Advent of Information And Inquiry Services

by GEORGE SCHUSSEL

AMONG THE COSTLIEST and most time consuming problems facing industry and government today is the storage, retrieval and communication of knowledge. Human knowledge is expanding at such a tremendous rate that the documentation supporting this advance has concurrently mushroomed in a multitude of directions.

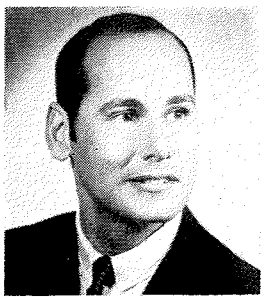
Scientific information output alone is doubling every eight and a half years and is expected to reach 120,000,000 pages annually before 1970. The task of storing and retrieving information so it can be easily and readily used to assist researchers, scientists, engineers and managers has become a significant problem.

To give a further understanding of the volume involved, estimates have been made that in the over 100,000 technical and scientific journals currently printed in 60 languages, the new information available each 24 hours would fill seven sets of a 24-volume encyclopedia. Several Government studies have shown that the lack of needed information for engineers and scientists at the right time in the right way represents annual losses in the billions of dollars and results in tremendous delays to the average development cycle of new systems.

Various estimates have shown that the average professional person spends from 25% to 75% of his time attempting to keep abreast of major developments in his field. The majority of this time is exhausted trying to locate the information in his interest area. Similarly, a vast amount of the Nation's research and development budget is spent on duplication of work that has been completed, but the results of which are not readily accessible. An inquiry by Hirsh, Milwitt, and Oakes into the activities of members of aeronautical and electronic societies in Southern California indicated that anywhere between 30% and 85% of scientific manhours were wasted because of the fact that previously generated information was unavailable to the right person at the right time.

Estimates have been made that the annual cost to the United States economy of the duplication of effort required to remedy this information lack runs between $200 million and $1 billion per year. In England, Sir Frank Francis, addressing the Parliamentary and Scientific Committee of the British Government, quoted an estimate prepared by the Scientific Attaché in Moscow that, "a working scientist spends up to one-third of his time searching for information and that the cost of this search represents one-fifth of all the money allocated to science."

To make this vast amount of knowledge available and readily accessible, it must be so organized and identified that the user will know exactly where and how to search for his answer. A new, multi-million dollar industry is being developed to cope with this explosion of information. This industry's primary goal is the establishment of information storage and retrieval systems. Such systems can be either manual or automated on both a large or small scale, depending upon the size and complexity of the area of interest.



GEORGE SCHUSSEL *is vice president of financial and banking systems for Informatics, Inc. He was previously general manager of the Information Systems Division at Brown Engineering, Huntsville, Ala., and also served as associate professor at the University of Alabama, teaching courses in management information systems. At Northrop Corporation in California, he recruited and directed the work of the corporate management science staff. He was also an assistant professor at the University of Southern California teaching courses in computers and simulation methods. Dr. Schussel holds a BA from U.C.L.A. an MA and a PhD from Harvard.*

Of Interest to Computer Specialists

The problem of information storage and retrieval is of interest to computer specialists because such machines are very valuable in the storage, manipulation, and retrieval of large volumes of information. For example, the speed of third generation computers allows not only for wide-ranging se-

quential or random searches, but permits the use of unstructured files that are very useful for storing certain types of data. Since the unstructured file does not require a given piece of data to be stored in a particular location, it is not necessary to predefine and pre-structure such a file. As a result, one has the ability to extract and integrate data from unique or nonstandard format documentation, in spite of the fact that most of the data entered into computer storage is repetitive format information from tabulations or standard forms.

Using an Unstructured Data Base

By using such an unstructured data base, many various types of data can be placed in one system. In most second generation systems, the economics of such an unstructured file were very poor and, therefore, it was not practical. With the advent of third generation equipment, the additional capabilities, especially in terms of access speeds to large storage, have made it possible to identify each data element (field) within each record individually. Instead of defining a field into which the same data element will always be placed, it has become possible to identify or to tag each data element as it is placed into the file. As a result, the structuring of the file is done during the input operation rather than during the programming or initial designing of a system. Therefore, it is not necessary to anticipate all possible needs for different fields.

Substantial economies are also introduced by the capability of third generation computer equipment to embody time-sharing and multi-programming concepts. The full power of rapid and remote access to randomly stored data and large storage files can be made available to users without requiring the dedication of a system. Modern techniques in this field also require the use of advanced applications in microphotography and magnetic recording.

The basic theory of an information storage and retrieval system is relatively simple. One has to make provision for abstracting, storing, and inserting information into the system and for quering and retrieving specific portions when necessary. Input from such sources as reports, drawings, and documents can be abstracted and indexed according to a pre-arranged categorizing scheme. The source documents can then be stored in some sort of graphic system which can be based on a photographic or magnetic process, or the hard-copy originals can be filed for ready reference.

Each piece of information as it is abstracted and indexed is usually assigned an address (i.e., a method of locating that peice of information). A set of key words which cross reference the appropriate graphic image address is also created. This key word indexing is usually stored in the memory of a computer. Retrieving the pertinent data from the computer requires only that the appropriate key words related to the desired information be identified. By matching these key words with the information stored in its memory according to various combinations of Boolean search schemes, the computer makes the information available to the requester, as well as the address of the stored graphic images. Duplicate copies of the graphic images should be made available on demand.

The growth in size and numbers of information centers (dealing in specialized information, as opposed to a general library) over the last few years has been spectacular. There are currently over 1100 such information activities in the United States.*

Library Services, Data Centers & Analysis Centers

Most information services can be classified into library services, data centers, or analysis centers. From the user's point of view, the library service requires the user to conduct further research into a document in order to obtain an answer for his specific question. Library services deal with para-

*There are several ways of locating information centers specializing in various types of data. One of the most general ways is to inquire to the National Referral Center for Science and Technology of the Library of Congress (telephone 202-967-8265). An inquiry will be answered with a list of potential sources of the type of information requested. A publication, **A Directory of Information Resources in the United States,** which lists hundreds of existing data centers, is available in the Library of Congress. Information centers especially relevant to United States Government programs are listed in a book entitled **The Information Systems Handbook** which can be acquired from Brown Engineering Company in Huntsville, Alabama.

Advent of Information & Inquiry Services continued

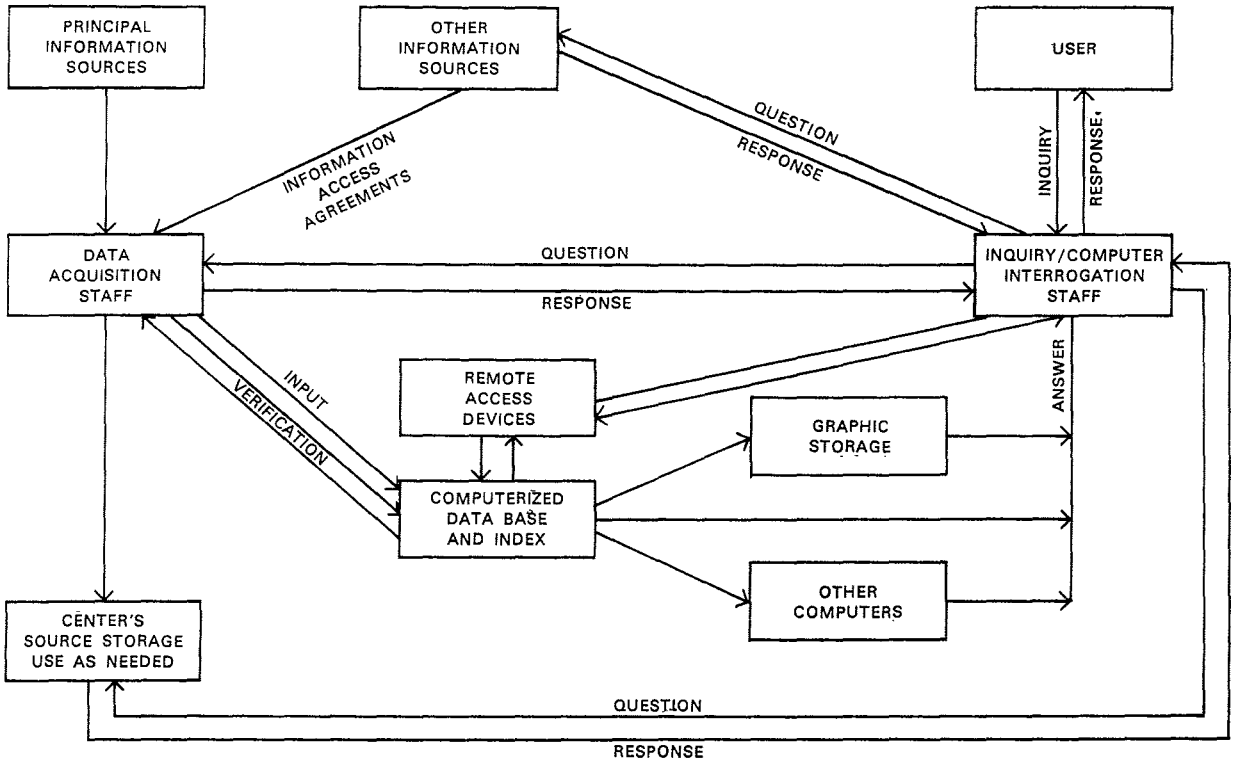
meter searches, abstract indices, and bibliographies and are concerned with documents which are stored, abstracted, listed, searched, et cetera. The output of such a service is usually in the form of a document, an abstract of a document, or a reference to a document.

Data centers, on the other hand, are more concerned with detailed data rather than documents and typically are oriented to give specific answers to specific questions. The data could be information on configuration mangement, research and development projects, financial controls, or marketing information. Special characteristics or properties of materials, chemicals, devices, processes, et cetera, which are useful to scientists and engineers are very amenable to storage by such systems. The output from this type of center is, therefore, detailed data that can be applied to a specific problem.

The analysis center acquires specific data which is analyzed to create new information. While the output of this type of center is also usually detailed data for application to a specific problem, this data is the result of an analysis function performed on input data rather than simple retrieval of stored information. The inquiry service which provides an interface between the information center and the user and is discussed below can be applied to any one of these three types of information centers.

Information Inquiry Service

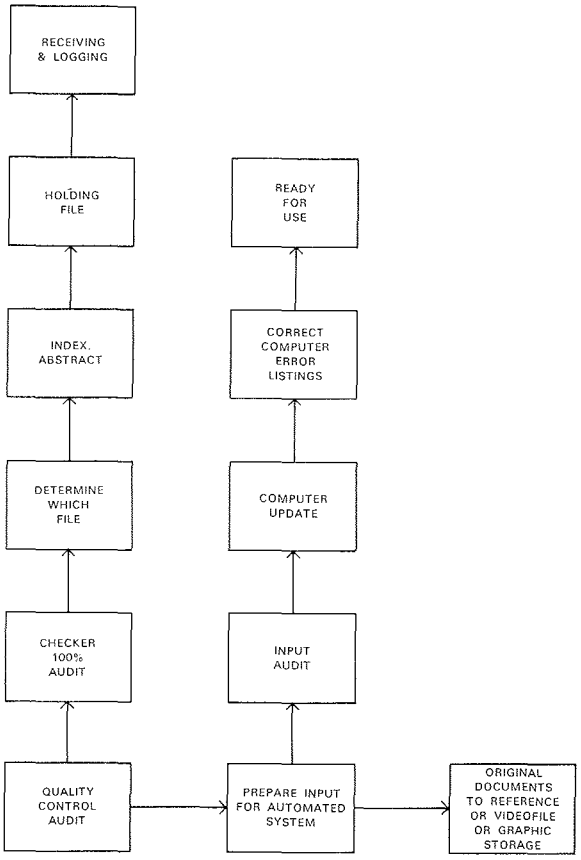
The use of large information storage and retrieval systems has developed into a science that can be considerably more complex than using a normal library. As a result, the average user needs help in locating information and the information in-



TYPICAL OPERATING CONCEPT OF AN INQUIRY SERVICE

quiry service has emerged. A center offering such an inquiry service employs technically trained information specialists to input, store, and retrieve information in a given area of knowledge. Users of the inquiry service typically will have need for specific knowledge and can communicate with the staff of the information center. Instead of simply directing the user to various reference documents, the staff undertakes to specifically answer the user's questions.

The inquiry service emerged from an effort to minimize the user's effort and required knowledge to use the system. The user-information center interface is simplified by employing the inquiry specialists to perform an analysis of the user's question. The specialists decide what data in the center is needed and then perform an analysis of the retrieval data to obtain the appropriate answer. A special characteristic of the better inquiry services is that their staff personnel will undertake to research an answer even when it requires going outside of the information files located within that particular center. Under such conditions, of course, it is necessary to have the service staffed with personnel capable of understanding the technical aspects of both the question and answer.



TYPICAL DATA INPUT OPERATION

Data Output Operation

One of the primary functions in running an information inquiry center is data acquisition and input. Information to be placed in such a center is obtained in numerous ways. Raw data, documents, vendor catalogs, magnetic tape, and punched cards are but a few of the formats in which data might arrive at such a center. Usually, primary emphasis is placed upon acquiring current data, although certain historical data may also be acquired.

Input control must be exercised to prevent the entry of questionable information. This is a serious problem since, if the data stored and later supplied by the center is questionable, its usefulness will be greatly diminished. Careful operation can often prevent the entry of suspect information. Such items as internal memorandums should automatically be excluded. Provision must be made for incorporating additional information on a particular report to supersede the preliminary data. Any preliminary data entered into the center must be labeled as such to provide any user with a clear indication of the validity of such data. Detection of other invalid information has to be basically controlled by the user of professional data acquisition personnel and the judgments of the data center's management.

In a different way, data acquisition may be a very difficult problem. Although selection of the specific input data for any specialized center may be relatively simple, practical tasks of physically acquiring it may prove to be quite a challenge. Persons having useful data might be reluctant to release it and such opposition must be overcome either by persuasion or directive. On Government programs, the acquisition function is more effective if all persons

having applicable data are directed to provide it.

Figure 1 depicts the flow of work through a typical data input process. The first step is a receiving and logging function. Here, a record is typically made of the document source, title, data and number. An accession file number is assigned for control of the document throughout its processing cycle and subsequent retrieval. After receiving and logging, a decision must be made as to whether to abstract the document. If it is decided that the document should be entered into the system, and therefore abstracted, then the preparation of the document for correct indexing follows next.

Indexing

One of the most important functions in the data input is the indexing of the material received. Proper indexing, of course, is tied not only to data input, but permits effective retrieval of the stored data. The most commonly used index schemes are key-word oriented. Important data should be indexed in detail while less important information receives less detailed indexing. There is no one indexing system identified as best for all applications.

Typical of indexing systems would be key-word in context indexing (KWIC), citation indexing, alphabetical subject indexing, bibliographical indexing, et cetera. A key-word is a term selected to describe the subject content of the input material and is usually drawn from the specific text or title. While matching or combining key words to locate stored data is generally not practical or efficient in manual search systems, a computer is ideally suited to this matching process. The use of key-word indexing allows the computer the flexibility of handling inquiries that can not be predicted during the preparation of the data base.

Most indexing schemes use permutations of the title, headings, text, and composition of indices. After the permuted terms have been selected from the subject matter, they are alphabetically aligned and accompanied where relevant by related terms, references, and notes to form the thesaurus of a system. The goal of all indexing techniques is essentially the same. That is to identify data in such a way that a particular bit of information can be found quickly, infallibly, and with a minimum of effort. In practice, the approach to this ideal varies, depending on the relative emphasis placed on the quantity and nature of data to be indexed. Any indexing scheme should at least have the following qualities:

- be as complete as possible, that is; cover the entire subject of the field
- be systematic, proceeding from the general to the specific
- be sufficiently detailed to represent all degrees of generality
- allow for the consideration of ideas and for classifying them from several points of view
- be logical and show a sequence of ideas
- be expansive and flexible in plan and notation.

One often hears the suggestion that the problem of communication between the originator, indexer, and inquirer of a specific document could be alleviated if the originator would only provide a thesaurus with his document since he has complete knowledge of all of the ideas and information he wishes to convey through the publication. This approach, however, has limitations because the author may not be aware of all the uses of his document. A more practical technique has been to provide an information center with an indexing staff thoroughly familiar with that particular field and related areas of interest. Such individuals are able to assign appropriate descriptors which often allows for a broader application of work reported in scientific papers.

As more effective indexing schemes replace current classification techniques, the amount of irrelevant material retrieved in response to a search request will be decreased. One of the general and most basic problems in information storage and retrieval and inquiry service responses is when the user is presented with more information on a particular request than is needed. This condition is acceptable only if the user desires to browse, and is essentially the situation prevailing in a library system. In cases where all of the material retrieved is only partially relevant to the search request, the inquirer is still left with the task of searching through a relatively large amount of material before locating the particular bit of information he originally sought.

Data Storage

Storage files can be either structured (formated) or unstructured (unformated). If possible, it is desirable not to require any given piece of data to be stored in a particular location. This allows the possibility of freedom in determining what data should be entered into the file and how it will be used. Data is stored often as combinations of data names and related values. The data name can be a key word that defines the content while the data item is the actual data value or reference associated with that particular data name. A unique accession number is associated with each data name/data item combination. These combinations are then entered into a file as they become available with no regard to order. Addition of new bits of data can be handled easily and effectively.

A certain amount of structuring into buckets and hierarchical storage is possible. For example, if a computer disc drive with several packs is used to store information and there happens to be a logical breakdown of the file into the number of disc drives, then it is possible to reserve each disc unit for a given classification of data.

Search statements are typically defined by using AND, OR and BUT NOT Boolean statements. These connectors are usually applied against the data names while the data items are screened by magnitude operators after the appropriate data name has been located. Magnitude operators could include such terms as EQUAL, GREATER, LESS, ALL, or any given RANGE.

The updating and purging of the data entered into such a system is very important if the data center is to be maintained currently without growing unnecessarily large. Updating, of course, can be performed on a batch or on-line basis, depending upon the system configuration. Purging, on the other hand, is typically done periodically on something analogous to a once-a-year basis. In some centrals, it is a good idea not to destroy updated and purged data, but to place it on magnetic tapes.

Videofile System

In addition to various microfilm, microfiche and aperture card systems, an exciting unique development in the graphical storage of large volumes of information has been the development of the Videofile by Ampex Corporation. Using video tape techniques, large volumes of graphical images can be stored on magnetic tape and subsequently very quickly accessed and reproduced when necessary. The tape drives that are associated with this system are similar to computer tape drives and each reel of two-inch wide magnetic tape loaded on one of these units is capable to storing one quarter-million optical images. If the correct tapes are mounted on the tape drives, almost random access is available to one million images within a maximum of 2-1/2 minutes.

Since much of the information that can be stored in an information inquiry center is more efficiently stored in the form of graphical images rather than abstracts, the use of advanced equipment such as Videofiles and the developments to come in facsimile transmission equipment promise to lend substantially greater capabilities to information centers.

Inquiry Service Operation

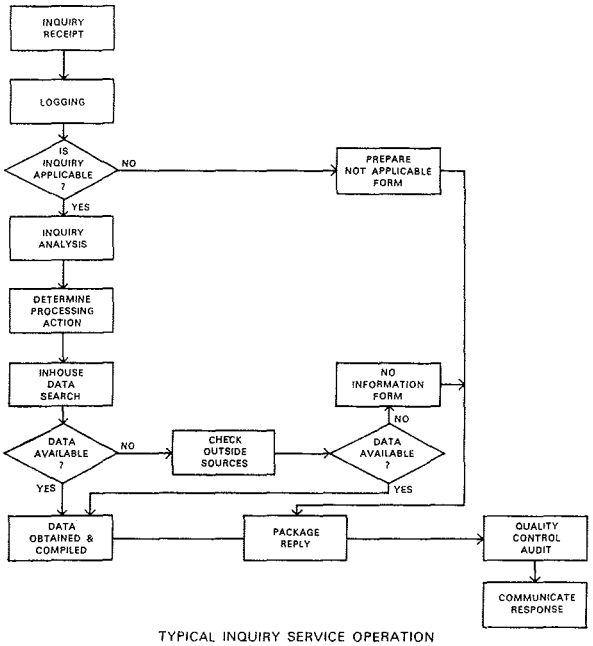
To a user, an inquiry service is a valuable addition to an information center. The mission of an inquiry service is to answer the questions submitted by all users. Turn-around time for many centers can range from several hours, to two or three days. Any information center can be most useful by not requiring a specific format for inquiries. Accordingly, the inquiry service must pattern each response to meet the needs of the user. One of the functions of the inquiry service would be to translate any questions into a retrieval language and develop search strategies. One of the most difficult problems facing an inquiry service is to determine exactly what the user needs. Inquiries are often ambiguous or misleading and a dialogue between the user and the inquiry service is often necessary to define the user need explicitly.

Using time-shared computers, an up-to-date inquiry service should have direct on-line random access to a data base. As soon as a search strategy for an inquiry has been formulated, the data base can be accessed through the remote terminal located in the inquiry service. By not requiring inquiries to be batched before making a search of the data base, the turn-around time to answer a user's question

Advent of Information & Inquiry Services continued

can be shortened. By not requiring a specific format for inquiries, the inquiry service can pattern each response to meet the needs of the user. Once data is retrieved by selecting the appropriate key words from a thesaurus and forming a search statement, a subjective analysis of the resulting data must be made in order to ascertain its applicability to the question. At the current state-of-the-art, there is no technique for automating this aspect. It is often necessary to reformulate the search strategy and repeat the retrieval operation several times before satisfying an inquiry.

After the proper data has been retrieved, it should be presented to the user in a usable and acceptable format. The response could take the form of a bit of data, a compilation of data, a document extract, a document copy, tabulation printout or any other format. System response can be disseminated by mail, telephone, TWX, or personal contact. If the user can not state data requirements exactly, the output will, of course, be less specific. After examining the initial output, the user may be able to restate his inquiry to better define his data needs.



TYPICAL INQUIRY SERVICE OPERATION

An important function for inquiry service personnel is to maintain detailed records of the usage of the data base. These records provide the basis for expanded capabilities and improved services. For example, when the usage of a type of data provided in documented form becomes sufficiently large, the data base should be expanded to contain this new data and handle the request more effectively. Efficient inquiry centers will also have a procedure for following up any service rendered with some form of questionnaire, which could be as simple as a postcard, requesting the user to state how satisfactory the service was and list any problems. Such a technique helps the inquiry service to maintain effective service.

Conclusion

It is clear that the use of inquiry services to answer specific questions will continue to grow substantially as more and more information in the scientific, engineering, and management field is generated. As the use of such services becomes more widespread, their cost effectiveness will be closely analyzed and such criteria as the cost per question answered will become important in measuring the efficiency of such systems. Other factors which will also be used to judge the effectiveness of information inquiry centers will be access and turn around time, the percent of satisfied questions and users, and the precision ratio, which is the ratio of relevant documents retrieved to total documents retrieved from the center.

Any center that is to be effective will also have to maximize it's recall ratio, which is the ratio of total documents in the system that were relevant to the inquiry.

Most current information centers of any size are government supported either directly or indirectly and are oriented towards scientific, engineering, and technical types of information. The Management Information System of the future, however, in a large organization could very well parallel the development of these scientific and technical information centers. Certainly the picture of high level managers sitting at computer consoles to directly access data bases is not realistic since most of the direct interrelationship with computers will be per-

formed through staff personnel. Is it not possible that the information inquiry service could provide the same function in a more efficient fashion? Management will certainly have to face up to this question.

In anticipation of the answer, we might look to see why most of today's information inquiry centers are government supported. At various times, proposals have been brought forward to charge each user of such a system the cost of answering his question. Such proposals have never worked because the individual engineer usually does not have the authority to easily make available the necessary funds to pay for such a service without going through an inordinate amount of red tape. Therefore, such centers have had to be supported by top level management decisions concerning cost effectiveness, and market place forces of supply and demand have not been brought to bear on the existence of information inquiry services.

As a corollary to this, because such services appear as budgetary line items, they are among the first to be cut when government expenditures must be reduced. This is often unfortunate since such services are justified by their cost savings. The cost savings, however, are hidden in that the existence of such an information inquiry service makes each scientist, engineer or manager more productive. When the service is lost, the cost is simply spread out as an opportunity loss over all of the work that is not, and could have been accomplished, and as such, is effectively hidden as compared to the obvious clarity of the costs of running the center.

In spite of such problems, however, it does appear as though the use of these centers will continue to grow. One problem already encountered is that many possible users do not even know of the existence of an information center which can specifically help them to solve their problems. This situation has led to proposals for linking all of the major inquiry type services in the United States in such a fashion that an inquiry to any one of them will be correctly channeled to its proper destination. Another alternative, of course, could be some sort of "super inquiry service" that would tell you which inquiry service to go to – to get your question answered!

In any case, one thing is clear. The continued vast growth of information in various fields of specialization will require more and more library, data, and information services and the information inquiry field will probably become an important information science specialty by itself. END

IIA Adopts Resolution On Relationship Of Copyright-Proprietary Rights—Information Usage

A two-part resolution on the relationship between copyright, new proprietary rights and the optimal use of information was adopted by the Information Industry Association, headquartered in Washington, D.C.

The IIA resolution, adopted unanimously recently by the member firms assembled for the meeting on Copyright and Related Proctections for Information Age Products, concludes that proprietary rights are comparable with and essential to the widest dissemination and use of information.

It also urges that a national commission be established to consider not only new technological uses of copyrighted works but also, more broadly, the impact of new information technologies on the optimal development and use of proprietary information products and services.

The resolution also notes that the new information technologies may require different legal concepts and the national commission should not be limited to considering only copyright solutions.

The panel of speakers for the program emphasized the new and different problems and relationships created by the efforts to apply new information and communication technologies to the efficient dissemination of information products and services.

Speakers included Barbara Ringer, Assistant Register of Copyrights, who observed that the alternative to copyright is patronage and subsidy which carries with it control of information and culture. She acknowledged that copyright legislation has not been prospective in nature in the past and that four pressures will effect changes in the copyright of the future: pressure for access; business cannot wait for paperwork; increased government use of the technologies; and the need for artistic independence being asserted by the young.