GEOGRAPHIC INFORMATION SYSTEM DEVELOPMENTS WITHIN THE U.S. GEOLOGICAL SURVEY

David A. Nystrom
U.S. Geological Survey
521 National Center
Reston, Virginia 22092
U.S.A.

ABSTRACT

A geographic information system (GIS) is a computer hardware and software system designed to collect, manage, analyze, and display spatially referenced data. Geographic information systems are emerging as the major spatial data handling tools for solving complex natural resource planning problems. The use of GIS technology has revolutionary implications for the way the U.S. Geological Survey, State, and other Federal agencies conduct research and present research results. Advanced GIS technologies have the potential to greatly enhance the Survey's ability to perform its traditional missions of earth science data collection, research, and information delivery.

INTRODUCTION

As the lead Federal agency in digital cartography and other earth science topics, the U.S. Geological Survey (USGS) must maintain a knowledge of advanced techniques in all aspects of data processing that relate to geographic information systems (GIS's). Sufficient advanced software and hardware capabilities must be available to earth scientists if GIS technologies are to be used for accomplishing the data collection and research missions of the Geological Survey.

Considerable data are being digitized at numerous locations throughout the country by, or at the request of, local, State, and other Federal agencies. With the proliferation of GIS software and methods for inputting and representing spatial information, the exchange of data from one system to another has become an increasingly difficult problem. Standard digitizing specifications and recommended procedures should be established and implemented to bring uniformity and standardization to these efforts. In addition, methods for transferring data among various earth science data bases and available GIS's must be developed. Successful cooperative projects among USGS divisions and between the USGS and other Federal and State agencies have already been completed. Additional projects

Publication authorized by the Director, U.S. Geological Survey.

are being actively pursued to facilitate transfer of data, knowledge, and capabilities to scientists and managers attempting to apply GIS technology.

At the present time, four USGS divisions are actively involved in GIS activities.

Geologic Division

The Geologic Division (GD) is using GIS capabilities for geologic/geographic information management and computer mapping capabilities. Those capabilities typically include gridding, contouring, feature extraction, overlay, linking locations to attribute files, and flexible output formats using both raster and vector presentations. Those functional capabilities match closely the general requirements for processing geologic data, including the need to manage multiple large data sets having diverse attribute data and the facility to update contour and boundary data as new observations are collected.

Water Resources Division

The Water Resources Divison (WRD) is by far the largest user of GIS's in the Geological Survey. WRD has installed GIS hardware and software in ten locations across the United States and has linked all their facilities together with an Earth Science Distributed Information System called GEONET. WRD applications include analysis and production of their National Water Summary; mapping boundaries for irrigation water rights; support of Ground Water Modeling Studies; and analysis of aquifer and water use data bases.

National Mapping Division

The National Mapping Division (NMD) is mainly concentrating on basic GIS research.

Basic Research

Basic research in GIS technology is being conducted by NMD research personnel in the following areas.

Spatial Data Base Design.--One objective of current NMD digital data production plans is to ensure that the digital spatial data will be responsive to the needs of present and future users, especially within GIS environments. To this end, and to fulfill the responsibility for developing technical standards governing digital spatial data, NMD is pursuing a series of studies on spatial data base content, structuring, and management.

The first task in these studies is to define a comprehensive model of the spatial data of concern. An inventory of existing structures used to represent spatial data is being compiled. These structures will be analyzed with respect to access orientation and specific methods of representing the components of the spatial data model. Further, possible new structures might be developed and analyzed. From existing and new structures, candidate physical implementations for this study will be identified. Feasibility studies will be conducted for each identified candidate physical implementation.

Dynamic restructuring requirements will be investigated, followed by the specification of search/access/retrieval paths and mechanisms (based on topology, space, and attribute data). Benchmark and testing procedures will then be designed to evaluate the ability of the candidate implementations to handle cartographic features and the specified access-related requirements.

Modeling of Cartographic Entities and Objects.--Current research topics include the definition of cartographic objects and their spatial and nonspatial attributes and relationships, as well as the use of these concepts in queries of spatial data. Cartographic objects represent actual features such as rivers, roads, political units, and airport complexes. The spatial attributes include information about the location and orientation of an object. Spatial relationships include both metric (distance) and topological (adjacency and connectivity) aspects of the relationships between features. Since the objects representing the cartographic features are separate from their spatial description, the method developed will allow various methods of encoding space (vector and raster) to be used. Nonspatial attributes include the nonspatial descriptions (such as name, classification attributes, or population) of the encoded feature. Nonspatial relationships include relationships established through the meaning of the object rather than by location of the object. An example includes the relationships between constituent counties of a State.

Based on these objects, attributes, and relationships, methods of querying the spatial data structure are being investigated. These queries include the ability to search the cartographic features (objects) on the basis of metric and relational spatial attributes and relationships, as well as by nonspatial attributes and relationships.

Prototype Spatial Operator Implementation.--Research activities have focused upon identifying fundamental manipulations of spatial elements which, in concert, underscore a wide range of geoprocessing. Generating buffers about features, computing separation, determining overlap or containment, delineating elements of intersection, union, and differencing for pairs of features, and other basic capabilities are represented by a candidate set of spatial operators that may support geographic inquiries of increasing complexity. A prototype vector implementation is under development that attempts to minimize the dependence of the software on any single data structure, and the potential tradeoffs of the approach are being investigated. Additional consideration is being given to the interdependence of operators for spatial entities and common relational operators in a data base environment.

GIS Evaluation.--A number of GIS's have been evaluated as to their functional capabilities. These tests were undertaken so that NMD has a clear understanding of how USGS digital data can be entered, manipulated, analyzed, and outputted by different systems and differences can be intelligently explained to users of our data.

Vector Data Interface.--A systematic approach to vector data interfacing problems, including a definition of basic concepts and development of data structures and prototype software, has been completed. Tests have been run to exchange data between a number of different GIS's.

Information Systems Division

The Information Systems Division (ISD), a supporting organization, is actively assisting the operating divisions in GIS activities. In cooperation with the operating divisions, ISD is working with and assessing the following: advanced computer hardware capabilities like optical disks and concurrent processors; GIS applications on existing microcomputer technology; an Earth Science Information Network to link a series of information data bases with a number of public contact points; the Earth Science Data Directory which provides online access to the Survey's central repository of detailed information about earth science data; and investigation of the feasibility of combining artificial intelligence with existing GIS technology.

COOPERATIVE APPLICATIONS PROJECT

NMD is working with other divisions, bureaus, and Federal and State agencies on many cooperative GIS projects, but perhaps the best example to date of demonstrated GIS capabilities is the USGS/Connecticut GIS project.

In 1984 the U.S. Geological Survey and the Natural Resources Center (NRC), Connecticut Department of Environmental Protection, agreed to jointly develop a demonstration project to evaluate automated GIS technology. A major goal was to demonstrate how a GIS could be incorporated into ongoing data collection, maintenance, and analytical programs of the Natural Resources Center. Of interest was the development of mechanisms for interagency cooperation in the creation, transfer, and use of digital geographic and cartographic data. The GIS Connecticut chose to evaluate was ARC/INFO, a two-part proprietary system of the Environmental Systems Research Institute, Redlands, Calif. USGS had installed ARC/INFO and Connecticut had already been using INFO as a data base management system for several years.

The project was a logical outgrowth of extensive hydrologic, geologic, and topographic inventories that the State and the USGS had compiled in Connecticut over the past 50 years. The project's digital data base and procedures developed for organizing, creating, and verifying it were to serve as a model for eventual implementation of a statewide GIS. This cooperative effort marked a significant commitment by the NRC and the USGS to convert existing mapped Federal and Connecticut data to digital format and to develop techniques to apply these data to environmental and natural resource programs at all levels of government.

The project was interdisciplinary in scope and was developed by a working group with members from the Natural Resources Center and the Geologic, Information Systems, National Mapping, and Water Resources Divisions of the USGS. The primary tasks of this group were to identify the geographic and attribute data to be entered into the GIS; to digitize, edit, and verify the appropriate data sets; and, to select and develop applications that demonstrated the effective use of GIS functions in NRC programs.

Project Area

An area encompassing the Broad Brook and Ellington 7.5-minute quadrangles in north-central Connecticut was selected as the project study area. It includes a wide variety of northern Connecticut landscapes. Going westward from the gently rolling hills typical of eastern Connecticut, the land levels to the broad lowland of the Connecticut Valley. Within the valley, elevations gradually decrease westward in a step-like manner from a series of low, rolling drumlinoid hills, to a nearly level expanse of glaciolacustrine terraces and post-glacial sand dunes, to the lowlying floodplain of the Connecticut River. Forests are the predominant feature of the eastern hills in contrast to the open landscape of the Connecticut Valley, where residential and industrial developments rapidly are changing the nature of the landscape.

Data Base Development

The project data base, listed in table 1, consisted of 28 primary data sets containing geographic features and the associated attribute files for each quadrangle. The data sets included land surface and cultural information and basic data from Connecticut's geologic, biologic, and water resources inventories. The geographic portion of the data base was developed from existing digital data and by digitizing published maps and open-file materials. Attribute data were entered into the system from existing computer files and manually via keyboard entry.

GIS Applications

Four applications scenarios were designed to evaluate the effectiveness of GIS technology in integrating data from a variety of sources for use in research, planning, management, and regulatory programs:

1. Industrial site selection model.

2. Determination of ground-water availability for water utility development.

3. Data base generation for ground-water model.

4. 7-day, 10-year low-flow model.

A brief description of each application follows. In the cases of the industrial siting and the ground-water availability applications, decision criteria were established to determine the GIS functions and procedures to be used. The decision criteria included such things as what land uses are compatible with ground-water development, how far wells should be located from pollution sources, and which slopes would make industrial site development too costly. The project's goal was not to validate the decision criteria that were used, but rather to demonstrate the GIS processes and outputs.

Industrial Site Selection

The State of Connecticut assists businesses considering relocating in the State by identifying the natural resource characteristics and probable limitations of sites available for industrial development. The assessment process requires integrating and analyzing natural resource, cultural, and public facilities data. The State wanted to investigate the use of GIS software in automating this process. A number of selection criteria were formulated and applied to the data base, and sites of high potential were derived. The GIS was used to evaluate all sites that were zoned for industrial use in the two-quadrangle area. A series of steps were performed in which unsuitable areas were eliminated from further consideration. Factors that determined unsuitability included excessive slope, poor soil characteristics, the presence of inland wetlands, seasonal flooding, the presence of sensitive environmental areas or endangered species, inconsistency with the State's water-quality classifications, and insufficient acreage.

Water-supply, sewer, and transportation data were merged with the potential sites to determine the availability of these public services. The product of the assessment was a map identifying potential sites and a tabular summary of the characteristics of each location.

Determination of Ground-Water Availability for Water Utility Development

This application scenario was designed to demonstrate the use of a GIS in State water-supply planning. The identification and protection of potable ground-water supplies suitable for future development is a problem facing many areas of the country. In Connecticut, sites that are used or are proposed for public water supply are protected through the State's water-quality classification program by assigning a classification that precludes the issuance of waste-water discharge permits in the area.

A single water utility was selected for the analysis based on projected need for increased water supply. The GIS software was used to create a one-half mile buffer around the water utility's service area. This buffered area served as the geographic window through which other spatial data sets were analyzed. Areas which were incompatible with ground-water development were removed as potential sites. These included all land uses/land covers other than forested and forested wetlands; areas with poor water-quality classifications; areas within 500 meters of pollution sources; areas within 100 meters of waste receiving streams; areas within 100 meters of existing water supply wells; and areas that were zoned for incompatible land uses. Geologic and hydrologic data were evaluated to further isolate those portions of the remaining suitable areas having coarse-grained materials with greater than 40 feet of saturated thickness. The product was a map depicting all suitable sites greater than 10 acres, overlaid for reference purposes with roads, streams, lakes, and geographic names.

Data Base Generation for Ground-Water Model

This application was designed to demonstrate the use of a GIS in extracting and manipulating data for use in a numerical simulation model of ground-water flow. The NRC uses 2-D and 3-D models to predict impacts from ground-water development within stratified drift aquifers and to identify those aquifers with a high potential for development. Typically, data are manually extracted from specially prepared maps and are manually prepared for input to these models - a very time-consuming task.

"The USGS three-dimensional finite-difference ground-water flow model," USGS Open-File Report 83-875, was used for this application. The Broad Brook subregional drainage basin, lying completely within the two-quadrangle study

Table 1.—Basic data sets for U.S. Geological Survey/ Connecticut GIS Project

Data set	Resolution/ scale	Source	Data type	Density
Land use/land cover	1:24,000	USGS	Polygons	High
Hydrography	1:24,000	USGS	Polygons, lines.	High
Transportation network	1:24,000	USGS	Lines	Med. high
Political boundaries	1:24,000	USGS	Lines	Low
State and Federal lands.	1:24,000	USGS	Lines	Low
Drainage basin boundaries.	1:24,000	USGS	Lines	Medium
Geographic names file	1:24,000	USGS	Points	Low
Digital elevation models.	30-m grid	USGS	Grid	High
Census boundaries	1:24,000	USGS	Lines	Low
Public water supply wells.	1:24,000	NRC[1]	Points	Low
Pollution sources	1:24,000	NRC	Points, polygons.	Low
Sewered area	1:24,000	NRC	Polygons, lines.	Low
Public water supply service areas.	1:24,000	NRC	Polygons, lines.	Medium
Water quality classification.	1:24,000	NRC	Points, lines, polygons.	Medium
Municipal zoning	1:24,000	NRC	Polygons	Low
Biophysical land classification.	1:50,000	NRC	Polygons	Low
Landsat multispectral scanner scenes.	80-m grids	USGS	Grid	High
Surficial materials	1:24,000	USGS/NRC	Polygons	High
National wetlands inventory.	1:24,000	NRC	Polygons, lines.	High
Bedrock elevations	1:24,000	USGS/NRC	Lines	Low
Soils	1:15,840	SCS/NRC[2]	Polygons	Very high
Endangered species	1:24,000	NRC	Polygons, points.	Low
Bedrock geology	1:24,000	USGS/NRC	Polygons	Low
Water table elevations	1:24,000	NRC	Lines	Med. high
100-Year flood zone	1:24,000	NRC/FEMA[3]	Polygons	Med. low
Mining inventory	1:24,000	NRC	Polygons, points.	Low
Water utility lands	1:24,000	NRC	Polygons	Low
Aeromagnetic survey	1:24,000	USGS	Lines	Low

[1]NRC = Natural Resources Center, Connecticut Department of Environmental Protection

[2]SCS = Soil Conservation Service, U.S. Department of Agriculture

[3]FEMA = Federal Emergency Management Agency

area, was selected as the area to be modeled. The model requires 2-D arrays of basic data as input. These include land surface, water table, and bedrock elevations, boundaries between layers one and two of the model area (two layers were used for the Broad Brook basin - stratified drift vs. till and bedrock), location of streams, basin boundaries, and hydraulic conductivity.

The GIS was used to create a gridded "map" of the model area. This consisted of a regularly spaced 200- by 200-m grid overlay of the basin. An INFO file was used to store the attributes of each grid, such as land surface elevation, that would be used as input to the 3-D model. Data were extracted from the INFO file, reformatted into the necessary 2-D arrays, and entered into the model. Model results consisted of 2-D arrays of water table elevations and relative changes at each time step. These were read back into the INFO file, and ARC/INFO plotting software was used to display the results. The use of the GIS significantly reduced the time required to generate the input data arrays for the 3-D model.

Seven-Day, Ten-Year Low-Flow Water Model

The State of Connecticut uses the 7-day, 10-year low flow of streams in water resource planning and management programs. These include evaluating waste-water discharge permit applications, determining minimum release criteria below impoundments, siting waste-water treatment plants, and similar applications. A methodology has been established to calculate the 7-day, 10-year low flow of ungauged streams in Connecticut (Connecticut Water Resources Bulletin no. 34, 1982) that is based on the area of the stream's drainage basin and the area of the basin underlain by coarse-grained stratified drift deposits. At the present time areal data required for estimating 7-day, 10-year low flows is acquired manually.

ARC/INFO, which calculates and stores the areas of all polygons, was used to generate the areal figures necessary to calculate the 7-day, 10-year low flow for the Broad Brook subregional drainage basin. First the surficial materials map was aggregated and reclassified to produce a surficial texture map. The Broad Brook drainage basin was generated from the detailed local drainage areas map and was overlaid with the surficial texture map to calculate the area of the basin underlain by coarse materials. An INFO program was written to use these figures in calculating the low flow value. Using the appropriate data, similar calculations could be performed for any drainage basin in the project area.

CONCLUSIONS

The U.S. Geological Survey is in a unique position because of existing programs, capabilities, and expertise to: (1) conduct research and development and implement appropriate GIS technology to support the bureau's mission to collect, store, retrieve, analyze, and disseminate earth science and other digital spatial data; (2) improve understanding of the nature and management of the Nation's energy, mineral, water, and land resources by conducting innovative interdisciplinary research that promotes application of GIS technology to new areas or activities; and (3) continue development of GIS technology for more efficient and effective future earth science applications.

The USGS/Connecticut GIS Project has allowed the USGS to gain substantial experience in a cooperative applications environment. This experience will enable USGS research personnel to better respond to digital data user needs in the future. Experience in the use of digital line graph (DLG) data was of particular importance. The DLG data structure has been used for a number of years and has been purchased for various types of uses. As USGS gains experience with this data structure in an applications environment, issues related to current types of coding schemes, vertical and horizontal integration of data, accuracy of data, and data distribution formats can be addressed with better insight into user needs.

After the USGS/Connecticut GIS Project, the Director created a Bureau GIS Policy Task Force. In November 1985, the Director signed a GIS Policy Statement which outlined a strategy to build a sound GIS research base, conduct cooperative applications projects, exchange digital data bases, and provide a funding mechanism for encouraging USGS scientists to use this powerful tool in their investigations.

Response to the GIS Policy Statement and competetive funding mechanism has been outstanding. Thirty-five cooperative applications project proposals were submitted to the GIS Policy Task Force for review. Many of those projects have been initiated and the cooperative multidivisional nature of these efforts is having a very positive impact on how Survey scientists go about accomplishing their investigations.