A SIMPLE INFORMATION MANAGEMENT THEORY

FOR

COMPLEX GEOGRAPHIC DATA

Helmut P. Moyseenko
Assistant to Vice President for Science
The Nature Conservancy
1800 North Kent Street
Arlington, Virginia 22209

I. Introduction:

The Nature Conservancy is a private non-profit conservation organization dedicated to the preservation of ecological diversity. We have already preserved and protected through outright acquisition over 1,500 ecologically significant tracts of land totaling more than 1.5 million acres in 50 states, Canada, and the Caribbean. Some examples of lands preserved by the Conservancy are: the Virginia Coast Reserve (undamaged coastal ecosystem); Bluff Mountain, N.C. (endangered plant species habitat); Big Spring Desert Pupfish Sanctuary, Nevada (endangered fish); and Hutchinson Memorial Forest New Jersey (virgin hardwood forest).

II. Heritage Inventories:

The approach described here has mainly been developed within the Conservancy's state oriented Natural Heritage Programs, although several aspects have been refined through a variety of national inventory efforts conducted by the Conservancy (Jenkins and Moyseenko 1976, Federal Committee on Ecological Reserves 1977). Natural Heritage programs are conducted in cooperation with

agencies of state government, usually under a 1- to 2-year contract. During this period, a systematic process for the collection, management, analysis, and use of pertinent ecological data is established. After the contract period these programs are transferred to the state agency for continuous operation and use. In the interest of brevity and because of the nature of this symposium this report will not discuss all the components of the Heritage program. For discussions on the classification system, inventory and protection planning methodologies, interested readers are referred to Jenkins (1977) and may also contact The Nature Conservancy directly.

III. Grid Cell and Digitized Line Systems:

Over the last few decades, geographical information systems have advanced steadily in technical sophistication. However, despite all the impressive hardware/software advances, it is questionable whether they will ever have the ability to deal effectively with the complex environment, (Voelker 1975).

A problem then in conducting an ecological inventory is to specify a data structure which represents a complex system; as a first step we apply a reductionist approach to classifying the landscape.

Because we cannot afford what we need (i.e. very detailed data), we attempt to inventory through coarser and coarser methods which condense and homogenize the actual landscape into imaginary and unnatural mosaics (grids).

Many states are using the cell-by-cell approach and have found it invaluable in cataloguing a vast amount of data needed for land-use planning. It is essential for them to collect and maintain data on areas with such different land uses as recreation and agriculture. Nonetheless this cellular methodology is not equipped to handle the complexity of ecological data. In the cellular system, a grid composed of cells is superimposed on a total region. The individual cells represent land areas and resolution varies from less than one acre to several hundred acres per cell. The actual ecological makeup of the land coverage for any individual cell is defined by the dominant feature found within the cell. The

cell-by-cell method attempts to catalogue the land areas covered by the grid. It has been our feeling that the rarest ecological features are sometimes lost in the cellular structure because of the gross scale. When the cell size is reduced in its attempt to capture more details, the process may become too expensive and time consuming. Smaller grid size increases the size of the computerized files, but may not necessarily contribute more valuable information.

Other more precise recordation methods exist, among them the digitized line technique. This methodology offers the best resolution but with large amounts of data (e.g. topographic, and soils map), this has proven to be an expensive process, either because the hardware and software requirements are extensive or the manual digitization prohibitively expensive, (Willcott, 1977).

IV. Text and Quantitative Systems:

The objectives here are to centralize, systematize, and store facts usually amassed by site investigations. As in the Grid Cell System and digitized line system, a reductionist approach is also used here to classify the land. Usually site specific, this type of information system must first establish the objectives of the data base and anticipate what kinds of data must be collected in order to meet the objectives. Then a framework must be designed for housing the facts to be recorded, maintained and analyzed. A problem then facing all large scale ecological inventory projects has been how much data can be collected within a specified time and budget versus how much is actually needed for analysis and decision-making. Common sense suggests that a vast majority of ecological data may actually be irrelevant. In the text and quantitative system, we will never be able to collect and manage all the actual microscopic details contained in the real world. Data requirements, day-to-day operations, human factors and systems have their limitations and restrictions (Lee 1973, Lucas 1973, Moyseenko 1977).

V. Our Approach to Managing Ecological Data:

What we have discovered is that whatever strengths most geographic management systems possess, no single computer-mediated program can ever satisfy every purpose. In

designing the Heritage program system, we have found it necessary to design an extensive set of multiple file structures.

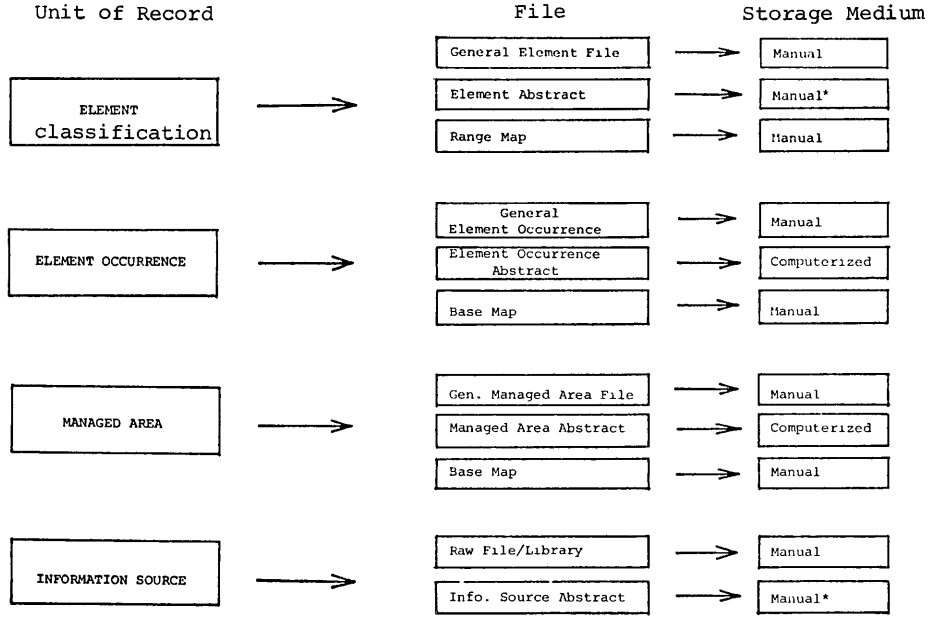
The information system so called is designed to store, retrieve, and combine various kinds of data derived from a variety of sources. For efficiency and cost effectiveness, we have relegated the great preponderance of data to non-automated methods and file structures. No part of the total system needs to be thought of as computerized although we have selectively identified and computerized key data categories for which automated storage and retrieval are especially convenient and useful.

It is our contention that a balanced approach to managing complex ecological data must involve the use of manual files and computerized files. We have therefore designed a hierarchical, balanced information management system. The system is referred to as balanced because it combines manual, map and computer techniques. One way of looking at the information system is in terms of repeating file organizations diagrammed in Table 1. The basic data is applied over and over to four units of record. The units of record are: (1) element; (2) element occurrence; (3) managed area; (4) information source. For each unit of record there exist three basic file structures: a computerized file, a map file and a manual file.

1. Element (the classification): The approach applied to ecological data inventories in Heritage programs focuses primarily on individual features or elements of natural diversity. The keystone of The Nature Conservancy's Heritage inventory approach lies in the critical process of reorganizing site data into the component elements of each site. This aggregation allows us to compare like features of the landscape that occur in different locations. The classification system then gives us the target for which to inventory things and places.

The element unit of record then consists of three components: the general element file, the element abstract and the range maps.

The general element file contains unformated general



*Computerization of these files is being tested.

Table 1: Conceptual Structure of the Balanced Information System.

information on the specific elements--this includes all published and non-published materials which add to and elaborate on the information contained in the element abstract.

The element abstracts contain specific categories of information including description, nomenclature, range, special data management conventions and the like. At present, standardized formats have been created for plant communities, special animal species, and special plant species.

The 7½ minute topographic base map is used for plotting the range of animal and plant species, etc. We have developed coding convetions for plotting and indexing the boundaries, plotting, and cross-referencing the data. These quad maps are organized by the universal map code location system.

2. Element occurrence: The element occurrence unit of record also consists of three components: the general element occurrence file, the element occurrence abstract and the base map.

The general element occurrence file contains information on each element occurrence, including field survey material, pre-existing records, etc. All general background material applying to a specific element occurrence is filed manually here.

The element occurrence abstract incorporates standardized data for analysis of the occurrence, including exact location, number and status of the elements defined by the classification. Each record in the file contains information on a separate element occurrence.

The 7½ minute topographic quadrangle base map is used for plotting element occurrences.

3. Managed Area: The managed area unit of record also includes three files: the general managed area file, the managed area abstracts, and the base map file.

The general managed area file contains individual files aggregated by managing agencies and includes protection information, management plans, and ecological surveys.

A managed area is an area which merits the protection status "preserved" or "protected" under guidelines based on ownership, special status, legal restrictions, etc. Among the managed areas are National Parks, Federal Research Natural Areas, Wilderness Areas, Nature Conservancy preserves, university-owned natural areas, and, in many states, National Wildlife Refuges and State Parks. The scope could be expanded to include less well-protected areas, such as Bureau of Land Management lands, recreational areas, national forests, and wildlife management lands. These areas should be part of the managed area file if reliable information is readily available. Eventually it may be possible to extend our attention to all public lands.

The managed area abstract file contains data on managed areas. Each fact retrieval format provides specific data describing a managed area: its location, size, ownership, the level and means of protection afforded the managed area, and the general characteristics of natural diversity contained on the site.

The same 7.5 minute Map series as for the element occurrence file is used to plot the managed areas.

4. Information source: In order to keep tract of all sources of information, a Raw File and Library and an information source abstract file have been created. In keeping with the data paradigm, the raw file and library maintain undifferentiated, unassimilated data of many types while the information source abstract contains standardized categories of data on individual and other sources of information; the computerization of this file is being tested.

VI. General Description of Our Data Processing Methods

The Nature Conservancy uses a collection of two dozen application-specific software programs to maintain and output its computerized ecological data. Our current record structures are single fixed-length records.

We identify each record by an indexed sequential method which is a technique used to place records alphabetically on disk storage devices. We have also created a X-Y mapping system which works on latitude and longi-

tude data accurate to seconds. The user can specify points on lines and create plots for overlay on any given scale and common geometric projection.

We are currently investigating the advantage of automating more of the files listed on Table 1. To test the effectiveness of a potential nationwide telecommunication network unifying all the Heritage programs and logically interlocking all manual and automated files, we are acquiring a Hewlett Packard System 3000, series 3. This will provide The Nature Conservancy with a more detailed, composite picture of our ecological resources in need of protection nationwide.

VII. Conclusion:

The potential of any Geographic Information System, whether manual or automated, grid based or literature based, are restricted by limiting factors. It is virtually impossible to exercise complete bibliographic or quantitative control of the real world data. Managing ecological data forces us to make some hard choices about which categories of data to emphasize and which to disregard; and no single set of choices can ever satisfy every purpose.

Our Heritage programs have notable advanges over most other ecological inventories. First: we have limited objectives, we deal with only rare entities, and we are not concerned with creating an atlas of information on every acre of land. The great preponderance of data is actually redundant and we make no attempt to tackle such a task. Much of the landscape is highly redundant in the sense that each site contains common species and forest types. A more general point to be made here is that such data clogs up the processes of data collection and processing and wastes money. When conducting ecological inventories one must carefully establish the information specifications and priorities and also place the limits on the amount of data assemblage. Secondly, Our Heritage program inventory process is committed to perpetual operation and this permits us to have a maturing dynamic ecological data base.

Literature Cited

Federal Committee on Ecological Reserves, A directory of Research Natural Areas on Federal lands of the United States of America, published by the Forest Service, USDA (1977).

Lee, D.L., Jr., "Requiem for Large Scale Models." American Institute of Planning Journal, (May 1973)

Lucas, Henry C., Jr., Why Information Systems Fail. Columbia University Press, New York and London, (1973).

Jenkins, R.E., H.P. Moyseenko, Creation of a Data Bank on Established Scientific Ecological Reserves. Final Report for the National Science Foundation. (contract BMS-75-20321). 299 pp. (1976).

Jenkins, R.E. Classification and Inventory for the Perpetuation of Ecological Diversity. The Proceedings of a National Symposium, Classification, Inventory, and Analysis of Fish and Wildlife Habitat, sponsored by the Fish and Wildlife Service, United States Department of the Interior held in Phoenix, Arizona, January 24-27, (1977).

Moyseenko, H.P., Limiting Factors and Pitfalls of Environmental Data Management, the Proceedings of a Symposium on Geographical Data Organization for Rare Plant Conservation, sponsored by the New York Botanical Garden and the National Park Service, held at the New York Botanical Garden, Bronx, New York, November 14-15 (1977).

Willcott, J.C. A Review of Existing Geographical Information Systems and Some Recommendations for Future Systems, The Proceedings of a National Symposium. Classification, Inventory and Analysis of Fish and Wildlife Habitat, sponsored by the Fish and Wildlife Service, United States Department of the Interior held in Phoenix, Arizona. January 24-27 (1977).

Voelker, A.H., "Some Pitfalls of Land Use Model Building." Oak Ridge National Laboratory publication (ORNL-RUS-1), Oak Ridge, Tennessee, 37830 (June 1975).