



GIS DATABASE DESIGN AND IMPLEMENTATION – THE MAUTECH EXPERIENCE

MUSA. A. A¹, IDOWU. T. O² & ZEMBA. A. A³

^{1,2}Department of Surveying & Geoinformatics, Yola, Nigeria

³Department of Geography, Modibbo Adama University of Technology (MAUTECH), Yola, Nigeria

ABSTRACT

The myriad of problems associated with storing and managing spatial data in MAUTECH had made it imperative to develop a Geographic Information System (GIS) database. The urge to split the database into two packages – one, a conventional database handling the attributes data and two, a GIS database handling the spatial data – was borne out of the fact that a database was already in existence which stores the attribute data. However, the need to keep the DBMS simple in order to make it user friendly made the option of storing both attribute and spatial data in one single DBMS an attractive option. The latter option was implemented in this paper using ESRI's ArcGIS 10.1. The design went through three stages - the conceptual, the logical and the physical design stages.

The main features used to link the numerous other related tables were the buildings which were digitized from a satellite image as polygon features. Though most of the problems associated with analogue record keeping were eliminated the objective of making it user friendly was not achieved. The design successfully eliminated data redundancy however; querying the database had new challenges for users that are unfamiliar with the GIS. The split arrangement whereby the spatial data is stored in a GIS and the attribute data is stored in a conventional database is considered a much better option.

KEYWORDS: Geographic Information System, MAUTECH Yola, Scale Factor – 0.9996

INTRODUCTION

A database is a collection of one or more data files or tables stored in a structured manner, such that interrelationships which exist between items or sets of data can be utilized by the database management system (DBMS) software for manipulation and retrieval purposes (Healey, 1991). The GIS database is similar to the conventional database. However, the ability of the GIS database to store, manipulate, retrieve and display spatial data (i.e. maps) in addition to the normal functions of the conventional database is the main characteristics that set it apart from other databases. Database management systems (DBMS) on the other hand are specially designed application software packages that interact with the user, other applications and the database itself to capture, query and analyze data (Wikipedia, 2013). There has been much debate in the GIS industry about the suitability of standard commercial DBMS for use with the GIS. Batty (1990) and Seaborn (1992) had tried to explain the advantages derivable from such arrangement.

They opined that the GIS vendor should be able to take advantage of functions provided by the DBMS vendor and concentrate on developing GIS functions. In particular, functions such as security and backup and recovery are well proven in conventional DBMS and the GIS can take advantage of these. Newell & Batty (1994) however, are of the view that spatial data types are not explicitly supported by most conventional DBMS, the GIS vendor typically has to use complex record structures and retrieval algorithms for geographic data, which means that it cannot be easily read or updated by external applications. The external applications also need to understand any special locking mechanisms which are used by the GIS to handle long transactions, since the standard DBMS locking does not handle this.

The Modibbo Adama University of Technology (MAUTECH), Yola, Nigeria already has a database designed in MS Access. Access, being a conventional DBMS, has no GIS capabilities. The first ‘gray area’ to sort out therefore was whether to create the spatial database in a GIS and link the GIS data to MS Access; or alternatively to design the GIS database completely within a GIS DBMS.

The latter was adopted, not just because of the reasons enumerated in the previous paragraph, but also because users of the database are not expected to be knowledgeable in database technology. Thus, having to connect and disconnect between two or more applications is considered too cumbersome for the end users. A counter argument however, could be that data entry and edit for the attribute data could become error prone since it will have to be done twice – one, in MS Access and the other in the GIS. In recognition of the validity of this argument, this paper will try to assess the suitability of this design to the peculiarities of the study area.

Apart from technical problems of design that this paper wish to sort out, there are non-technical problems associated with analogue record keeping. The problems associated with analogue techniques of keeping records in MAUTECH are not different from what is obtained in other similar institutions of higher learning in Nigeria. Spatial data are best stored as maps. The importance of these maps has increased tremendously over the years as society became more and more sophisticated. In MAUTECH Yola, in particular, maps are frequently used by the Physical Planning Unit (PPU) to plan out where to locate new structures and facilities.

Likewise the works and maintenance unit need maps not only to locate existing facilities but also to determine the age of certain facilities so as to arrange for repairs and replacements when necessary. Academic units also require maps for their research activities. Yet getting spatial data (maps) of the distribution of features (e.g. roads, pipe-lines, houses, generators e.t.c.) in MAUTECH, Yola campus is very difficult. Where they are available, they are usually not very useful. This difficulty can be linked to several issues:

- The few maps created for certain utilities, were created to address specific problems. Once the problems were solved, the maps were tucked away in the drawers of the departments that initiated them and forgotten there.
- Some features were distributed, not based on pre-planned designs, but based on an ‘on-the-site-assessment’. After laying the features spatially, no effort was made to document the various locations by means of a map. Where such features lie underground (like in the case of water pipe-lines) their locations become known only to older artisans who either participated or witnessed when they were buried.
- Where some relevant maps are available, they are unusable for a number of reasons:
 - They have become so old and worn-out that many details have been obliterated.
 - They are drawn in a scale that is incompatible with the user’s needs and can’t be changed to the required scale
 - They do not have enough details to meet user requirements since adding too many details in a map can overcrowd the map and make it illegible.
 - They could become very outdated and updating such analogue maps is usually very difficult.

Putting all these myriad of problems into consideration, coupled with the dry academic treatment that database design is given by most authors (Litwin et al, 1994), the objectives of this paper can be summarized thus:

- Demonstrate and document the step-by-step design of a typical relational database.
- Provide comprehensive, consistent, easily maintained GIS database for the university.
- Assess the feasibility and practicability of ArcGIS 10.1 for managing university database.

Among the different applications which the GIS could efficiently handle, database management has grown to take a place of prominence. Progressive realization of the importance of this tool has translated into the emergence of a variety of databases modelling different concepts of reality. For instance, NCI (2012) reported the development of a database for breast cancer for a place known as Long Island. The database known as the Long Island Geographic Information System (LIGIS) is designed to study the potential relationships between environmental exposures and breast cancer in the island. Data used in the database consisted of cancer rates and trends, socio-demographic, medical resource, behavioural risk factors etc.

Vardhan (2013) also reported how the state of Goa (India) developed the Land Records Information Systems (LRIS), a land record database developed with an Indian GIS product – VISION mapmaker. The state government undertook the ambitious project, which involved 100% computerization of maps and Record of Rights of the state, business process automation for the survey department and dissemination of information through a Web GIS portal. The initiative, called Dharnaksh (www.dharnaksh.com) makes land records available anywhere, anytime at the click of a mouse. The initiative was not just to computerize information but to change the way records were managed and delivered to stakeholders, leading to accurate storage, faster delivery and easy and quick update process. Data is shared by everyone but updated only on authentication of fingerprints of the authority concerned.

Winterford (2013) described how Montreal's local transit authority – The Societe de transport de Montreal (STM) developed an application that is hoped will drive use of public transport system for needs beyond the daily commute. The application splits traveler's data into two separate databases. For privacy purposes, key personal details are 'de-identified' to the point that access to any one of the two databases cannot identify an individual. Among the most clever uses of the application are: A book store that offers commuters moving toward its store the opportunity to download a sample chapter of a book that is on sale, in the hope they'll get hooked and buy it; the Montreal Opera House which offers special prices on unsold (and thus perishable) seats at 4pm the afternoon of the performance; Supermarkets that offer their specials on bus routes for commuters soon to pass by who are likely to be thinking about what to cook for dinner; Food trucks can let patrons know when they are parked close to their office, should they be up for a lunchtime treat.

Souleyrette (1998) developed a GIS database for the Iowa dept of transport (DOT) maintenance division. Before the creation of the database, inventory data were referenced using a route and milepost offset referencing method. (Data are only collected on the primary highway system.) Data are maintained in attribute tables without a direct relationship to graphic element(s) representing their physical location and extent along a highway. All data, including distributed attribute tables, were maintained in an Oracle database, while spatial data were maintained in GeoMedia 2.0

Other application areas of database creation include DNA databanks for criminal investigation (Kimsky and Simoncelli, 2010) and database for soil profile data and attributes of soil map units (Burrough, 1991). There is no doubt that Africa is lagging far behind in this technology. Even when the public is aware of the existence of a database, e.g. Nigeria's national identity card scheme (NICS) or the Abuja Geographic Information System (AGIS), it is so shrouded in secrecy that the academic and/or professional world knows next to nothing about its technical details. This has made it difficult to criticize or analyze its functions with a view to improving on its capability.

METHODOLOGY

Overview

A new map document was created in ArcMap. A Garoua UTM zone 33N coordinate system was specified for the data frame with properties as follows: Projection – Transverse Mercator; False Easting – 500000; False Northings – 0.0; Central Meridian – 15.0; Scale Factor – 0.9996; Latitude of Origin – 0.0, Linear unit – Meter(1.0). Two base maps were added to the map document in ArcMap - the first was the QuickBird image of the study area. Most of the ground features were digitized from this image. The world file attached to the image made georeferencing unnecessary. The second base map used was the scanned boundary map of the university. From the scanned map, survey features (namely property beacons and survey boundary lines) were digitized. The boundary map was georeferenced using the combination of GPS coordinates from the field and the ArcGIS georeferencing tool (Customize/Toolbars/Georeferencing).

With the base maps in place, creation of the database began. The design went through three stages - the conceptual, the logical and the physical design stages. The conceptual model provides a global view of the entire database. It describes the main data objects, avoiding details. The logical design matches the conceptual design to the requirement of a DBMS. Entity relationship diagrams (ERD) are the visual representations of the first two stages. The last stage i.e. the physical design, maps the data storage and access characteristics of the database.

Data/Equipment

Data used for this study include Quick Bird images of study area; the original boundary map of the study area produced in 22/07/1983 by Surv F. C. Okoli; and GPS coordinates of selected boundary beacons. Equipment used includes a promak III GPS unit and a Hewlet Packard laptop, while software used was ESRI's ArcGIS version 10.1

Conceptual Design

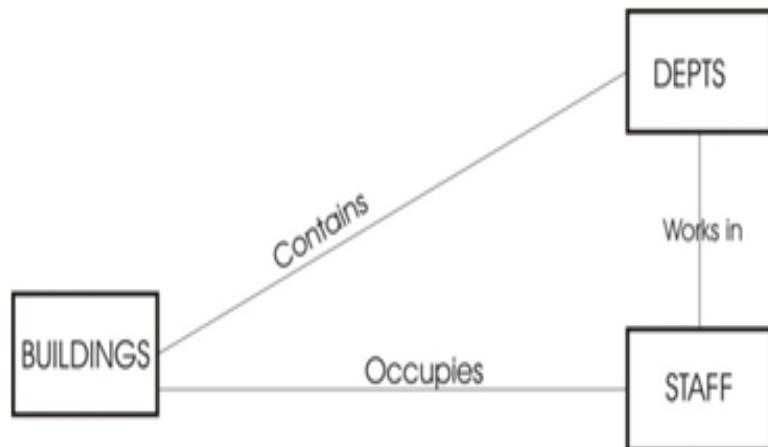


Figure 1: Conceptual Design

The concept of reality imagined was that every staff in the university is attached to a building. Staff living in staff quarters within the campus therefore will be attached to two buildings – the work place and the residence. Therefore it should be possible, not only to access information about specific buildings, but also possible to identify buildings based on information about the occupier. The design was kept simple with the aim that it could be elaborated upon in future. Thus only three entities were used – Buildings, Department and Staff. Buildings contain the departments; while staff works in the department and lives in buildings (Figure 1). Though limited data on road network, boundary lines and bore holes were captured, they were only used to aid visual recognition and interpretation of the buildings. Once the underlying concept behind the database was determined, the design moved to the logical stage.

Logical Design

In the logical design, tables were created out of the entities. Tables were normalized to eliminate data redundancy and the fields within each table were determined. Also, it was at this stage that primary keys and foreign keys were determined. Where necessary, indexes were inserted and relations between the tables were determined. The ERD diagram makes it possible to put the entire logic together in one comprehensible whole. However, the difficulty in getting the diagram right at the first instance makes the process of drawing the diagram an iterative one. With each iteration, the tables and relations were edited until the final diagram was designed.

In the first normalization form a relational table was created for the ‘Building’ entity (as shown in Figure 1). However, an assessment of how the data were repeating (redundancy) informed the designer as to how the second order normalization form should proceed. Thus, in the second order normalization form the ‘Building’ table was split into four tables – Hostels, Staff_Qtrs, Schools (i.e. Faculties) and Others (Figure 2). The ‘Others’ table stores data of buildings not captured by the other three tables e.g. the division of student affairs (DSA), computer centre, Chevron biotechnology centre e.t.c.

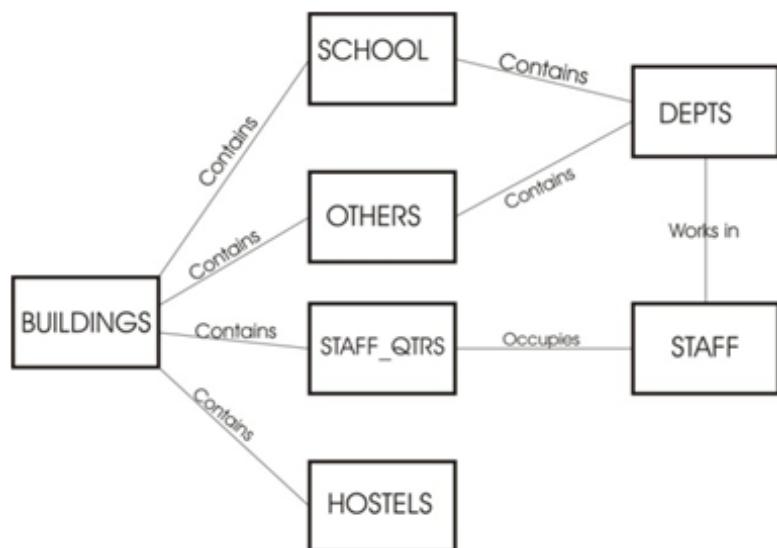


Figure 2: ERD Based on First Order Normalization

With the creation of the four tables – School, Others, Staff_Qtrs and Hostels, redundant data were greatly reduced. But despite this, data redundancy was still discovered in the ‘Schools’ table. The redundant data were traced to how the academic buildings were constructed. While many departments are housed within schools (a case of one structure representing several departments) e.g. the school of pure and applied sciences (SPAS) and the school of management and information technology (SMIT), many other departments occupy separate buildings of their own.

The third order normalization process therefore dictated that another table named academic units should be created which was later related to the schools and dept tables. Using the more current Crows foot notation model (Carlos et al 2011), as against the original Chen model, the final entity relation diagram was created (Figure 3). The completion of this process marks the beginning of the physical design.

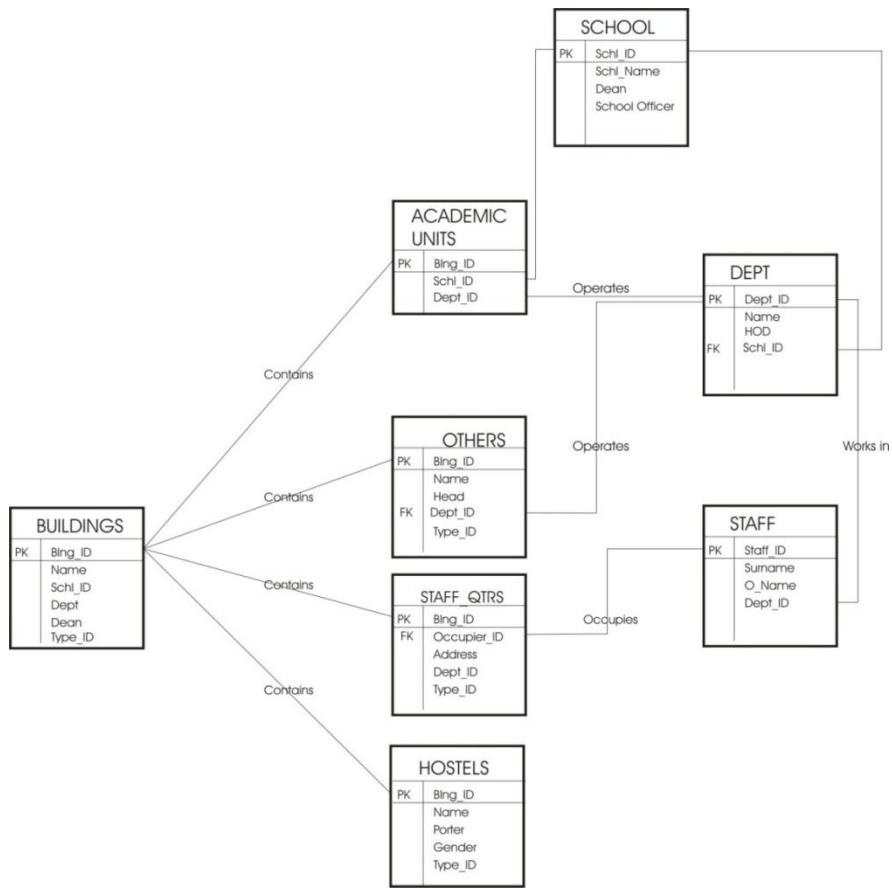


Figure 3: Final Entity Relationship Diagram

Physical Design

The term physical design is the phrase given to the entire process of encoding the ERD into the software. The physical design therefore is dependent on the software used. ArcCatalog was first opened from where various feature classes and tables were created. A total of seven feature classes were created. A polygon feature class was created for the buildings, three line feature classes were created for Routes, Streams and Survey Lines and two point feature classes were created for bore holes and survey beacons. Similarly seven standalone tables were created for Schools, Others, Staff_Qtrs, Hostels, Academic_Units, Dept and Staff. In ArcMap, all the created layers and tables were added to the map document. All the stand alone tables were linked with the Buildings feature class as depicted in Figure 3. The theme tables (tables directly attached to map layers) were automatically created as the layers were edited. The next stage was creating the fields within the tables. The theme table ‘Buildings’ had four tables inserted by default – OBJECTID, SHAPE, SHAPE_length and SHAPE_area. Additional three fields were inserted – Name, Type and Type_ID. Two field formats were used throughout i.e. the Text and the Short Integer. Text field format were used for most of the descriptive fields e.g. Name of HOD, Address etc. The short integer were used mostly for the indexes e.g. Deptal_ID, Sch_ID, Blng_ID etc. In order to create the relations and also enforce referential integrity between the related tables, relationship classes were created for each of the relations. This was done by right clicking the geodatabase icon in ArcCatalog and selecting New > Relationship Class. Majority of the attribute data used were derived from the database of the Management Information System (MIS), which is a MS Access based database.

Queries

Two tools were used to search for needed information within the database – the ‘Related Tables’ tool and the ‘Identify’ tool. The query builder, an SQL based query tool, could only be used when information from a single table is

required. This is because it operates only within a single table. The query builder was therefore considered an irrelevant tool for this database. After selecting a record in one table, the ‘related tables’ tool allows a user to browse through all the other related tables to view records linked to that initial record. The ‘related tables’ tool is best used when using attribute data to search for required features in a layer, for instance when a lecturer’s name is used to search for the house he lives (Figure 4). The ‘identify’ tool on the other hand provides the attribute data for any selected feature. Using the ‘identify’ tool, data from the theme table is displayed by default. Related records can however be displayed by manipulating the feature in the left panel of the identify result dialog box. Tables related to the theme table are listed in this left panel. By expanding any of the table names more tables related to the expanded table are listed. By clicking any of the table names, information about the selected table is displayed in the identify dialogue box. Figure 5 shows a selected feature (i.e. School of Pure and Applied Science - SPAS). The identify dialogue box is divided into two. The right panel of the identify dialogue box, shows records in the Schools table for that feature. The left panel shows other related tables e.g. ‘Dept’ tables. By expanding a particular Dept table, information about the table is again displayed. Thus, even though one cannot view all the attribute data concerning a selected feature at once, one can gradually browse through the records one table at a time.

DISCUSSIONS

The database was created in a personal geodatabase. The created database successfully solved most of the problems associated with analogue record keeping highlighted earlier. By arranging the maps in digital layers, maps were easily accessed, and can be converted to any scale. The data could be stored in CDs and external hard drives thus eliminating the problem of wear and tear. There were enough tools to allow the designer replicate the logical design in the software with relative ease. Attribute data entry has not changed much since the days of ArcView 3.2a. The tools used for spatial data entry however has become richer. The polygon data entry, for instance has tools allowing creation of squares, circles, ellipse and free hand. With additional tools allowing splitting, merging and demerging of features a smart designer can reconstruct almost any kind of architectural design. The line tool has about ninety two symbology types which in this writer’s assessment is far more than is required. Only four symbology types were used for this project. They were used for roads, fences, streams and boundary lines. It is clear that recording spatial data as accurately as possible is the main concern of ESRI.

Querying data was done using the Related Tables tools and the Identify tool. The problem with both the ‘Related Tables’ tool and the ‘Identify’ tool is that the user is conscious of the fact that data is stored in different tables. A user must be fairly familiar with the software to query the database. Lastly queries cannot be customized to ones specification. The tool that will naturally solve this problem is known as the query layer. The query layer is a layer that is defined by an SQL query. Query layers allow both spatial and non-spatial information stored in a DBMS to be easily integrated into GIS projects within ArcMap. Using the query layer therefore, the full power of SQL can be used to filter out information that would have been impossible to get otherwise. The query is executed every time the layer is displayed or used in ArcMap. Unfortunately, the query layer tool cannot work without connecting the geodatabase to an external database. Therefore, whether the GIS database is connected to an external database or created within a GIS, familiarity with the DBMS is inevitable. Databases supported by ESRI’s ArcGIS include IBM DB2, IBM Informix, MS SQL server, Netezza, Oracle and PostgreSQL. Surprisingly, MS Access is not among the list, but another Microsoft product MS SQL Server is included. To make matters worse, connection to an external database cannot be done from either a personal geodatabase or a file geodatabase. It can only be done in an enterprise geodatabase or an ArcSDE geodatabase. Apart from the fact that personal geodatabases become unstable or start slowing down after inputting 500MB of data, this is the

singular most significant reason why the ArcSDE geodatabase is considered a better option than the personal geodatabase. This is an area for further examination by database professionals. Moreover ESRI claims that ArcSDE geodatabase support long transactions.

The question of suitability of the database can no longer be viewed from the perspective of familiarity with the DBMS. It is clear that providing a database for users that are unfamiliar with the DBMS is a near impossibility. In view of the fact that policy makers of the university are not likely going to have enough time to study the DBMS, it is recommended that a professional team is mandated to make queries on the database on the bases of user's requests.

CONCLUSIONS

The relational database system still remains the best option of storing and accessing the kind of data described in this paper. However the database must be well designed to take advantage of its processing power. Geographically referenced data is better stored, maintained and manipulated in a split system i.e. two DBMS – one for the attribute data and the other for the spatial data. This allows specialized tools within each DBMS concentrate on the kind of data it was designed for. It also allows for large memory space. An appropriate link between the two DBMS allows it to work like a single system when desired. There is a limit to which the database can be made user friendly. Therefore professionals should be made to handle the day to day manipulation of the database.

REFERENCES

1. Batty P.M. (1990), Exploiting Relational Database Technology in a GIS, *Mapping Awareness* magazine, July/August 1990.
2. Burrough. P. A (1991); Soil Information Systems, in D. J. Maguire, M.F. Goodchild & D. W. Rhind (Edit) *Geographic Information Systems* (First Edition). Vol 2. Section III. Applications. John Wiley & Sons.
3. Carlos Coronel, Steven Morris & Pete Rob (2011) *Database Principles – Fundamentals of Design, Implementation and Management* Cengage Learning, pp 106
4. Paul Litwin, Ken Getz and Greg Reddick (1994) *Fundamentals of Relational Database Design*, Microsoft Access 2 Developer's Handbook, Sybex
5. Healey. R. G. (1991) Database Management Systems, in D. J. Maguire, M.F. Goodchild & D. W. Rhind (Edit) *Geographic Information Systems* (First Edition). Vol 1. Section II. Chapter 18. John Wiley & Sons.
6. NCI (2012): National Cancer Institute (US National Institute of Health – www.cancer.gov), http://gis.cancer.gov/overview/database_development.html#ligis (downloaded 5:20pm 06/10/2013)
7. Richard. C. Newell & Peter. M. Batty (1994): GIS Databases Are Different. AM/FM (1994), 279-288 <http://libraries.maine.edu/Spatial/gisweb/spatdb/amfm/am94029.html> (downloaded on 12/10/2013)
8. Reginald. R. Souleyrette (1998) *Geographic Information System (GIS) Database Development, Integration, and Maintenance* for the Iowa Department of Transportation Maintenance Division Center for Transportation Research and Education Iowa State University: http://www.intrans.iastate.edu/reports/maintenance_gis.pdf (downloaded 10/10/2013)
9. Seaborn D. (1992): The Year GIS Disappeared, AM/FM XV Conference Proceedings, San Antonio, April 1992 pp 822-826

10. Sheldom Krimsky & Tania Simoncelli, (2010); Genetic Justice: DNA Data Banks, Criminal Investigations, and Civil Liberties: Columbia University Press.

11. Vardhan R. Mihir (2013): Towards e-ffective governance: Geospatial World

<http://www.geospatialworld.net/Paper/Application/ArticleView.aspx?aid=30558> (downloaded 12/10/2013)

APPENDICES

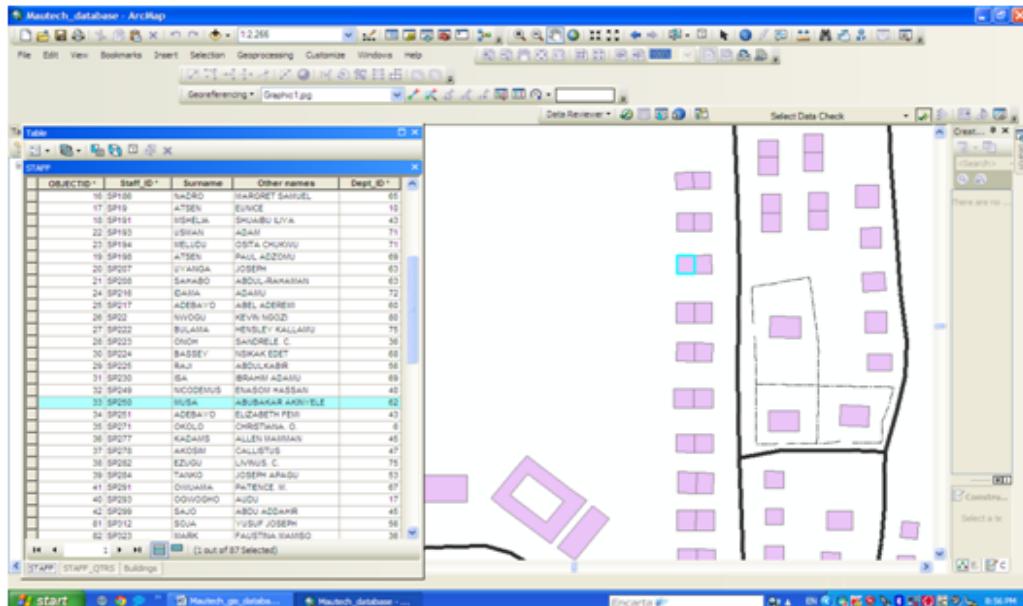


Figure 4: Using Selected Record to Identify a Feature (Using ‘Related Tables’ Tool)

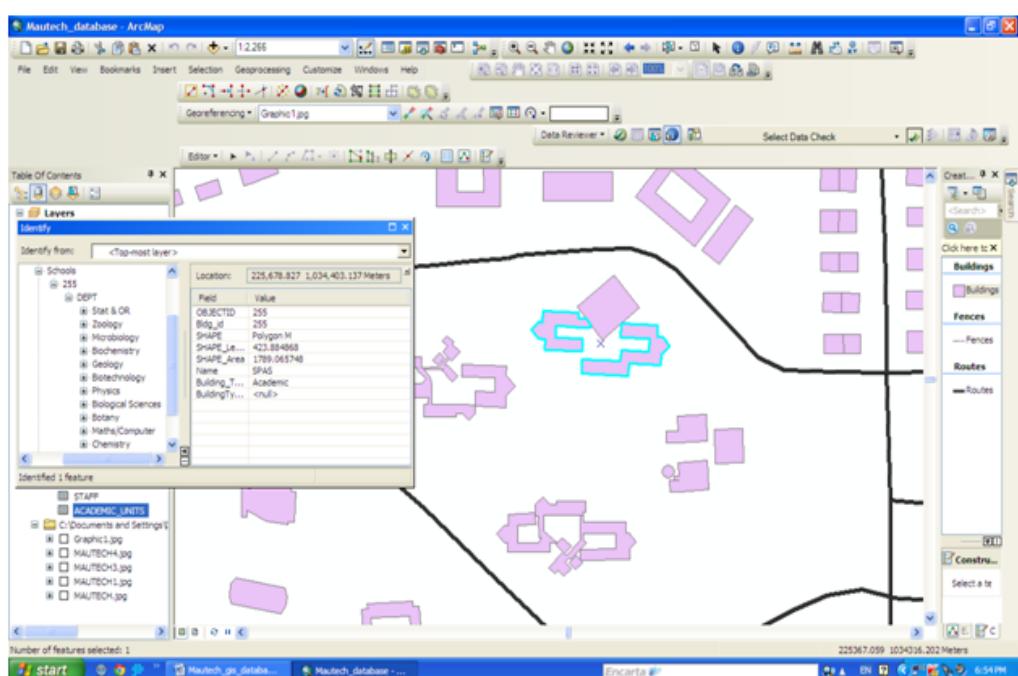


Figure 5: Identifying a Feature to See its Attribute Information (Using ‘Identify’ Tool)

