112191

Design and Implementation of a Land Ownership Database for a GIS/LIS at County Level[1]

An Extended Entity-Relationship and the relational models were used to design and implement a rural land ownership database, which is part of GIS/LIS.

Authors: *R. Norberto Fernández[2], Marek Rusinkiewicz[3], Lucia Morais da Silva[4], and Chris J. Johannsen[2]*

[2]Purdue University, Laboratory for Applications of Remote Sensing (LARS), 214 ENTM Building, West Lafayette, Indiana 47907.

[3]University of Houston, Department of Computer Science, 4800 Calhoun Road, Houston, Texas 77004.

[4]Purdue University, Department of Agronomy, Lilly Hall of Life Sciences, West Lafayette, Indiana 47907.

Submitted to: *Photogrammetric Engineering and Remote Sensing*

[1] Contribution from the Agronomy Dep., Purdue Univ. Agric. Exp. Stn., West Lafayette, IN 47907. Journal paper no. 12883.

ABSTRACT: The Laboratory for Applications of Remote Sensing (LARS) at Purdue University is conducting a pilot project in Miami County, Indiana, with the purpose of developing a geographic information system/land information system (GIS/LIS) to solve rural needs, particularly agricultural reassessment. Land ownership is one of the basic layers of this GIS/LIS. As the amount of data to be included in a database, and the number of users of the database increase, the need for a structured approach to database design arises in order to ensure an efficient data processing. We have designed a rural land ownership database using a conceptual model, the Extended Entity-Relationship (EER) model, and implemented it in a microcomputer using the relational model. The EER model was an effective design tool that permitted several modifications during the design process, and is capable of accommodating future changes in the database without substantial modifications of the basic design. The relational model was adequate for the implementation of this database. Several programs were written to allow a user, with minimum knowledge in computers, to perform different operations on the database such as data input, record update, database query and record delete. These interactive programs were organized from a main menu, and with different alternatives for each option. Appropriate database design is necessary to provide data integrity and consistency, as well as good database performance.

INTRODUCTION

Local governments are under increasing pressure to become more efficient in handling their usual operations and services. Recent developments in hardware and software technology allow users to analyze large quantities of data and to execute routine operations in relatively short time. These operations would be, otherwise, time-consuming and prone to error. Spatial data can now be combined with descriptive databases within a geographic information system (GIS) environment for planning, management and modeling purposes. GIS technology is an effective way of handling and manipulating large amounts of spatially-referenced and descriptive data for rural resources management (Niemann Jr. *et al.*, 1987; Ventura *et al.*, 1988; Ventura, 1990).

Land ownership is one of the basic layers of any GIS for rural programs. The importance of land property information has been stated by the National Research Council (1980) when defining the basic concept of a multipurpose cadastre as a system able to provide a comprehensive land-related information at parcel level.

As the amount of data to be included in a database, and the number of users of the database increase, the need for a structured approach to database design arises. The goals of the database design process are to ensure an efficient data processing through the elimination of redundant information, and the minimization of update and deletion problems (Jackson, 1988).

The objective of this study was to design a rural land ownership database using a conceptual model, the Extended Entity-Relationship model, and to implement it in a microcomputer using the relational model. This database, which is part of a land information system (LIS), is intended to be used by county officials in tax assessment of agricultural lands.

MOTIVATION

The Laboratory for Applications of Remote Sensing (LARS) at Purdue University is conducting a pilot project in Miami County, Indiana, with the purpose of developing a GIS/LIS to solve rural needs, particularly agricultural reassessment. Layers of this GIS/LIS include rural land property, soils, land cover/land use, roads and surface hydrology. Descriptive attributes for each layer are stored in relational databases (Johannsen *et al.*, 1990).

Currently, the information on rural land property is kept in a microcomputer and managed with a commercial database management system (DBMS). All data are included in one table. This approach to database design is known as the universal relation. The universal relation has all the attributes placed into one relation which could store additional data in a future time (Jackson, 1988; Table 1). Although this approach looks like a straightforward methodology for database design, several problems were identified with this database. Among the most obvious are: the database contains redundant information; for example, the attributes SEC-CODE and SECTION have exactly the same data. Data redundancy creates a problem during update operations because of the possibility of modifying only part of the data. Also, repetition of data increases the volume of data to be handle by the DBMS with a consequent decrease in processing speed; and increases the amount of storage requirements.

The attribute PARCEL-ID repeats the information contained in SECTION, TOWNSHIP, RANGE, and PARCEL-NUM. The attributes ADDITION, LOT, and BLOCK are used to describe land property in urban areas, but they are not necessary for rural parcels. Normally, no data are input in these fields (Table 1). The fields designed for the owners' names and address (LAST-NAME, FIRST-NAME, and ADDRESS) do not provide with the flexibility that is necessary for efficient queries. For example, only one last name can be input, and no distinction between owners' names (FIRST-NAME) can be done. This database cannot keep separate records for each owner without a substantial repetition of the information. Overall, whenever a query is executed the DBMS has to search the entire database in order to retrieve the information requested by the user. This results in an increased search time.

THE DATABASE DESIGN PROCESS

Database design is the development of the structure of the database, the definition of its contents, and the validity of the data which are to be placed in it (Marble, 1988). The goals of database design are: to store all pertinent data in the database in order to satisfy the requirements of the users; to eliminate redundant data; to provide a way to understand the organization of the data; and to support processing requirements and performance objectives (Jackson, 1988; ElMasri and Navathe, 1989).

The typical steps in the design process are: *a) requirements collection and analysis*, which involves the identification of all users and applications in order to formulate all data and processing requirements; *b) conceptual design*, where a conceptual schema or interpretation of users needs is attained (This step requires a good understanding of the users applications, and it is best achieved by applying a high-level data model which is independent of the DBMS where the database will be implemented); *c)* the *data model mapping* involves the mapping of the conceptual schema into the data model of the DBMS (This phase is DBMS dependant); *d)* the *physical design* relates to the process of selecting specific storage structures and paths, based on the options offered by the DBMS, in order to achieve good efficiency; and *e) implementation* of the database (Navathe and Schkolnick, 1978; ElMasri and Navathe, 1989) (Figure 1).

The design process may require modifications of an early phase while working on a later phase. This is an iterative process that will loop back as many times as needed (ElMasri and Navathe, 1989).

CONCEPTUAL DESIGN: THE ENTITY-RELATIONSHIP MODEL

In order to remove the anomalies found with the current implementation of the database, we have redesigned the land ownership database using a conceptual model, the Entity-Relationship (ER) model, and implemented it under the relational model. The ER model is a high-level data model which describes the elements of a database in terms of *entities*, things or objects in the real world with an independent existence; *relationships*, which describe the associations or linkages among entities; and *attributes*, which are the properties that describe the entities or the relationships (Chen, 1977). There are two kinds of attributes: *identifiers* (or key attribute/s), which are used to uniquely identify each entity; and *descriptors*, used to describe each entity. *Weak entities* are entities that do not have key attributes of their own; and they are related to some other specific entities by *weak relationships* (Chen, 1977).

The ER model is closer to the user's perception of data and applications; it is independent of the DBMS to be used for the implementation of the database (ElMasri and Navathe, 1989); and it provides flexibility for modifications during the design process (Marble, 1988). The ER model can be graphically expressed by ER diagrams, where entities, relationships, and attributes are represented with different geometric forms (Figure 2).

Several modifications and extensions to the original ER model has been proposed in order to accommodate new abstraction concepts that are needed for more complex databases (Navathe and

Cheng, 1983, ElMasri *et al.*, 1985, Navathe *et al.*, 1986). These modifications and extensions are included in the Enhanced-ER (EER) model, and their graphic representations in the EER diagram. We have used the EER model to design this database because of its flexibility to incorporate various forms of subclasses and superclasses, as well as generalizations and specializations.

THE EER SCHEMA FOR THE MIAMI COUNTY LAND OWNERSHIP DATABASE

Location and description of property in the State of Indiana is done with the U.S. Rectangular Survey System (McEntyre, 1978). Any location can be defined with reference to two survey lines: an east-west "base line, and a north-south "principal meridian". A *survey township* is an area of approximately 36-square-mile (93.2 km^2) within a set of survey lines. Each survey township is divided in 36 *sections*, where each section has an area of, approximately, 1-square-mile (2.6 km^2) and contains 640 acres (259 ha). Each section has an unique designation based on section number and township identification (McEntyre, 1978; Steinhardt and Franzmeier, 1981). A *civil township* is a political and arbitrary division of a county. The number of civil townships per county is variable, as well as their areas. Survey and civil townships are not related.

The following specifications and assumptions were considered for the design: the database should contain information about parcel location within a section and within both a survey and a civil township. A survey township can have a minimum of one and a maximum of 36 sections; while a civil township can have a minimum of one and a variable maximum number of sections. A parcel can be owned by one individual, a partnership, a corporation, an organization, or it could be temporarily administrated by the State or Federal Government. The database had to keep information about real estate transactions, land records, and future tax coding system. Finally, this database had to be linked to the spatial database of the GIS/LIS for interactive queries.

After several iterations, we produced an EER schema for this database, that considers the users' needs (Figure 3). Participation constrains on relationship sets are represented by an integer pair *min:max* on each participating entity set. The value *min* gives the minimum number of relationship instances in which an entity of the participating entity set must be included, while the value *max* gives the maximum number. These participation constraints are more general than the cardinality constraints used with basic ER diagrams to indicate the type of the relationship set, i.e. *one-to-one*, *one-to-many*, or *many-to-many* (Czejdo *et al.*, 1990)

The entity type MAP-POLYGON represents all polygons on the digitized property map. The attribute <u>Polygon-ID</u> is unique for each polygon, and it is the link between the spatial database and the attribute database. Each instance of MAP-POLYGON contains information about a parcel, as indicated by the set of relationships REPRESENTS, where each instance relates one entity from each of the participating entity sets (**MAP-POLYGON** and **PARCEL**). The participation constraints for this relationship imply that a **MAP-POLYGON** entity participates once in the relationship (1:1), i.e. one polygon contains information about one parcel, and the information of only one parcel is contained in each polygon.

To identify a parcel it is necessary to know first the section, township and range in which the parcel is located; this is indicated by the weak entity set **PARCEL** which takes the key attributes from the entity sets **SECTION** and **SURVEY-TOWNSHIP**. In term, the entity set **SECTION** takes the key attributes of **SURVEY-TOWNSHIP** (Figure 3).

The participation constrains for the relationship type **IS_IN** imply that a **PARCEL** entity participates one or more times in the relationship (1:N), i.e. one or more parcels are in one section; whereas a **SECTION** entity participates exactly once in the relationship, i.e. only one section contains those parcels. Similar constraints apply for the relationship type **IS_LOCATED_IN**. On the other hand, the constraints for the relationship type **IS_INCLUDED_IN** (1:N and 1:M) indicate that a civil township might have a variable number of sections in it; and, a section can be part of one or more civil townships.

We have defined the category **OWNER** to represent the different kinds of ownership that are known to this database. Therefore, **OWNER** is a subclass of the union of the superclasses **INDIVIDUAL**, **PARTNERSHIP**, **CORPORATION, ORGANIZATION** and **ESTATE**, as denoted by a circle with the **U** symbol (set union operation) in it. This implies that an entity that is a member of OWNER must exist in at least one of the superclasses, but does not have to be a member of all of them (Elmasri and Navathe, 1989). The category OWNER is total because every member of the superclasses must be a member of OWNER. This type of generalization/specialization is denoted by connecting the entity set OWNER to one side of the circled-union symbol with a double line, and by joining arcs emanating from each subclass to the other side of the circled-union symbol (Figure 3).

The situation where a "partnership" is the proprietary of land is indicated first by the attribute Percent in the relationship type OWNED_BY. Members of that partnership are represented by the PERSON entity set, where each instance relates to PARTNERSHIP through the relationship type PARTICIPATES. Each instance of PERSON also represents those people that own land individually, as indicated by the superclass INDIVIDUAL.

The relationship type OWNED_BY holds attributes that are descriptive of the transaction process, such as transaction date, deeds registration number and name of the parcel's former owner(s). The primary key of OWNED_BY is the combination of Parcel-ID and Owner-ID, which are the primary keys of the participating entity sets PARCEL and OWNER (Figure 3).

MAPPING OF THE EER SCHEMA INTO THE RELATIONAL MODEL

The relational model, first introduced by Codd (1970), represents data as a collection of relations. Informally, a relational database is perceived by the user as a collection of tables, and nothing but tables (Date, 1988). All data in the database are then strictly organized in tables, and all database operations work on these tables. The relational model provides flexibility for model implementation and system development (Armstrong and Denshman, 1990). For a formal definition of the model the reader is referred to Codd (1970) and Date (1988).

We have derived a relational schema from the conceptual schema shown in Figure 3, according to existing rules for conceptual-relational mapping (ElMasri and Navathe, 1989). In general, for each entity set and relationship type with attributes of the EER schema, we have created a relation that includes all the attributes of the original entity or relationship. The following initial relations were generated from the EER diagram:

MAP-POLYGON (Polygon_ID, Area, Perimeter)
SURVEY-TOWNSHIP (Township, Range)
CIVIL-TOWNSHIP (Township Name, Unit #, Tax_Rate)
SECTION (Section#, Township, Range, Area)
PARCEL (Parcel-ID, Section#, Township, Range, Q_Section, QQ_Section, D_Acres, Old_Tax#, New_Tax#, Zoning)
OWNED_BY (Parcel-ID, Owner-ID, Percent, T_Date, D_Record, D_From)
OWNER (Owner-ID)
PERSON (SS#, Name, Address)
CORPORATION (CName, CAdrress)
ORGANIZATION (OName, OAddress, RTaxable)
ESTATE (EName, EAddress, ExName, Status)

Three relations can be dropped from this initial set: *SURVEY-TOWNSHIP* and *SECTION* hold no useful information because their attributes are key attributes that are, also, part of *PARCEL* as a whole. The exception is the attribute Area in the *SECTION* relation, which represents the area of each section. But, this information can be easily obtained, on request, by adding the areas of the individual parcels that are in each section, through the *PARCEL* relation. Similarly, the *OWNER* relation contains only one key attribute, Owner-ID, which can be included as a foreign key in the relations *PERSON*, *CORPORATION*, *ORGANIZATION*, and *ESTATE* as a link to *PARCEL* through *OWNED_BY* (see Figure 3).

Dropping these relations, the final set of relations with their attributes is:

MAP-POLYGON (Polygon ID, Area, Perimeter, Polygon#, $Recno)
CIVIL-TOWNSHIP (Township Name, Unit #, Tax_Rate)
PARCEL (Parcel-ID, Section#, Township, Range, Q_Section, QQ_Section, D_Acres, Old_Tax#, New_Tax#, Zoning)
OWNED_BY (Parcel-ID, Owner-ID, Percent, T_Date, D_Record, D_From, More_Owner)
PERSON (SS#, Name, Address, Owner-ID)
CORPORATION (CName, CAdrress, Owner-ID)
ORGANIZATION (OName, OAddress, RTaxable, Owner-ID)
ESTATE (EName, EAddress, ExName, Status, Owner-ID)

The attributes Name and Address can be further decomposed into: Last_Name, First_Name, Middle_Initial; and, Street_Number, Street_Name, City, State and Zipcode, respectively.

PHYSICAL DESIGN AND IMPLEMENTATION

The last two steps in the design process involves the definition of the structure and the implementation of the database (Figure 1). The selection of specific storage structures and paths is based on the options offered by the DBMS, and with the objective of achieving good efficiency in the database performance.

The LIS was developed and implemented using PC ARC/INFO[1] . Attribute databases were implemented in PC INFO[2]. Several programs were written to allow the user an interactive update and retrieval of information.

The relations ***PARCEL, CIVIL_TOWNSHIP, PERSON, CORPORATION, ORGANIZATION*** and ***ESTATE,*** from the final set of relations, had a straightforward implementation using the data definition language (DDL) provided by INFO (Henco, 1984; Table 2). The relation ***OWNED_BY*** needed an auxiliary table in order to accommodate the multivalued composite attribute {**Owner, Percent**} not allowed in the relational model. This auxiliary table, ***OTOW,*** contains two attributes (**Owner-ID** and **Percent**) which represent the owner of the parcel, and the percentage of the parcel that is owned by that owner. This is for the case of parcels owned by more than two owners.

The attribute **More_Owner**, of the relation ***OWNED_BY,*** is set to 1 as a simulation of a logical field (not available in INFO),

---

[1] ARC/INFO is a trademark of Environmental Systems Research Institute, Inc. (ESRI), Redlands, California.

[2] INFO is a trademark of Henco Software, Inc., 100 Fifth Avenue, Walthman, MA 02154.

to indicate that ***OTOW*** must be searched when a parcel has more than two owners.

In INFO, tables (relations) are physically organized in sequential files ordered by key attributes; and no additional organization by index is allowed by this DBMS. The file order is handled by the programer and the M:N and 1:N relationships must be performed on ordered files, otherwise the results will be incorrect.

The relation ***MAP-POLYGON*** implemented by ARC has the same format used by INFO; therefore, all information from map polygons can be related to parcel information (in the attribute database) through **Polygon_ID**.

Several programs were written to allow the user, with minimum knowledge in computers, to update and retrieve information interactively. These programs allow to perform different operations on the database such as data input, record update, database query and record delete. Operations were organized from a main menu, and with different alternatives for each option. For example, the user can request information on a particular parcel (database query) using different attributes, such as parcel-id, owner's social security number or name, etc. (Figure 4).

The nonprocedural data manipulation language (DML) of INFO can be embedded in a high-level language, provided also by INFO. However, this language has several constraints that restrict the programming flexibility. Such restrictions refer to *set-a-time* and *record-a-time* operations performed on tables within specific sections, limited number of memory variables, and register pointers not controlled by the programmer.

SUMMARY AND CONCLUSIONS

We have designed and implemented a rural land ownership database as part of a land information system to be used for agricultural reassessment in Miami County, Indiana. The Extended Entity-Relationship model was used in the conceptual design, and the relational model for the logical design.

The EER model allowed to represent graphically all database concepts. This helped to visualize the nature of the data and the different relationships among data. The model permitted several modifications during the design process, and is capable of accommodating future changes in the database without modifying substantially the basic schema. The fact that conceptual design is independent of the DBMS helps the designer/s to concentrate on data and relationships, during the early stages of the design, rather than on implementation problems and solutions.

A disadvantage of the EER model is that it does not provide with an unique solution to the design problem, i.e. in most cases there is more than one set of entities. Different designers can arrive to different solutions for the same database.

Our final design removed the update anomalies that were observed in the original database; however, the data are now located in several places (relations) rather than one, as it was

REFERENCES

Armstrong, M.P. and P.J. Densham, 1990. Database Organization Strategies for Spatial Decision Support Systems, *International Journal of Geographical Information Systems* 4(1):3-20.

Chen, P.P., 1977. The Entity-Relationship Model: Toward a Unified View of the Data, *ACM Transactions on Database Systems* 1(9):9-36.

Codd, E.F., 1970. A Relational Model of Data for Large Shared Data Bank, *CACM* (13)6.

Czejedo, B., R. Elmasri, and M. Rusinkiewicz, 1990. A Graphical Data Manipulation Language for an Extended Entity-Relationship Model, *Computer* March:26-35.

Date, C.J., 1988. *An Introduction to Database Systems,* Addison-Wesley, Reading, MA. 639 pp.

ElMasri, R., A. Hevner, and J. Weeldreyer, 1985. The Category Concept: an Extension to the Entity-Relationship Model, *Data and Knowledge Engineering* 1(1): 75-116.

ElMasri, R. and S.B. Navathe, 1989. *Fundamentals of Database Systems,* Benjamin/Cummings Publ. Co. Inc., Redwood City, CA. 802 pp.

Henco Software Inc, 1984. *pc INFO Reference Manual.*

Jackson, G.A., 1988. *Relational Database Design with Microcomputer Applications,* Prentice Hall, Englewood Cliffs, New Jersey. 207 pp.

Steinhardt, G.C. and D.P. Franzmeier, 1981. Indiana Land Surveys, their Development and Uses, *Agronomy Guide*, Cooperative Extension Service, Purdue University, West Lafayette, IN 47907, Pub. AY-237.

Teory, T.J., D. Yang, and J.P. Fry, J.P., 1986. A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model, *Computing Surveys* 18, June:197-222.

Ventura, S.J., B.J. Niemann, Jr., and D.D. Moyer, 1988. A Multipurpose Land Information System for Rural Resource Planning, *Journal of Soil and Water Conservation*, May-June:226-229.

Ventura, S.J., 1990. Conversion of Automated Geographic Data to Decision-Making Information, *Photogrammetric Engineering and Remote Sensing* 56(4):511-516.

TABLE 1. EXAMPLE OF A RECORD FROM THE CURRENT RURAL LAND PROPERTY DATABASE FOR MIAMI COUNTY, INDIANA*.

FIELD	DATA	DESCRIPTION
AREA-CODE	6	Township and Range code
SEC-CODE	3	Section number (1-36)
QUARTER-SEC	0	Location of parcel by quarter section
UNIT-NUMBER	18	Tax unit number (identifies civil townships)
PARCEL-NUM	8	Parcel number on map
MIATAX	0183201300	Existing tax number from Miami County
SECTION	3	Section number (1-36)
TOWNSHIP	28	Congressional Township -North
RANGE	5	Congressional Range - East
BRIEF_LGL	PT NW	Indicates general parcel location
LAST-NAME	WEST	Owner(s) last name
FIRST-NAME	SMITH JOHN & MARY	Owner(s) first name
ADDITION		
LOT	0	
BLOCK		
DEED-AC	122.6300	Legal acreage of parcel
ADDRESS	R R 1 MIAMI IN 12345	Owner(s)' mailing address for tax statement
DEED-RCRD	WD 263/774	Instrument of Conveyance (Type Book/Page of Deed)
DEED-FROM	BROWN PETER F & SUSAN	Previous owner(s) of parcel
PARCEL-ID	328,508	Parcel Identification number on map

*Names and address have been changed for publication purposes.

LIST OF FIGURES