Conference on

Machine Processing of

Remotely Sensed Data

October 16 - 18, 1973

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

ON THE MANAGEMENT AND PROCESSING OF EARTH RESOURCES INFORMATION*

C. W. Skinner and R. C. Gonzalez

Department of Computer Science, North Carolina State University, Raleigh, North Carolina; and Electrical Engineering Department, University of Tennessee, Knoxville, Tennessee

I. ABSTRACT

The basic concepts of a recently completed large-scale earth resources information system plan are reported. Attention is focused throughout the paper on the information management and processing requirements. After the development of the principal system concepts, a model system for implementation at the state level is discussed.

II. INTRODUCTION

The past two decades have witnessed unprecedented growth in our ability to generate and record information in all fields of human endeavor. The much-discussed information explosion with its attendant figures on the enormous quantities of data generated each day is a fact of life with which we must continually deal. In spite of the growth and development of digital computer technology and the significant strides made in techniques for information storage and retrieval, our ability to generate information far exceeds our capability to intelligently process and assimilate it.

An area of recent national emphasis where the discrepancy between information generation and utilization has become increasingly apparent is the area of earth resources, where critical decisions are made time and again without adequate information. In many cases the basic data necessary for intelligent decision-making have been collected and stored in the files of an agency or corporation--the missing link is a system for informing the user of the existence of the data and for retrieval and processing into a useful form of application.

The need for earth resources information systems has existed to some degree for many years. Soil and mineral surveys, oil well core samples, and stream flow information, for example, have been gathered and used for specific purposes for generations. The capability to bring the relevant information from many disparate sources to bear on complex problems has always been lacking. However, the development of remote sensing technology and the advent of large-scale remote sensing projects has given a new sense of urgency to the problem. The data rates and quantities have increased by orders of magnitude. Not only must better processing and storage techniques be developed, but comprehensive earth resources information systems must be designed and implemented if we are to avoid being inundated by obviously useful--but unused--information.

This paper discusses basic design considerations and problems of fundamental importance in the establishment of such an earth resources information system, with particular emphasis on the information management and processing functions. The conclusions presented here are based to a large extent on the authors' participation in the recent system design project at the Marshall Space Flight Center which resulted in the design of the ERISTAR (Earth Resources Information Storage, Transformation, Analysis, and Retrieval) system (Vachon, et.al., 1972).

---

*The work reported in this paper was sponsored by the National Aeronautics and Space Administration under grant NGT-01-003-044.

III. DESIGN PHILOSOPHY AND CONSTRAINTS

There are dozens of information processing systems in existence today which deal with one or more specialized areas of earth resources information. Some of them have been operational for many years, while others have arisen as a result of the design and subsequent launch of the Earth Resources Technology Satellite (ERTS). Several of the systems are computer-centered and make use of sophisticated techniques for automatic information organization and retrieval. Others are primarily manual in nature with information stored in hard-copy form (photographic imagery, maps, printed documents). A brief list of some of the systems and operating agencies is given below:

United States Geological Survey (USGS) - Aerial photographs, catalogs, browse files
Jet Propulsion Laboratories (JPL) - Image processing, data processing, storage (archives of data)
NASA Aerospace Safety Information System (NASIS), Lewis Research Center - Interactive search and alphanumeric display capability, dial-up capability, geographic search and terminal on-line capability, and is expandable
Earth Resources Observation System (EROS), United States Department of Interior, Sioux Falls, South Dakota - Satellite and aircraft imagery, USGS photographs, computer tapes
National Space Science Data Center (NSSDC), Goddard Space Flight Center - Archives of space data
Earth Resources Image Processing System (ERIPS), Manned Spacecraft Center - Aircraft imagery, processing (scaling, enhancing, mosaicking, contouring, magnification, integration, convolution, multispectral analysis, image correlation, color enhancement, noise removal, smear removal, area identification, Fourier transforming, image registration)
National Oceanic and Atmospheric Administration (NOAA), United States Department of Commerce
Environmental Science Services Administration (ESSA), United States Department of Commerce
National Oceanic Data Center
National Climatic Center
National Geophysical Data Center and Aeronomy and Space Data Center
Environmental Science Information Center

A more detailed description of these centers is available (IBM, 1972).

New national priorities with respect to the environment have created a need for earth resources information which goes beyond individual disciplines and specialized areas of interest. Just as the ecologist is concerned with the total environmental system, so must the information scientist be concerned with developing systems which provide comprehensive information spanning the entire field of earth resources. If, for example, an agency or corporation is charged with preparing an environmental impact statement, information from a variety of sources and in a variety of forms must be drawn together to assure that all the relevant environmental factors have been considered.

In addition to treating information from a multitude of disciplines, the system should accept data from different sources and integrate them into a total data base from which questions can be answered. Aerial and satellite remote sensing obviously plays a major role, but surveys and measurements made on the ground also constitute significant elements.

With this in mind, we can state the first principle of design which is used throughout our analysis:

> An earth resources information system should accept data from many different sources, dealing with many different subject areas; process the data and manage the resulting information to effectively answer a broad spectrum of queries relating to earth resources and the environment.

None of the earth resources information activities listed above nor any of the many similar activities in federal, state and private agencies throughout the United States constitutes a total earth resources information system in this sense. Yet, together they represent a vast but diffuse information processing capability. Any total earth resources information system should provide for tapping this enormous reservoir of current activity. This leads to our second principle of design, which may be thought of as a constraint on the system.

> A comprehensive earth resources information system must provide for interaction with, or inclusion of, existing specialized systems with useful capability.

An approach to the overall system design which combines both design principles is the establishment of a network of these specialized information analysis centers interconnected with a network of centers designed to interface with the users of earth resources information. It is this concept of system design which we have followed in defining the requirements and approaches for the management and processing of earth resources information.

The information network approach, as opposed to a centralized system, is by no means unique to this study. Other important planned and prototype systems also make use of the network concept.

The Environmental Information Systems Office (EISO) of the Oak Ridge National Laboratory presents an example of an operating information network which makes use of existing data bases (Combs, 1971). EISO is connected with some 200 environmental information centers around the nation, including a number in the Environmental Protection Agency. Data bases are largely maintained in computer storage. When a user presents an inquiry, an information specialist interprets the question, accesses the appropriate data base, and performs any detailed analysis required to satisfy the user's needs.

An activity in the planning stage is the Resource and Land Information (RALI) program proposed by the Department of the Interior (Hanshaw, 1972). As conceived in the initial design, there would be three levels of contact: (1) national, (2) regional, and (3) state. The program would be established at the national level immediately. The implementation of the total system would stretch into the 1980's. The focal point for information storage and analytical expertise would be at the regional level, while state centers would largely provide a reference service. The user's requirements would be satisfied by raw, interpreted or derivative data, by analytical predictions, or by analyses of alternative policies and their impacts. The data and information would be provided as tabular or point data, maps, reports, and computer tapes.

Less ambitious but related projects are found at the state level. For example, the North Carolina Environmental Information and Education Network (NCEIEN), which was begun last year, has as its goals

> "...to identify environmental information and education needs at all levels of our community and then to apply existing and potential resources to these needs." (Slaughter, 1973).

In this case, the emphasis is on education and the dissemination of information rather than on sophisticated information processing systems.

The information processing requirements of each of these systems differ due to differences in emphasis and scope. A total earth resources information system, such as the ERISTAR system, will have requirements which overlap each of these to a large degree.

IV. INFORMATION PROCESSING REQUIREMENTS

The requirements for processing within an earth resources information system may be generally divided into four categories:

1. Organization
2. Storage
3. Retrieval
4. Transformation

The interrelationships among these requirements relative to the sources and users of information are shown in Figure 1.

A. <u>ORGANIZATION</u>

The information and information products acquired by the system must be placed in a logical organizational structure for proper and efficient storage and retrieval. The storage requirement discussed below deals with the physical form and format of the information; organization deals with the logical structure of relationships imposed upon the information prior to storage.

The major elements of information organization are:

1. Content analysis
2. Classification and indexing
3. File structure design

Content analysis is the determination, for each information item, what that item "is about." For a book or report, this might consist of extracting a set of keywords or terms which describe the content of the document. For remote sensing imagery the process might be as simple as listing the area covered by the image and the time at which it was taken or as complex as attempting to identify every significant feature in the image. Clearly, these are problems worthy, in their own right, of a great deal of study and research.

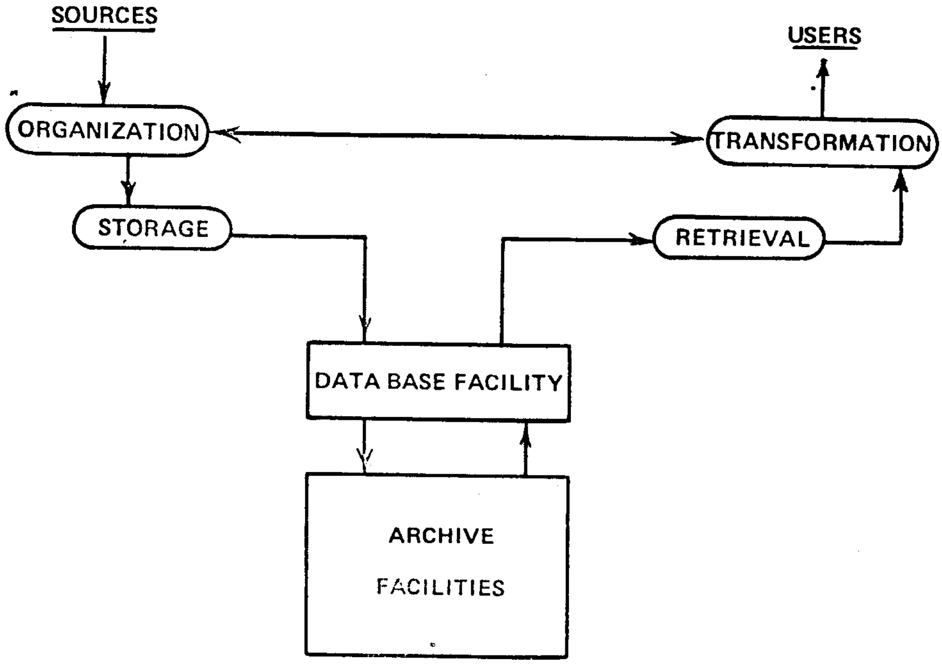


Figure 1. Primary Data and Information Flows

Once the content of an information item has been determined to an acceptable degree, classification and indexing can take place. Here we mean classification of the entire set of documents, images, or whatever, into categories rather than classification of individual data points for identification purposes. Indexes provide "names" for each of the categories and aid in effective retrieval.

File structure design relates the information organization to the mode of storage, since file structure decisions cannot be made independent of storage considerations. The choice of a file structure for representing the information strongly influences the efficiency and effectiveness of retrieval. File structures may be as simple as the sequential direct files traditionally associated with alphabetized folders in a filing cabinet; they may be inverted file structures like a card catalog in a library; they may be much more complex, using linked structures with multiple pointers, when the data base is being processed by computer.

B. <u>STORAGE</u>

To fulfill the storage requirement, the system must take into account the implications of the form and format of the information, and the methods of access which are required. The most distinguishing features of requirements on storage facilities for earth resources information are huge volumes of data at high transfer rates to be kept for long periods of time such that any part may be accessed quickly for a diverse community of users.

The physical form of the information has immediate implications for its accessibility, permanence, bulk and incremental cost. Except for special cases, the form is oriented toward man or machine, depending upon its expected use.

Human-oriented forms are the most bulky, most permanent, and least accessible in large quantities. Books, reports, and other hard-copy printed text constitute a significant burden on present storage facilities, although reduction to microform has been instrumental in reducing the physical space requirements. Photographs comprise a major portion of earth resources data and represent the fastest-growing segment of the storage requirement. Photographs and photographic maps and overlays have particular demands concerning reduction in size and storage for long periods, since any distortion in the physical form is a distortion of the data stored.

Machine-oriented forms include both analog and digital data. Analog forms include video tape and other continuous signal recording media. Digital forms encompass the traditional punched cards and paper tape; the various magnetic surface devices such as tape, disks and drums; higher-speed devices such as magnetic core and semiconductor memories; and the many ultra-high speed or ultra-high volume storage media under research and development.

The storage requirements for any type of data, whether oriented toward man or machine, and the storage devices which can satisfy those requirements can be compared in terms of the quantity of information involved and the rate at which the information must be stored and retrieved. The Information Flow vs. Information Quantity diagram (Figure 2) which was suggested by W. O. Clark (Vachon, et.al., 1972), provides a convenient means to illustrate the various data products which are relevant to an earth resources information system. The log-log plot allows comparison of items which differ by several orders of magnitude and spans the spectrum of current technology.

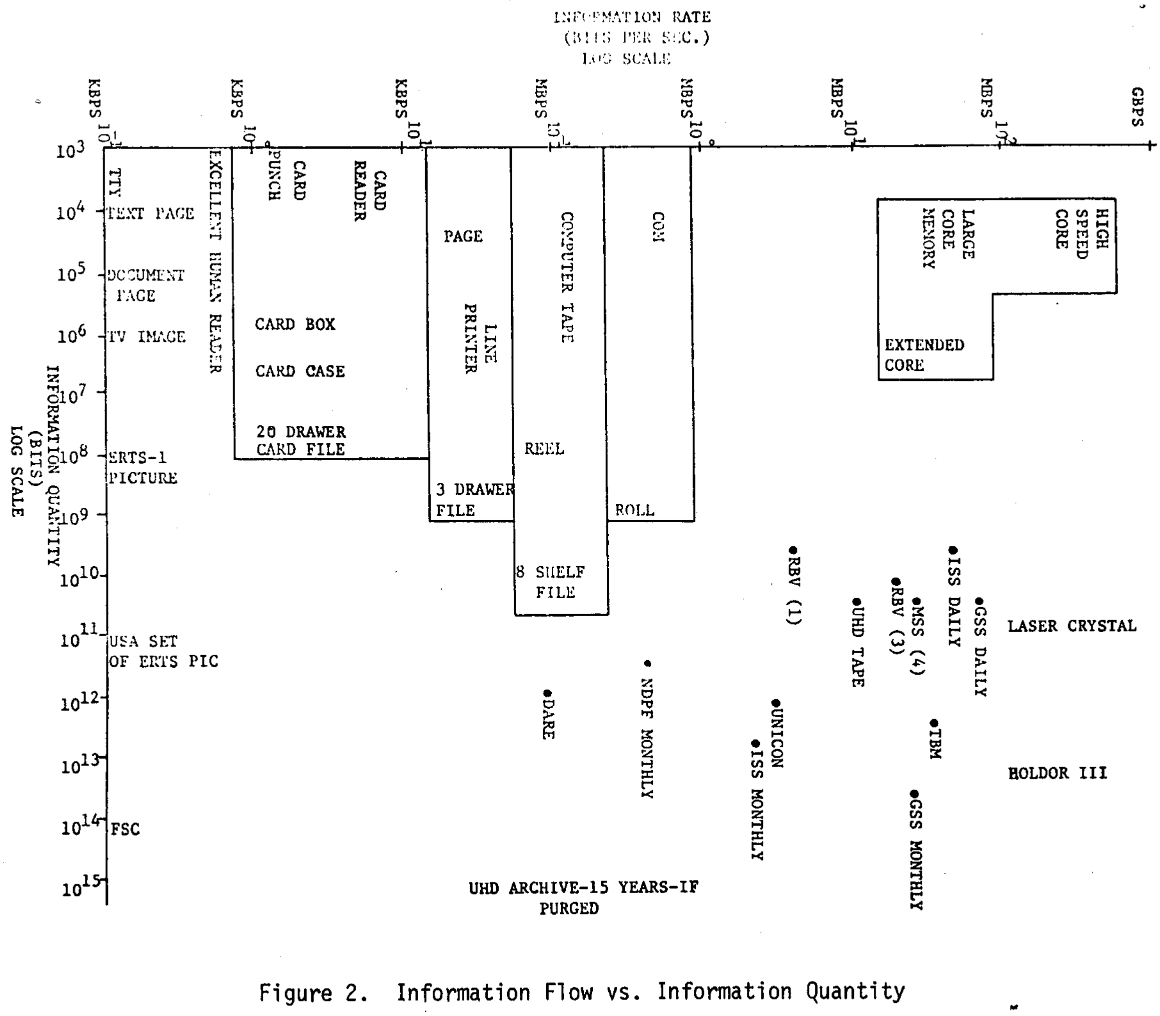


Figure 2. Information Flow vs. Information Quantity

Information quantity boundaries give the minimum and maximum storage size for a particular technique. Information flow boundaries show the minimum and maximum rates at which the information can be stored and accessed. These boundaries depend upon many factors and in reality have considerable overlap. However, the intent is only to illustrate the magnitude of the storage parameters in the developing imagery technology. Note for example that to fully support the proposed Growth Space Station (GSS), it would require memory capability and processing rates each two orders of magnitude greater than the NASA National Data Processing Facility (McDonnel Douglas, 1972).

C. RETRIEVAL

How well the system satisfies the retrieval requirement is, to a great extent, a function of the information organization and storage facilities which have been used in the design. In response to a query concerning earth resources, the system should be able to recall and retrieve any relevant data available to it and should also be able to recognize when pertinent data are not available. The retrieval function is basically one of matching the formalized user query to the data base in order to draw out those items which are relevant to the query. There is little chance of user satisfaction with the system unless the system responds in a timely manner and unless the user is allowed to interact with the system to help match his real information needs to that which the system can supply.

In addition to the high degree of recall required, the system must be effective in eliminating noise from its responses. Since the largest source of noise to be eliminated is irrelevant material, the system should have a high degree of precision in its retrieval performance. Other noise can be introduced through photographic reproduction of text and imagery, and through the communications links which are an integral part of a network system. Some kinds of noise are, of course, much more tolerable than others. Random errors introduced into a bit stream transfer between computers may wreak havoc with subsequent computation, while a significant degradation in the microform image of a page of text may be of little concern as long as the characters are recognizable by the human reader.

D. TRANSFORMATION

The transformation requirement is the one most often ignored when designing an information system. It expresses the need for conversion of the information from the form in which it was entered and retrieved to the form most appropriate for the user's needs. The transformation involved may be both logical and physical in nature, and the extent to which this requirement is satisfied will strongly influence the system's acceptance by the user community.

Physical conversion refers to changing the physical appearance of the information. This may involve the conversion of tabular data to charts, graphs or maps, or perhaps the digitization of a photographic image for subsequent computer processing. In some cases it may mean simply providing enlarged hard-copy reproductions from microfilm or microfiche.

Logical conversion includes the entire process of analysis and inference--deriving results which are not explicitly stored in the data base. Obviously any information system implementable in the near future will be limited in the degree to which it can perform logical conversion. Both deductive and inductive inference procedures are required. Prediction, using simulation and other modeling techniques constitutes an important part of these. For example, a user of the system who is anticipating construction near the flood plain of a creek or river is interested not only in the soil types, rainfall distribution, and topographic features in his area, but also in estimates of the damage to his structure by a 100-year flood.

The various elements of the transformation requirement deal with user service and user satisfaction. As such, the exact capabilities required depend strongly upon the system environment and the composition of the user community. However, based upon experience with other types of information systems, one fact stands out--the system must be interactive. Whether it is man/machine interaction or man/man interaction will vary in the different components of the system, but if the user is to receive the type of response he needs, in the form in which it is most useful to him, there must be adaptive system feedback.

V. INFORMATION MANAGEMENT

There is no clear-cut distinction between information processing and information management. For the purposes of this paper, we have used processing to refer to the detailed operations performed on the information within the system, while management refers to the overall information flow from sources to users via processing.

An example of information flow in the System for Earth Resources Information proposed to NASA by the National Research Council (NAS-NRC, 1969) is shown in Figure 3. In this case aerial, satellite, and ground truth information is used to produce a product such as land-use maps. The ERISTAR plan has a similar flow with respect to its remotely sensed data, but it also provides for tapping the vast reservoirs of earth resources information gained by other means and allows processing and interpretation by humans as well as by machines.

In the world of information systems the terms "user-oriented" and "cost-effective" hold unquestionable positions as guiding principles of design. The trick, of course, is to determine exactly what "user-oriented" means in a particular system, and to design the system to be truly cost-effective. It is well recognized that the success of a large-scale earth resources information system is considerably dependent on the system's ability to serve a population of users whose backgrounds and educational levels cover a spectrum of staggering proportions.

A successful information system is one that is used by the group for which it was designed but, historically, information systems have not attained a high level of user acceptance. Because of the glamour of new hardware and new processing techniques, system designers have tended to underestimate the problem of satisfying the user of the system. An interesting analysis of this situation has been developed (Katz, 1969) and is illustrated in Figure 4. The degree of cross-hatching indicates the relative emphasis placed on the various elements of an information system. The least emphasis is found in the lower right corner, which might be interpreted as representing "presentation of results to users."

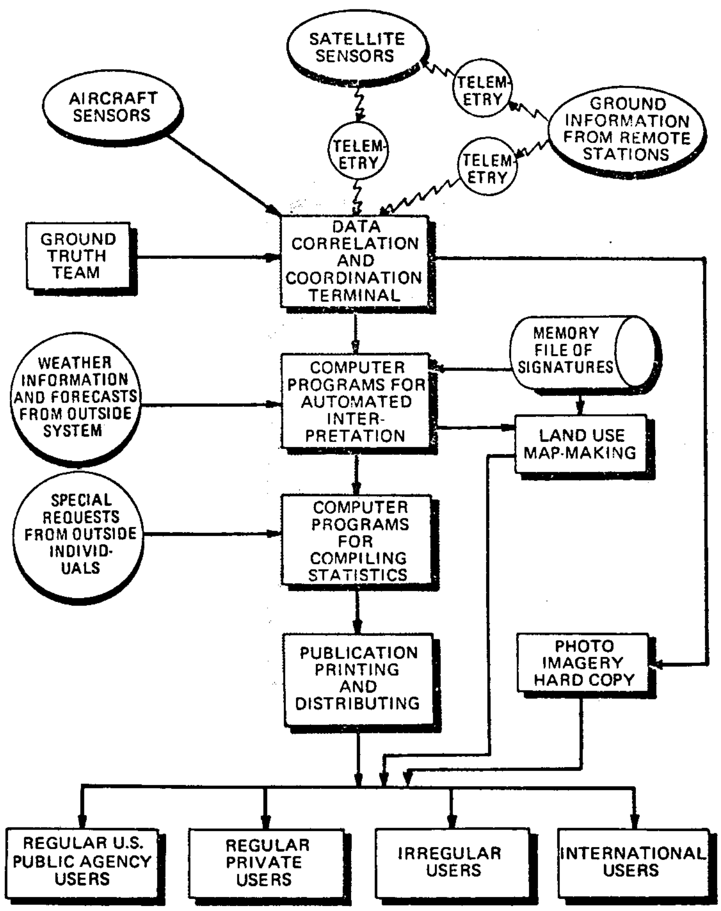
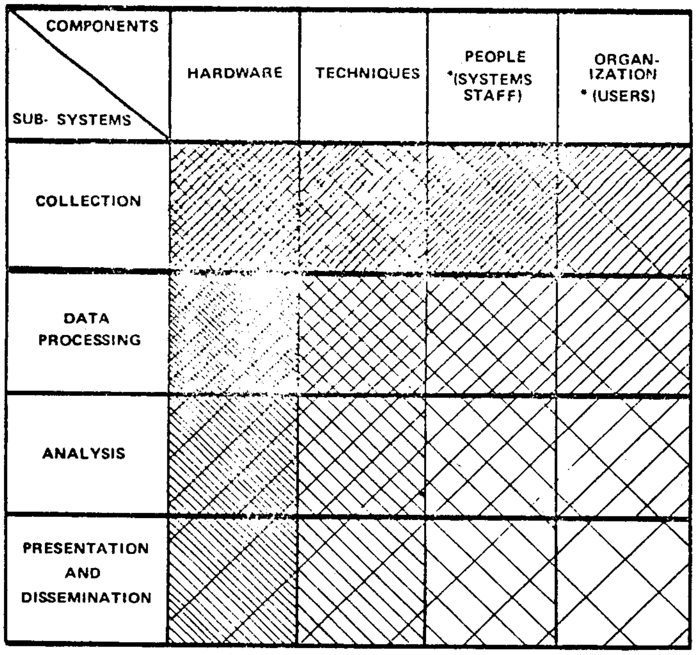


Figure 3. Information Flow Diagram--
System for Earth Resources Information



DEGREE OF EMPHASIS PROPORTIONAL TO INTENSITY OF SHADING

*ADDED BY ERISTAR DESIGN GROUP

Figure 4. Placement of Emphasis in Information Systems Development

It is clear that the driving force in the system design should be the information needs of the users. Data should not be acquired simply "because it is there." Nor should processing be performed only because the capability exists. Yet, an analysis of the users' needs today is likely to be of little value ten years from now. The membership of the user community will certainly change over a period of time, and the needs of a given user will also be modified. The system itself must be able to identify the needs of its current and potential users, and to adapt its structure and operation as conditions change.

A similar conclusion was reached in a report concerning a United Engineering Information Service (Battelle, 1969).

> "There seems to be little choice but to recognize the study of user needs as a continuing requirement. It should, perhaps, be viewed as a monitoring function rather than as a series of studies with definite termination dates."

Thus we can establish some of the requirements for the overall earth resources information system:

1. Adaptability - The system must be able to respond to changes in the user community and in its operating environment.
2. Evolutionary Potential - The system must be able to evolve over time in scope, capability and structure.
3. User Satisfaction - The overwhelming consideration, for which the other requirements are necessary but not sufficient.

The ERISTAR plan proposes to meet these requirements with a decentralized network of centers distributed around the country, each highly responsive to a given user community. The proposed ERISTAR system in its full form is a nationwide network of functionally similar information centers interconnected by a comprehensive communications system. The centers in the network are oriented toward detecting and fulfilling the earth resources information needs of a broad user community. The communications network extends to a complex of subject-oriented information analysis centers, numerical data centers, and to other external data and information acquisition and management systems. The network also includes a set of communications and information technology support facilities. These aid the main ERISTAR Centers in keeping abreast of new technological developments in data acquisition techniques and methods for storage, transformation, analysis and retrieval. The support facilities conduct basic and applied research and development in both hardware and software.

A network system can respond to a changing environment by adding or deleting centers, by altering the communication links, or by shifting the functional emphasis of any center. A modular network system can grow in scope and capability by simply adding additional centers into the network. By having centers in each state, responsive to the state governments, the system can provide services tailored to the local user community. It is anticipated that during the early phases of system operation some of the major users will be state and local planning agencies who can be served better by state and local centers connected into a national network than by a centralized federal system.

The ERISTAR system is organized into a hierarchial structure based upon geographical subdivisions. There are centers at four levels in the hierarchy:

1. National
2. Regional (multi-state)
3. State
4. Local (sub-state)

Not all levels need actually be present in all parts of the system.

Since the complete system is a wide-ranging, ambitious program, a time-phased development plan has been devised which includes three phases:

1. Initial - a single point of contact and referral service for existing systems.
2. Intermediate - a linked network of existing systems.
3. Advanced - complete earth resources information network.

In this way an initial version of the system can be established at once by making use of existing state and federal information programs and establishing simplified state-level centers to provide a referral service. This rather simple creation would immediately make existing programs more accessible to the users. The uncertainty, time delays, and attendant frustration which accompany current efforts by the uninitiated in locating and utilizing existing information programs would be greatly alleviated. At the same time, the cornerstone of a comprehensive, highly capable information system would be laid.

VI. A MODEL SYSTEM AT THE STATE LEVEL

The recent passage of the Land-Use-Policy Act, along with the numerous environmental impact statements required of practically all large-scale projects affecting the quality of the environment, has created considerable pressure at the state level for the development of a comprehensive earth resources information system. In this section, a plan for the implementation of the ERISTAR concept at the state level is discussed in some detail.

Since the ERISTAR design project was housed at the Marshall Space Flight Center (MSFC) in Huntsville, Alabama, a plan was formulated for the development of a model center in Alabama. This plan involves the active participation of MSFC and the Alabama Development Office. Similar plans could be developed for other states which would involve the principal state planning agency and a center of technical expertise within the state.

Figure 5, step 1, shows the model center established at MSFC. A solid line is used to indicate a strong line of communication with the Alabama Development Office and the Alabama Earth Resources Data Committee, a committee established by the Governor to study the uses of ERTS data in Alabama. Solid lines are also shown to represent close liaison with the planning and development groups in other states in the southeastern region, and to other state and federal agencies whose cooperation would be required. Dashed lines are used to indicate lines of communication with state agencies that constitute users and/or sources of earth resources information.

Step 2 of Figure 5 shows the state center fully developed and transferred to state control and administration at Montgomery. The model center remains at MSFC to serve as a test bed for innovation and as a training facility. (In some states the center of technical expertise selected to house the model system may be under state control and located in the capital, which would greatly facilitate the transition from model center to operational state center.)

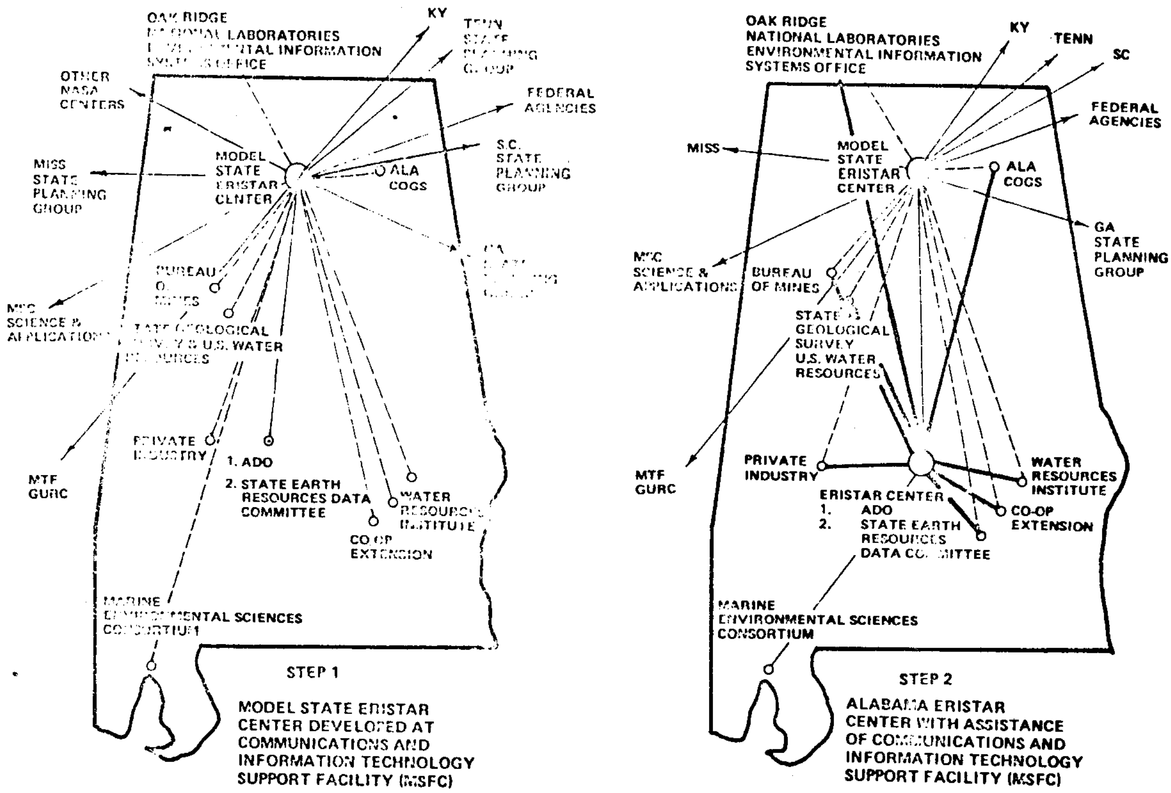


Figure 5. ERISTAR Plan--Model Center and Alabama State Center

Some of the functions of a model center would be to:

1. Serve as a stimulus and catalyst for the development of an operational state center.
2. Provide a focal point for the collection and development of information technology as it applies to an earth resources information network.
3. Establish communications links with sources and users of data and information.
4. Coordinate educational activities on remote sensing, earth resources information processing, and related technology.
5. Establish a system of charges and a marketing mechanism to be used by the operational state center.

By beginning the system with a model center located at a center of technical expertise in the state, then proceeding to an operational state system controlled and administered by the state, the impact and interference of new technology with the daily operations of state government are minimized. At the same time, by having the system develop through implementing the state centers first, then building the network and implementing regional and national centers, first priority is given to making the capability of the system available to one of the most vital areas of our society--state land and resource planning.

Personnel in the Environmental Applications Office of the Marshall Space Flight Center are actively cooperating with the Alabama Development Office in the implementation of the ERISTAR concept at the state level. Although this implementation is still in its very early stages, it is expected that this system will be one of the first comprehensive earth resources information facilities to become fully operational.

VII. REFERENCES

1. Ad Hoc Study Group on the Earth Resources Survey Program, "Report on the ERSP," Washington, D. C.: National Aeronautics and Space Council, March 10, 1971, Vol. III, pp. II-47 through II-54.

2. Carver, K. R. (Editor-in-Chief), "TRIAD: Preliminary Design of an Operational Earth Resources Survey System," Technical Report NASA-CR-106275, 1969.

3. Combs, Z., Trubey, D. K., and Buchanan, J. R., "Directory of Environmental Information Sources, Report ORNL-EIS-71-5, Oak Ridge National Laboratory, 1971.

4. "COSATI Panel on Information Analysis Centers: Directory of Federally Supported Information Analysis Centers," COSATI-70-1, PB 189300, 1970.

5. Hanshaw, B. B., "Resource and Land Information Program Description, Executive Summary," USGS, Department of the Interior, 1972.

6. Hearle, E., "A Data Processing System for State and Local Governments--Revisited," Conference on Information Systems Coordination, Washington, D. C., June, 1968.

7. IBM Corporation, Federal Systems Division, "Space Station Data Flow," IBM Summary Report No. U2W-00213, Electronic Systems Center, Huntsville, Alabama, 1972.

8. Katz, A. M., "Comments on Aircraft and Space Survey System," Useful Applications of Earth-Oriented Satellites, National Academy of Sciences--National Research Council, Vol. 6, 1969.

9. Kruzas, A. T., Ed., Encyclopedia of Information Systems and Services, Ann Arbor, Michigan, Edwards Brothers, 1971.

10. McDonnel Douglas Astronautics Company, "Space Station User's Handbook," MSFC-DPD-235/DR-No. 5E-08, West Huntington Beach, California, 1972.

11. National Academy of Sciences--National Research Council, Useful Applications of Earth-Oriented Satellites, NAS Panel 1 Report, p. 20, Washington, D. C., 1969.

12. Slaughter, Eric, "Environmental Information Network Established," The University in Action, NCSU Extension, Raleigh, North Carolina, June 1973.

13. Vachon, R. I., Skinner, C. W., Gonzalez, R. C., Clark, W. O., et.al., "ERISTAR--Earth Resources Information Storage, Transformation, Analysis, and Retrieval," Technical Report NASA CR-61392, September, 1972.