J. Kej. Awam Jil. 7 Bil. 2 1994

AN OVERVIEW OF ARC/INFO A TOOL-BOX APPROACHED GIS

by.

A. Hamid Mohamed
Faculty of Civil Engineering
Universiti Teknologi Malaysia

ABSTRACK

ARC/INFO is an integrated system in which database and graphics are combined together in one system, and the relationship between them is maintained. In the graphics section, geographical data can be processed in several modules which have different roles. This means that ARC/INFO splits its GIS functions (tools) in several places which intersupport each other. This new method of data handling and manipulating is known as a tool-box approach. The purpose of this article is to investigate the capability and suitability of ARC/INFO and the significance of the tool-box approach in order to provide a very outstanding system to fulfill the user's ever changing needs.

INTRODUCTION

ARC/INFO is a geographic information system (GIS) used to automate, manipulate, analyse and display geographic data in digital form. It was designed and produced by Environmental Systems Research Institute (ESRI), California, United States and began to be widely used in early 1980's. ARC/INFO is a good example of a modern GIS which uses a tool-box approach, based on ESRI's work in some 400 actual application projects over a period of fifteen years. These studies of user needs and requirements helped ESRI to design and develop important tools in order to fulfill and support a variety of GIS applications efficiently (DANGERMOND 1986).

In order to cope with the developments and variety of other GIS equipment, ARC/INFO was designed in such a way as to make it compatible with various types of computers, graphics display, data loggers, digitisers, plotters, etc. Even though it provides a large number of tools, there is the consideration that a user may need to create his own tools for specific applications. For this reason, ARC/INFO provides a powerful macro language known as ARC MACRO LANGUAGE (AML). It is a fully functional programming language with facilities to use variables, performing logical branching and loops, manipulating character strings and text, etc.

This language can be used for creating new tools and other applications, such as to design and create custom screen menus, by manipulating existing commands and functions to improve ARC/INFO operations, etc.

ARC/INFO consists of a number of modules which are used to hold several types of tools. In general, these modules can be classed into two sections : ARC and INFO. The ARC section (ARCEDIT, ARCPLOT, LINEEDIT, SHADEEDIT, MARKEREDIT, TEXTEDIT, MAPLIBRARY) is used for graphics management, and INFO section (INFO) for database management. The combination of these two sections gives the name of ARC/INFO.

The following sections will discuss the data model used in ARC/INFO, the functions of every module, and the ARC/INFO basic operations.

ARC/INFO DATA MODEL

The ARC/INFO data model is based on the idea that geographic data can be represented as a set of features. Each feature has both associated locational and thematic data. For example, if a feature is a land parcel, its locational data might be the xy coordinates defining its boundary, and the thematic information might be referred to parcel number, owner name, annual tax, etc. The locational and thematic data are equally important. However, it is proven that data structures optimal for the analysis of locational data are not optimal for thematic data. To overcome this problem, ARC/INFO uses a hybrid data model in which a topological data model is used to represent locational data and a tabular or relational data model to represent thematic data. Both these two data models are designed in such away to meet all the goals of a data model.

The ARC/INFO data model obviously refers to how data is stored and arranged in ARC/INFO. The arrangement of the data will then determine how it can be accessed and processed. The accessibility of data affects the important stage of GIS, i.e performing queries. The performance of query facilities which strongly depend on the data model will determine the overall ability of the system.

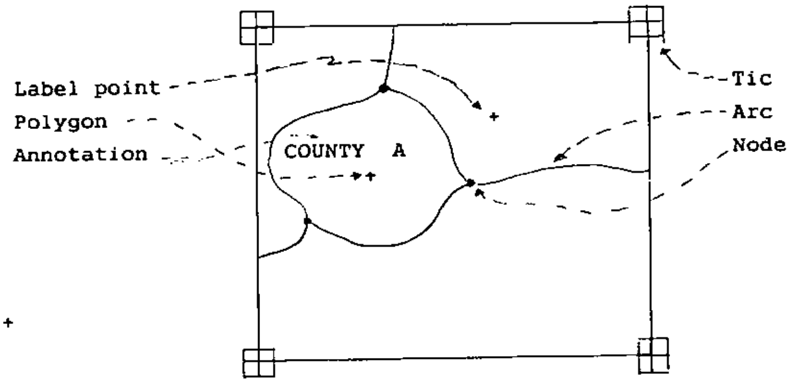
Coverage : The basic storage unit in ARC/INFO

The basic unit of storage in ARC/INFO is called coverage. It represents a single map sheet layer in digital form and generally only describes one type of map feature such as buildings, roads, land parcels, etc. A coverage contains both locational and descriptive information of map features of a selected area. Map features of a coverage are stored as simple points, lines or polygons. The locational data of these features is represented explicitly as a series of x, y coordinates, or topologically as a combination of other features. For example, rivers might be represented by a set of arcs and stored as sets of ordered x,y coordinates which define each river. A set of rivers can then be used to topologically define a farm area.

The descriptive data of map features is store in a special data table called a feature attribute table. There are three types of feature attribute tables which store different types of data; the Polygon Attribute Table (PAT) for polygon information, the Arc Attribute Table (AAT) for arcs and the Point Attribute Table (PAT) for point information. There tables are automatically created when the coverage is cleaned or built (section 5.4). The tables can be appended to, therefore, users can add any new item to them in order to provide extra attributes for each map feature. The main functions of these tables are to associate attributes with map features and to connect a coverage with external data in order to provide more information for map features. The external data can be in several types of database systems, such as info (ARC/INFO database system), oracle etc.

Coverage Features

Features which exist in a coverage can generally be grouped into two classes; primary and secondary. The primary coverage features include arcs, nodes, label points and polygons, and the secondary features are tics, the map extents, links and annotations. These features are shown below:



The primary features represent points, lines and areas on maps. Arcs represent linear features, the borders of polygons or both. A line feature may be made up of one arc or many arcs. Each arc is assigned a reference number called a User-ID. A series of x,y coordinates is used to define the location and shape of arcs. Using topology, arcs can be linked to their end points and to the polygons on each side of them. Descriptive data of arcs in a coverage are stored in the AAT. There is one record in the AAT for each arc in the coverage. This record is related to the feature via its User-ID. Nodes represent the beginning and end points of an arc. They also represent the connection point of two arcs. Their coordinates are stored with each arc as the first and last points. A node may be topologically linked to the set of arcs which connect to each other at the node. The third primary features are label points which represent point features or assign User-IDs to polygons. Each label point is described by a single x,y coordinate and a unique User-ID number. In the first case where label points are used to describe point features, their x,y coordinates describe the location of the points such as well site, bench marks, lamp posts, etc.

However, in the latter case where a label point is used to identify a polygon, it can be anywhere within the polygon and its exact position is not important. The label point User-ID is used to associate attributes describing what the point is or to associate attributes to describe the polygon in which the label point falls. It is clear that label points can represent point features or can identify a polygon in a coverage, but not both. In other words, point features and polygons are not allowed to be in the same coverage. When representing point features, their data will be stored in Point Attribute Table (PAT), and there is one record in the PAT for each label point. The record is related to the point map feature via point's User-ID.

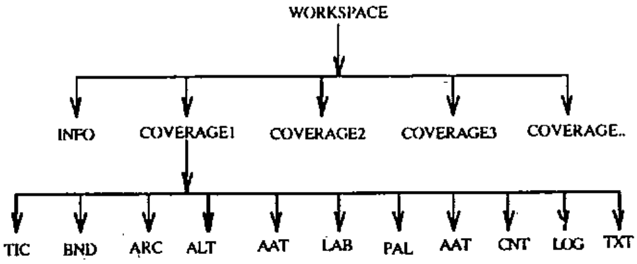
The last primary coverage features are polygons which represent area features. A polygon is defined topologically by the series of arcs which compose its border and by a label point positioned inside the polygon. The polygon ID number which is represented by the label point ID number is used to associate attributes describing the polygon.

The secondary features which include Tics, Coverage extent, Annotation and Links are still important in order to provide more general information about a coverage. For example, Tics are used as geographic control points of a coverage. These control points allow all coverage features to be registered to a common coordinate system. In general, TICS are used to orientate a coverage during digitising, for map merging and overlay, and for plotting. At east four tics must be created for a coverage. Using more than four tics is better and will increase the accuracy of map registration when digitising. Links are a special feature type created and used in arcedit to perform rubber sheeting. Each link has two points used for adjusting coverage coordinates, a from-point and a to-point. The to-point may be a feature coordinate in the snap coverage, or simply a control point entered by the user. Coverage extent is used to define map extent, and Annotation for annotating maps of the coverage.

How a coverage is stored

A coverage is stored in the computer as a directory of a drawing in a workspace which normally represents a particular job. The directory name is the coverage name. Each coverage has a set of files in which every file contains a particular feature class. This set of files varies depending on the feature classes present in a coverage. Basically these files are used to store information of Arcs (ARC, AAT,ALT), Points features (PAT,LAB,PRF), Nodes (ARC), Polygons (PAT,PFF,PRF,CNT,PAL), Annotation (TXT), and Coverage General Information (LOG,BND,TICS,TOL). Please refer to appendix A (ii) for descriptions of every coverage file.

The general organisation of a coverage directory is as shown below:



STORING DATA IN ARC/INFO

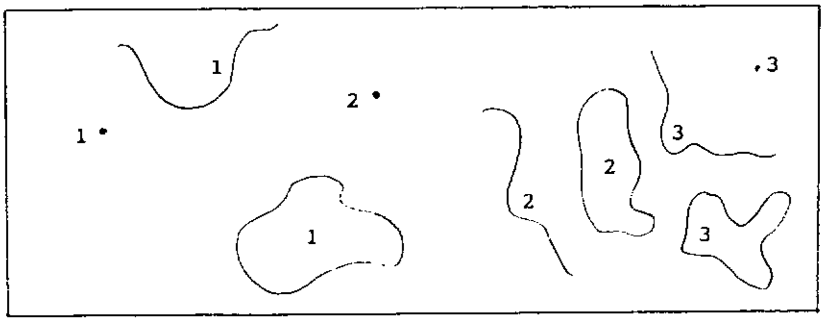
a) Locational data

The main condition of a GIS is the locational (graphical) data which should be transformed and saved in digital form. The map features are saved as one of three types of map elements i.e points, lines or areas. The x,y cartesian coordinate system is used to reference map locations to around locations. Each points feature is recorded as a single x,y coordinate, lines as series of ordered x,y coordinates and areas as series of x,y coordinates defining line segments which enclose each area. A Closed area is referred as polygon which means a many sided figure. Therefore, map features are represented as lists of x,y coordinates instead of as pictures or graphs. This is the concept of digital form.

The Concept of User-IDs

The coordinate of one feature can easily be stored in the form of x,y coordinate. When representing multiple features, it is useful and helpful to give a reference number for every feature and the coordinate can be recorded for each feature. In ARC/INFO, there are two types of reference numbers that are User-ID and Internal Sequence numbers. The User-ID is a reference number given by users to a map feature and it can be any number. However, User-IDs must be unique, one feature is only allowed to be assigned one User-ID, but two or more features can share a same number. In case there are more than one User-IDs are assigned for a feature, only one of them will be accepted and applicable as a reference number. this number is not affected by any coverage processing such as clean, build, overlay, etc. If the coverage is edited and some of the features are deleted, this does not affect the User-ID number of the rest of the features as long as the new numbers are not assigned.

By contrast, the Internal-ID number is internally assigned by ARC/INFO to every feature once the coverage is processed (Clean or Build). This number is sequentially fixed according to the number of features, if there are 30 features, the internal-ID number will be from 1 up to 30. these numbers will be re-assigned every time after coverage processing or editing. If the coverage processing or editing affects the number of features such as subtracting some features or inserting new features, this will change the sequence of the features. This means the Internal-ID number will also change to the new sequence which is the same as the feature order.



Point Feature number, X, Y Pairs

1	x1	y1
2	x2	y2
3	x3	y3
4	x4	y4

Line Feature number, X, Y Pairs

1	xal yal, xbl ybl, xcl ycl
2	xa2 ya2, xb2 yb2, xc2 yc2
3	xa3 ya3, xb3 yb3, xc3 yc3

Polygon Feature number, X, Y Pairs

1	xal yal, xbl ybl; xcl ycl,....xal yal
2	xa2 ya2, xb2 yb2, xc2 yc2,...xa2 ya2
3	xa3 ya3, xb3 yb3, xc3 yc3,...xa3 ya3

Figure 1 How Locational data is stored in the computer

b) Descriptive Data

Descriptive data is used to describe the attributes of map features. They are stored in the computer in a manner very similar to the way coordinates are stored; as a set of numbers and characters which are arranged according to the pre-defined format. The pre-defined format of each item; such as item name, item type, item input and output width, and decimal places if the item type is not an integer.

For example, the attributes for a polygon which represent a river might include name, river depth, river width and river length and be stored in the computer as a list of values as below:

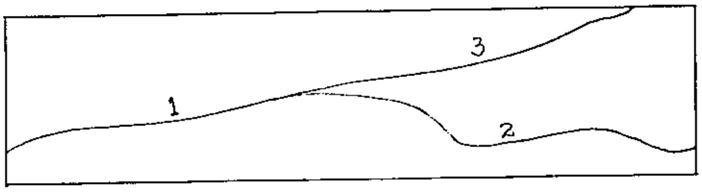
Items >	Name	Depth(m)	Width(m)	Length (km)
Values >	River Tyne	30	250	50

If there are many features, one set of attributes are stored for feature as shown in figure 2

Feature-No	Name	Depth (m)	Width (m)	Length (km)
1	Red	30	250	30
2	Green	50	350	50
3	Blue	25	150	45

Figure 2 storing descriptive data in the computer.

This descriptive data is only usable if it is relatable to map features. The relationship between the attributes data and locational data can only be established if there is a common item between both data. This item is known as relate item.



Computer Coordinates (spatial data).

Feature No.	X,Y Pairs
1	xal yal, xbl ybl, xcl ycl..
2	xa2 ya2, xb2 yb2, xc2 yc2..
3	xa3 ya3, xb3 yb3, xc3 yc3..

Figure 3 : Relating Feature coordinates and Attributes

Figure 2 and 3 show that both locational and descriptive data share a common item i.e feature number which is the easiest item to be used as a relate item. Therefore, the attributes of feature number 1 in the locational data table is reffered to the feature number 1 in the attributes data table. This is the basic concept of how points, lines and polygons are stored in ARC/INFO and how they can be described using attributes. However, the applications of this concept are limited to the description of feature location, feature types and other attributes, as well as in associating symbols to each feature for map display. It is obvious that the relationship is limited to the extent of direct connection between a map feature and associated attributes. However, in GIS we have to extend the scope of relations to the relationship between one map feature and other map features. For example, the relationship between a road and farm areas or one land parcel and another land parcel and so on.

DEFINING SPATIAL RELATIONSHIP USING TOPOLOGY

On a map, spatial relationships can be represented graphically and depend on the map reader to interpret them. For instance, a reader can estimate the relative distance along road between two cities, as well as the shortest path to take, identify the nearest school or clinic, the farm on the left and right side of the road and so on. Our mind can derive or interpret the spatial relationship from the map graphics. This is the traditional method of defining spatial relationship. In the case of a GIS which uses digital form to represent spatial data, these relationships are depicted using topology.

Topology is simply referred to as a mathematical procedure for explicitly defining spatial relationships which are generally divided into three types. These types are area definition, connectivity of lines and the contiguity of areas. This is the highest level of generalisation at which geographic features can be stored. By storing information about the location of a feature relative to other features, topology provides the basis for many kinds of geographic analysis without having to access the absolute location held in the coordinate files.

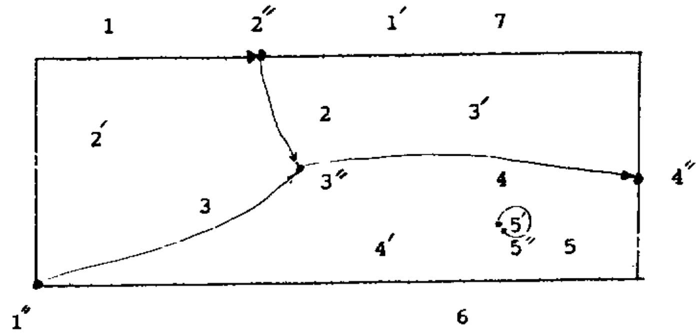
a) Area Definition

A closed area (polygon) is made up of a set of arcs which consists of lists of coordinates. Topologically, a polygon can be defined by listing all the lines which enclose it instead of listing all the coordinates in the polygon loop (figure 4). The polygon number 4 shown in figure 4 is defined by the 4 arcs that are arc numbers 3, 4, 6 and 5. The '0' in the arcs list is used to denote that the next arc represents an island (hole). The sign of arcs represents their directions in which '-' is used to show a reversed direction. This sign is important in order to build a closed loop for a polygon.

This type of topological data structure is knows as Polygon-Arc topology. The polygon/arcs list is automatically created and updated when the coverage is cleaned or built using a poly option. The advantage of using this type of data structure is that the arcs coordinates need only be listed once for each arc. It leads to the reduction of the amount of required storage and overcoming the limitation problem of how many coordinates can be used to define a polygon.

b) Contiguity

Contiguity is referred to the identification of polygons which touch each other or are connected (adjacent). The contiguous polygons are defined by a set of shared arcs between them. Every arc has direction which determines the left and the right side of the arc. It was mentioned earlier that every polygon and arc is assigned an internal number every time a coverage is built or cleaned. The polygon internal number can be used to refer which polygons is on the left and the right side of the shared arcs and list of contiguous areas for all arcs can be structured as shown in figure 5. The left-poly and right-poly items for each arc can be accessed from the Arc Attribute Table (AAT). This table is automatically created and updated when a coverage is cleaned or built.



Node
1, .. 7, Arcs Number
1' .. 5' Polygons Number
1" .. 5 Nodes Number
Arc direction
Polygon/arcs list

Poly#	No. of Arcs	List of Arcs
1	-	(outside map)
2	3	1, 2, -3
3	3	2, 4, -7
4	3	3, 4, -6, 0, 5
5	4	5

Arc/coordinates list

Arcs	X, Y Pairs
1	xa1 ya1, xb1 yb1, xcl
2	ycl..
3	xa2 ya2, xb2 yb2, xc2
	yc2..
	xa3 ya3, xb3 yb3, xc3
	yc3..

Figure 4 : Polygon-Arc Topology

Arc#	Left-Poly#	Right-Poly#
1	1	2
2	3	2
3	2	4
4	3	4
5	4	5
6	4	1
7	1	3

Figure 5 Arc Left- Poly and Right-Poly list.

Note: Refer to figure 4 for diagram.

c) **Connectivity**
This is another spatial relationship which refers to the identification of interconnected arcs. Connectivity defines how arcs are connected to each other and it can be represented by a list of arcs which meet at each node. It was mentioned earlier that nodes represent the beginning and ending point of an arc. It was also mentioned that all arcs are sequentially numbered after building or cleaning the coverage. The nodes of every arc are also sequentially numbered after the same process. If two nodes which represent the beginning or ending points of two arcs have the same value of x,y coordinates, these two nodes will receive the same reference number. The first nodes of arcs are known as Fnodes (From-node) and the last as Tnodes (To-nodes). Therefore, every arc must have from-node and to-node node. From-node node and to-node of all arcs of a coverage are arranged and stored in the AAT. This topological data structure is known as arc-node topology and Figure 6 below shows how arc-node topology is defined.

ARC#	FNODE#	TNODE#
1	1	2
2	2	3
3	1	3
4	3	4
5	5	5
6	1	4
7	2	4

Figure 6 From-node To-node list.

Note: Refer to figure 4 for diagram.

The connectivity is determined by comparing arc node numbers. If there is the same node number, it means that node is a common node and the arcs which share this common node are connected to each other (ESRI 1987).

ARC/INFO MODULES

ARC/INFO supports the variety of GIS applications by providing a huge of number of commands which are applied for various purposes. These commands are grouped together based on their main functions and saved into several sub-systems which are accessible from ARC/INFO/INFO main directory. For example, all the cartographic editing functions are organised in one module called ARCEDIT, while all the map output and drawing functions are grouped into another module named ARCPLOT. Thus, ARC/INFO main directory. For example, all the cartographic editing functions are organised in one module called ARCEDIT, while all the map output and drawing functions are grouped into another module named ARCPLOT. Thus, ARC/INFO contains a number of modules or sub-systems, each having its own set of commands and logical functions.

There are six main modules which consist of groups of commands for performing different tasks of GIS functions. The most important module is ARC from which the other modules such as ARCEDIT,ARCPLOT, INFO, MAPLIBRARY, ADS are accessed. The ARC module can be considered as the main environment system and the rest are the sub-system. A part from these modules, there are three other related software modules which are Network, Tin and Cogo which are used for network analysis, three dimensional surface analysis, and for design and layout purpose respectively. The brief discussions below concentrate on the functions of every module.

a) ARC : The ARC/INFO Main System

This is the main program environment in ARC/INFO from which the other sub-systems start off. It also consists of a large number of commands to perform several GIS functions which can be summarised as follows:

- Data conversion
- Data capture and automation
- Error refinement and verification
- Workspace and file management operations
- Transformation and projection of coordinates
- Management and manipulation of feature attributes
- Analytical operations

It is important to note here that all the above functions are executed without a graphics display. Thus, these tasks can be performed using a batch environment facility without requiring any graphic devices. This is the main difference between the ARC module and others which are interactive in nature and use graphic terminals and digitisers.

b) ARCEDIT : Edition Section

ARCEDIT is a unique graphics and database editor. the capabilities of CAD functions and the power of a geographic database are combined together in this module. This combination is very important not only for creation of high quality maps, but for the creation and maintenance of sophisticated geographic databases upon which the rest of ARC/INFO can operate. ARCEDIT combines all of the facilities for digitising map coverages, by a comprehensive set of editing commands.

The functions which operate in this section can be summarised as follows:.

- Editing Map features, and Texts.
- Creating a snapping environment.
- Performing Rubber sheeting
- Transferring attributes and producing statistics.

c) ARCPLOT: Mapping section

ARCPLOT is the interactive cartographic and mapping subsystem of ARC/INFO. This module provides facilities for drafting to produce a great variety of map graphics from simple screen displays to high quality cartographic plots for reports and presentations. ARCPLOT comes with an extensive symbol library, plus facilities for customising user's cartographic symbols, including a front editor, symbol editors for lines, points, shades, and text strings.

The scope of ARCPLOT applications can be summarised as follows:

- creating graphic multiple displays.
- Performing database queries
- Creating, composing and displaying high quality maps.
- Managing databases.

d) INFO : Database section

This is the database section of ARC/INFO. It uses the relational database model to manage the tabular data associated with map features. This directory provides facilities for creating, manipulating, and managing both types of databases. The INFO directory is integrated with other ARC/INFO modules (ARC, ARCEDIT, ARCPLOT, etc). Any changes made onto any coverage in the ARC directory will automatically change the spatial database of that coverage. This means that the relationship between each coverage feature and a corresponding tabular record is maintained.

INFO also permits the creation of tables or charts of various forms. These are flexible options which can be used for generating reports in several styles.

e) Other Modules

The ARC, ARCEDIT, ARCPLOT, and INFO are the main modules in ARC/INFO and they are used for almost all type of jobs. There are, however, several other modules which are only used for special purposes. The modules are Map Library, ADS, Network, TIN, COGO, LINEEDIT, SHADEEDIT, MARKEREDIT, and TEXTEDIT.

MAP LIBRARY is specially used for providing a unique system for managing large cartographic databases, such as those covering a region, state, or county. It uses a unique spatial library system for efficient insertion, storage, and extraction of geographic data. While the ADS (ARC Digitising System) subsystem provides full digitising capabilities for efficient data capture. The main function of NETWORKS, TIN, and COGO have already been described in the introduction of this section. The LINEEDIT, SHADEEDIT, MARKEREDIT, and TEXTEDIT are used to design line symbols, shade symbols, point symbols, and text symbols.

DISCUSSION AND CONCLUSION

The application of a hybrid data model gives several very important advantages of topological data structure and can be summarised as follows:

a) Polygon boundary data is efficiently stored as structured networks of line segments or arcs rather than as closed polygon loops conventionally used in polygon database structures. This arc-node structure cuts cartographic data storage almost in half.

b) The arc-node data structure also increases the speed of retrieval and processing of data. It is very important when performing some complicated functions such as polygon overlay, buffering etc.

c) Certain types of spatial analyses which simply are not practical without topological structure can be performed, including polygon redistricting, network simulation, optimal path determination and so on.

d) The conventional data file limits the maximum number of lines, points and polygon features and also the number of coordinates representing them. As a result, the size of tasks which are able to be performed is limited. By contrast, topological data structure removes the limits of the above constraints and allows for storage and processing of very large continuous map coverage (ESRI 1987).

References

[1] ARCNEWS, 1989, Using ARC/INFO for Engineering Applications, ESRI News buletin Spring 1989.

[2] BORKIN, S.A, 1986, Database models: A Semantic approach for data base systems (LONDON; MIT)

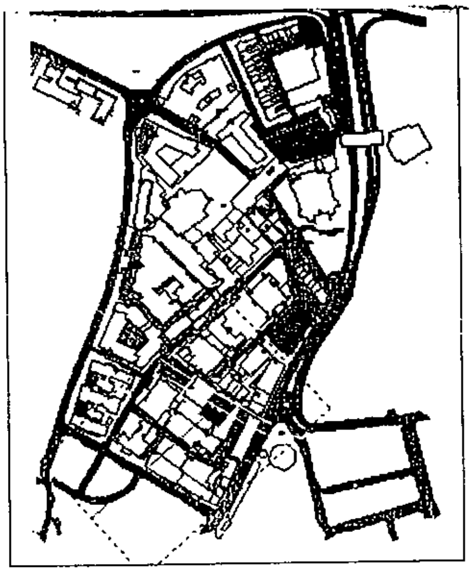
[3] DANGERMOND, J., 1986 The software toolbox approach to meeting the user's needs for GIS analysis. Proceedings of GIS workshop, Virginia 1986.

[4] DEEN, S,M., 1985 Principles and Practice of Database System (LONDON; Macmillan)

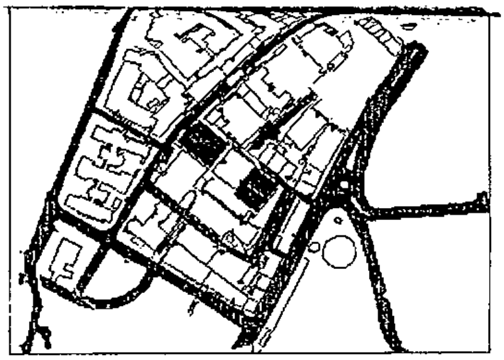
[5] Mohamed, A.H 1989, The operation, application and development of Geographic Information System, M. Phil Thesis, NewCastle UN.

TEST PROJECT

This project was carried out in order to test the efficiency of the ARC-INFO in handling various types of query. Some of the result are as follows : (test is the Newcastle University area).



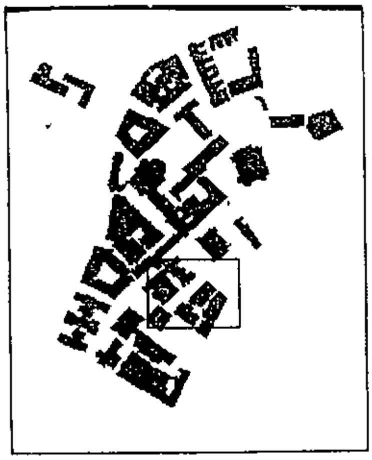
Creating a line buffer
Buffer distance is 100 metres



Features inside buffer zone

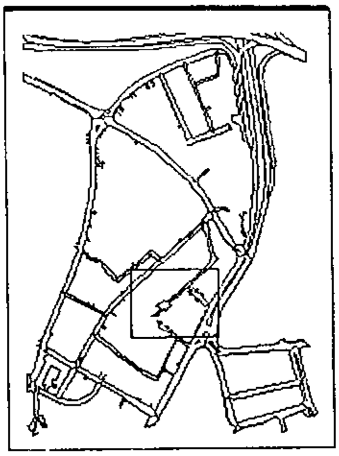
Fig. 7 : Application of the Instant Buffer program

Creating a buffer zone with any fixed size. All the affected features inside the area can be clearly highlighted and a detail report can be produced. (Report is out attached here).



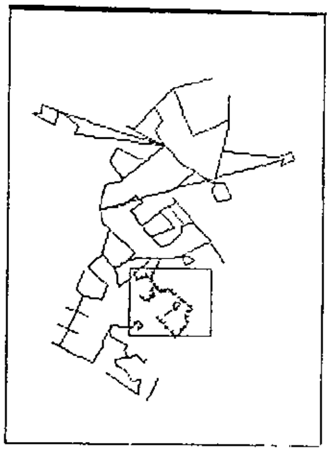
Current land use : Building

Inside area
Outside area



Existing roads

Inside area
Outside area

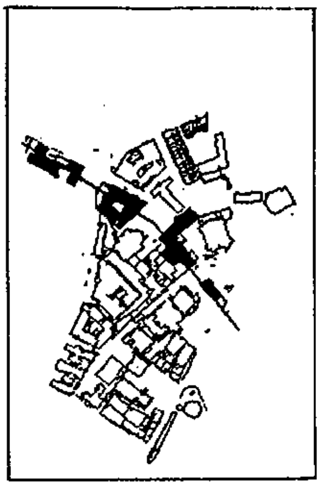


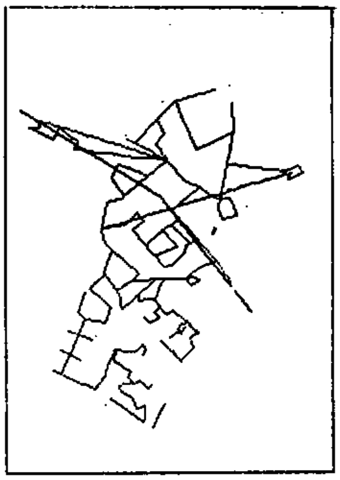
Existing heating-lines

Inside area
Outside area

Fig. 8 : Examples of existing land use and utilities

The above fig. shows a proposed route crosses an area. All features which are crossed by this proposed route will be shown in different colour. A detailed report of the affected features can be sent to printer or plotter for hard copy production.



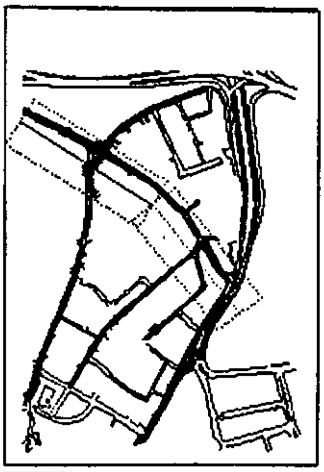


—— Proposed routes

Buildings crossed by proposed routes

—— Proposed routes

—— Heating lines crossed by proposed routes



Roads within 50m from the proposed routes

Fig. 9 : Examples of routes planning

An existing pipeline passes through an area. We are able to list down detail information about each feature crossed by this pipeline. Graphically, these features can be displayed in different colours.

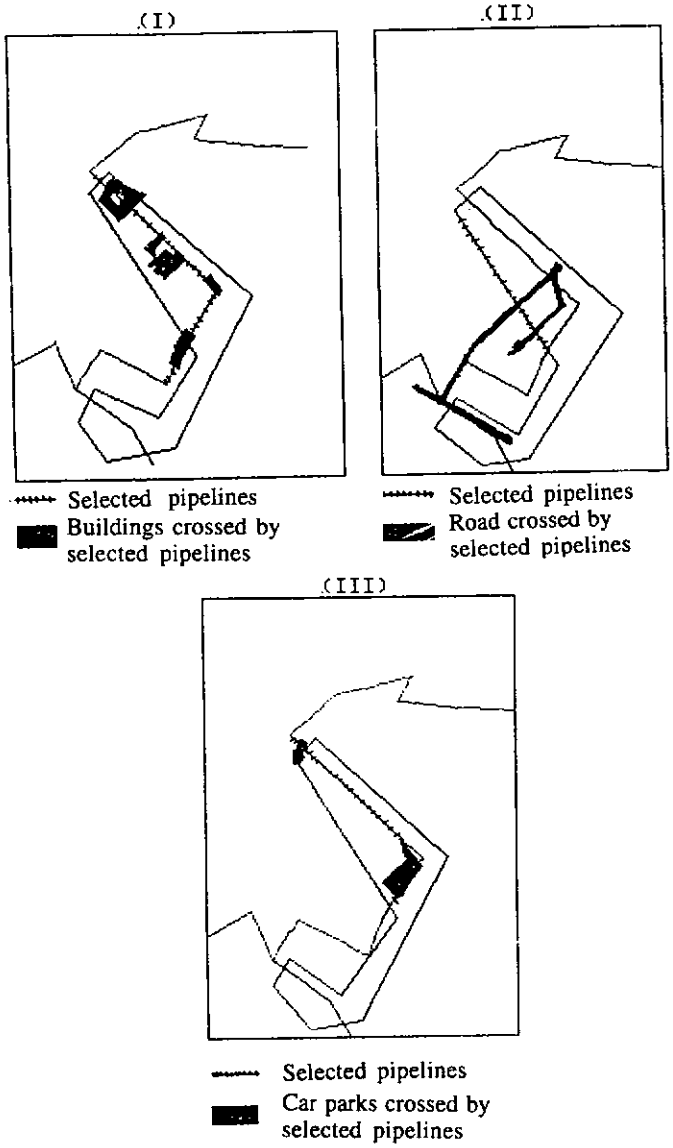
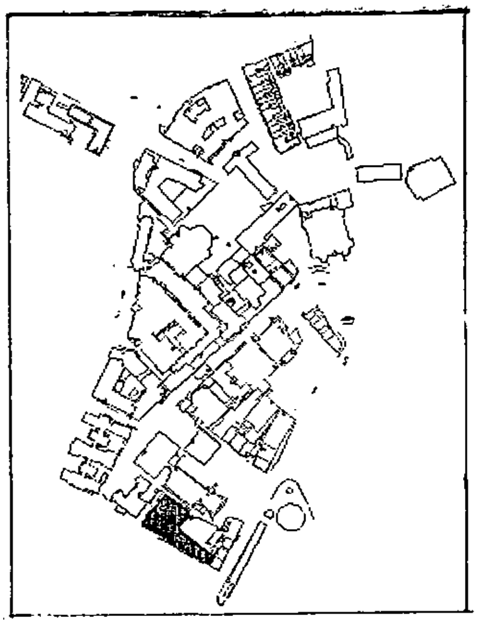
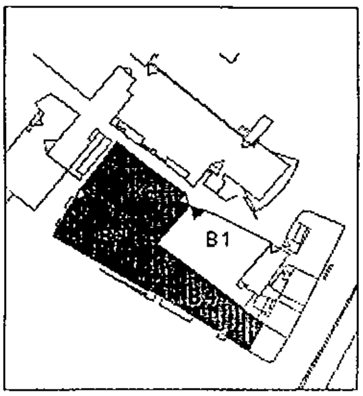


Fig. 10 : Examples of pipelines maintenance

The application of adjacency concept is GIS is very important especially for designing and planning purposes. B_1 is the selected building. Users can extract detail information of all adjacent buildings i.e B_2, B_3, B_4. This concept can be applied for land development project. Users can easily determine the status and other important information of the adjacent lot.

The above figures (7 -11) show some of the capabilities of the ARC-INFO in handling various types of quearies in order to cope with a number of applications.





B1 The selected building
B2, B3 & B4 The adjacent buildings

Fig. 11 : Application of the NEIGHBOUR program