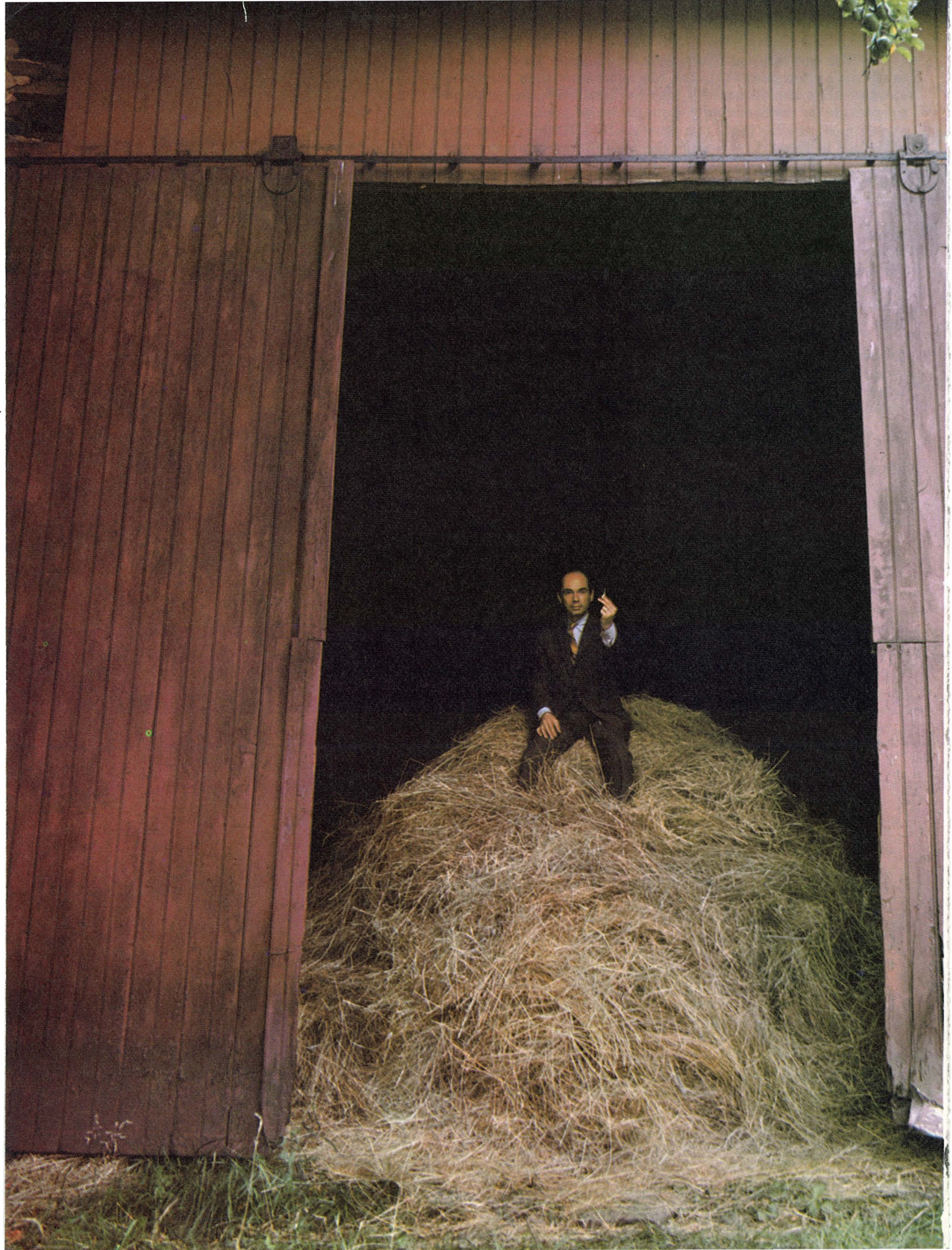
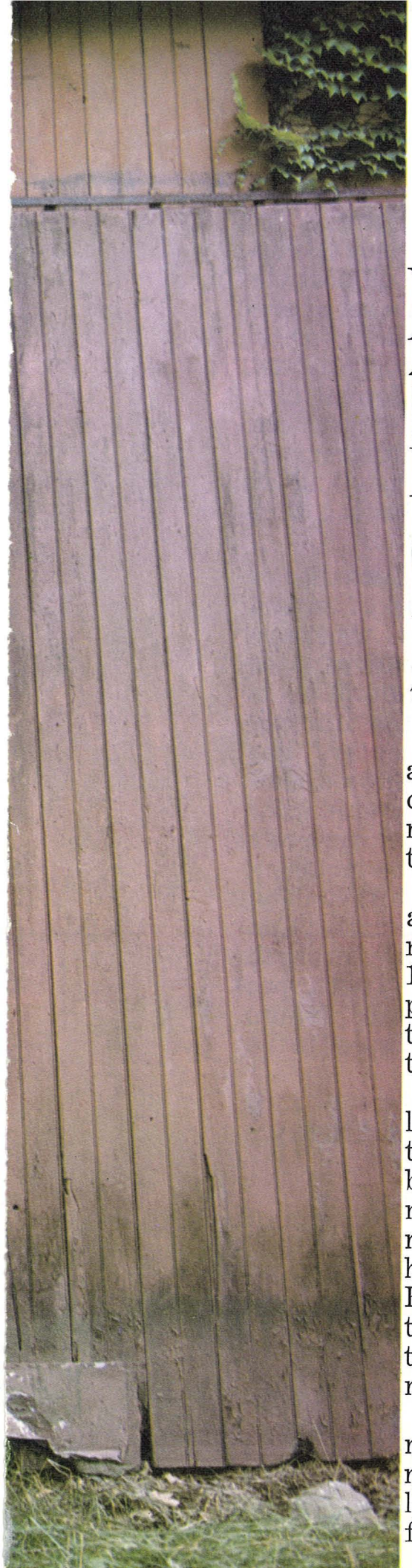
forum69

THE MANAGEMENT MONTHLY ON DATA PROCESSING OCTOBER 1969

Mini-computers: Evolution of a bold new breed.





Finding the buried nugget in a mountain of information.

By George Schussel

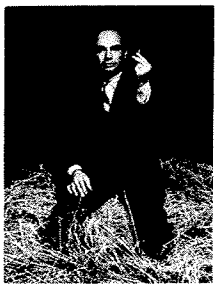
New information services help you extract what is relevant to your problem.

The current explosive expansion of knowledge or, at any rate, information may be the pride of twentieth century technology, but it is also the despair of those who must cope with it, whether to put it to some use or merely to sort and store it.

For example, the output of scientific information alone is doubling every eight years and is expected to reach *120 million pages annually* before 1970. Over 100,000 technical and scientific journals are currently published in sixty languages, and the new information that becomes available every 24 hours would easily fill the equivalent of seven sets of a 24-volume encyclopedia.

Several government studies have shown that the lack of needed information in the right place at the right time causes annual losses in the billions of dollars. It has been estimated that a working scientist in an industrial nation spends up to one-third of his time looking up information and that the cost of this search represents perhaps one-fifth of all the money allocated to science. Besides search for specific information, the same scientist must spend, according to other estimates, from 25% to 75% of his time just keeping abreast of major developments in his field.

Business management is in a similar predicament—overwhelmed by vast masses of technical, commercial, and manufacturing data, yet frequently at a loss to find the specific piece(s) of information required for a given situation or problem.



It is fitting that the computer—one of the primary causes of this information explosion—should be brought to the rescue of those who have to deal with it. In fact, third generation computers with their large storage capacities and high-speed access are the only devices that *can* deal with it. The fact that they permit storage in unstructured files is a further convenience since this obviates the necessity of trying to anticipate all possible kinds of input and requests for retrieval in a set of preselected categories into which all data must fit (or be excluded). An unobstructed file will accept any datum, however odd or unpredictable in content or format, and store it at any available vacant "address". Furthermore, the modern computer's time-sharing and multi-programming capabilities provide for economical handling of miscellaneous types of data in one system, instead of requiring a separate system for each.

Hence, it is not surprising that most discussion of management information systems has centered around computers – computer hardware, software, programs, and services. Still, the average user, or seeker of information, is no data processing expert, and he still needs help. In many scientific and technical fields, the "information inquiry service," or "information center," has emerged to provide it. Such a center employs a staff trained both in a given area of knowledge and in information handling. They collect, evaluate, and store relevant data and retrieve it as required by the inquiring client. For example, the PRINCE/APIC (Parts Reliability Information Center/Apollo Parts Information Center) is a 45-man service set-up by NASA's Marshall Space Flight Center to provide information on tests, specifications, vendors, comparative performances, failure rates, and other relevant matters for all parts used in the Saturn Apollo Lunar Project.

The increase, both in size and numbers, of such centers in recent years has been spectacular. There are over 1100 of them in the United States today.[†]

TYPES OF INQUIRY SERVICES

Most information centers can be classified as library services, data centers, or analysis centers. *Library Services* deal primarily with bibliographical information, and in response to inquiries will furnish documents, abstracts, or references to documents in which the inquirer must search farther to get the answer to his specific question.

Data centers, on the other hand, deal with detailed data rather than documents and provide specific answers to specific questions—e.g., about the special properties

continued on page 42

continued from page 22

and characteristics of materials, chemicals, devices, processes, etc., or about configuration management, research and development projects, financial controls, or marketing methods. Beyond this, *analysis centers* will permute, combine, and analyse data to create new information, again in answer to specific questions.

DATA ACQUISITION

The first order of business for an information center is the acquisition of data, which can come from many sources (government agencies, technical publications, manufacturers, etc.) and in a variety of formats (documents of all kinds, punch cards, catalogs, magnetic tape ...). Some of this comes easily – some of it perhaps too easily; and some must be extracted from reluctant possessors by persuasion or, where possible, as in some government programs, by directive. In many fields, emphasis is necessarily placed on *current* data, though historical data may also be required.

Once acquired, the data must be logged–typically by source, title, date, and number. It must then be carefully scrutinized to prevent the entry of questionable information. For example, preliminary data must be clearly labelled as such, to give a clear indication of its validity (and should be superseded by more solid information as soon as it is available). Other criteria for validity depend upon the judgment of the acquisition personnel and the center's management. Figure 1 depicts the flow of work through a typical data input process.

INDEXING

Proper indexing is, of course, essential to effective data retrieval. No single system is unquestionably best for all applications. Among the most commonly used are Key Word In Context indexing (KWIC), citation indexing, alphabetical subject indexing, and bibliographical indexing, of which the first is one of the handiest for use with computers. Key words are terms selected from the title or the text of a document, and when matched or combined (a process for which the computer is ideally suited) with each other or with previously stored terms, they provide rapid retrieval of data precisely tailored to the inquiry at hand.

The more important the information or the greater the range of its possible application, the more detailed the indexing should be. The goal of all indexing

†To locate information centers specializing in various types of data, consult the National Referral Center for Science and Technology, Library of Congress, (202) 967-8265. Or see **A Directory of Information Sources in the United States**, which lists hundreds of existing data centers.

techniques is to identify data so that a particular bit of information can be found quickly, infallibly, and with a minimum of effort. The approach to this ideal varies, of course, but any indexing scheme should, at least:

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

In passing, it should be noted that though the originator of a document may supply a thesaurus of key terms for indexing it, this is often insufficient for the broadest application of the document's information. The special staff of an information center may also assign other descriptors to the document, making it accessible to workers in related fields where it may be relevant.

As indexing procedures are refined and improved, the amount of irrevelant material retrieved in response to a request will decrease. Although the superfluous material may be acceptable enough if the inquirer wants to browse, it is a nuisance to the man who wants a specific piece of information.

DATA STORAGE

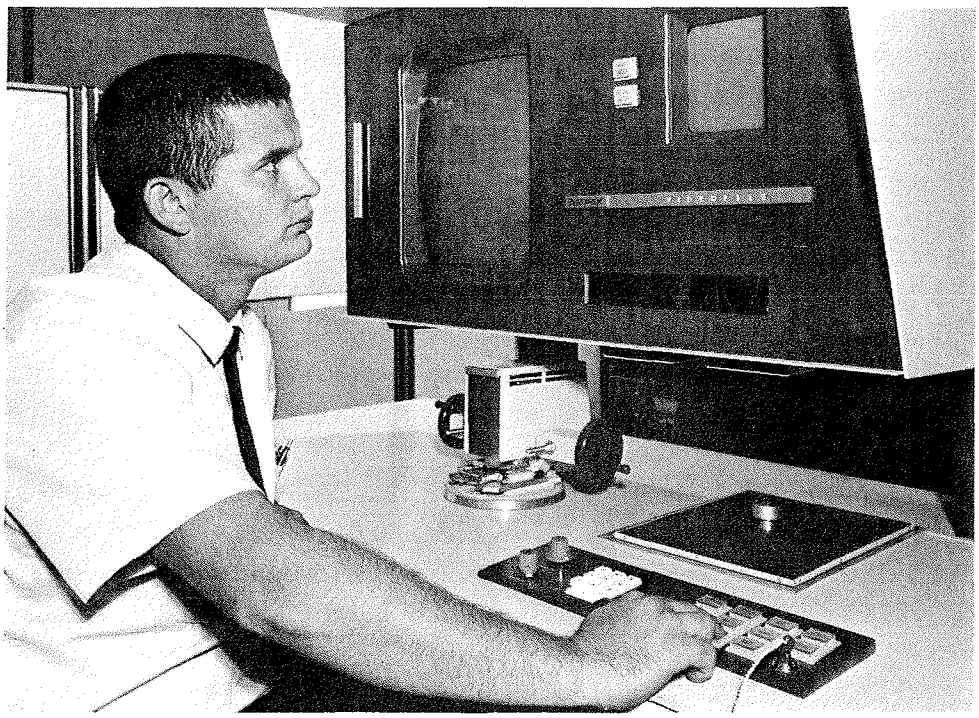
The unstructured file, in which it is not required that a particular piece of information be stored in a particular location, is to be preferred, as much more flexible, to a structured file. Data is stored often as combinations of data names and related values. The data name can be a key word that defines 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.

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 remain current 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 every so many months—typically twelve. In some data centers it is a good idea not to destroy updated and purged data, but to place it on magnetic tapes for dead storage so it can be recalled if the need arises.

VIDEOFILE SYSTEM

In addition to various microfilm, microfiche, and aperture card systems, one interesting new development in the graphical storage of large volumes of information has been the development of the Videofile by Ampex. Using video tape techniques, large volumes of graphical images can be stored on magnetic tape and subsequently very quickly found and reproduced when necessary. The tape drives 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 of 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 2½ minutes. The first Videofile was installed at Brown Engineering Company's facilities in Huntsville, Alabama to service NASA's PRINCE/APIC information inquiry system. This 1967 installation proved unsuccessful because the TV camera that photographed the images had insufficient resolution to pro-



duce acceptable copies when the information was retrieved from the file. The second, but first successful, installation of the Videofile appears to have been at the Southern Pacific Company in San Francisco, California. The installation and use of this large scale information storage and retrieval system was described in the October, 1968 issue of *Business Automation*. The total storage capacity of the Southern Pacific system exceeds 20 million documents, which can be filed on the average of one every 10 seconds and retrieved about one every six seconds. 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 future developments in facsimile transmission equipment promise to lend substantially greater capabilities to information centers.

INQUIRY SERVICE OPERATION

The distinguishing feature of an information center is its inquiry service—the human interface it provides between the inquirer and the data bank in which he thinks the information he wants may be found.

Its principal functions are two:

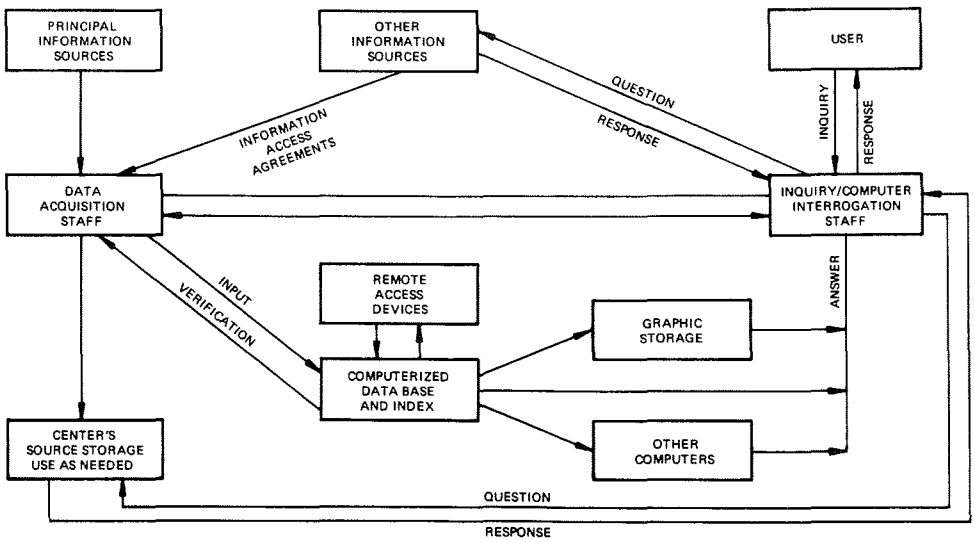
1. To define the query and translate it into the most efficient and precise retrieval language.
2. To evaluate the information retrieved for completeness and relevance to the inquirer's needs and to present it in the format most useful to him.

In addition, some information centers, if unable to provide the requested data from their own resources, will undertake to find it elsewhere.

Inquiries are often incomplete, vague, or ambiguous, and a dialogue between inquirer and inquiry service is often necessary to define his needs precisely, and formulate an explicit request. (To remain as flexible as possible at this step, the service should not require a specific format for queries. Nor should it require queries to be batched before search, since this only lengthens turnaround time.) Subjective analysis of the applicability of the data retrieved is also required as a further check on the appropriateness of the query. (After examining the initial response, the inquirer may be able to restate his question more precisely.) At the current state of the art, there is no technique for automating these two functions of the inquiry service.

In order to remain as efficient as possible, an in-

TYPICAL OPERATING CONCEPT OF AN INQUIRY SERVICE



formation center will maintain detailed records of the use of the data base. If any type of data is in frequent demand, special efforts should be made to augment it, and keep it up-to-the-minute and thoroughly indexed. For its own information, the center will maintain a follow-up procedure—a questionnaire (which need be no more than a post card) asking comments about the adequacy of services rendered, problems encountered, etc.

COSTS

Before information inquiry services become widely used in business and industry, their cost effectiveness must be closely analyzed. Some of the factors involved in this analysis are:

- Cost per question answered (about $20 in the PRINCE/APIC system mentioned above).
- Access and turnaround time.
- Precision ratio (proportion of relevant data to total data retrieved).
- Recall ratio (proportion of documents retrieved to total documents in system relevant to inquiry).

Because most information centers are directly or indirectly supported by government and have not yet been subjected to market place forces of supply and demand, not much private cost information is yet available. Actual costs are further obscured by the fact that such services appear as budget line items (to be cut whenever government expenditures must be reduced), but their cost savings do not. The savings are hidden in the increased productiveness of the engineers and scientists using the services; when the services are cut, the cost simply spreads out over lost opportunities to accomplish

work that could have been done, but is not.

Most government supported information centers have been set up to provide scientific, engineering, and technical data. The Management Information Systems, or centers, of the future could very well parallel the development of these technical centers. However, they will have to be supported by top level management decisions based on cost-effectiveness analyses. Subordinate management or technical personnel usually do not have the authority, without going through an inordinate amount of red tape, to commission funds to pay for such information services.

Despite such problems, it appears that such centers will continue to proliferate and expand and their users to multiply. The continuing vast increase of information in various fields of specialization will require more and more library, data, and analysis services, and the information inquiry field will itself probably become an important information service specialty. Indeed, a kind of "super inquiry service" that would tell you which information center to go to has already been proposed.

Dr. George Schussel is Vice President, Financial and Banking Systems, of Informatics, Inc. He has taught courses in computers, simulation methods, and information systems and has published numerous articles in Management Science and other journals.

glos·sa·ry

Access Time—The time between a command to obtain information from a memory and the actual receipt of that information.
Address—Designation indicating the location of information in a memory.
Address Register—A memory register in which an address is stored.
Algol—ALGorithmic Oriented Language. An international programming language.
Base—Also called Radix. A number, most commonly 2 or 10, which is used for the representation of numbers by indicating only the exponents to their base. For example, the exponents to the base 10 are "0" (units), "1" (tenths), "2" (hundredths), "3" (thousandths), etc.
Binary Coded Decimal System—A number system in which each digit of a decimal number is represented by four binary bits.
Binary System—A number system with a base of 2. It contains only two symbols, often designated as "0" and "1," the absence or presence of a mark, hole, voltage, etc.
Character—A decimal number (0 to 9), letter (A to Z), or any specially designated symbol, such as carriage shift, upper or lower case, etc.
Cobol—Abbreviation for common business oriented language. A programming system adaptable to many different computers.
Data Processing—Often used interchangeably with "computing," but referring primarily to non-mathematical transformation of data.
Digital—Pertaining to the use of numbers. Often contrasted with "analog."
Digital Computers—Devices in which information is represented in coded form and manipulated in discrete steps.
Electronic Data Processing—The application and use of computers or other data handling devices for business, scientific, and engineering problems. ("Electronic Data Processing" is the term usually applied by private industry and by computer manufacturers. See also Automatic Data Processing.)
Hardware—The physical devices used in a computer.
Hexadecimal System—A numbering system based on 16. Synonymous with sexadecimal system.
Interface—Characteristics, programs, or equipment involved in the joining of two or more information handling devices or systems.
I/O—Abbreviation for Input/Output.
Macro Instruction—An instruction in a source language that represents several instructions in machine language.
Magnetic Core—A small ring made out of magnetic material, usually ferrite, which is stable in either of two magnetic states and therefore can be used to store one binary bit.
Microfilm—A 35mm or smaller film used for photographic copying in reduced size.
Microsecond—One millionth of a second (10^{-6}).
Millisecond—One thousandth of a second (10^{-3}).
Multiprocessing—The execution of several processes at the same time by a single computer or computer group.
Nanosecond—One billionth of a second (10^{-9}).
"Off-Line"—The operation of data devices which are not under direct control of the main system.
"On-Line"—The operation of data equipment under control of the main system.
Pert—Program Evaluation and Review Technique. A management tool for scheduling major research, development, or construction projects.
PL/I—A programming language that can be used for both business and scientific applications. The language was developed by IBM primarily for use with the 360 system.
Program—The operating instructions furnished to a computer from an "outside source," usually in the form of punched cards, tapes, or keyed-in instructions.
Random Access—1. The ability to "instantaneously" reach any part of a memory; 2. The characteristics of a memory whose access time is independent of storage location.
Simulation—The dynamic representation of a large physical system or process by more convenient means.
Telemetry—The technique of measuring a quantity at one point and transmitting the information to a distant point for further use.
Time-Sharing—The use of a computer by several operators for different problems, with the computer coordinating the work and the individual operators normally not being aware of the other programs.
User Terminals—Those devices in a communications or computer system with which the user is in direct operational contact.
Real-Time Computer—A computer that is controlled by the requirements of an outside source or an outside event. In many cases, but not always, it is a time-dependent control.
Register—A temporary memory for a limited amount of data.