A DESCRIPTION OF THE TEXAS NATURAL RESOURCES INFORMATION SYSTEM GEOGRAPHIC INFORMATION SYSTEM

by CHARLES TUCKER,
TNRIS SYSTEMS CENTRAL

April, 1976

A Description of the TNRIS Geographic Information System

The basic requirement of the GIS is to provide a unified set of techniques and procedures for dealing with any kind of geographic data. All computer applications using geographic input have certain procedures in common, e.g. digitization, plotting, editing, merging, coordinate transformation, etc. It is the purpose of GIS to provide a set of standard, modular routines to handle these tasks and to avoid the rewriting of similar software for every application. By having a single graphics file structure which is independent of the logical content of any particular geographic file, a high degree of standardization and modularity can be achieved for graphic and/or geographic information processing.

Ordinarily, a digital geographic file is entered into GIS by digitizing (scanning, etc.) all relevant points and lines from a map sheet. No requirement is made of the digitizer operator to record areas as such or to assign any sort of meaning to any configuration of lines; also, he need not worry about aligning the map sheet in any special way. Software has been developed to "register" any digitized map sheet to the geodetic (spherical) coordinate system, provided the scale and projection of the sheet are known, and provided a control network of known points can be established. Another routine has been written to break up digitized lines at points of intersection with other lines. This allows for greater accuracy than can be obtained when a human digitizer operator has to visually detect and digitize such points; such a visual detection procedure also requires that the same point of intersection be digitized two or more times, thus introducing even more error.

Ideally in a geographic file there is no redundancy of graphic information. For example, a file of county boundaries should contain the single boundary between two adjacent counties only once. The direct approach would be to digitize every county individually, storing the complete boundary polygon for a county as a single record. Such a file clearly contains approximately twice as much information (i.e. storage) as necessary, and problems of line matching also tend to be introduced this way. A more efficient and satisfying technique is to define an area (polygon) by listing identifiers of lines which comprise the polygon. This is essentially the technique suggested in the proposed USGS Digital Cartographic Data Base design, as well as other geographic information systems.

The fundamental graphic units in GIS, then, are points and "chains", or lines which intersect each other only at their endpoints. In an initial digitized file, only points and chains are stored. Each record has the same structure: record-id, n, x_1, y_1, x_2, y_2, ..., x_n, y_n.

Each record is assigned a unique identifier which is used in all later processing. A point record requires four words of storage, and a chain record requires 2 (n + 1), where n is the number of points which define the chain. An arbitrary maximum of 2000 points (not words) per chain record was established. Note that in such a file, called an "intermediate graphic file" (IGF), no special distinction is made between points and lines, except for the value of n - if n = 1, the record represents a point, if n > 1, it represents a line (chain).

No attempt is made in GIS to represent lines using polynomial fits or any technique other than straightforward digital approximation. All purely graphic manipulations are done using IGF's. These include editing (graphic editing only), plotting for verification, coordinate transformations, and merging of or selection from graphics files. Before any logic (definition of areas, features, points, etc.) is assigned to a graphics file, it must be "clean", and in a final coordinate system, usually, but not necessarily, geodetic coordinates. GIS has the capability to transform a file between any two map projections defined to the system, as well as to geodetic coordinates. Presently available projections are Mercator, UTM, State Plane, Lambert's Conformal Conic, Alber's Equal Area, and Orthographic. Other projections will be added in the future, as needed.

A three-level coding scheme is used in GIS to attach externally defined labels to graphic data. The top level code represents a "feature" or broad category of data, such as "county", "SMSA", "Stream System", etc. The low level code uniquely identifies an instance of a feature; for example, "Travis county" is a particular item in the feature category "county". The middle level code is used to distinguish different classes within the same feature category. For example, within the feature "island" there may be a need to categorize named and unnamed islands. Each particular island has a unique "item" code (lowest level), but some may have a "class" code indicating unnamed.

A feature code and item code must always be specified. The class code is available as an added measure of flexibility in defining geographic files. Internally a two-word packet is used to store these codes: a whole 36-bit word is used to store an item code, while the class and feature codes are stored in two 18-bit fields. A feature dictionary is established by a user of GIS to relate feature, class and item codes to external, textual labels.

To complete the process of defining a geographic file, a more complex file structure is obviously needed. A final GIS file uses three physical files, all of which reside on a random access disk device, with backup on tape. A system has been developed which catalogs all permanent GIS files and automatically handles transfers between tape and disk, so that a program using GIS files only needs to know the logical names of the files and need not handle the routine chores of setting up and manipulating many different physical files.

The largest, but simplest, component is the chain file. Each chain record, in addition to a list of x, y coordinates, contains feature/class/item codes for the left and right areas bounded by the chain. Inherent in a chain is direction or orientation, so that "left" and "right" are well defined. If a chain is not used in an area definition, but in a line definition, only the "right" feature/item code is specified.

The next component of a GIS file is the definition file. This file contains the definitions of the four data types recognized by GIS: points, lines, areas, and networks. A single GIS file may contain any combination of these four data types.

Point definitions are simplest and require only four words: Feature/class, item, x, y. Point definitions are stored contiguously within the definition file, on fixed-length "pages". All GIS files are structured using an index and paging system that allows for rapid access to any particular record with a minimum of file searching.

A "line" is any logical collection of chains, for example a highway or river. A line definition consists of a feature/class/item code, length of line, number of sub-lines, and lists of pointers to the chain file for the constituent chains which define the line item.

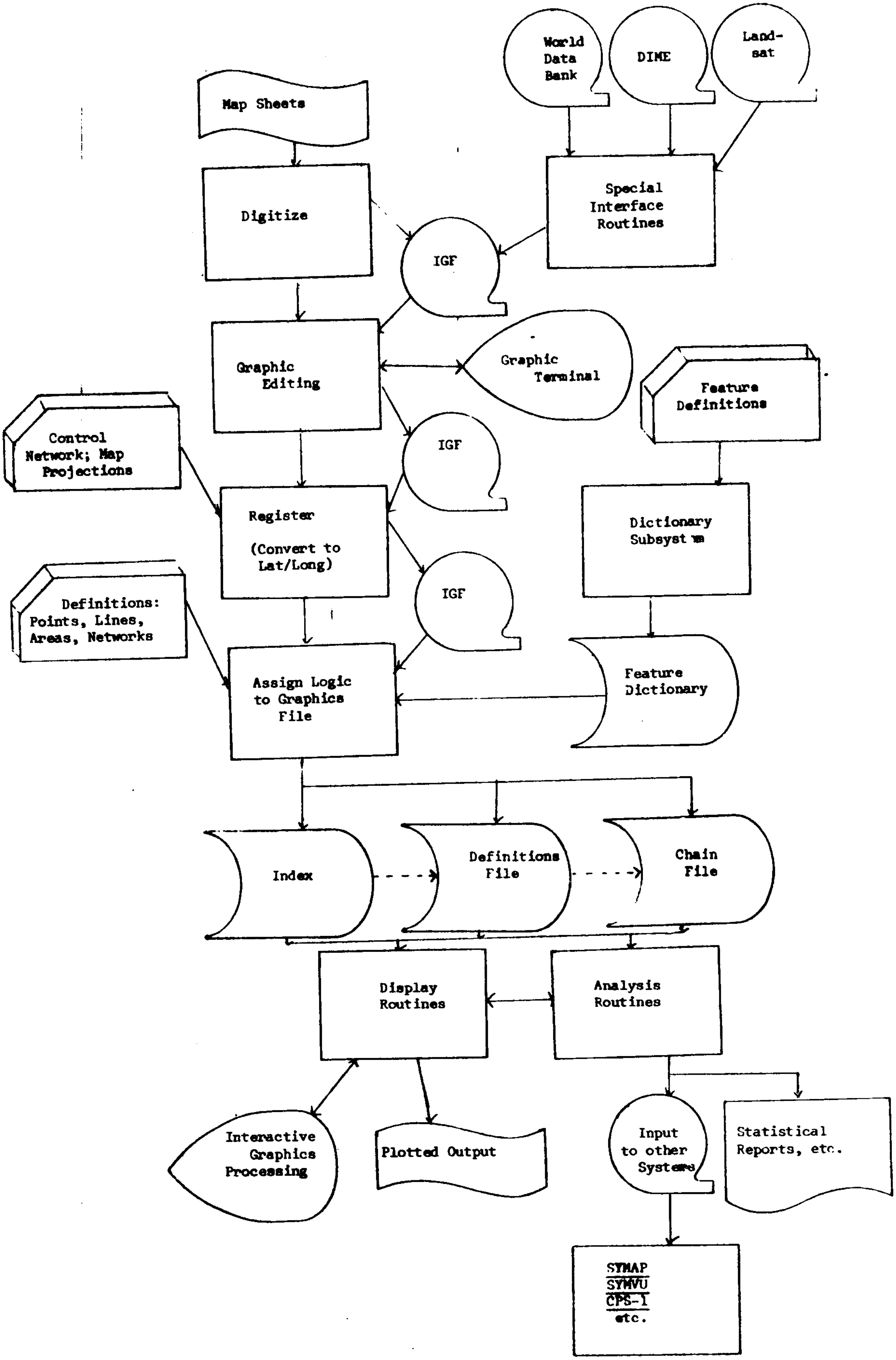
An "area" can be a simple polygon, or it can be a complex set of polygons, some of which can contain "holes". The area definition is similar to the line definition: each area definition has a feature/class/item code, a value for total area enclosed (in hectares), and lists of chain pointers to define all enclosing polygons and holes. A "hole" polygon list always immediately follows its enclosing polygon list, so that no special structure is required to show containment relationships in complicated area definitions.

A "network" definition can be used to represent transportation systems (e.g. highways, railroads, pipelines), river systems, communication networks, etc. Networks require a somewhat more complicated data structure. A network consists of two types of record: edges and nodes. Either type of record may have feature and item codes attached to it to relate the structure to a particular network model. For example, the nodes may be defined as towns and the edges as highways, using class codes to distinguish the various types of highway (interstate, federal, state, county, etc.). Such a model could then be used for analysis of interurban traffic flow, or other similar studies.

All definitions (points, lines, areas, networks) reside in a single physical file. Definitions are retrieved on data "pages", which are indexed within a third physical file, the GIS definition index. Access to a GIS file can therefore be made via three distinct files: 1) the definition index, 2) the definition file, and 3) the chain file. The figure on the next page shows an overview of GIS.

Work now in progress includes routines for automatically defining polygons from an IGF (containing only chains), cartographic plotting, logical combinations of geographic files (thematic overlays), interactive display and analysis, and interfacing with LANDSAT digital imagery files.

OVERVIEW OF GIS



The polygon extraction routine will be used to speed up the process of defining areas to GIS. It will determine all "elementary" polygons created by a set of chains; these polygons can then be used to define more complex areas to the system.

The cartographic plotting capability is a very important component, since it provides a direct interface between GIS and a human user of the system. Maps can be quickly and inexpensively produced by computer according to the exact specifications of the user, which include geographic coverage, scale, projection, and feature selection.

Interactive display and analysis will be available to more sophisticated users who are interested in performing statistical or other mathematical analysis of geographic files or who are interested in simulation or modeling. This aspect of GIS will include the creation of new files by combining existing files, using both the logical content (feature codes) and the geographic content of the input files.

Much interest has been expressed in using LANDSAT digital imagery files and available pattern recognition/classification software to locate and inventory surface water in the state; in particular, the investigation of shallow, transitory "playa" lakes in West Texas will be greatly aided by the use of LANDSAT data and GIS software.

Future additions to GIS will allow for the processing of three dimensional geographic files. Existing software provides for storing a "z-component" in points and chains, but little work has been done on any analytic or manipulative software for three-dimensional files.