Multiresource Inventories Incorporating GIS, GPS, and Database Management Systems

Loukas G. Arvanitis, Balaji Ramachandran, Daniel P. Brackett, Hesham Abd-El Rasol and Xuesong Du

Abstract.—Large-scale natural resource inventories generate enormous data sets. Their effective handling requires a sophisticated database management system. Such a system must be robust enough to efficiently store large amounts of data and flexible enough to allow users to manipulate a wide variety of information. In a pilot project, related to a multiresource inventory of the Suwannee River Water Management District (SRWMD) holdings in Florida, a new temporal database model will be implemented. This model is a relational one with an enhanced entity relationship between component tables. It consists of a superclass, the spatial database index, and a number of subclass component tables related through a unique identifier. This superclass may be thought of as a central repository, where all the sample locations with their rectangular coordinates are stored. In addition, a unique identifier denoting various aspects of the sampling designs is associated with the coordinates. The global positioning system (GPS) will be used to reference sample locations for periodic updates. The model will be implemented via Oracle's database server. ArcView's Event Theme option will be used to integrate the processed data with existing spatial data sets for further analysis. This approach will have the advantage of working with smaller sets of data as opposed to the traditional geographic information systems (GIS) spatial analysis. The ultimate objective is to provide timely and reliable information for making intelligent decisions.

The State of Florida is divided into five water management districts. As state agencies, the didtricts' mission is to protect streams, freshwater springs, lakes, wetlands, underground water supplies, forests, wildlife, biodiversity, and other natural resources within each district for the benefit of the public. Intelligent management and conservation of these often fragile resources require a multitude of reliable multiresource data in a timely and cost-effective manner. For the past few months, we have been working with the Suwannee River Water Management District in designing an inventory system that will provide the information needed for future planning. The district currently owns and protects close to 100,000 acres (40.5 thousand ha) of river frontage and wetlands, just over 2 percent of the total land area of 4.9 million acres (1.98 million ha) located within its boundaries.

Professor, School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA; Research Assistant, Department of Civil Engineering, University of Florida, Gainesville, FL, USA; GIS Programmer/Analyst, School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA; Visiting Scientist, Department of Geology, Suez Canal University, Ismailia, Egypt; and Graduate Assistant, Department of Computer Information Systems, University of Florida, Gainesville, FL, USA, respectively.

The district has an aggressive program of buying sensitive lands to secure future supply of water while protecting other valuable natural resources in its jurisdiction. Our task is limited to forest and vegetation sampling. However, the database that was developed provides also for wildlife; rare plants; shrubs; recreational, cultural, and geological features; water structures; access points trails; conservation areas; and many other details of the landscape profile and its various ecosystems. A number of powerful tools will be used in conducting this project: GIS, Relational Database Management System, GPS, and Spatial Statistics.

LITERATURE REVIEW

Geographical Information System

GIS has come a long way from the initial simple cartographic tool for automated map production and display. It is viewed now as a powerful component of information systems for spatial analysis of complex knowledge-based systems. Such systems have embraced the spatial component that GIS brings to various organizations. It is an integral part of decisionmaking in various fields ranging from natural resources, environment, and utilities to mining, planning, and many others. As Berry (1995) puts in his article, "Ask not what you can do for GIS, but what GIS can do for you." The potential of GIS is being realized in every field, and its acceptance is increasing.

Before too long, even ordinary persons will use some type of GIS without realizing it as such.

The major challenge facing GIS users is to integrate it effectively with decisionmaking algorithms as well as with numerous existing and emerging technologies, such as external relational database management systems (RDBMS), GPS, and Remote Sensing Imagery. In our case, focus is on the integration of RDBMS to handle descriptive data, GPS to georeference the locations of interest, and geostatistics to model spatial relationship in a multiuser environment (fig. 1).

In a good GIS, about 80 percent of the resources are related to database development and management. The spatial data sets become rather static once created. The frequency of updates is considerably low compared to that of descriptive data. In most multiresource inventories, descriptive data are more dynamic than spatial data. A good GIS design is one that incorporates a powerful database management system for efficient data storage and fast retrieval and manipulation of data.

Relational Database Management System

Database management systems have progressed from simple flat files to hierarchical, network, relational, and object-oriented systems. The most popular and powerful one is the relational database management system. It is based on a simple high-level language, known as Structured Query Language (SQL). A relational model represents the data in a database as a collection of relations. Each relation can be thought of as a table of values. Each row in the table represents a collection of related data values. These values can be interpreted as a fact describing an entity or a relationship instance (Elmasri and Navathe 1989).

Object-oriented database management systems are promising, especially in handling spatial data. However, they require a high-level language such as C++, SmallTalk, or JAVA. Some of the major advantages in using RDBMS, as opposed to the tabular management module that comes with traditional GIS software, are the following:

Data redundancy.—Often in multiresource inventories, large amounts of redundant data are collected. For example, foresters collect data on tree species, diameters, heights, site quality, health, regeneration, groundcover, and others. Similar data may also be collected by ecologists in the same spatial domain. This can be avoided by storing relevant information as individual tables in a central RDBMS and retrieving what is needed for specific studies in a cost-effective manner.

Performance Efficiency.—Studies have been conducted to evaluate the query and update efficiency by linking GIS with an external RDBMS (Sawicki 1995, Hejazi

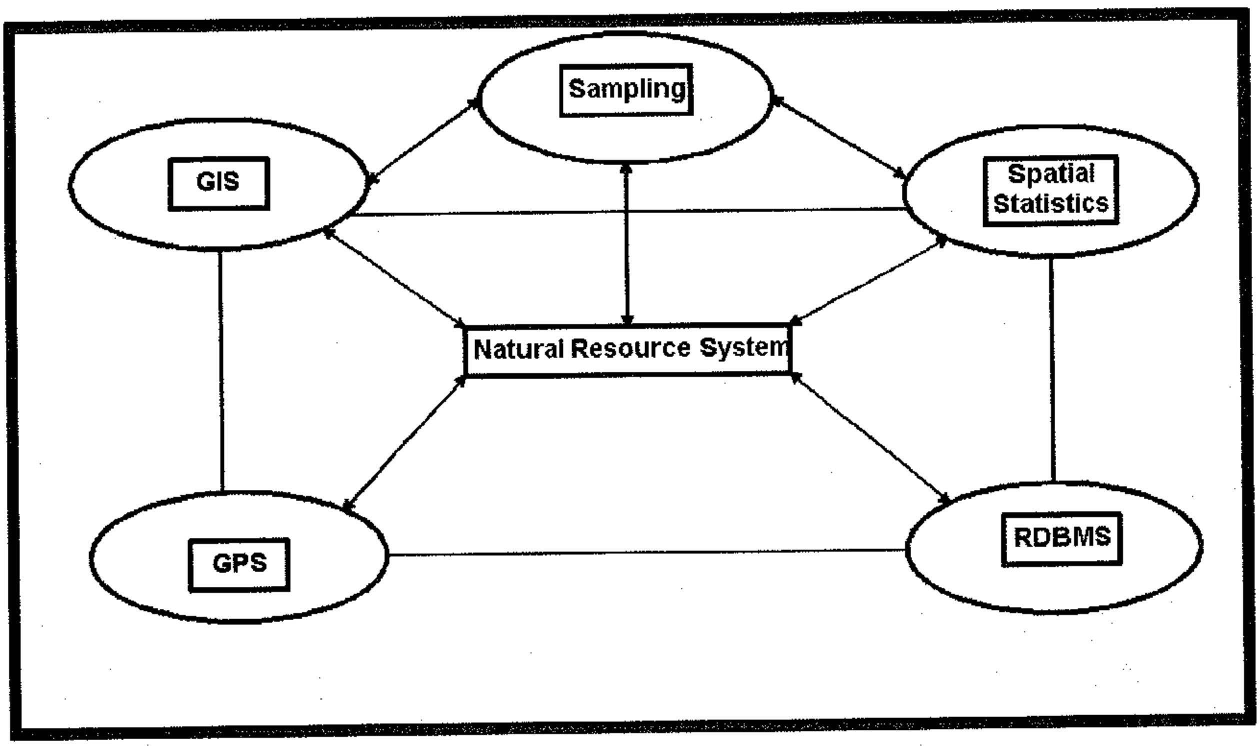

Figure 1.—*Components of a natural resource inventory pilot study for the Suwannee River Water Management District.*

1997, Guan 1997). In these studies, the authors have used external RDBMS as a substitute for the tabular management module in existing GIS software. It has been shown that the database could be managed about 95 percent more efficiently for data entry, update, and querying when stored in an external database (Hejazi 1997). Our model is unique in the sense that the data exchange is bi-directional.

Data Integrity.—SQL allows users to develop databases by employing integrity rules. For example, if one is maintaining an employee database, the Social Security Number (SSN) is a unique value that represents a specific person. SQL will not allow users to enter two records with the same SSN. Similarly, domain constraints, table constraints, column constraints, row constraints, check constraints, system constraints etc., to name but a few, could be applied accordingly to prevent corruption of good data by transcription errors and/or negligence of end users who are populating the database. This kind of rules and constraints also contributes to efficient database administration.

Multiuser environment.—The biggest asset of RDBMS in a network environment is to provide multiuser accessing capabilities. Users can update or retrieve data from the same table simultaneously. RDBMS provide concurrency control mechanisms to handle multiuser transactions. Any organization will benefit by sharing the same information from a single source. This reduces the overhead associated with maintaining databases at each terminal.

Security.—This feature allows selective access to various parts of a database. Some users may be allowed to access certain tables of the database for input only, but may not have update and/or delete privileges. Depending on a user's needs and the organizational hierarchy, individual user authorization profiles can be created. This prevents users from performing operations that corrupt or destroy the data.

Database Backup and Recovery.—Usually, backup refers to copying the whole database on to a disk once a week or at the end of the day, depending on the frequency of transactions. Any failure between these backup periods results in loss of data. With RDBMS, the database can be recreated from the start to the last completed transaction without loss of data.

Client-Server Architecture.—This model allows clients and servers to be placed independently on nodes in a network. It provides the opportunity for cross platform access to the database through PC, NT, or Mac computers over the network.

Cost-effective Application Development.—Since database is based on SQL with integrity rules, the time required to develop new applications is reduced considerably, to about one-sixth of the effort required to develop traditional file system applications (Elmasri and Navathe 1989).

Global Positioning System

The Navstar Global Positioning System (Navstar GPS) was employed in the late sixties by the US Department of Defense (DoD) for strategic military navigation. The system provides accurate, continuous, worldwide, three-dimensional position and velocity information to users (Kaplan 1996). The GPS has three main sectors: space segment, control segment, and user segment. The space segment consists of a constellation of 24 satellites, including three backup satellites, orbiting the earth at high altitudes (around 20,200 km). The control segment is the heart of the GPS system and monitors the space segments. It is operated by DoD from Falcon Air Force Base, Colorado Springs, CO, USA. This control segment monitors the satellites' transmission of navigation messages and sends adjustments, if required. The third component is the user segment where several applications use GPS signals for mapping, navigational, and other tasks. Some of the civilian GPS applications are in the area of agriculture, such as precision farming, transportation for aviation and vehicle tracking, emergency response services, forestry, wildlife, and other natural resource studies. It is estimated that the civilian users outnumber those of the military (Trimble 1998).

In terms of accuracy, GPS can be broadly classified into three categories: survey, mapping, and recreational. The sources of error in GPS signal are selective availability (S/A), obstruction, multipath, and atmospheric delay. The major error component is S/A, which is controlled by DoD to prevent hostile forces from using GPS for malicious purposes.

In 1992, the USDA Forest Service GPS Steering Committee identified more than 130 possible forest applications (Kruczynski and Jasumback 1993). These include area estimation for timber sales, inventories, archaeological sites, red-cockaded woodpecker sites, genetically superior trees, delineation of forest stands, location and navigation to permanent plot points, roads and trails, fire perimeter mapping, wildlife monitoring using GPS based telemetry system, water sources and structures, as well as property corners (Rodgers *et al.* 1996, Evans *et al.* 1992). The biggest obstacle in the implementation of GPS in forestry is canopy interference and mulitpathing of signals. Studies have been conducted under variable forest canopies, terrain, and satellite availability to evaluate GPS accuracies (Deckert and Bolstad 1996). The results from these indicate a postprocessed accuracy ranging from 3 to

6 m, the latter under the worst conditions. Ongoing navigational studies at the Austin Cary Memorial Forest, School of Forest Resources and Conservation, University of Florida, have resulted in a submeter accuracy when navigating to sample points (Brackett. D, personal communication 1998). This can be attributed to the relatively flat terrain, close proximity to Coast Guard real-time differential correction signals, canopy structure, and good mission planning.

METHODOLOGY

Sampling Design.—The elementary sampling unit is a cluster of four circular subplots, recently adopted by the Southern Research Station of the USDA Forest Service (fig. 2). The radius of each subplot is 24 ft (7.32 m), and the distance between subplots is 120 ft (36.6 m). Around the center of each sublot, a microplot of 6.8 ft (2.07 m) in radius will be established. All trees of 5.0 in. (12.7 cm) or larger dbh within each subplot will be measured. In each microplot, all live trees between 1.0 and 4.9 in. (2.54 and 12.69 cm) in dbh, as well as the four dominant seedlings will also be measured.

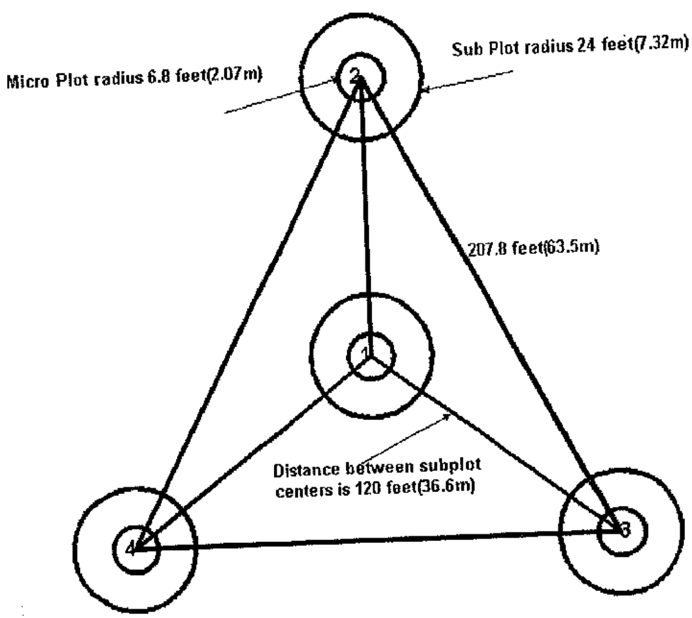

Figure 2.—*Plot configuration.*

Individual clusters will be located systematically along transect lines. To minimize the adverse effects of periodic variation, the center of each cluster will be located along a random azimuth with a differentially corrected GPS receiver at a fixed distance from the transect line (fig. 3). This approach will also facilitate subsequent spatial predictions (kriging). The estimated from pre-sampling average basal area per acre for all trees larger than 5.0 inches dbh and its corresponding variance will be used to compute the required sample size, assuming a 0.95 confidence coefficient and an allowable error of estimate

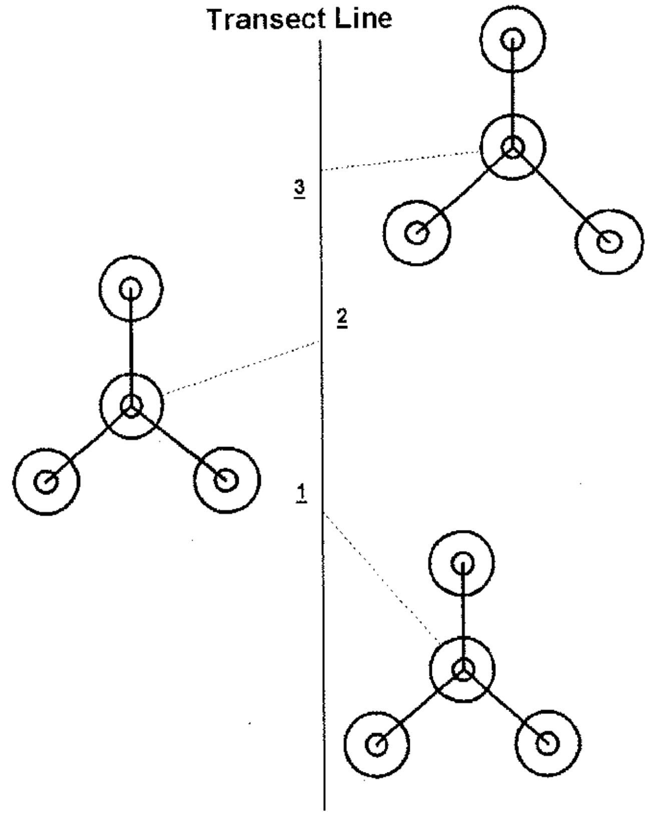

Figure 3.—*Allocation of clusters.*

±0.10 of the mean. These were the requirements set by the SRWMD, following numerous discussions. Obviously, the precision of the rest of the many variables of interest will vary considerably.

Percent ground cover by vegetation and its species composition will be estimated from four 50- x 50-cm quadrats located systematically on a random strip, 50 cm wide and 2.07 m long, through the center of the four microplots of each cluster (fig. 4).

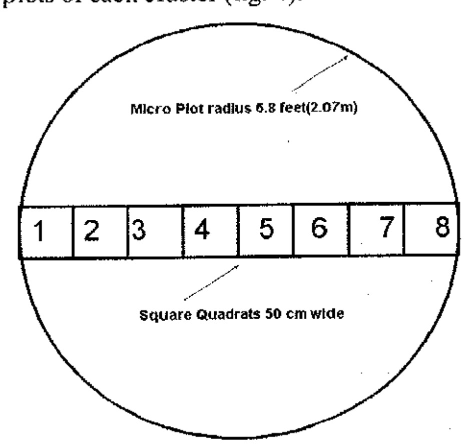

Figure 4.—*Ground cover sampling design.*

System Implementation.—The architecture of the database model is depicted by figure 5. On the server side, Oracle

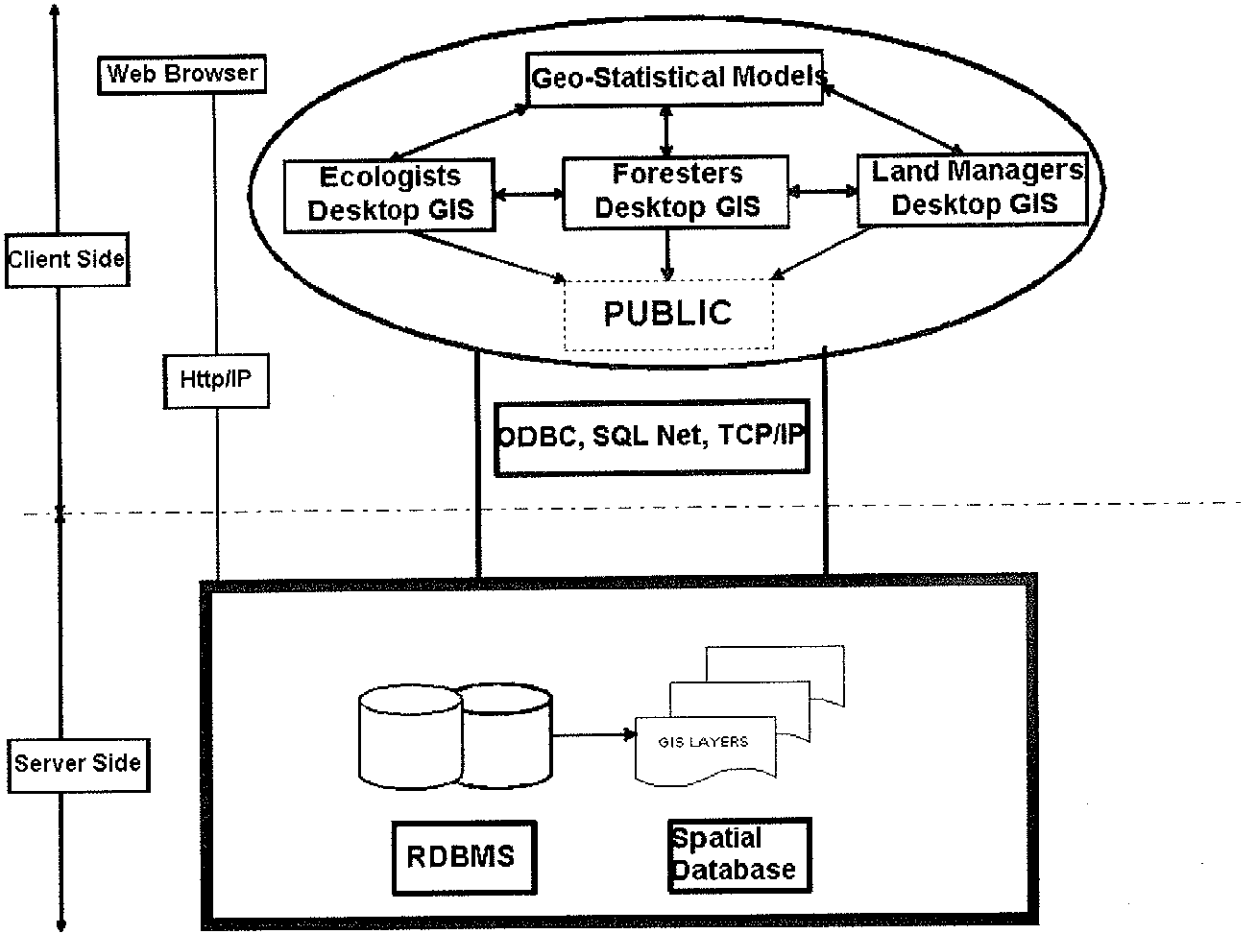

Figure 5.—*Natural resource inventory: System implementation.*

Corporation's *Oracle7* workgroup server for SPARC Solaris and the Environmental Systems Research Institute (ESRI) *Arc/Info* Librarian are used. On the client side, ESRI's desktop GIS, *ArcView* version 3.0a, and corresponding applicable models will be installed in a PC network environment.

The client, the *ArcView* application in this case, is allowed to access data in an *Oracle7* database through the Open Database Connectivity (ODBC) interface. ODBC basically consists of a driver manager and a set of ODBC drivers that enable applications to access data using SQL as a standard language. It communicates with the network through *SQL*Net*, which is an Oracle product that works with the Oracle server and enables two or more computers that run the Oracle server to exchange data through a third-party network. *SQL*Net* is independent of the communications protocol, such as TCP/IP or Microsoft LAN Manager. Users can interface it with many network environments. Similarly, the centralized database can be accessed using web browsers through the Internet protocol.

In this multiresource inventory, a wide variety of entities need to be captured. It is a real challenge to develop a database that incorporates diverse entity types ranging from static features like water structures and access points to highly dynamic natural resource data. The proposed temporal database model is a relational one with enhanced entity relationships between component tables (fig. 6). It consists of a superclass, called spatial database index, and a number of subclass component tables related through a unique identifier. The superclass may be thought of as a central repository, storing all the sample locations with their rectangular coordinates for entities associated with sampling. Entities that are linear or area features like trails or water structures will have only a unique identifier without coordinates. This unique identifier will be later used to join with its GIS coverage identifier for analysis. The spatial database index serves as a link between the spatial domain and the descriptive data associated with the domain. A field name 'Type' will be added to this table, which keeps track of the entity type like wildlife, rare-plant component, water structures, and others.

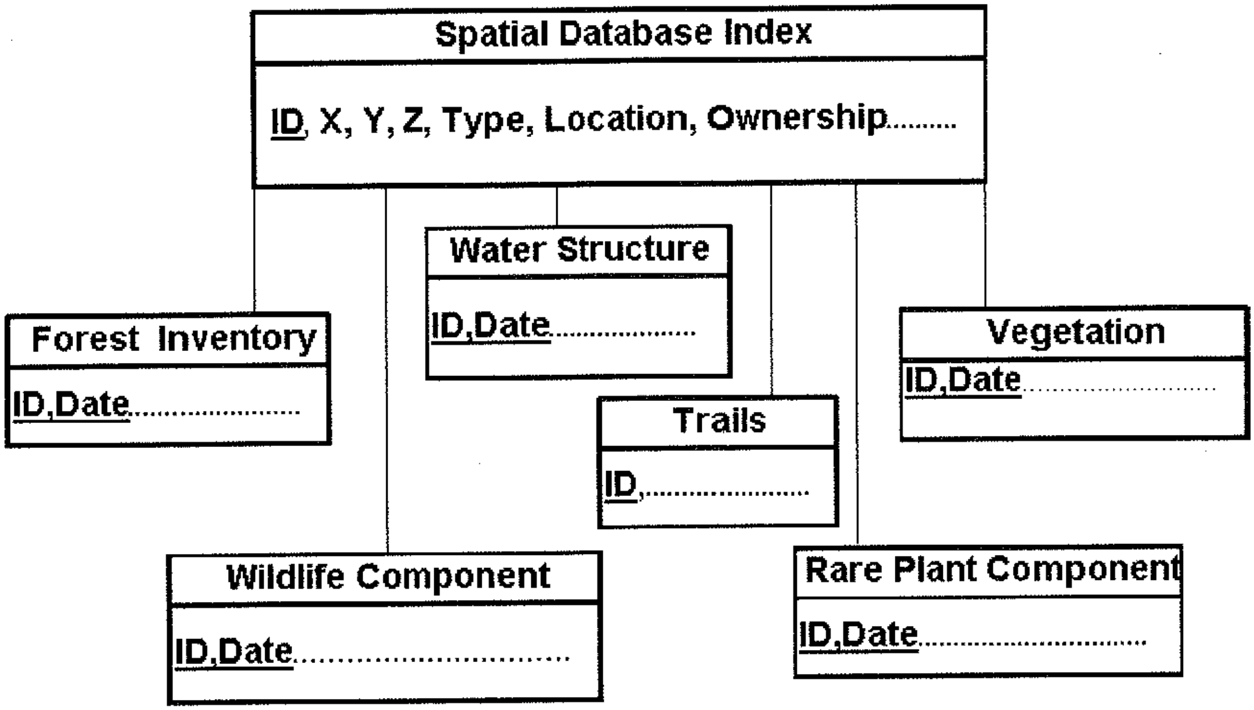

Figure 6.—*Partial entitiy relationship diagram showing the database structure.*

The temporal nature of the database is implemented by adding a date attribute as a primary key in the required subclass component tables. Thus, the inventory-related entities associated with sample points will be revisited using GPS for periodic updates. As an example, each individual tree with a diameter greater than 10 cm at breast height is stored for trend analysis and modeling. In this phase of the project, the temporality of the database consists entirely of snapshots. Any changes that occur between two time snapshots, if inferred through satellite imagery, aerial photography, or other means, will be properly accounted for in the models. Attempts are being made to model this database to make it truly temporal in the future.

Once the appropriate number of sample clusters is determined, a systematic grid will be superimposed on the pilot study area using *ArcInfo*. A program is written to generate sample clusters, which will be systematically located along transect lines at random azimuth and constant distance from transects. The coordinates of the center of each cluster will be used to monument the plot centers using GPS. Later, the coordinates will be stored as waypoints in the GPS units for field personnel to navigate to the cluster center. Other entities like water structures, access points, and archaeological features will also be georeferenced with GPS.

The final phase of this project will be to integrate processed data from the database with various spatial statistical models and for spatial analysis in *ArcView*. There are many options for accessing the database.

Novice users can employ the customized *ArcView* application, which uses SQL connect option to activate pre-compiled stored procedures and SQL scripts in the Oracle database. This would allow users to perform predefined summaries, reports, and other spatial analysis functions.

Users with advanced knowledge will be given the option of changing parameters to fit their own analysis objectives. For example, an ecologist may query the main spatial database index table in the field 'Type' just for wildlife and ground cover. With *ArcView's* event theme option, the resultant table from the query may be added to a point theme for further analysis with the existing spatial datasets like soils, hydrology, topography, roads, and others. Subsequently, foresters may debate on the use of a management action, such as prescribed burn, and assess its anticipated impact on wildlife in the area of interest. Similarly, the user can bring the whole Spatial Database Index table as a theme and drag a bounding rectangle or select a polygon from the existing base themes to obtain the points falling within his/her selection. Based on the selection, other information could be retrieved for those points from the Oracle database. The same methodology applies while running the models; they could be run either for the whole area or just for the data falling within oneís area of interest.

There is a concern about integrating highly accurate GPS information with existing GIS coverages that vary in accuracy. In this project, most of the coverages are generated from digital orthoquads (1-m resolution) with accuracy comparable to the one obtained by GPS in forested areas. In cases where it is required to use data from other sources the different accuracies will be dealt properly in the models and reporting results from such an analysis.

In summary, all indications suggest that future forest and other natural resource inventories will increasingly employ some type of a relational database. In addition, use of existing and emerging new technologies, including high-resolution remote sensing imagery, will continue to play an important role in providing timely, reliable, and cost-effective information for policy formulation and intelligent decisionmaking.

ACKNOWLEDGEMENTS

We gratefully acknowledge the cooperation of Charles H. Houder , III, and Robert Heeke of the Suwannee River Water Management District, Live Oak, FL, USA.

LITERATURE CITED

Berry, J.K. 1995. Spatial reasoning for effective GIS; with a collection of mathematical formulae by Nigel Waters. Fort Collins, CO: GIS World Books. 208 p.

Deckert. C.J.; Bolstad, P.V. 1996. Global positioning system(GPS) accuracies in eastern U.S. deciduous and conifer forests. Southern Journal of Applied Forestry. 20(2): 81-84.

Elmasri, R.; Navathe, S.B. 1989. Fundamentals of database systems. Redwood City, CA: The Benjamin/ Cummings Publishing Company, Inc. 802 p.

Evans, D.L.; Carraway, R.W.; Simmons, G.T. 1992. Use of global positioning system (GPS) for forest plot location. Southern Journal of Applied Forestry. 16(2): 67-70.

Guan, W. 1997. Evaluating ARC-to-Oracle linkages. Geo Info Systems. 7(2): 38-44.

Hejazi, F. 1997. Implementation of ORACLE in an ARC/ INFO-ARCVIEW GIS environment. In: ESRI's user conference proceedings; 1997 July 8-11; Redlands, CA. Environmental Systems Research Institute Inc. http://www.esri.com/library/userconf/proc97/ prochome.html

Kaplan, E.D. 1996. Understanding GPS principles and applications, mobile communication series. Boston, MA: Artech House, Inc. 554 p.

Kruczynski, L.R.; Jasumback, A. 1993. Forestry management applications, Forest Service experiences with GPS. Journal of Forestry. 91(8): 20-22.

Rodgers, A.R.; Rempel, R.S.; Abraham, K.F. 1996. A GPS-based telemetry system. Wildlife Society Bulletin. 24(3): 559-566.

Sawicki, J. 1995. ARC/INFO and ORACLE: an approach for feature storage and representation from oracle. In: ESRI's user conference proceedings; 1995 May 22-25; Redlands, CA. Environmental Systems Research Institute, Inc. http://www.esri.com/library/ userconf/proc95/prochome.html

Trimble. 1998. Surveying and mapping product training, ProXR/XRS with pathfinder office classroom manual. Sunnyvale, CA: Trimble Navigation Limited. 210 p.