let's start with an important philosophical definition: information is the substance which makes society an organism; otherwise we are just so many billion isolated cells with all the limitations which that implies. It is also - of all the resources at our disposal - the one with the most leverage. That is why they say that information is power. So if you feel your work is unappreciated remind yourself that what you handle is more precious than diamonds.

There are, of course, other definitions. The Bureau of Labor Statistics defines information in such a way that 1 of every 2 people in the work force is in its industry. This appears to include entertainers, e.g., me and Dolly Parton - a distinction that I personally have no trouble making.

To us, however, as a practical matter information refers to the published record of society's doings and thinkings, particularly (for this group) in science and technology. Those who manage any aspect of it only number in a few tens of thousands, not a few tens of
millions. This small group of us is charged with massaging that published record. We identify its elements, document them, and somehow get them to the people who need them. Here for your information are a few dimensions of the task which I've come across in recent reading:
- 15,000 books/year (in science and technology alone)
- 30,000 journals and serials currently published (in science and technology alone)
* 60 million pages per year
* 400 thousand articles per year (science and technology, U.S. alone)
* 250 million readings of these per year
- 1500 abstract or index publications (in English alone)
- 4-5 million computer searches per year (expected by 1985)

INFORMATION MANAGEMENT

Remember that the main tasks of information management are to take that mass of information I just itemized, identify and document its elements, organize its subsets, and call those subsets to the attention of various subsets of users. These jobs are to my mind still more art than science (which is why I entitled my talk as I did). There will always be a strong scholarly component to information management, regardless of its high-tech tools. Yesterday's appropriate grouping of information is not tomorrow's. Yesterday's subsets of users are not tomorrow's. And yesterday's classification schedule and index terms are not tomorrow's. Even so, there are some general concepts that are likely to aid the information management art. There is not time to fully discuss these concepts or their ramifications, and
I wouldn't expect everyone to agree with them, at least as they are stated. Nevertheless, they may stimulate your thinking as they have mine.

(1) Information is best considered a natural resource, perhaps the resource which humans make available to the planet Earth the way plants make oxygen available. I prefer to call it a 'natural' resource in order to emphasize that it is generated everywhere, occurs in such profusion and bulk, and can be put to so many diverse uses that it will probably always defy any single universal scheme for its management. The "crisis" behind the so-called "information explosion," for example, is not new at all. It has always been a chronic problem for the individual. Even before the time of Gutenberg there was more information available than anyone could possibly absorb. Editing, reducing, and summarizing this unmanageable bulk is the job of educators and other cultural translators and interpreters. Indeed, it is a group activity in which all humans participate. And (to iterate again) the universal documentation of all the pieces does not, in itself, help much in the reduction process. Indeed, the more universal and comprehensive a subset of information becomes, the less relevant it will be for any individual user - and the more poorly organized.

(2) Information in its natural state is too diffuse to be usable. It must be refined and concentrated the way other natural resources are refined and concentrated. In the process, value is added to the pieces so collected, and the collection itself has a value which previously didn't exist at all.

(3) The main mechanisms of adding value to information can loosely be described as reduction tasks (textbooks, review articles,
and summaries are all associated with reducing what must be examined in order to eventually know something), and channelization tasks which involve aggregating, re-sorting, and re-distributing information. Selecting, collecting, cataloging, classifying, indexing and alerting activities, for example, fall largely in the area of channelization.

Most of these tasks we describe as the activity of "documentation." A book is a document, but it has associated with it a body of documentation. Representations of the document within the documentation are sometimes called surrogates. An abstract, e.g., is a surrogate for the original document, as are a set of index terms which describe the document. The purpose of the documentation, of course, is to increase awareness of the original document's existence and availability.

(4) The most promising organizational means for accomplishing these tasks are via nested (hierarchical) schemes and systems. For example, the task of documentation involves "sending" a document (via surrogates) into realms larger than itself (a discipline and a subdiscipline or two) and realms smaller than itself (its abstract, its title, its set of index terms). But note that while we all agree a discipline is "larger" than a document which it subsumes and an index term "smaller" than the document from which it came, the index term provides entree to moving up or down other nested systems (via the narrower and broader terms of the indexing metalanguage) which may either parallel or cut across the first scheme, and which exist in great numbers.

(5) The possible nested schemes which pertain to a given document are innumerable and not mutually exclusive.

Because the ways in which information can
be organized are innumerable there is no point in one institution attempting to organize it in all possible ways, any more than there is in finding one universal way. Even so, we should be able to use our computer technology to liberate us from the physical constraints of our actual documents. At the moment our documentation products are too often like horseless carriages. We employ a new technology but the products are the same old-fashioned ones. We still, e.g., catalog a book within the framework of only one scheme, and then assign it a unique number within the scheme. Why must we stick to one scheme and why must the book have only one number? Why not as many schemes as may be helpful to some group of users which we serve, and why not as many classification designations as make sense? Why, e.g., must a book on natural resources law be filed either with natural resources or with law rather than with both? The physical record - the book itself - must still be filed in only one way, but why not by date of acquisition, or alphabetically by author? Then, any number of surrogate 'logical' records of the book can exist side-by-side and simultaneously. The user may object that he can no longer browse, but was he that well served by browsing when segments of interest to him were widely separated anyway, perhaps without his knowledge? Today, he could be browsing document surrogates such as abstracts, tables-of-contents, whatever - pulled off the computer. Such surrogates need not be so blandly alphanumeric either. Facsimile devices should, and no doubt will, allow sampling the graphic and typographic characteristics of the book as well. Naturally, we're getting far ahead
of our own story - too far, in fact, into the vision of the bookless library. Long before that day comes we have to organize our information more flexibly, map it in more than one dimension, and find ways to successfully permit multifaceted, multileveled searches of it. Before we are finished, we must draw on much more than computer and software engineering. Our answers must be teased out of linguistics, psychology, sociology, general systems and operations research, and even epistemology which, like linguistics before it, is coming to have applications. In the process, we are beginning to define an information science which you're sure to realize is under-represented and poorly organized (in the context of the growing scope of the field) by either Dewey or Library of Congress, illustrating the constraints of the rigid schemes to date.

INFORMATION MANAGERS

In spite of the talk about the paperless society I don't think the local library is about to become a thing of the past. There are sound political and managerial reasons for numerous redundant and widely dispersed repositories and information facilities. To the extent that publishing goes the paperless route, we must insist on widely and redundantly distributed paperless media. I haven't time to discuss this further now, but because we know so well that information is power we ought to be the last to permit centralization of its storage or distribution facilities.

Still, the role of the library is clearly changing, even as the library has fallen on hard times. How these factors should be handled is discussed more by yourselves
than by me, but because of the general and increasing flux in the ways information is handled and distributed, and because of the new sense of what an information manager is, the time is ripe to seize some initiatives for yourselves. In point of fact, you already have been doing so: look, e.g., at OCLC. And the points I'm about to make you have no doubt heard, read, and acted on before. Nevertheless, to reiterate briefly:

1. New products and new services can lead to new income, but the initiative in developing them must come from within the information community.

2. Beyond every special library's sponsoring clientele is a distant clientele. The publishing of documentation literature (like indices and catalogs) for special collections may, in itself, be a direct source of revenue, as can republication of special collections on microfilm. But it can also increase the library's use, thus its profile, and ultimately its budget. I think, e.g., of Mary Jane Beardsley's (URI-ICMRD) collection on artisanal fisheries and on how useful that facility is to areas incapable of supporting a similar library. But this is just one of dozens of possibilities right in this room.

3. As everyone in the business knows, there is a tremendous payoff in promotion; libraries need to do more of it. Go read some books on it, and don't think of it as unseemly. If you can't budget for it overtly, do it covertly, but do it. Information centers are simply too much taken for granted and far too underutilized.

4. The time may have to come to expand the librarian's role in a truly scholarly direction - into those areas we call reduction tasks - the writing of abstracts, reviews and overviews. Such special librarians might
be called "readers" (to borrow the British term for an assistant professor); they will need at least a masters degree in a subject area and perhaps a Ph.D. But reading as we've defined it has scholarly importance equal to that of education and research in spite of the fact there is not a formal slot for it. Its importance can only increase as the increasing bulk of information demands more than its characterization - it demands a new scholarly tier (with a literary and verbal bent) whose task is to reduce it and put it in context.