



GIS for the **CURIOUS**

A Hands-On Guide to Geographic Information Systems

Y.C. Lee

Geomatics Canada - Universal Systems Limited - University of New Brunswick

G
70.2
L43
1995
omgre

This document was produced
by scanning the original publication.

Ce document est le produit d'une
numérisation par balayage
de la publication originale.

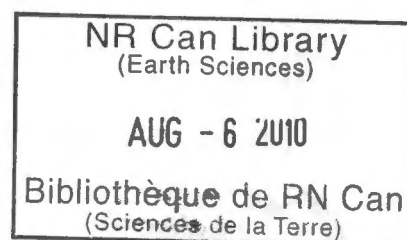
GIS for the Curious

A Hands-On Guide to Geographic Information Systems

Y. C. Lee

Copyright © 1995 by Geomatics Canada

All rights reserved. No part of this publication can be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.



To the Team

Albert, Jamie, Jennifer, Kasia,
Marta, Wendy.

To the Models

Caroline, Jonathan, Michael.

Table of Contents

PART 1 INTRODUCTION

Chapter 1	What is a GIS	1
Chapter 2	Components of a GIS	2
Chapter 3	GIS Applications	4
Chapter 4	Starting the Demonstrations	7
Chapter 5	Starting the GIS	8
Chapter 6	The CARIS Information Manager	11
Chapter 7	Getting Help	13
Chapter 8	Viewing the Geographic Database	14

PART 2 COORDINATE SYSTEMS

Chapter 9	Geographic and Ground Coordinates	19
Chapter 10	Coordinate System and Distance Measurement	26
Chapter 11	Coordinate System and Area Measurement	35
Chapter 12	Resolution	37

PART 3 REPRESENTATION OF FEATURES

Chapter 13	Different Perspectives of Reality	41
Chapter 14	Vector Representation	44
Chapter 15	Raster Representation	51

PART 4 RETRIEVAL

Chapter 16	Retrieve Information on Features	59
Chapter 17	Display Selected Features	70
Chapter 18	Feature Attributes	75
Chapter 19	Vertically Linked Databases	88
Chapter 20	Select Features Within an Area	90
Chapter 21	Thematic Mapping	93

PART 5 TOPOLOGY

Chapter 22	Topological Relationships	97
Chapter 23	Use Topology to Identify Polygons by Pointing	106

PART 6 ANALYSIS AND MODELLING

Chapter 24	Overlay Analysis	111
Chapter 25	Raster Conversion	117
Chapter 26	Select Features within a User-Defined Zone	121
Chapter 27	Retrieval Using Buffer Zones	127
Chapter 28	Reduce a Selection	133

Chapter 29	Underground Gas Tank Leakage: Two Variations	136
Chapter 30	Area Affected by Underground Gas Tank Leakage	142
Chapter 31	Overlay Analysis Involving Points	147
Chapter 32	Generate Buffer Zones Around Lines	152
Chapter 33	Manipulate Two Topological Layers	155
Chapter 34	More For You to Do	157
Chapter 35	Network Analysis	158

PART 7 DATA COLLECTION AND EDITING

Chapter 36	Data and its collection	163
Chapter 37	Accuracy and Data Editing	170
Chapter 38	Cartographic Enhancements and Data Editing	172
Chapter 39	The GIS Editor	174
Chapter 40	Move Features	175
Chapter 41	Modify Shape of a Line	181
Chapter 42	Line Splitting and Resolution	190
Chapter 43	Modify Text	195
Chapter 44	Add Point Features	198
Chapter 45	Add Contour Labels	203
Chapter 46	Add Line Features	206
Chapter 47	Topology and Editing	214

PART 8 DATA SHARING

Chapter 48	Data Conversion and Integration	219
Chapter 49	Transformations	231

PART 9 WRAPPING UP

Chapter 50	Create a Geographic Database	237
Chapter 51	A Survey of Concepts	244
Chapter 52	Where to Go From Here	256

APPENDICES

Appendix A	Glossary of Terms	259
Appendix B	Data Descriptions	269
Appendix C	Database Table	274
Appendix D	Tips for Running CARIS	276

INDEX 277

About this Guide

This guide on GIS is for the curious. It is for those who want to learn GIS concepts, use actual GIS software to obtain hands-on experience, and gain an insider's view of how GIS software operates.

We have provided animated demonstrations, exercises, questions, and technical notes to help you.



The demonstrations are animated movies that are packaged independently but are designed to be used in conjunction with the guide. Some of them explain concepts, while others illustrate operations. They complement explanations in the text but often provide alternative examples.



The exercises are for you to follow using the GIS software that accompanies this Guide. Some of them have been animated.



The questions encourage you to use what you have learned to explore the other capabilities of the GIS software and learn new concepts.



The technical explanations probe deeper into the design and operation of the GIS software. They are placed at the end of the chapter.

A ♦ terminates the text for the exercises, questions, and technical explanations.

Many GIS concepts are covered, and a survey of them is provided at the end of this Guide. When you encounter a new concept while going through the text, you might want to relate it with others by consulting the chapter on A Survey of Concepts.

You can use the guide in conjunction with the accompanying software or you can read it by itself. We have designed this guide to be used more than once. During the first time, follow the text and the exercises to build up the necessary concepts and skills to operate the software. Run the demonstrations to supplement what you have read. You might then read the guide by itself and find that you have a much better grasp of the concepts the second time after you have gained experience using the GIS software. When you are familiar with the concepts and the software, explore on your own to discover the vast world of GIS.

About the Software

The animated demonstrations were created using Macromedia Director and were packaged to run by themselves. There is no need for you to purchase the Director software.

The exercises are based on CARIS, a GIS developed by Universal Systems Limited. A customized version of it has been created for this Guide having identical interfaces as the production version but with somewhat reduced functionalities. In addition, the customized version can only work with the data files supplied with this Guide. You can also use a production version of CARIS with this Guide.

About the Data

This Guide comes with all the data that you need to run the demonstrations and the exercises. Natural Resources Canada, the New Brunswick Geographic Information Corporation, and the District 18 School Board of New Brunswick have kindly granted us permission to use their data for this Guide. Some additional data were obtained from the United States Geological Survey. A description of the files and their contents are provided in Appendix B.

The municipality database is for the City of Fredericton, New Brunswick, Canada. Although we have based the data on public records, we have modified some of the entries as a courtesy to the parties involved.

GIS for the Curious Software and Documentation Order Form

To order software and documentation for GIS for the Curious, please complete the form below.

1. Please enclose:

- Completed order form.
- Payment in Canadian funds (plus applicable freight and taxes) for orders within Canada, or
- Payment in US funds (plus applicable freight) for orders outside of Canada.

2. Print your name, shipping address, and phone number:

Name _____

Company Name _____
Shipping Address _____

City _____
Province/State _____ Postal Code/ZIP _____
Daytime Phone () _____
Fax () _____
E-Mail _____

Calculate your total cost:

Interactive GIS Software & tutorial dataset \$ 100 x () = _____
(distribution on CD)
Additional GIS handbooks & animation \$ 90 x () = _____
Freight * _____
Sub-Total _____
GST for Canadian Orders (cost and freight by 7%) _____
Sub-Total _____
New Brunswick PST for NB residences (sub-total by 11%) _____
Total Cost _____

*Payment method is cheque or money order. Do not send cash or purchase order.
Make cheques payable to Universal Systems Ltd.*

* Canadian Freight: Handbook/Animation/CD FEDEX \$ 14.00 each
US Freight: Handbook/Animation/CD FEDEX \$ 37.00 each
International Freight: Handbook/Animation/CD FEDEX \$ 92.00 each

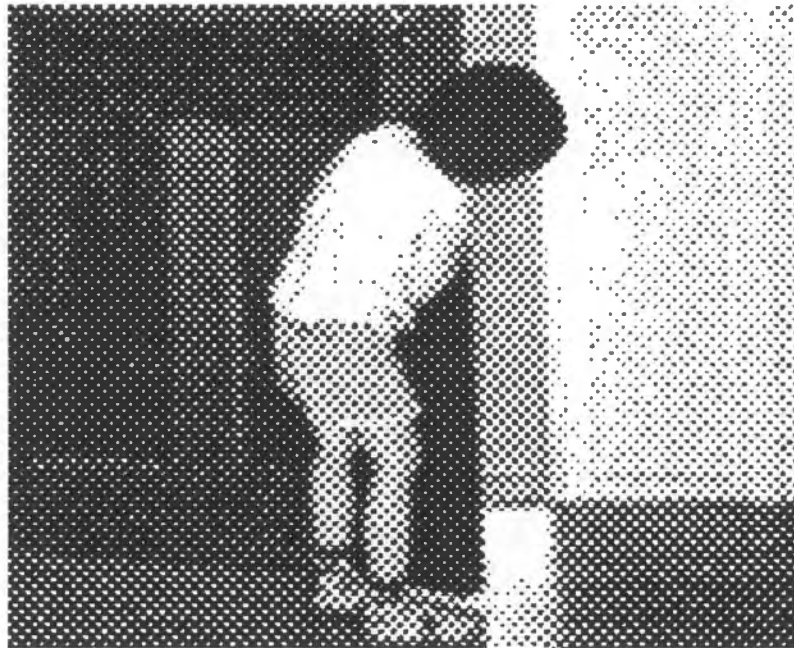
4. Please send all items to:

GIS for the Curious
270 Rookwood Avenue
Fredericton, New Brunswick, Canada
E3B 2M2

If you have any questions about this offer, call Universal Systems Ltd. at (506) 458-8533,
fax (506) 459-3849 or E-Mail sales@universal.ca.

Please allow 2 - 4 weeks delivery upon receipt of this order form.

Part 1
Introduction



Chapter 1

What is a GIS?

A Geographic Information System (GIS) is a computer system that provides geographic information to support decision making. Similar to other information systems, it offers functions to input, manage, retrieve, analyse, and output data.

A GIS is unique in that geographic data are rather different from traditional data used in business and similar applications. Most geographic data are multidimensional. For example, it takes at least two coordinates to specify a location. When height is involved, a point becomes three dimensional. Moreover, geographic objects might change with time. The addition of this temporal parameter makes geographic objects four dimensional.

The unique characteristics of geographic data require specialized functions which are not normally provided by traditional Database Management Systems (DBMS). Among them are cartographic functions, geographic analysis, and geographic modelling.

Cartographic functions are important to human interface design in a GIS. We often interact with a GIS through a map display and visualize geographic data in graphic forms.

Geographic analyses, such as finding the quickest route from one city to another, are more complex than most of the statistical functions found in DBMS.

Geographic modelling, the ultimate tool in supporting decision making by allowing us to simulate real world situations, is highly specialized. This kind of modelling, such as predicting the effect of an oil spill, is particularly useful when a corresponding real life undertaking is prohibitively expensive or risky.

Although GIS are very different from DBMS, data management is an important module of a GIS since the operation of an information system depends heavily on the availability of data. Hence many GIS, such as Arc/Info, Intergraph, Vision and CARIS do incorporate off-the-shelf DBMS as part of their systems. This is mainly to reduce the development cost of the GIS software.

Chapter 2

Components of a GIS

A GIS is composed of modules to manage, process, analyse, and model geographic data.

Data Management: This module organizes data in the database and retrieves them from the database for use in other modules. It has the following subcomponents.

Organization: Data are organized to maximize efficiency and minimize storage space. A collection of related data collected with a purpose and serving a specific application is called a *database*. Geographic databases, particularly those of national coverages, are often very large. Hence the demand for efficient data storage on a GIS is particularly stringent.

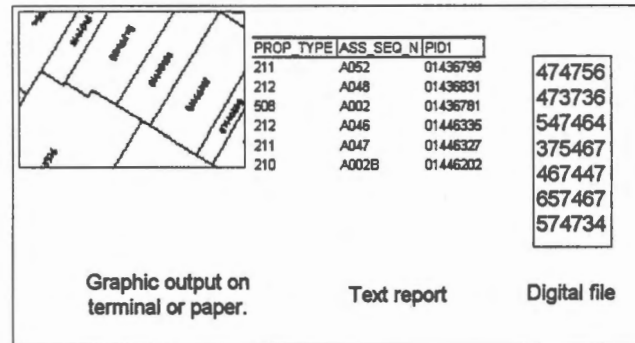
Retrieval: Data are retrieved from the database for use in other operations. A retrieval is invoked by the user entering search conditions that could include spatial or nonspatial components. When we search for all properties for sale within one kilometre of a certain school, we are using a spatial condition "within one kilometre" and a nonspatial condition "properties for sale". Depending on the interface used, spatial conditions could be entered graphically. For instance, to specify the condition "within one kilometre", the user might be asked to draw a circle of one-kilometre radius over the display. Nonspatial conditions are entered as text or numbers.

Input and conversion: Data are converted to a digital form that can be stored and processed by computers. Some land surveying instruments record data in digital form. If we collect data from aerial photographs, satellite imageries, and existing maps, special equipment has to be used to convert analogue graphics to digital data. Data can also be imported from other systems, often requiring the help of a format converter to make it compatible with data in your system.

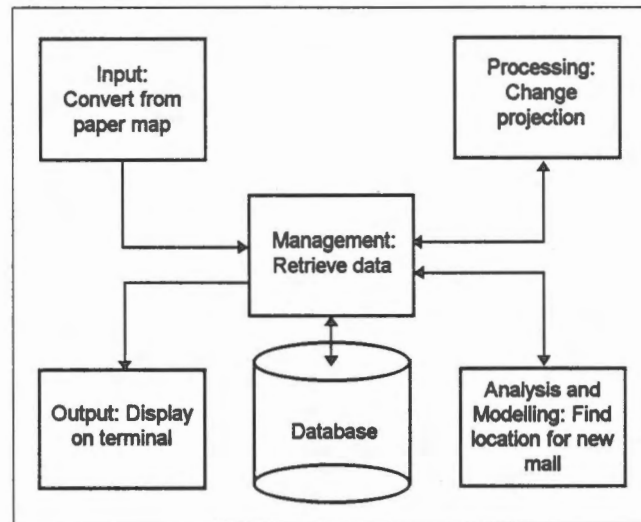
Processing: Data are updated, corrected, and modified to suit different applications. Common geographic processing includes scale changes, map projection changes, clipping, and cartographic enhancements.

Analysis and modelling: Analytical and statistical operations are performed on data to derive new information; present and future situations are modelled and simulated to support decision making. There is such a wide variety of geographic analysis and modelling functions that no GIS can yet support all of them. As a result, many of the more specialized functions must be custom-built into a GIS by the user.

Output: Results of the operations are displayed or exported to another system. Examples of output are graphic displays, text reports, and digital files.

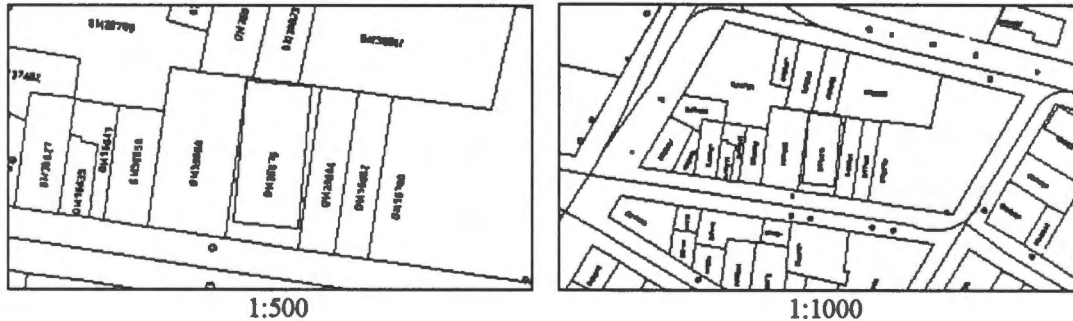


Data: Data are a very important element because without them a GIS is like a car without fuel. The flow of data through the GIS software, with an example given for each module, is shown below.

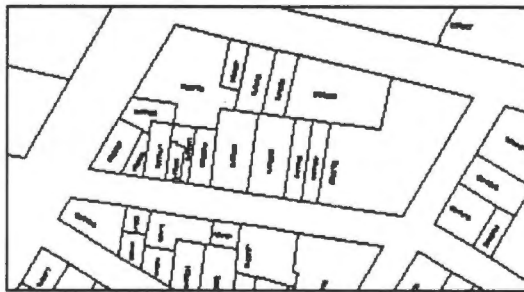


Chapter 3 GIS Applications

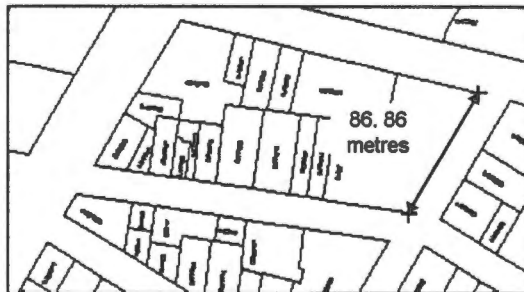
The power of a GIS lies in the geographic database that can support many different applications. It can be used to generate maps of different scales with different contents. For example, the database can generate a display at a scale of 1:500 or 1:1000.



It can also generate a map consisting only of land parcels or other features of your choice.



You can make measurements such as length, area, and angle.

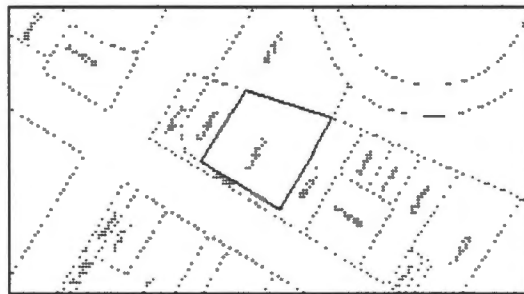


You can retrieve information from the database, such as a description of the property highlighted in the diagram.

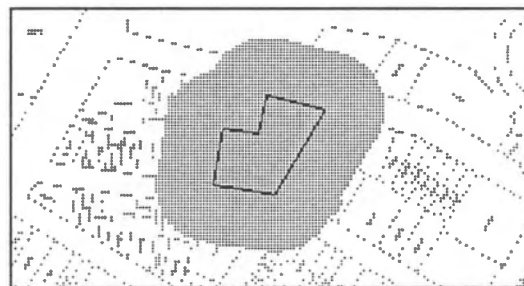


Owner Name: Irving Oil
Street: 181 King St
Postal Code: E3B 1C8
Description: Gas Station

You can enter a search condition, such as POLICE STATION, and let the GIS retrieve and highlight its location.



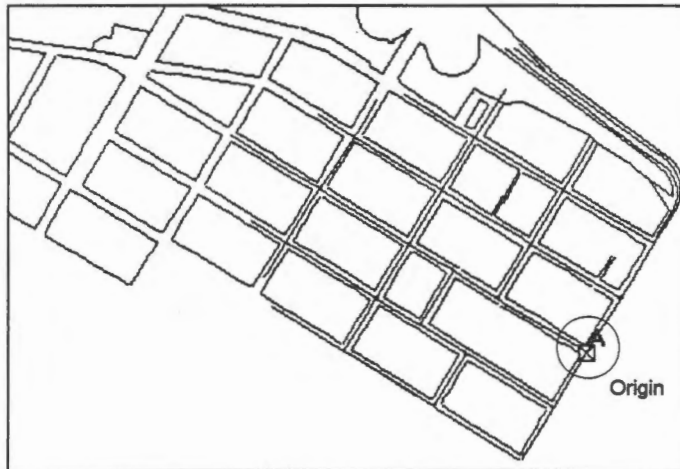
You can simulate a situation, such as the extent of an underground gasoline tank leakage from this gas station.



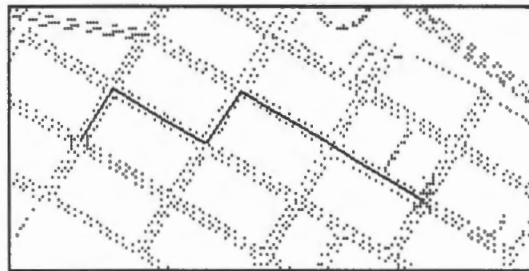
From this, you can then retrieve all properties that are affected by the leakage.



You can calculate the streets you can cover within a driving time of 10 minutes from any given origin.



You can determine the shortest route from one point to another.



You can do many more tasks. The challenge of a GIS application is that there are many ways to perform a task. On the other hand, no GIS can provide the tools to solve all your problems. However, the tools they do provide are general enough to solve most of them, and if you do want something special, you can always customize or extend a GIS to suit your needs.

Chapter 4

Starting the Demonstrations

 This symbol indicates an animated demonstration. To run the demonstration, double click on the GISADV icon to open it.



Then double click on the Demo icon



The following controls are available in the demonstration. STOP stops the animation, REWIND plays the animation from the beginning, BACK backs up to the previous pause, and PLAY continues with the animation.



Chapter 5

Starting the GIS

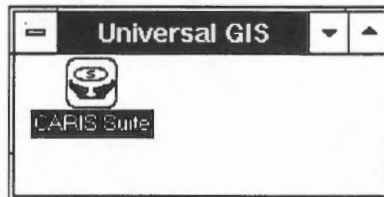
We will illustrate a number of GIS features in the following chapters. First of all, we will show you how to start the GIS software called CARIS, which we have used for all our exercises and examples.

To start CARIS, first locate the Universal GIS icon on your desktop, and open it by double clicking.



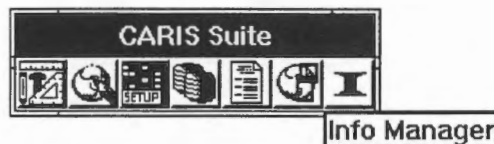
Universal GIS

This will expand the Universal GIS window showing the CARIS Suite icon.



Double clicking it will show the CARIS suite of programs.

If you move your cursor over a Suite icon and leave it there momentarily, the meaning of the icon will show. Double click on the Info Manager icon to open the CARIS Information Manager. This icon is specially designed for this Guide.

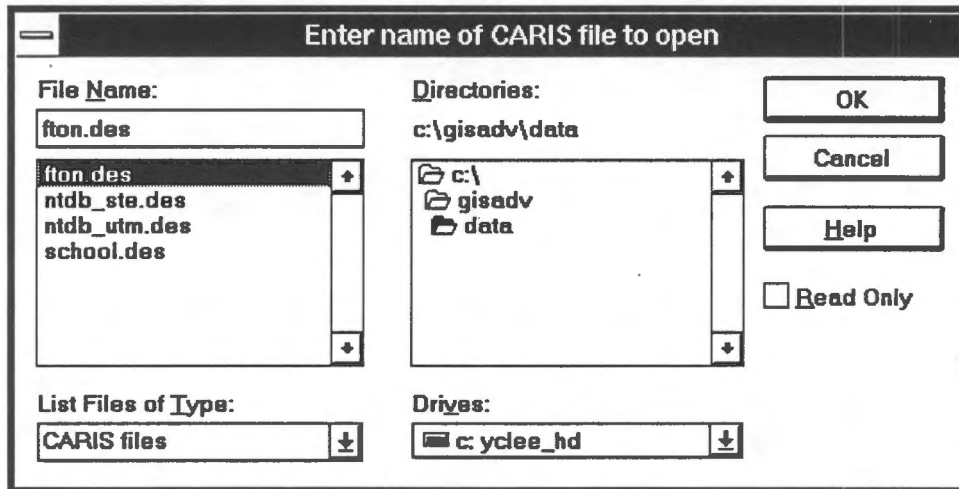


If you are using a production version of CARIS, the Info Manager icon will not show. You should use the following procedure to invoke the GIS.

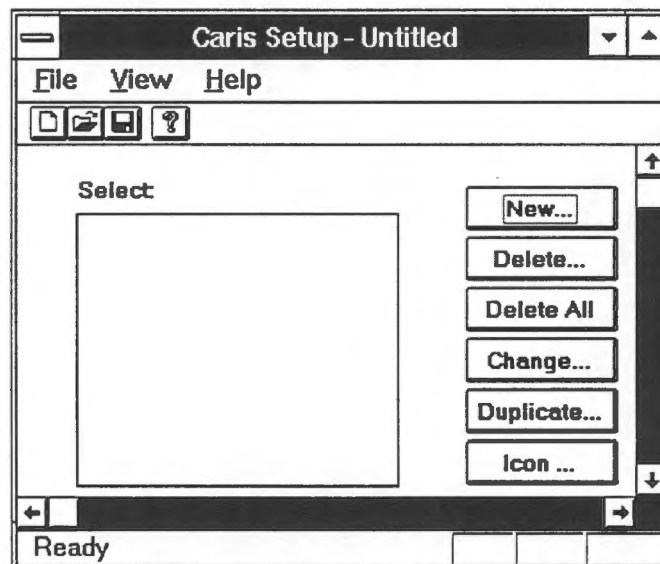
1. Click on the CARIS Manager Zones icon on the CARIS Suite.

2. Select FTON.DES on c:\gisadv\data and click on the OK button.

NOTE Your file might be on a different directory.

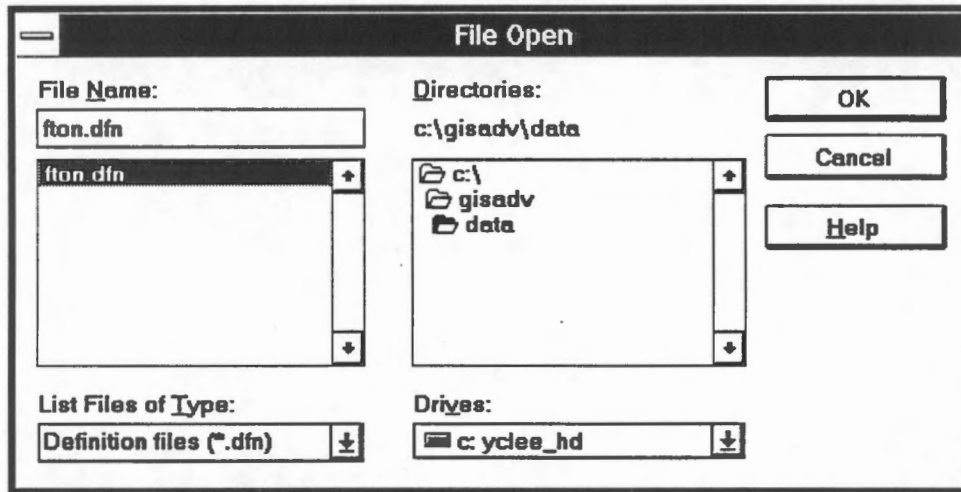


3. The CARIS Information Manager window opens.
4. Click on the Database Manager icon on the CARIS Suite. This opens the CARIS DB Manager window.
5. Click on the CARIS Setup icon on the CARIS Suite. This opens the CARIS Setup panel.



6. Select the File menu item and select the Open ... menu item under it.

7. In the File Open panel, select FTON.DFN in c:\gisadv\data and click on the OK button.



NOTE Your file might be on a different directory.

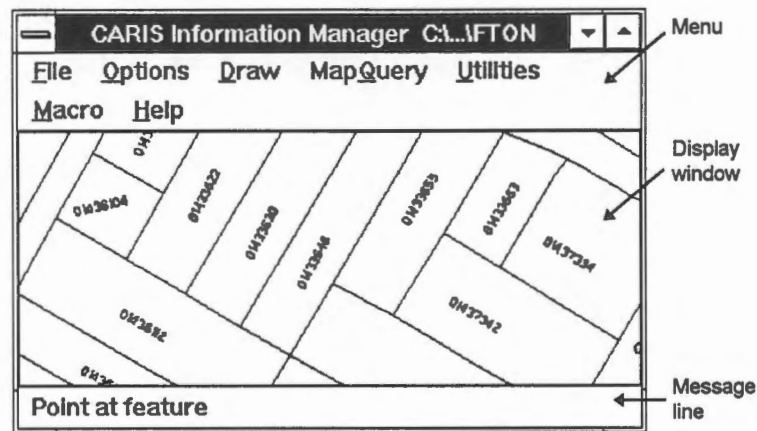
8. Close the CARIS Setup panel.

Chapter 6

The CARIS Information Manager

The CARIS Information Manager is a module for retrieving and analysing geographic data. It provides two windows and a panel for interaction, which are opened automatically when you invoke the software through the Info Manager icon mentioned in the previous chapter.

The *CARIS Information Manager* window displays the map. We will sometimes refer to it as the *Map Window*.

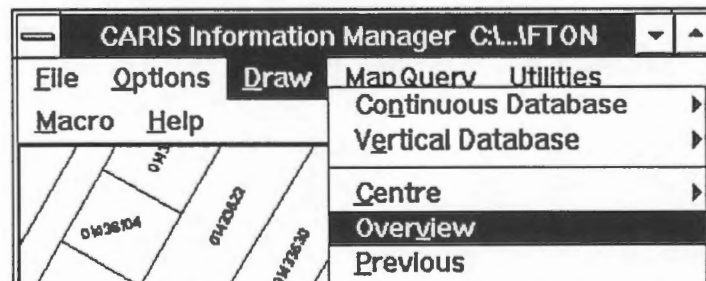


NOTE At the bottom of the window is a message line. One message remains until it is replaced by another. To avoid confusion caused by leftover messages, read the messages just before or after an operation.

Menu items are often nested. For example, to select the Overview menu item, you first have to select the Draw menu item. We will use the notation

Draw > Overview

in this Guide to show nested menu selections.

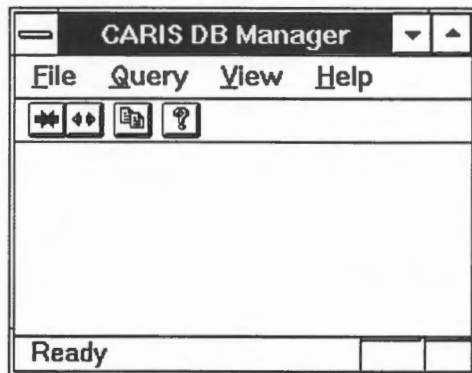


The *CARIS Panel* contains a set of the more commonly used commands for the map window. For example, the Zoom In icon (first one) corresponds to the Draw > Zoom In

menu item in the Map Window. Some of the icons, however, do not have corresponding menu items. They will be pointed out in the exercises.

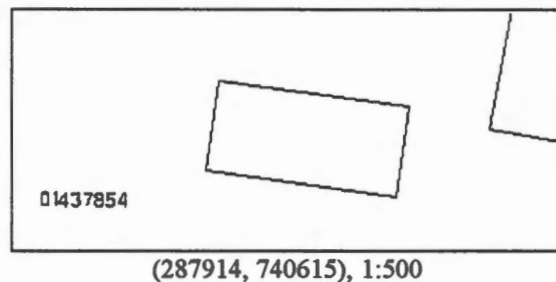


The *CARIS DB Manager* window is for interaction with the database which contains textual and numerical information complementary to that shown in the map window. This window and the map window together allow the user to query the GIS database using both spatial and non-spatial conditions.



We will use both the left and the right buttons of the mouse to interact with the GIS. The left one is used to select, enter, and confirm. It is equivalent to the Y key (for Yes) on the keyboard when used in that context. The right one is used to terminate, cancel, and reject. It is equivalent to the Q key (for Quit).

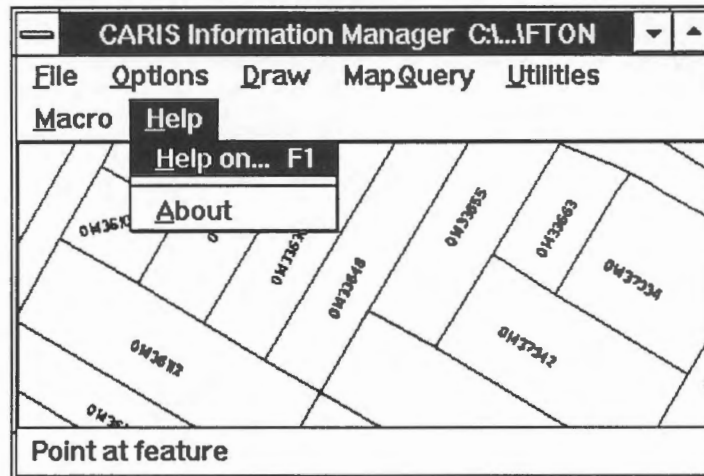
When we show a map of the database, we will indicate the location of its centre and its scale. They provide a guide for you to display the same area on the screen.



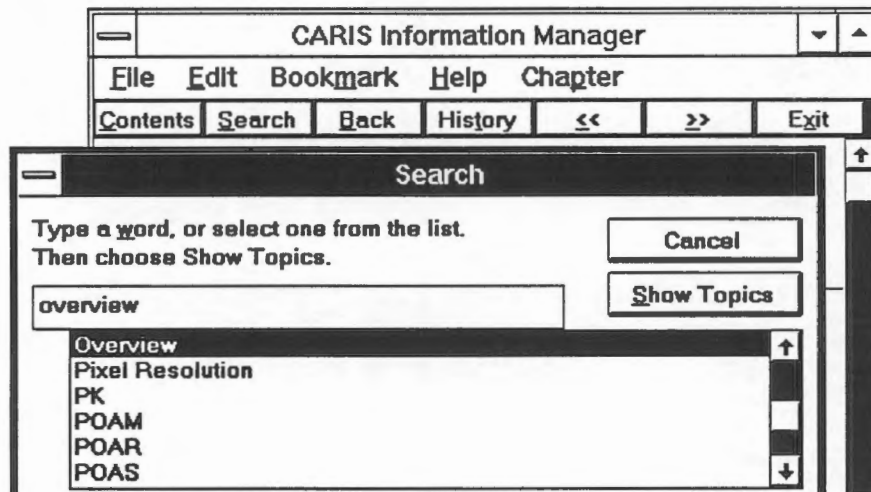
NOTE The maps might have been clipped or reduced to fit the page of this Guide.

Chapter 7 Getting Help

You can get help on how to use the GIS commands by selecting Help on the menu.



If you know the name of the command and need more information about it, you can search for that command to find the topic you want.



NOTE You *cannot* invoke Help when you are in the middle a command.

 The demonstration Using the Help Facility shows how to get help about the command to measure the distance between two points.

Chapter 8

Viewing the Geographic Database

The geographic database stored on the computer disk consists of numeric codes to represent the location, shape, and other information about geographic features. A printout of the database files will not yield anything comprehensible but will appear to be a collection of seemingly random numbers.

```
38478398923738943734289327468546754654484854948484389
43745834754897967959058688975787532847547589476857695
76957857689758937237465456746575657576857669767608909
20292182397438743856746548979685960796706706976704594
93232734734845765868898003291238122347836416477546858
65978698796849859845747387454666356373845847349754960
65486956859685849484856759903849384583984593845348599
```


The output modules of a GIS can interpret the format with which these numbers are organized, and it will display the numeric information in a form understandable by us.



The map display is a window into the database. It is in many ways similar to a traditional map except that it is much more dynamic — we can easily change the scale of the display, exhibit different part of the database, or show only selected features.

There are two basic operations we perform to modify the display: zooming and panning. Zooming changes the scale; panning moves the display window to show another part of the database.

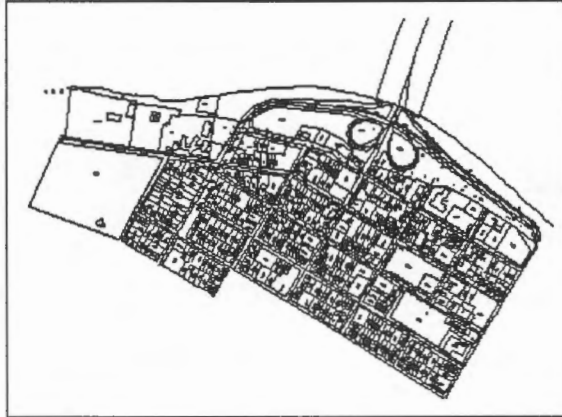
Three commands can be used to zoom in or out. The *Zoom In* command enlarges the scale, the *Zoom Out* command reduces the scale, and the *Draw Overview* command selects a scale that would fit the entire database on the display window.

 Draw the entire database.

To practice zooming, first use the *Draw > Overview* menu item or the Overview icon on the CARIS Panel.

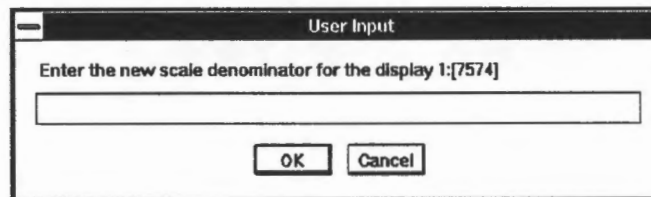
NOTE You can stop the drawing on the screen anytime by clicking on the left mouse button. You would end up with a screen half filled with the map, but all GIS functions can perform as usual.

When you first start the system, the entire database as shown below might have already been displayed. In that case, the Draw Overview command would create an identical image.




 Show the scale.

You can find the scale the GIS had used for this display through the Options > Display > Scale ... menu item. The scale is shown within brackets on the User Input panel. Since we are not entering a new scale here, click on the Cancel button.




NOTE The scale for your display might not be 1:7574 if you have resized the map window. The Draw Overview command chooses a scale that will make optimal use of your window.



 Change the display scale.

Follow the same procedure as in the previous exercise, but enter a scale of 2000 (meaning 1:2000), and click on OK to exit. The display will be refreshed automatically. ◆


 Enlarge a selected region.

You can enlarge a display by using the previous procedure and entering a larger scale. You can also select a rectangular region on the display to enlarge it. To do this:

1. Mentally select the rectangular region to enlarge.
2. Select the *Draw > Zoom In* menu item.
3. A crosshair will appear. Move it to one corner of the desired rectangular region and click on the left mouse button.
4. Move the mouse to drag to the diagonal corner of the rectangle. When the rectangle is of the right size, click on the left mouse button.
5. The area within the rectangular region will be enlarged to fill the map window.

NOTE The display might show an area a bit larger than the rectangular region you have selected. This is to produce a display that matches the width-to-height ratio of the display window.




 The demonstration *Change Scale and Display Window: Zoom in Operation* shows how to use the *Zoom In* command.


 What is the effect of the *Zoom Out* command?

Use the *Draw > Zoom Out* menu item to reduce the scale. You will notice that each invocation of the *Zoom Out* command reduces the scale by a factor. Repeat the *Zoom Out* command a few times to determine this factor. ♦

 **Panning.**

Panning is done by redefining the centre of the next display. To practice panning, zoom into an area and select the *Centre* icon on the *CARIS Panel* or the *Draw > Centre > By Pointing* menu item. Move the crosshair to the spot you would like to be the new centre and click on the left mouse button. The screen will refresh automatically. ♦

 The demonstration *Change Scale and Display Window: Pan by Pointing* shows how to change the display centre by pointing.

 Speed of screen refresh.

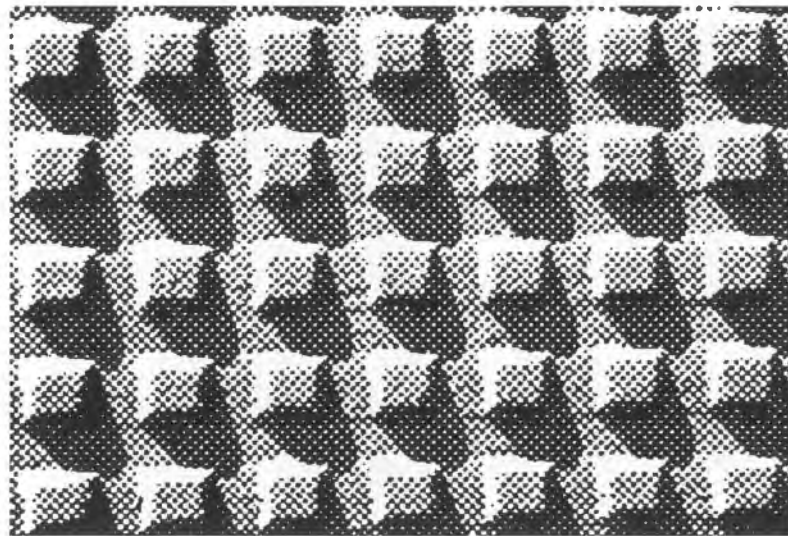
Notice the speed at which the screen refreshes after a scale change. It is faster to refresh a window containing a smaller number of features. Hence the Draw Overview command, which draws the entire database, will take the longest time to refresh.

It would seem obvious that it takes less time to draw a smaller number of features. However, a computer system such as a GIS must spend time to examine the features in the database and decide whether they should be shown in the display window. A simple but very inefficient strategy is to examine each and every feature in the database. When you are displaying very little from a very large database, you will notice significant delays in the display caused by the system spending time stepping through features that fall outside the window.

One reason that you do not notice such long delays here is because the database you are using is not that big. The major reason, however, is that the GIS makes use of a spatial indexing system to help reduce the number of features that need to be examined. *A spatial indexing system*, like the index at the end of a book, is a way to help a GIS look up spatial data in the database as quickly as possible.

Part 2

Coordinate Systems



Chapter 9

Geographic and Ground Coordinates

Underlying each geographic database is a coordinate system from which the position of every point in space can be measured. We will explore the characteristics of coordinate systems in this chapter.

 Read the coordinates of a point.

1. Select the MapQuery > Point Location menu item.
2. Place the crosshair on the spot for which you want coordinates and click on the left mouse button.
3. The readout shows three different measurements: geographic (Lat, Long), ground (XG, YG), and scaled ground (X,Y). The meaning of them will be explained later in this chapter.
4. Click on the OK button to exit from this panel.
5. The crosshair will reappear. Click on the right mouse to exit from this command.

Coordinates of a Point		
X: 5770460	XG: 288523.00	LAT: 45-57-46.427N
Y: 14806267	YG: 740313.35	LONG: 66-38-53.057W
	ZG:	<input type="button" value="OK"/>

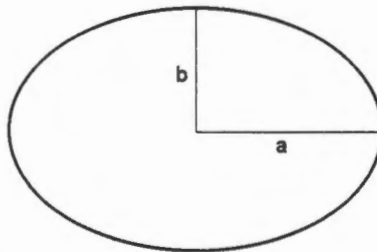
NOTE After selecting a spot to obtain its coordinates, a cross will be left behind as a reminder. You can refresh the screen to get rid of it by selecting the Draw > Refresh menu item or the Refresh icon on the CARIS Panel.

 The demonstration Specify Position: Reading the Coordinates of a Point shows how to obtain the coordinates of a point.

Geographic coordinates (Lat, Long) are measured over an ellipsoidal earth, created by rotating an ellipse about its shorter axis through which the North and the South Poles are connected.



Such an ellipsoid can be completely defined by two measurements, a and b , where a is called the *major semi-axis* and b the *minor semi-axis*.



Many such mathematical bodies, called *reference ellipsoids* in mapping, have been defined to form the basis of measuring degrees of latitudes and longitudes. Different ellipsoids are used for different parts of the earth because none can represent the earth perfectly, although they all do a very good job. Over time, new ellipsoids are also being created as we improve our knowledge about the true shape of the earth.

Most mapping in North America has been based on the Clarke Ellipsoid developed in 1866, which provides a good fit for North America. Recently, the GRS (Geodetic Reference System) ellipsoid was developed in 1980 to provide world-wide coverage. The two ellipsoids are different in shape as indicated by the different a and b values (rounded to the nearest metre) used:

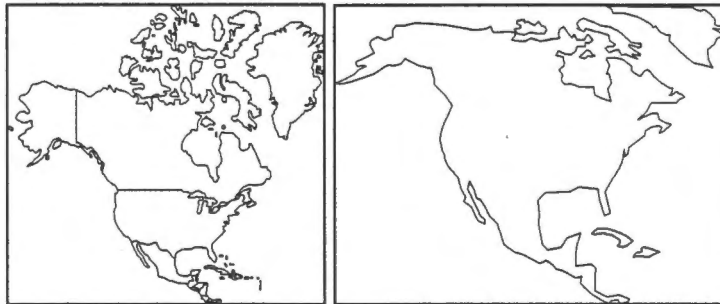
	a	b
Clarke 1866	6 378 206 m	6 356 584 m
GRS 80	6 378 137 m	6 356 752 m

Ground coordinates (XG, YG) are not measured in degrees but usually in metres. Although the two coordinates are labelled as XG and YG here, a more common way is to label them as Eastings (E) and Northings (N). For instance, (40 000 mE, 30 000 mN)

denotes a point measured 40 000 metres from the origin towards the east and 40 000 metres towards the north. The origin is usually placed at the southwest corner of the area to be mapped, so that there is no need to measure south or west of the origin.

Ground coordinates are not measured on the curved surface of an ellipsoid but on a flat surface like a map. The techniques of transforming an ellipsoidal earth to a flat map, called *map projections*, involve complicated calculations. There are many different kinds of map projections being used for mapping, and the two shown here are just two examples.

It is mathematically impossible to preserve the true shape, size, length, and angle on a flat map. When a map covers a large area, these distortions can be significant, as shown in the two maps below. Different kinds of map projections have been invented to serve specific requirements. For example, some map projections preserve angular measurements better than others and these are more useful for land surveying.



The map that you see on the display is not exactly correct since it is depicted on a flat surface. The ground coordinates are therefore only approximations of true positions since they are calculated on a flat surface. Over a small region like the municipality database that we are using in the exercises, the approximation is very good indeed.

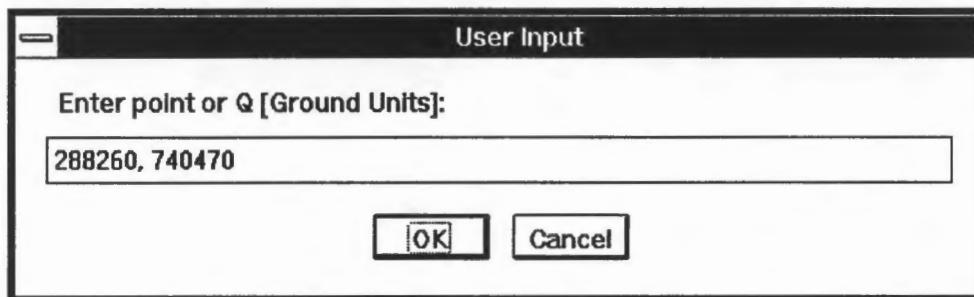
NOTE Ground coordinates are used most of the time in this Guide because they are more convenient

Scaled ground coordinates (X, Y) are integers related to the ground coordinates. They are specially designed for use by this GIS because they can be stored more efficiently.

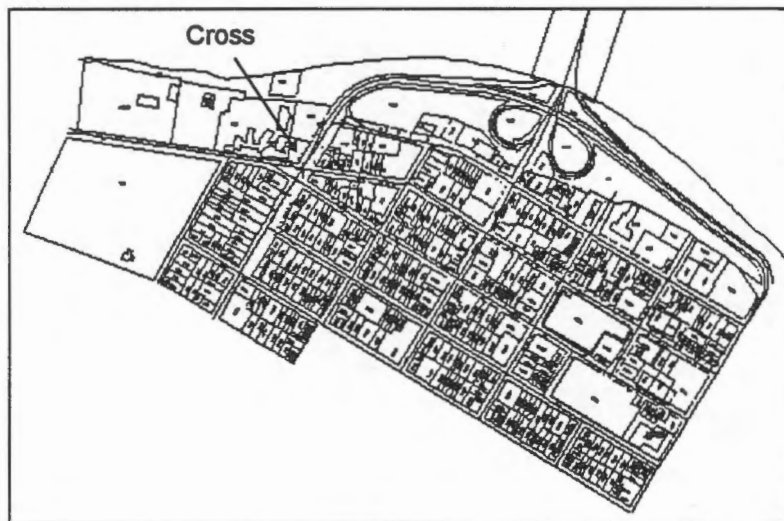
 Show the location of a point given its ground coordinates.

The opposite of reading the coordinates at a point is to show the location of a point given its coordinates, such as (288260, 740470).

1. Select the MapQuery > Point Location menu item.
2. When the crosshair appears, *do not* click on the mouse button but instead enter K on the keyboard. This is to inform the GIS that you want to key in the coordinates. A User Input panel appears, and you should type in (288260, 740470).

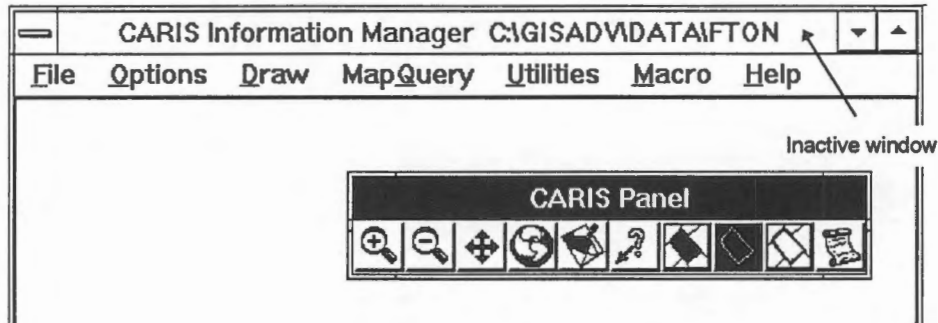


Clicking on OK will cause a cross to appear at that location. If the map is crowded, you might have to look carefully to find it.

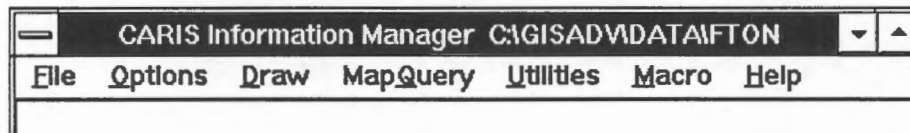


You can also click on the Centre icon from the CARIS Panel to perform this operation. However, doing this would inactivate the map window as shown by the top bar turning from black to white in this case. The GIS would not pick up the letter K you enter.

NOTE Other colours might be used to indicate active and inactive windows, depending on how you assign the colours in the Windows Control Panel.

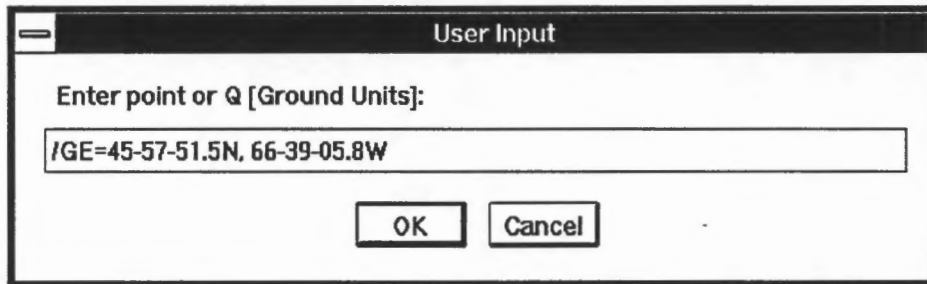


Click at the top bar to make it active before entering K on the keyboard.

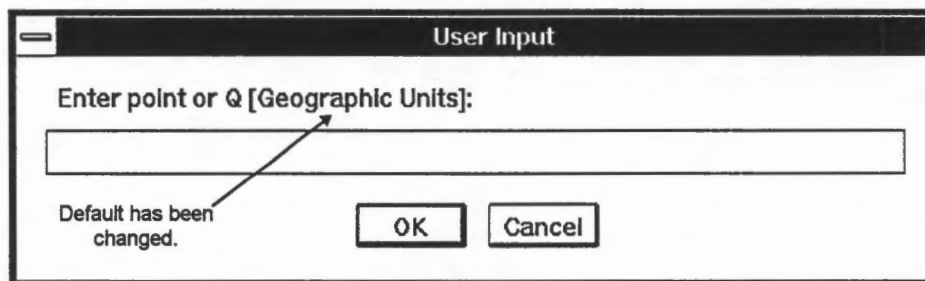


 Enter geographic coordinates.


You can specify a location by entering its geographic coordinates. To do this, enter the K key when you see the crosshair as usual, and prefix the coordinates with /GE=.



After you have entered geographic coordinates, the default will be changed. If you enter another set of geographic coordinates again, there is no need for the /GE= prefix.

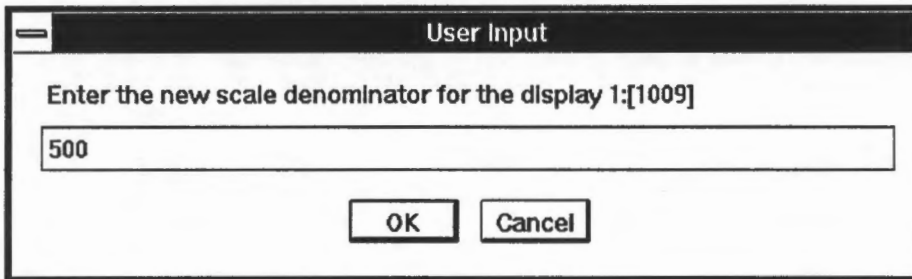


To enter a ground coordinate when the default is something else, use the /GR= prefix. ♦

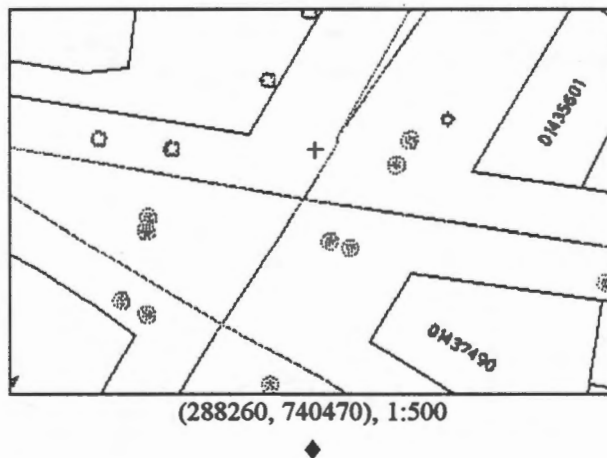
 Centre display at a known location.

It is difficult to see the cross on a crowded map. A better way is to draw an enlarged map centred at the location wanted.

1. Use the Options > Display > Scale ... menu item to set the scale to 500 (meaning 1:500).



2. Select the Draw > Centre > By Pointing menu item and enter the coordinates of the centre (288260, 740470).
3. Use the MapQuery > Point Location menu item; enter the coordinates and you can see the cross better.




 Datum.

We need more than a reference ellipsoid to measure coordinates. We must also know the origin of the coordinate system, the orientation of the axes, and so on. We refer to all these parameters that allow us to define the location of points as a *datum*. One that allows us to measure two-dimensional positions on a surface is a *horizontal datum*, and one that allows us to measure elevations or depths is a *vertical datum*.

The horizontal datum traditionally used for North America is the NAD 27 (North America Datum of 1927) based on the Clarke 1866 ellipsoid. The more recent NAD 83 is based on the GRS 80 ellipsoid. The redefinition of the datum has caused the coordinates of points on the earth's surface to change.

The mean sea level is a popular vertical datum for measuring elevations and depths. Traditionally, the vertical datum and the horizontal datum have been defined independent of each other. The newer NAD83 has changed this, and the same ellipsoid is used for both the horizontal and the vertical references. ♦

 Reference ellipsoid and map projection for this database.

The reference ellipsoid used for this database is the Average Terrestrial System of 1977 (ATS77) with $a = 6378135.0$ m and $b = 6356750.305$ m. The map projection used is the stereographic projection. It is a *conformal projection* which literally means one that preserves angular measurements. This is in fact true only within an infinitely small area, and it is impossible for any map to preserve angular measurements over a wide region. For large scale conformal maps, angular distortions are negligible for most applications.

The origin of this projection is located where latitude 46° 30' North meets longitude 66° 30' West. It is labelled not as (0,0) but as (300000, 800000) mainly to avoid the presence of negative numbers in the coordinate system. For this reason, the origin is called a *false origin*.

You can use the MapQuery > Point Location menu item to verify the above statement.

Coordinates of a Point			
X: 6000000	XG: 300000.00	LAT: 46-30-00.000N	<input type="button" value="OK"/>
Y: 16000000	YG: 800000.00	LONG: 66-30-00.000W	
	ZG:		

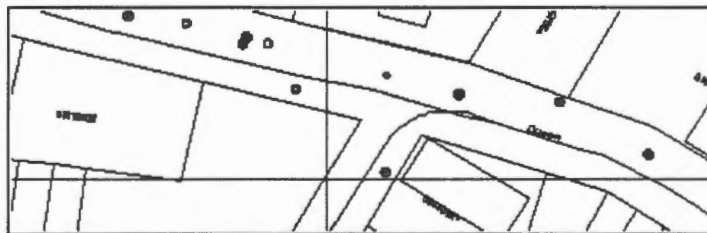


Coordinate System and Distance Measurement

In a GIS, distances are computed from coordinates. As explained earlier, the coordinate systems affect the coordinates of a point, and therefore also affect distance computations. We will illustrate this in the first exercise. We will then show that length calculation along a line is an extension of distance calculation between two points. In order to use the stored points for a line to compute its length, we are faced with the challenge of finding points in a large database efficiently. An approach to this search problem will also be discussed.

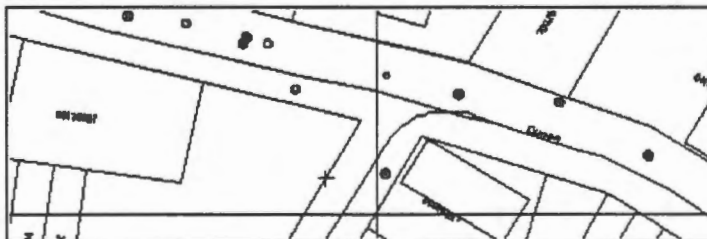
 Measure the distance between two points.

1. Centre the display at (288530, 740510) ground units, and choose a scale of 1:700.
2. Select the MapQuery > Distance Between 2 Points menu item.
3. Move the crosshair to the first point (on one side of the street) and click on the left mouse button.

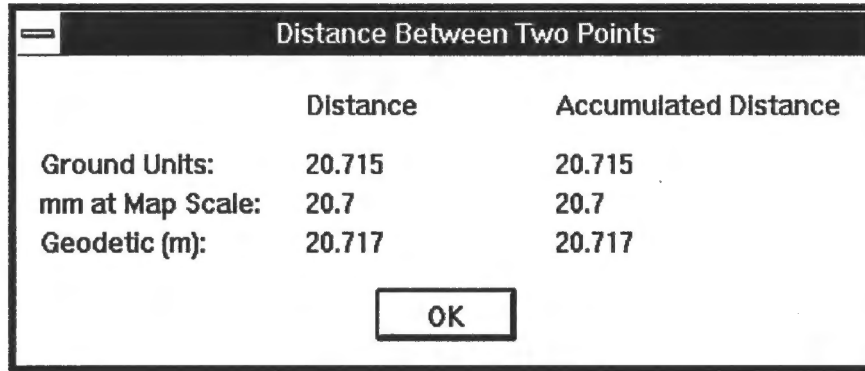


(288530, 740510) ground units, 1:700

4. This will leave a cross indicating where you have entered the point. Position the crosshair on the other side of the road, and click on the left mouse button.



5. The distance between the two points, hence the width of the road, will be displayed.



	Distance	Accumulated Distance
Ground Units:	20.715	20.715
mm at Map Scale:	20.7	20.7
Geodetic (m):	20.717	20.717

OK

6. Click on OK after you have finished reading the display.
7. The crosshair will reappear. Click on the right mouse button *twice* to exit from this command. ♦

The panel shows three different measurements:

- The first one (in metres) is based on the ground coordinates measured on a flat map.
- The second one (in millimetres) is measured at the map scale, which is 1:1000 in this case. This means that a distance of 20 millimetres on that map corresponds to a distance of 20 metres on the ground. This scale factor was entered when the map was created and it is stored with the database.

Note This scale is a property of the database and is not related to the display scale which is 1:700.


- The third one (in metres) is based on latitudes and longitudes measured on an ellipsoidal surface.

The difference of two millimetres between the geodetic and ground distances in the above example is due to distortion of the projected map. For a longer distance, this difference could be significant.

It is important to note that a distance measurement on an ellipsoidal surface is not always longer than the corresponding distance on a flat map. A map can stretch distances in some areas and shrink them in others.



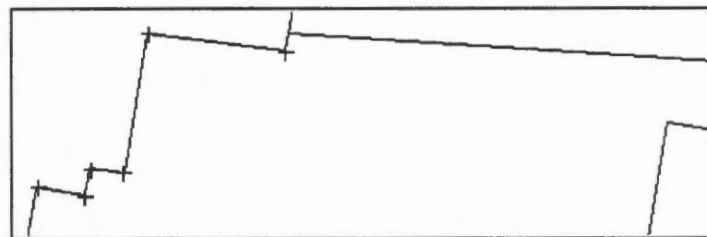
The demonstration Specify Position: Distance Between Two Points shows how to measure the distance between two points.

 Measure the length of a line with multiple points.

The same command can be used to measure the length of a line with multiple points. This is done by measuring the distance between two successive points at a time and building up an accumulated distance.

1. Measure the first two points as described before.
2. The distance between the two points is shown. Click on the OK button to continue.
3. You are now ready to enter another two points for the distance computation. In this case, we want to measure the distance between the second and the third point so that an accumulated distance from point one through point two to point three can be computed. You could point at the second point again, but a faster (and more accurate) method is to click on the *right* mouse button (or the Q key) *once* to automatically select it.
4. Point at the third point and press the *left* mouse button as usual.
5. Repeat from step 2 for the rest of the points and press the *right* mouse button *twice* after the very last point. ♦

The distance measurements shown are in two parts: the distance between the last two points (about 29.7 metres) and the accumulated distance for the six points marked by crosses (about 82.4 metres).



(288150, 740600), 1:500

Distance Between Two Points		
	Distance	Accumulated Distance
Ground Units:	29.739	82.433
mm at Map Scale:	29.7	82.4
Geodetic (m):	29.741	82.439



The demonstration Specify Position: Length Along a Line shows how to measure the length along a line with multiple points.

◆ Is map distortion uniform in all directions?

On a map, length distortion could change with direction. That is to say, an east-west line and a north-south line of the same length in ground distances could be different in geodetic distances. As an example, measure the distances between the follow pairs of points.

	Point 1	Point 2
East-West	(287562, 739529)	(289562, 739529)
North-South	(287562, 739529)	(287562, 741529)

Verify that both pairs yield a ground distance of 2000 metres but have different geodetic distances. Try a few of these measurements using lines of different lengths to investigate if distance measurements on a flat surface are distorted by a constant factor in one direction ? ◆

📏 Measure distance on an area where distortion has stretched the map.

Measure the distance between the two points (300 000, 1 000 000) and (400 000, 1 000 000). They are outside the range of this database, but we can still compute the distance between them because they are covered by the same map projection.

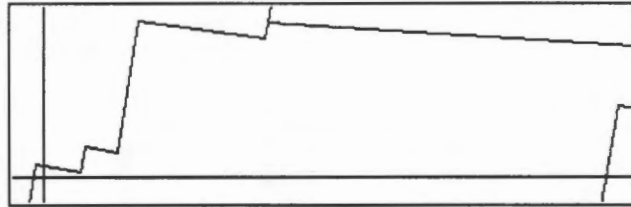
Distance Between Two Points		
	Distance	Accumulated Distance
Ground Units:	100000.000	100000.000
mm at Map Scale:	100000.0	100000.0
Geodetic (m):	99982.186	99982.186

OK

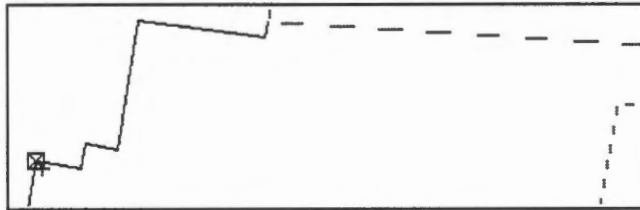
📏 Measure the length of a line using existing corners.

In the previous examples, the distances were computed from where the crosshair was placed. It is a challenge to place the crosshair at the exact corners on the line, especially when the scale is small. This exercise shows you how to use the existing corners for length computations so that you obtain a more accurate measurement.


1. Select the MapQuery > Distance Between 2 Points menu item.
2. When the crosshair appears, do *not* click on the mouse button, but enter S on the keyboard instead. This will invoke the snap mode allowing you to snap to the closest point in the database.



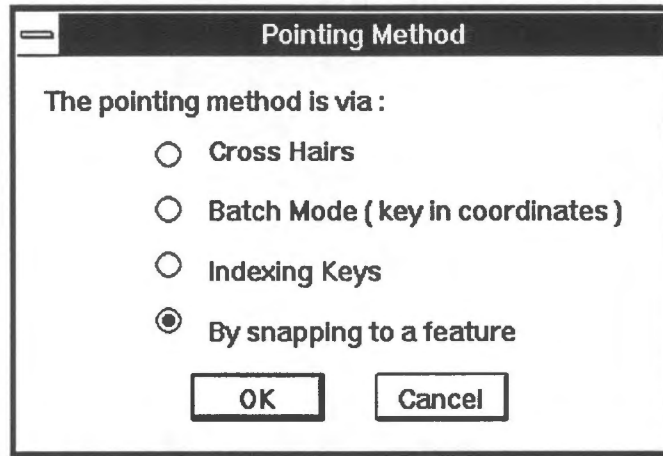
3. Move the crosshair close to the first corner of the line, and then click on the left mouse button. The corner closest to it (as well as the line containing it) will be highlighted. Since this is the point we want, click on the left mouse button to confirm.



4. Repeat step 2 to select the second end point on the line.
5. Select the rest of the points as described in the previous exercise. ♦

 The demonstration *Edit Graphic Data: Snap to a Point* shows how feature snapping can be used to enter a line starting and stopping at existing points. Although we have not introduced data editing yet, this demonstration shows another application of feature snapping.

To avoid pressing the S key all the time, you can tell the GIS to always snap to an existing feature when pointing. Select the Options > Pointing Method .. menu item, and click on the button for snapping to a feature.

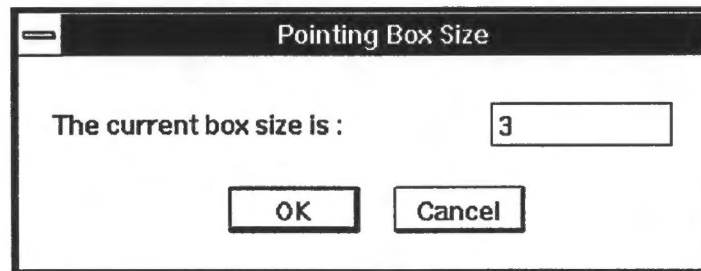


NOTE Turn off the “snapping to a feature” option when you no longer need it. Leaving it on affects the behaviour of pointing and could cause unnecessary confusions .

When you point at a corner, you will rarely place the crosshair exactly on top of it but will usually approach its vicinity. In order to know which corner you are pointing at, the GIS must search the database for a point close to where you have placed the crosshair. How close you must point at a feature before it will be selected is controlled by the *pointing box size*. The smaller the box, the closer you must point in order to pick up a feature. This is useful in a clustered place where precise pointing will avoid picking up the wrong feature. In a less crowded place, however, a larger box will make picking of features easier.

 Change pointing box size.

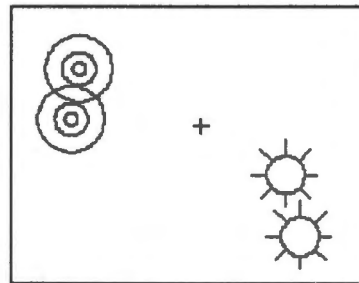
1. Select the Options > Pointing Box Size... menu item.
2. Enter a box size of three if it was set otherwise.
3. Click on OK to exit.



☞ See the effects of a small pointing box size.

To see the effects of the pointing box size, go to a denser part of the database and point at the features.

1. Centre at (288235,740455), choose a scale of 1:120, and use the MapQuery > Point Location menu item to place a cross at the centre of the display.

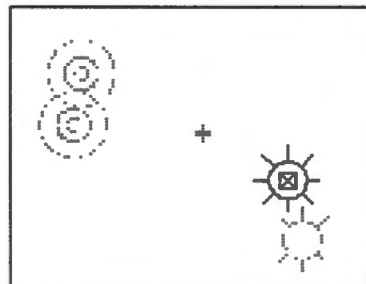


(288235,740455), 1:120

2. Turn on snapping to a feature.
3. Select the MapQuery > Point Location menu item.
4. Place the crosshair over the cross and click on the left mouse button.
5. The GIS will report that no point has been found in that vicinity. ♦

☞ See the effects of a larger pointing box size.

1. Set the pointing box size to 20 and repeat the previous exercise.
2. The GIS will pick up a symbol. The coordinates that are displayed indicate its location.



3. Enter N on the keyboard to reject this point. Another close symbol will highlighted.

If you keep on rejecting the points, all symbols in the vicinity of the pointing box will be highlighted in turn. ♦

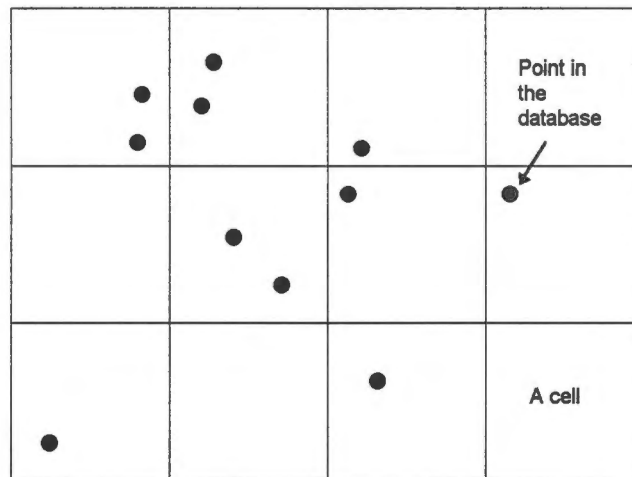
NOTE Return the pointing box size to 3 after you have finished with this exercise.

 Searching for the closest point within a neighbourhood.

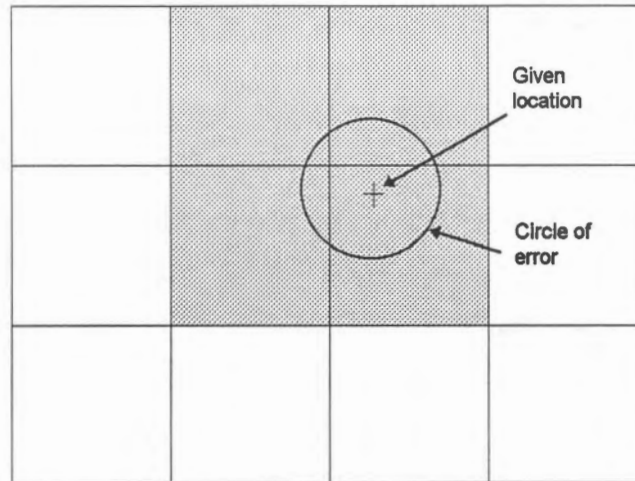
Searching for a feature closest to the spot you are pointing at is a common GIS operation. To simplify the case, we will assume that the database consists only of point features. The problem is this: given the location of the crosshair (the input location), find the closest point in its vicinity.

A simple but time consuming solution is to match the input location with all the points in the database. Realizing that a user normally does not point precisely at the feature, we allow for a small pointing error. Hence the feature within the circle of error around the input location and closest to it is the one we want. The pointing box mentioned earlier serves the same purpose as the circle of pointing error except that it uses a square to approximate the circle. This is just to simplify the calculations because a square is a less complicated geometric object than a circle.

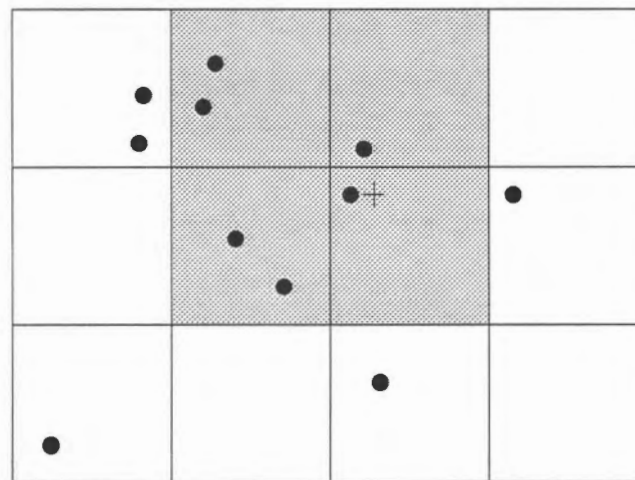
To help reduce the time it takes to find the closest point, this GIS uses a spatial indexing system involving a grid of regular cells. In this indexing system, a regular grid of cells is superimposed over the entire database, *logically* dividing it into smaller portions. Note that the division is logical in that the database is not actually partitioned into smaller files. Each cell, however, is associated with the features contained in it. Such a logical subdivision, allows the GIS to work on one subset of the database at a time.



During the search, the first thing the GIS does is to determine the cell containing the input location. This can be done very quickly using simple arithmetic. Then the circle of pointing error is used to determine how many cells need to be examined. In the example below, the four shaded cells overlapping the circle have to be examined.



Then all the six points contained within the four cells will be tested against the circle; those outside will be rejected. The point within the circle and closest to the input point will be selected.



Without this indexing system, all eleven points contained in the entire database would have to be examined.

It is possible that the feature wanted is not the closest one but instead is the next closest one. The GIS we use approaches this problem by first picking the closest one, highlighting it for confirmation, and if it is rejected, highlighting the next closest one within the tolerance circle and so on. ♦

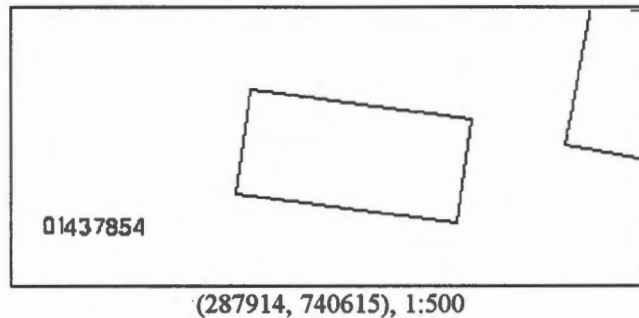
Chapter 11

Coordinate System and Area Measurement

As in distance measurements, area measurements are also based on coordinates. The coordinate system used has therefore an effect on the size of the areas computed. We will use the outline of a house in the database to illustrate this point.

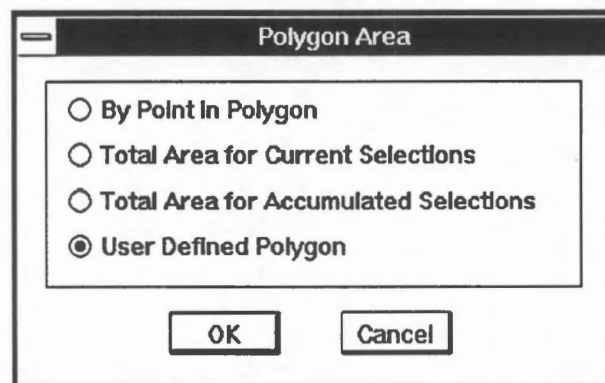
 Compute area.

1. Centre at (287914, 740615) and choose a display scale of 1:500.
2. Select the Options > Pointing Method .. menu item, turn on the snapping to a feature switch, and click on OK to exit.



3. Select the MapQuery > Polygon Area ... menu item, turn on the User Defined Polygon switch, and click on OK to exit.

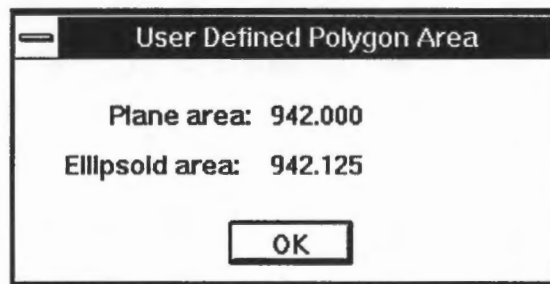
The switch allows you to define a polygon anywhere on the display window and have its area calculated. Since we have the snapping to a feature option turned on, we will be using the existing corners of the house and hence will compute the plan area of the house.



- Point at the four corners of the house in either a clockwise or a counterclockwise order. Then click on the *right* mouse button to terminate.

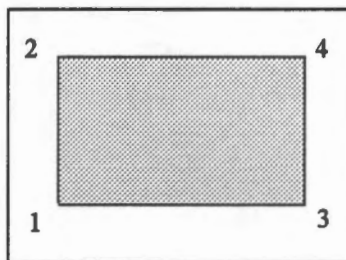
The GIS knows that you are measuring an area, and therefore will close off the outline of the house automatically. You could enter the fifth point if you wish, and click on the right mouse button to finish.

- Area measurement, like distance measurement, is performed on both the flat map surface and on the curved ellipsoidal earth. The slight difference of 0.125 square metres in this example is the result of area distortion introduced by the map projection used.



◇ Does point order affect area calculation ?

We have used either a clockwise or a counterclockwise ordering of points to measure the area of a polygon. What is the effect of the following point order on area the calculation?



Different map projections were developed to minimize geometrical distortions, but there are no map projections that can preserve angular or distance measurements in all directions. *Conformal projections* preserve angular measurements over a small area, and *equidistant projections* preserve distance measurements between certain points. However, there are certain kinds of map projections, called *equivalent projections*, that preserve area measurements over the entire map. This is achieved at the expense of angular and length measurements, which in this case are distorted.

Chapter 12

Resolution

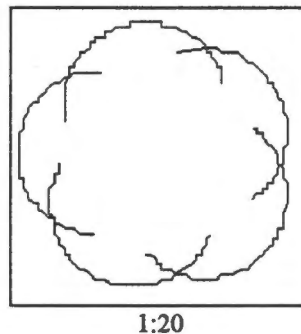
The resolution of a coordinate system is the smallest gap or the shortest line that can be represented. It affects the accuracy of the data stored, the shape of the features, and the results of analyses. In this chapter, we will explore the visual and other effects of resolution.

In the database we are using, the resolution has been set at 0.05 metre (or 5 cm) on the ground. This is called the *ground resolution* and is specified by the user when creating the database.

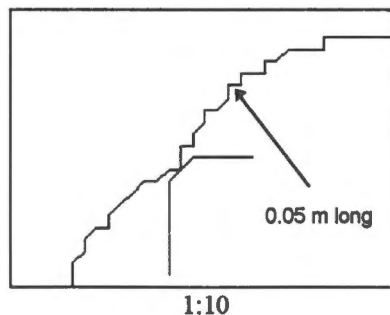
This fine resolution has little effect on the display unless a very large scale is used. The following tree symbol can be found at (289157, 740029), and the display scale is 1:80.



At a scale of 1:20, the effect of resolution on the image is quite noticeable.



When we enlarge the image to 1:10, the staircase effect due to resolution is large enough to be measured with accuracy.



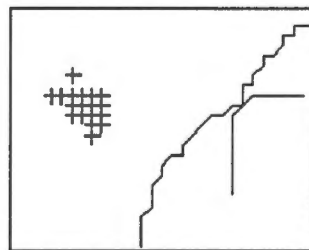
❖ Verify the length of the straight segments.

Use the MapQuery > Distance Between 2 Points menu item to verify that the short straight segments are indeed 0.05 metre long. ♦

📏 Visualize the resolution grid.

We mentioned earlier that you can read the coordinates at any location on the display. This is not entirely true, because there are locations that you *cannot* point at. To illustrate this, we have to zoom into the database at high magnification.

Centre at (289156, 740030), and choose a scale of 1:10. Select the MapQuery > Point Location menu item, and read the coordinates at various locations that are very close together. Recall that every time you select a point, the location at which the coordinates were read will be marked by a cross. At this high magnification, you will notice that where you point is not necessarily where the marker will appear. By pointing at several locations that are very close together, you would have left behind a series of markers that form a regular grid on the display. If you are patient enough, you could generate a grid that covers the entire database. This grid would indicate the only places at which you can read coordinates.



(289156, 740030), 1:10

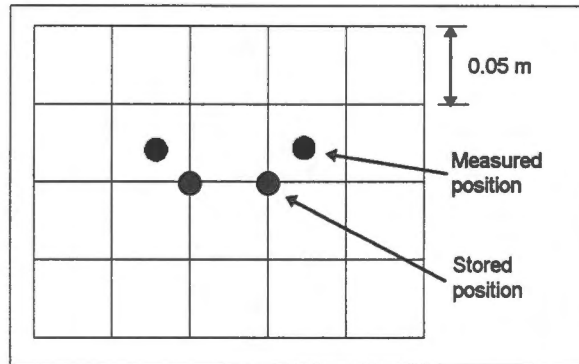


This is the visual effect of the underlying *resolution grid* which we know is 0.05 metre apart. You can verify that these crosses are indeed spaced exactly 0.05 metre apart by measuring the distance between them.

The implication of the resolution grid is that we can store coordinates only at its intersections. As a consequence, we can record a point in the database only at these intersections. That explains the staircase effect of the lines that we mentioned earlier.

We define the ground resolution of a database when we create it. This choice affects the accuracy of the data stored in the database. If we pick a resolution that is coarser than the accuracy of the data collected for the database, then the accuracy of the coordinates will be reduced. Suppose the database has a resolution of 0.05 metre (or five centimetres), but

the data collection device is accurate to 0.01 metre. All the coordinates must be rounded to 0.05 metre when stored in the database. Naturally, this affects the positions of the features and any subsequent measurements involving them.



It is possible to increase the ground resolution of a database afterwards to accommodate data of a different resolution. For example, if we have created a database with a ground resolution of 0.1 m but later find out that we had additional data available at 0.05 m resolution, we would have to increase the resolution of the database to 0.05 m before we could accurately store the data.



The demonstration *Concepts: Resolution* explains the concepts of resolution.

Part 3
Representation of Features

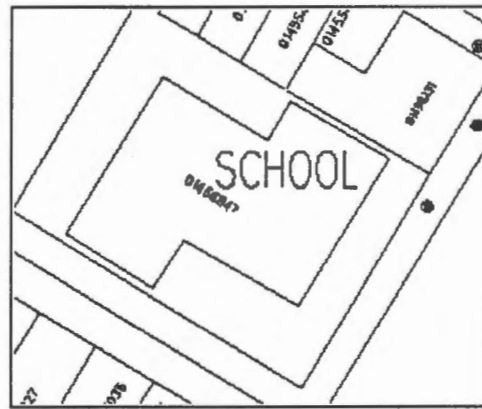


Different Perspectives of Reality

Geographic features are objects which we usually depict on a map. They could be real objects (such as houses, highways, and rivers) or abstract ones (such as political boundaries, migration paths of animals, and climatic regions). In a digital environment, these features are stored in geographic databases.

Other than geographic features, there are also *cartographic features*, which do not represent real world objects but are used purely for cartographic presentations. Annotation labels to show the name of the river or the name of a street are examples.

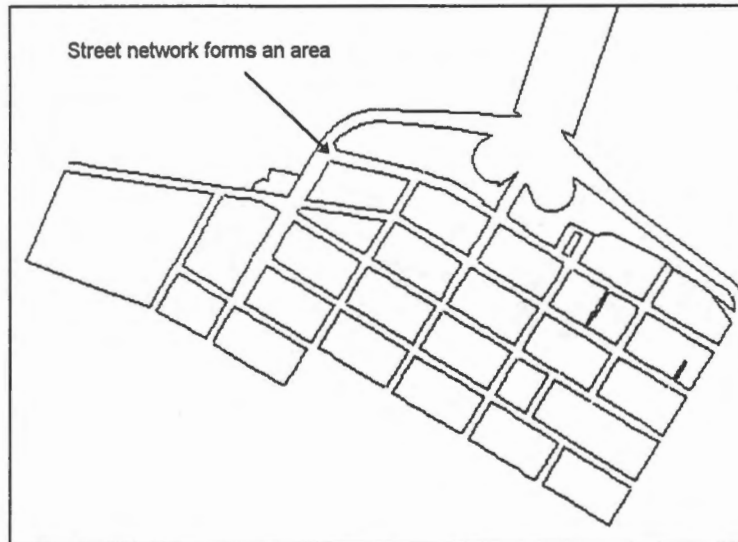
Geographic features might cover areas such as the school building in the map below. They might follow lines such as the edge of the pavement. They might be so small at the scale of representation that they simply occupy points such as the utility post in front of the school.



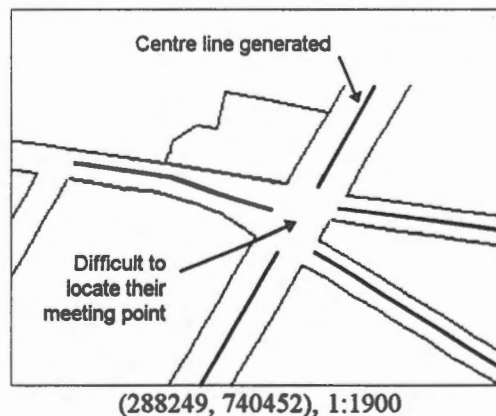
(289037, 739863), 1:850

Whether a feature should be represented by an area, a line, or a point is a matter of scale and compatibility. Scale is related to the level of detail appropriate for the application, and the representation must be compatible with the way a GIS stores and manipulates the features.

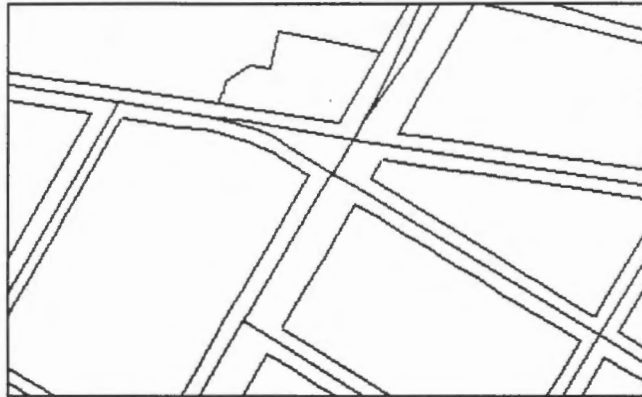
All geographic features occupy areas in the real world. They are reduced to lines or points because of the need to avoid unnecessary details in the database. A street network includes the surface of the road and hence is an area feature at large scale. At small scale, the streets would be too narrow on the display and would collapse to single lines.



Even at large scale, we might still want to represent streets using single lines because it is more convenient for a GIS to operate on them. For example, it is difficult for a GIS to automatically compute the length of a street using the area representation. In order to do this, the software must first locate the centre of a street and then measure the length along it. This is not too difficult if it is nicely bounded by two lines, although this process could be time consuming on a large database. The difficulty arises at an intersection where it is difficult to determine where the centre lines meet.



One way of getting around this problem is to create a network of centre lines manually and store it along with the double-sided streets.



The combination of the two representations gives the user two perspectives of the street network. The area perspective is suitable for applications that must view the streets as covering an area, and this perspective can be used to compute, for example, the width and surface area of the streets. The line perspective is suitable for applications that must view the streets as connectors. It is the more appropriate perspective for measuring lengths and finding the shortest route between two points.

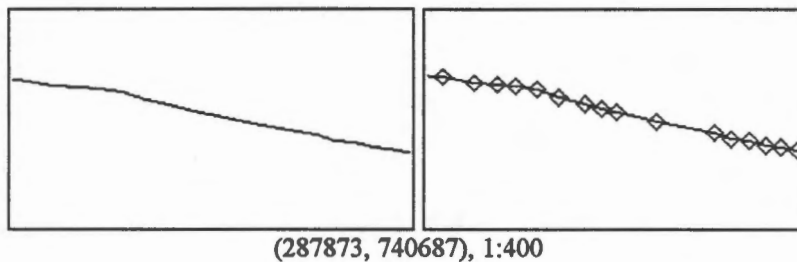
It would have been ideal if a GIS could automatically create and switch from one representation to another depending on the operations being performed. GIS are performing such conversions to some extent. However, the users must still make the major decision of which representations to incorporate when creating a GIS database. A wrong choice can later affect the performance of the system.

Chapter 14

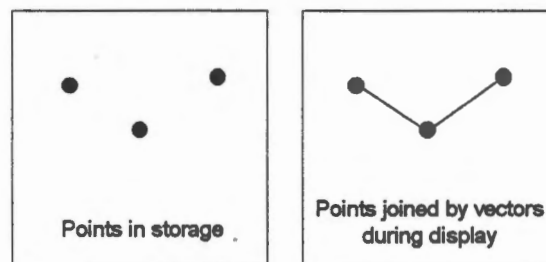
Vector Representation

A point in the database is stored using two or three coordinates, and lines are in turn composed of points. For example, the line in the left box below is composed of a series of points marked with squares, as shown in the right box. These points are spaced from two to eight metres apart on the ground in this example. In general, the spacing of the points is determined by factors such as the curvature of the line and the speed with which the line was traced during input.

NOTE You will not see the square markers on the display. They have been added here to show the points more clearly.

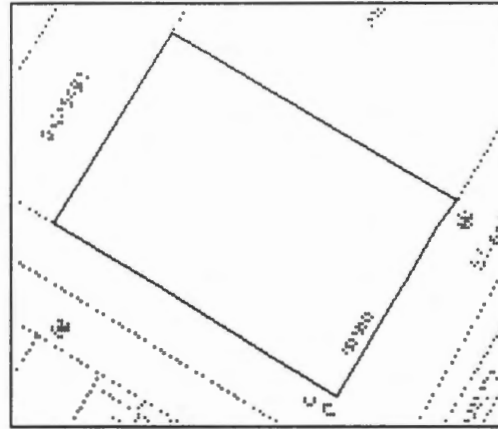


In the database, only the points forming a line are stored. The short straight segments, called *vectors*, joining them are created by the display module of the GIS when the map is being drawn. This is also the assumption used by distance and area calculation modules of a GIS. This method of representing the geometry of geographic features is called the *vector representation*.




Line Features

Lines form linear features as well as the boundaries of areas. For example, the following rectangular parcel is composed of three different lines, only two of which are straight ones. It could have been composed of four lines, but the operator has chosen three in this case to reduce the operational and storage overheads required to start and end a line.

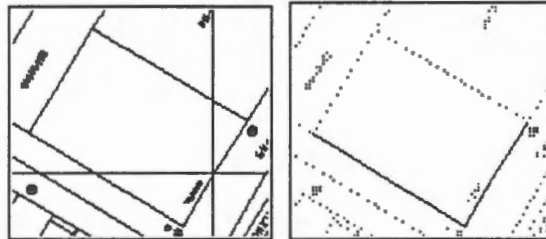


(288647, 740371), 1:400

 Show the lines forming the boundary of the parcel.

We will use the Information icon on the CARIS panel to show the individual lines forming the boundary of the parcel. This command can be used to display certain information about the features, but we will use it here only to show the extend of lines.

1. Click on the Information icon.
2. Point at one of the lines and click on the left mouse button. The line will be highlighted and the Line Information panel will appear.



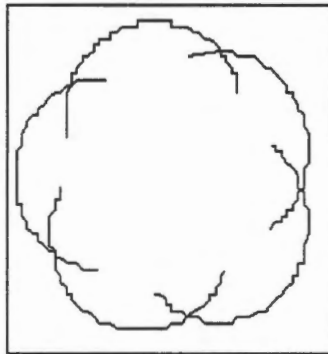
Line Information		
Feature Code: PARCEL	User #: 50	<input type="button" value="OK"/>
Key:	Theme #: 50	<input type="button" value="Cancel"/>
Source ID: 13L31SE	Type: LINE	<input type="button" value="More ..."/>

3. The panel shows that it is a line and it is a parcel boundary. We will explain the entries in this panel later. For now, just click on the Cancel button to exit.
4. Repeat from step 2 to highlight the other two lines.
5. Click on the right mouse button to quit. ♦

Point Features

A point feature occupies a single point and is not visible unless symbolized. Symbolizing a point feature is also a way of rendering its properties in a graphic way.

If you enlarge the tree at (289157, 740029), you will see that it looks very much like the other lines in that it is also composed of straight line segments.



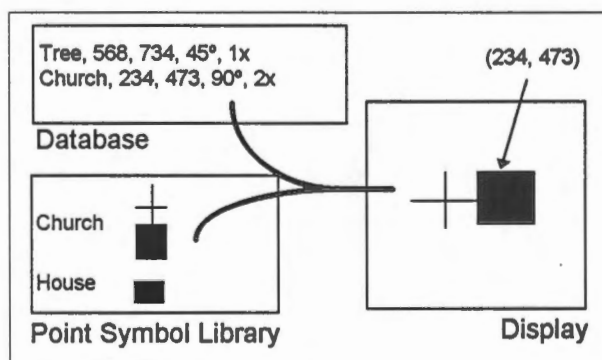
(289157, 740029), 1:20

However, this similarity is only superficial. Symbolized points and lines are stored in different ways. The points forming the lines are stored in the database, but the points forming the graphics of the symbols are stored separately in a point symbol library. To show them on the map, the display module performs the following operations:

1. Determines the kind of feature being drawn, which is a church in the example shown below.
2. Consults the point symbol library and looks up the graphics for the church symbol.
3. Refers to the database for the location, orientation, and size of the symbol.

NOTE The sizes of the symbols will be adjusted according to the scale of the display, and they do not reflect their real sizes because being symbolized points implies that their areas have already been reduced to points.

4. "Stamps" the symbol at the location wanted, and rotates and scales it if necessary.



There are three reasons for this arrangement.

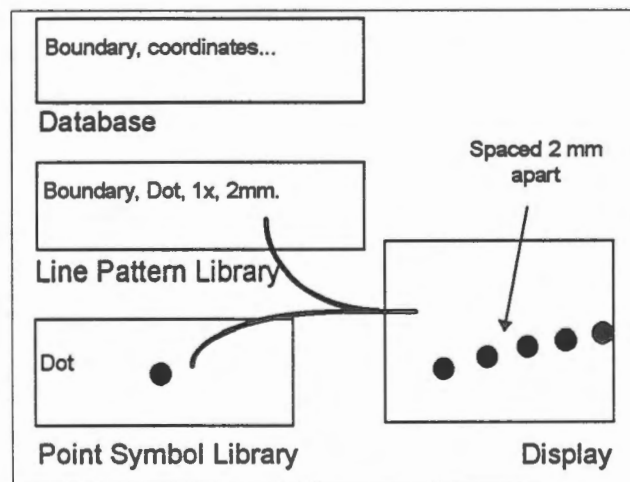
1. A common record for the graphics of symbols in the library guarantees symbols of the same type look identical to each other.
2. It is easier to change the shape of the symbols. If for some reason we need to use a simplified tree symbol, such as a circle, because we are producing smaller scale maps, we need only change its definition in the library once and the new shape will automatically take effect on all subsequent displays.
3. It helps to reduce the size of the database because there is no need for it to store the graphics of the symbol at every occurrence. All it needs to store is the location of each symbol plus a few parameters to identify its type and describe its orientation and size.

Symbolized Line Features

Lines are symbolized on the display to depict the characteristics of the features they represent. For example, provincial boundaries would be assigned a different pattern to distinguish them from international boundaries.

Visually, a symbolized line is a line composed of many smaller pieces. For example, a dashed line breaks down the original line into short straight segments with gaps between them. A dotted line is composed of many dots (symbolized points) spaced at regular interval.

The normal practice is to keep unsymbolized lines in the database and symbolize them at display time with the help of a line pattern library. This library contains pattern definitions for different classes of features. For example, a pattern of dotted lines could be assigned to provincial boundaries identified by a certain feature code. The line pattern library then consults the point symbol library for the graphics of the dots required.



We do not want to store lines in their symbolized form in the database for the following reasons:

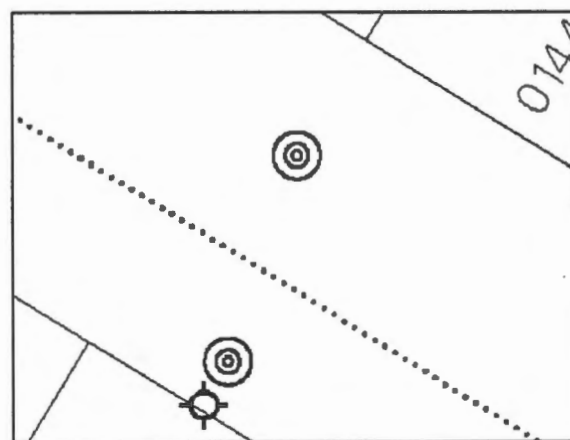
1. The overhead of storing the individual components is too high. Instead of a single solid line for a political boundary to be symbolized later as a dashed line on the display, we would have to store the individual dashes along it. Each dash is a short line by itself, thereby increasing the storage requirement because there are overheads involved in handling the beginning and ending of lines in the database.
2. A continuous line is more convenient for analysis than one broken up into pieces. Take for example the computation of the length of a line, the mathematics is much simpler for a continuous line than for a broken one.
3. It is easier to change the line pattern later.

Unlike symbolizing point features which can be done in real-time, symbolizing line features takes a considerable number of computations and is usually not done in real-time. Hence a display shows line features as solid lines which are distinguished from each other only by colour. In fact, colour is the simplest form of line symbolization.

In the GIS you are using, it is possible to display symbolized lines. Naturally, the display would take longer time to finish.

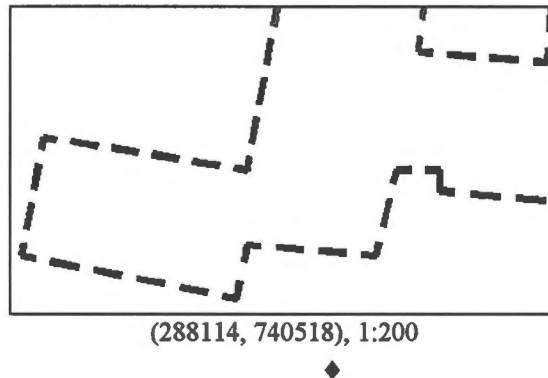
Display symbolized lines.

1. Select the Options > Display > Symbolized During Draw menu item.
2. Draw the part of the database shown below. You will notice that the road centre line has been symbolized. The point symbols have also been assigned the proper line weight as specified in the database.



 Special consideration of line symbolization.

1. Select the Options > Display > Symbolized During Draw menu item.
2. Display the house as shown in the diagram below.



Rules of cartography requires that a gap does not occur at the corner of a line. Simple line symbolization software may produce a dashed line around a rectangle like this:



A line symbolized in this way is not only less pleasing, it is also less accurate because corners of lines, which are usually well-defined points, are lost. This is not an important concern to a digital database because the lines are not symbolized in storage, but it is a concern to users of hardcopy maps who see only the symbolized version.

In order to achieve a cartographically acceptable symbolized line, the software must adjust the spacing and the length of the dashes to ensure that there is a symmetrical turn at each corner of the line. This and other considerations make line symbolization a rather complicated task.

 Effects of database and screen resolutions.

The resolution of the database coordinate system (measured in metres on the ground) and the resolution of the screen (measured in number of dots across and down the screen) both create the staircase effect on lines. Normally, the screen resolution is much higher than that of the database. Hence at high magnification, the effect of database resolution dominates; at low magnification, the effect of screen resolution dominates.

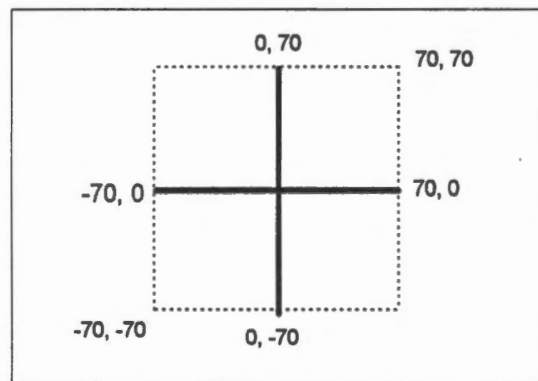
If you look closely at the lines on the maps in this Guide, which were captured directly from the screen, you can see the staircase effect caused by the screen resolution. This staircase effect, unlike that caused by the database resolution, does not get worse when enlarged. ♦

 Storage of symbols in the library.

In the GIS we are using, the points forming each symbol are recorded in the library in a local coordinate system. Such a coordinate system is arbitrary, and it is not related to the real world at all. This is reasonable because before we place the symbols on a map, they exist independent of the coordinate system of the database.

In this local coordinate system, the origin is located at the centre of the symbol. The following is an entry for the cross symbol, identified as ALAP, stored in the point symbol library.

ALAP Line 1 (-70, 0), (70, 0)
 Line 2 (0, -70), (0, 70)



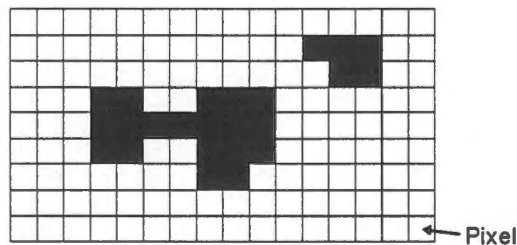
Chapter 15

Raster Representation

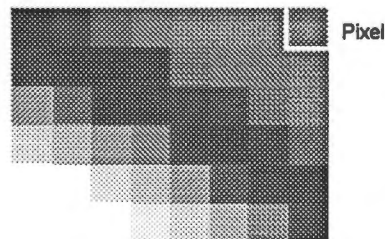
We have explained in the previous chapter the method of using points and lines to represent features, and the joining of the adjacent points by straight lines to form continuous lines at display time. These straight lines joining up the points are called vectors, hence this way of organizing graphic data in the computer is called the *vector representation*.

The vector representation is good for line maps where features are represented by points or lines. Images such as aerial photographs do not use lines. Traditional cartographic techniques break down a continuous tone image into dots of various sizes to show different intensities. A similar technique in a GIS is used for digital images, resulting in a *raster representation*.

A raster representation is based on a regular grid, very much like the resolution grid we have described, superimposed upon the image. Each cell of the grid, called a *pixel* for picture element, would be assigned a colour or a gray scale. The following shows a simple black and white image of two buildings in raster format. The grid is not visible on a raster image but is shown here to illustrate the pixel size used. As in the case of the resolution grid for vector representation, the pixel size in a raster representation determines the resolution and hence the quality of the image.



On a photographic image, a gray scale will be assigned to each pixel to represent the intensity at that spot. The following is a much enlarged portion of an aerial photograph. At this magnification, you will see the staircase effect due to resolution.



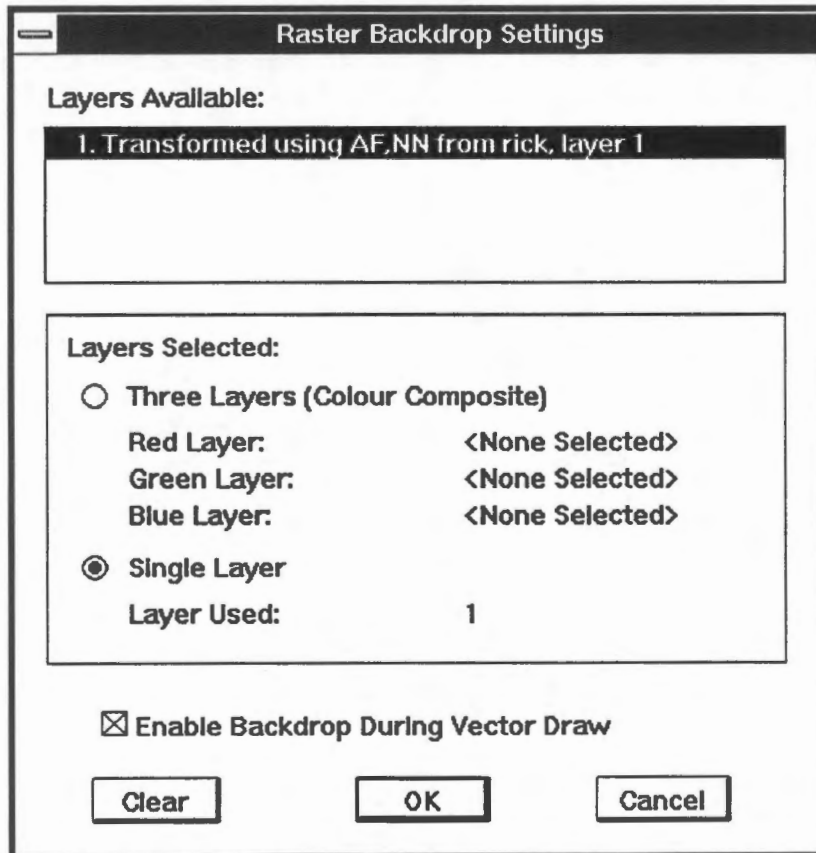
 Display the raster image of an aerial photograph.

An aerial photograph can be drawn over the vector map we have been using so far. The combined effect would be the line map superimposed over an aerial photograph as a backdrop. This produces something like a conventional orthophoto map, which provides both map and image information for the same area.

1. Select the Draw > Raster Image Backdrop > Settings .. menu item.
2. The Raster Backdrop Settings panel appears.

Turn on the Single Layer switch, click on the highlighted layer (in the Layers Available area) so that number 1 appears on the list of Layers Selected, and turn on the Enable Backdrop During Vector Draw switch. Turning on the last switch allows you to draw the aerial photograph and the line map at the same time.

NOTE Although layer 1 has already been highlighted in the list of layers available, you must still click on it to select it.



Raster Backdrop Settings

Layers Available:

- 1. Transformed using AF,NN from rick, layer 1

Layers Selected:

Three Layers (Colour Composite)

Red Layer: <None Selected>

Green Layer: <None Selected>

Blue Layer: <None Selected>

Single Layer

Layer Used: 1

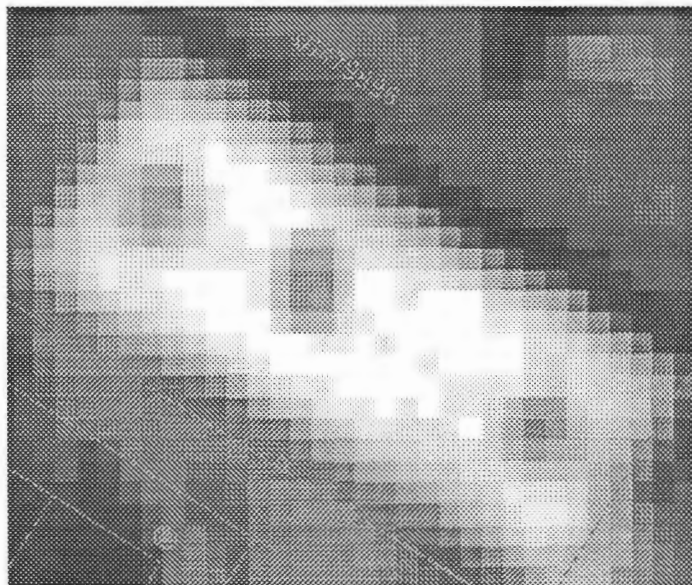
Enable Backdrop During Vector Draw

Clear OK Cancel

3. Click on the OK button.
4. Select the Draw > Overview menu item or the Overview icon on the CARIS Panel to draw the entire database.



5. Zoom into an area of the image, and you will start to see the effect of resolution on raster images.



(288875,740095), 1: 300



❓ What is the resolution of this image ?

Determine the resolution of this raster image by measuring the size of each pixel. ♦

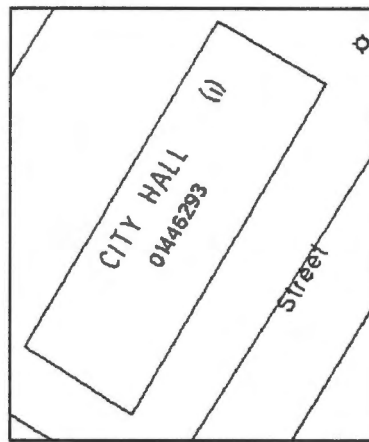
❓ How does this resolution compare to that of the vector map ?

Is this resolution coarser or finer than that used for the vector map ? ♦

Other kinds of raster images can be stored in a GIS besides aerial photographs. The example below demonstrates the linking of a photograph to a geographic feature in the database.

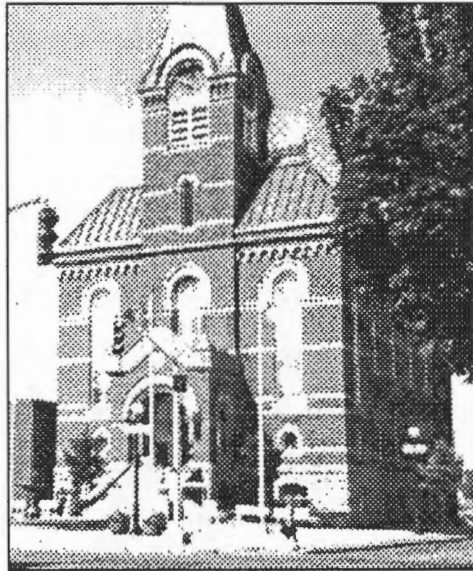
📷 View photographs.

1. Pick the MapQuery > Extended Attributes > By Pointing menu item and point at the name CITY HALL.



(288906, 740363), 1:350

2. Image processing software will be invoked to show a picture of City Hall.

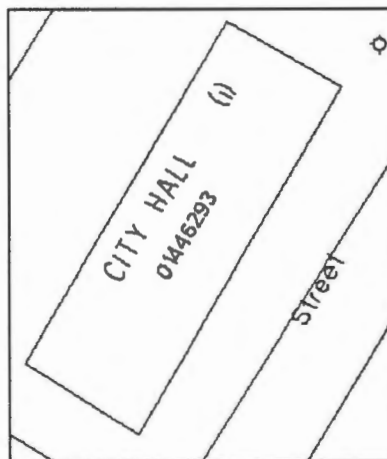


3. Pick the File > Exit menu item to close the picture. ♦

 View related text document.

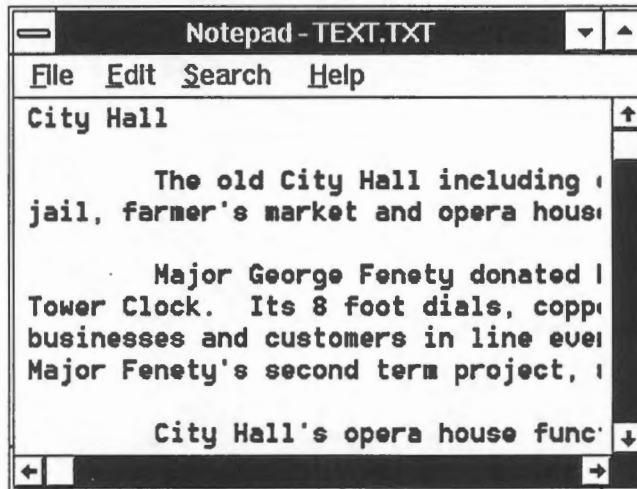
Photographs are part of the extended attributes that we can associate with a feature in the database. We can also attach text documents as extended attributes. In this exercise, we will retrieve a description of City Hall.

1. Pick the MapQuery > Extended Attributes > By Pointing menu item and point at the (i) after CITY HALL.



2. A text document describing City Hall will appear.

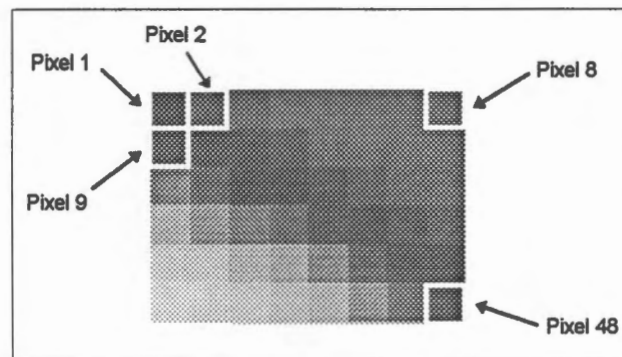
NOTE This is the Notepad editor supplied with Windows. Please do not change the text.



3. Close the Notepad editor and click on the right mouse button to exit.

 Raster data structure.

The simplest way to organize a raster image is to record the intensity of each pixel row by row. First of all, choose a systematic way of numbering each pixel, such as starting from the top left corner and following each row from left to right as shown in the diagram below.



The internal data structure could be regarded as a table with the following entries:

Pixel Number	Pixel Value
1	9
2	5
3	7
4	6
...	
8	5
9	8
10	9
...	
47	5
48	8

A raster image represented by such a table is not tied to any coordinate system unless we know the coordinates of one of its pixels. If we know the location of the top pixel (pixel 1), the size of the pixels, and assume that the image is on a flat and not ellipsoidal surface, then the coordinates of all other pixels can be derived by simple arithmetic. ♦



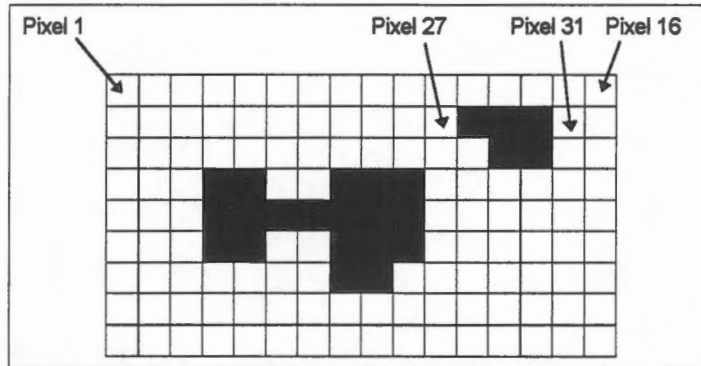
Raster data compression.

Unlike a vector map that stores points only at places where they appear, either as part of a line or as isolated point features, a raster image assigns a pixel to every spot regardless of whether it is occupied by a feature or not. In other words, the size of a vector file is dependent on the number of features, the complexity of the features, and the resolution of the coordinate system. The size of a raster image, however, depends mainly on the resolution. Hence the size of a vector map is generally smaller than that of a corresponding raster map covering the same area with the same resolution.

There are several methods of compressing raster images. One of them is called *run-length encoding*, which basically groups similar pixels together to save storage space.

Without compression, we assign storage for each pixel and its attributes as explained earlier. When many adjacent pixels are of the same attribute, we can represent them by a single count followed by their common attributes.

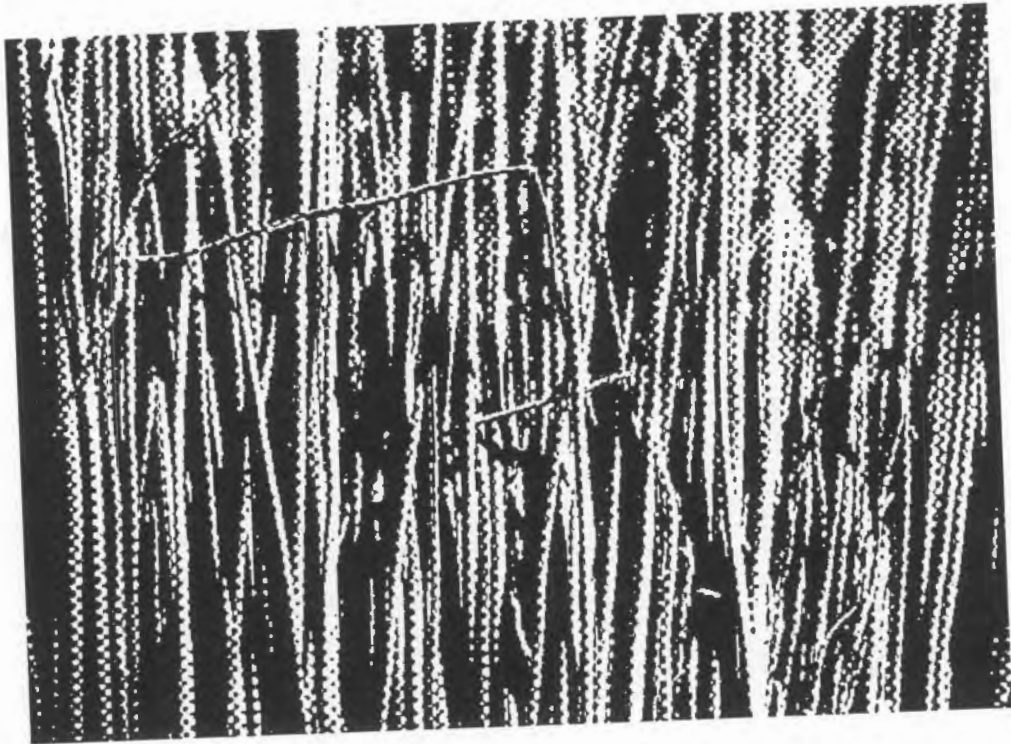
In the binary raster image shown here, the attributes are B (black for building) and W (white for empty space). The first 27 pixels are all white. This is followed by three black ones and so on. A run-length compressed raster image would start with 27 W, 3 B, 14 W, 2 B, 5 W, 2 B, 2 W, 4 B,



In a black-and-white case, there is really no need to store all the attributes except the first one because we know they alternate. An even more compressed version of the raster image is 27 W, 3, 14, 2, 5, 2, 2, 4, This scheme would not work when there are colours or grey colours involved in the raster image.

Run-length encoding compression does not work very well if the images are highly segmented and do not occupy blocks of neighbouring pixels. An extreme case is a checker board image where black and white pixels alternate. ♦

Part 4
Retrieval



Chapter 16

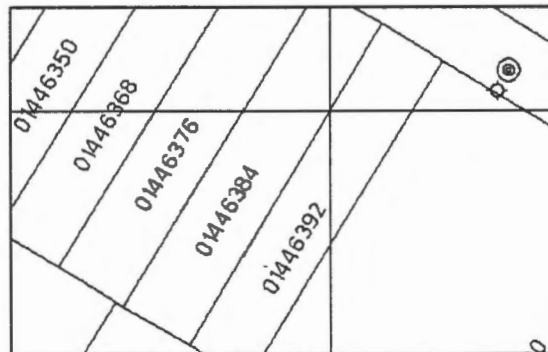
Retrieve Information on Features

Other than the geometric representation, we need to associate with each feature other information such as its length and area. In this chapter, we will take a look at some of the information stored with the features. It should be noted, however, that the amount and nature of this information are quite different from one GIS to another.

Earlier we used the Information icon on the CARIS Panel to highlight a feature. That command also displays information about features.

 Display information about a line.

1. Centre at (288658, 740417) and draw with a scale of 1:400.
2. Click on the Information icon on the CARIS Panel.
3. Point at the line.



(288658, 740417), 1:400

4. It will be highlighted.



- The Line Information panel will show immediately. Click on the More... button to show the full panel.

Line Information	
Feature Code: PARCEL	User #: 50
Key:	Theme #: 50
Source ID: 13L31SE	Type: LINE

- After you have finished reading the information, click on the OK button to return to the short panel. Then click on the Cancel button to exit from this command.

Line Information	
Feature Code: PARCEL	User #: 50
Description:	
Source ID: 13L31SE	Type: LINE
Line Weight:	Theme #: 50
Left Polygon: 01446392	Scale Indicator:
Right Polygon: 01446384	Key:
Elevation: 50.799 m	Length: 50.799 m
Arc (Y/N)?: Y	Elevation (Y/N)?: N
Compilation Status: Background	Header No.: 1



There are certain pieces of information shown in the panel which are useful for the design and operation of the database.

Feature Code is a code indicating the feature type. Each code can be up to 12-characters long for this GIS. For example, the Code PARCEL indicates that it is a boundary of land parcels.

This code was entered by the user when inputting the line. In this GIS, there is a feature code file which stores all the standard feature codes used for a project. Each organization or a group of users could design their own feature code files to enforce feature coding standards. Each feature code entered will be checked against those in the file for verification.

Source ID, another 12-character code, indicates the data source for this feature. This code was also entered by the user according to some convention specified by the project manager. In this example, the code 13L31SE indicates the map from which this feature was obtained.

User Number is a numeric code that can be used to group features of similar characters. The meaning of this number is left entirely to the user.

Theme Number is a numeric code. It is used to group features that are topologically related to each other. A topological relationship describes whether two lines meet, whether two regions share a common boundary, or whether a region is within another one.

It is possible for the GIS to determine automatically topological relationships for features of the same theme and store these relationships in the database. If the database has explicit knowledge about adjacent areas, for instance, then it would be easier to process queries involving topology such as: What are the parcels surrounding my property? Without that explicit knowledge, the GIS would have to establish a topological relationship from the coordinates every time a topological question was asked. That would involve considerable processing overhead.

The theme number was assigned by the user. In this example, theme number 50 is used to group data related to land parcels. This includes all the boundary lines and the parcel labels.

Left Polygon identifies the polygon to the left of a line, and it was assigned by the system with reference to the direction the line runs. The direction of a line is determined by the order of the points when they are entered into the system. In the above example, the line runs down from the northeast corner. Hence the left polygon is the one labelled 01446392.

Together with the Right Polygon identifier, these make up some of the topological information stored in the database.

Right Polygon identifies the polygon to the right of this line. It was also assigned by the system.

Type is a code assigned by the system to each feature to indicate its geometric type. LINE in this case indicates that this is a linear feature represented by a line that can be straight, angular, or smooth. The code was generated by the system according to the command used to create this feature.

Length is computed by the system to indicate the length of the line you have just selected. It is valid only for LINE type features.

Elevation indicates whether this line has height associated with it. An entry of N (for No) indicates that this is a two-dimensional line.

Verify the length of the line.

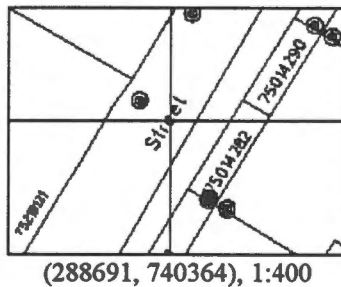
Verify that the length of the line is indeed computed from the coordinates of the two end points by measuring the distance between the two end points.

1. Select the Options > Pointing Method menu item to turn on snapping to a feature.
2. Use the MapQuery > Distance Between 2 Points menu item to measure the distance.

You should get the same length of 50.799 metres in ground units. ♦

Display information about a name string.

1. Click on the Information icon and point at the text string Street.



2. Show the full Name/Display Label Information panel.

Name/Display Label Information			
Feature Code:	STREET_NAME	User #:	1003
Description:			
Source ID:	13L31SE	Type:	NAME
Style/Point:		Size:	2.400 mm
Reference Position:	5773769 14807189	Font:	1
Label:		Line Weight:	
Polygon Area:		Polygon Theme:	
Polygon Key:		Scale Indicator:	
Polygon Perimeter:		Name:	Street
Theme #:		Key:	
Compilation Status:	Background	Header No.:	1


OK

The contents of this panel are similar to that for Line Information. The feature code shows that it is a street name, a type of cartographic feature. The type shows that it is a name for map annotations. It has a user number of 1003 but it does not belong to a theme, meaning that the database does not know any topological relationship between this text and the other features.

NOTE This text string “Street” is only part of a street name. The other part of it, “Westmorland”, is stored separately.

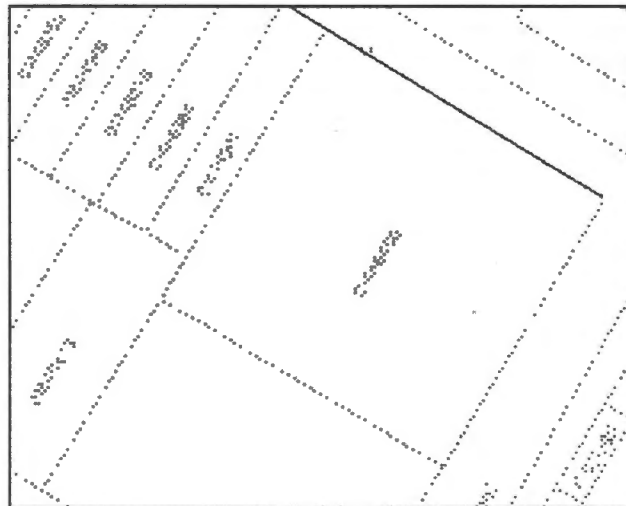
Here are a number of items specific to this type of data:

- Name contains the character string “Street”.
- Size indicates the height of the characters forming the label, which is 2.4 mm high at the database scale, which is 1:1000 in this example.
- Font indicates that it is using font number 1.

 Get information on features close by.

If in the previous step you selected OK on the short panel instead of clicking on More ... to get the full panel, then the GIS would have kept selecting features within a neighbourhood. This is useful when features are crowded around the one you want, making it difficult for you to point with precision.

1. Display information on this line.



- It shows that it is part of a city block. Click on the OK button to display information on a line close by.

Line Information		
Feature Code: CITY_BLOCK	User #: 60	OK
Key: 20411	Theme #: 60	Cancel
Source ID: CITY	Type: LINE	More ...

- The other line close by is actually overlapping part of the city block.



- It is the outline of a parcel. Click on the OK button and the crosshair will return since we have exhausted all features close by.

Line Information		
Feature Code: PARCEL	User #: 50	OK
Key:	Theme #: 50	Cancel
Source ID: 13L31SE	Type: LINE	More ...

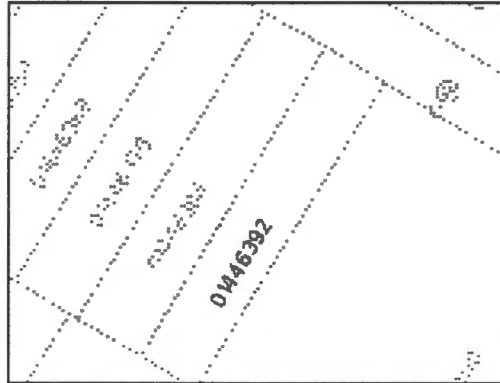
- Click on the right mouse button to quit this command. ♦

It is common for line features to overlap each other. In the previous example, city block outlines overlap parcel boundaries. Two line features are said to be *co-locating* when they overlap, and we can expect more co-locating line features in a database of coarser resolution.

 Display information about display label.

A display label is a special kind of name string that is topologically related to the area it annotates. Being topologically related means that the display label knows the area containing it, and the area knows about its display label.

1. Click on the Information icon and point at the text string 01446392.



2. Show the full Name/Display Label Information panel.

Name/Display Label Information			
Feature Code:	PARCEL_ID	User #:	50
Description:			
Source ID:	13L31 SE	Type:	DISPLAY LABEL
Style/Point:		Size:	2.000 mm
Reference Position:	5772786 14808500	Font:	1
Label:	01446392	Line Weight:	5
Polygon Area:	0.0477 ha	Polygon Theme:	50
Polygon Key:	01446392	Scale Indicator:	0
Polygon Perimeter:	120.40 m	Name:	
Theme #:		Key:	
Compilation Status:	Background	Header No.:	1

Names and display labels share the same panel. But because display labels are topological elements, they have theme numbers and other topological information as explained below.

Feature Code shows that it is a parcel identification code, which is cartographic feature. It is customary for a land records authority to assign unique codes to identify land parcels.

Type shows that it is a display label containing the unique parcel identification code.

Label contains the character string for the display label. In this example, it is 01446392, which is the unique code assigned to this particular parcel.

Polygon theme serves the same purpose as the theme number. It is used because a display label is topologically connected to other features. In this example, a number of 50 indicates that it belongs to the same theme as the boundary lines.

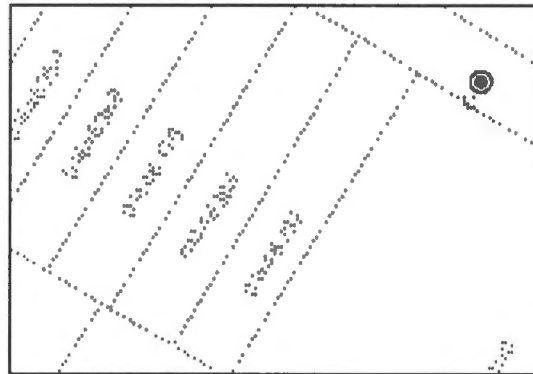
Polygon area is the area of the parcel to which this label is related. One hectare equals 10 000 square metres.

Polygon key is an identifier or name for this feature. Unlike the feature code showing what type of feature it is, the key identifies which feature. In this example, the polygon is called 01446392, which is its unique parcel identification code.

It would appear that the polygon key contains the same information as the display label. They do in this case. It is possible to later change the display label to show other information, such as the postal code, about the parcel. In other words, the key of the polygon will rarely change but its display label could be modified.

 Display information about a symbol.

1. Click on the Information icon and point at the utility pole symbol.



2. Show the full Name/Display Label Information panel.

Symbol Information			
Feature Code:	UTILITY_POLE	User #:	30
Description:	TRANS. LINE SY		
Source ID:	CG-WBS	Line Weight:	
Theme #:	30	Scale Indicator:	
Key:	560	Symbol ID:	
Disk #:		Elevation:	
Symbol #:		Node (Y/N)?:	Y
Optical Center:	45-57-51.006N 66-38-46.170W		
	X = 288671.55 Y = 740454.45		
Compilation Status:	Background	Header No.:	1

OK



The source ID of this symbol is CG-WBS, showing that it came from a source different from the other features we have seen so far. It also belongs to a different topological layer. In other words, the system does not know the relationship between this symbol and the parcels because they have different theme numbers. The key contains its unique pole number assigned by the utility company.

◆ ? Display information on features close to the symbol.

Use the method we have explained in the previous exercise to display information about features close to the symbol we have just selected. ◆

📏 Display information about a contour.

Contours are lines of constant height. Hence they are organized in a similar manner to lines except that they are each associated with an elevation.

1. Display information about this contour.



(287728, 739785), 1:1500

2. It shows that it has an elevation of 10 metres.

Line Information			
Feature Code:	CONTOUR	User #:	90
Description:	CONTOUR		
Source ID:	REFOAUTOPOPO	Type:	CONTOUR
Line Weight:		Theme #:	90
Left Polygon:	!70	Scale Indicator:	
Right Polygon:	!68	Key:	~20511
Elevation:	10.000 m contour	Length:	993.788 m
Arc (Y/N)?:	Y	Elevation (Y/N)?:	Y
Compilation Status:	Background	Header No.:	1

Note that the system has assigned a sequential key, ~20511, to this contour and sequential keys !70 and !68 to the areas on each side of the contour. ♦



Vector data structure.

In a raster data structure, pixels are stored together with their values indicating, for example, whether they belong to houses. In a vector data structure, each feature is stored together with other data describing its characteristics. For example, all information related to a house will be stored together.

Many GIS offer a number of geometrical representations with which real world features can be represented. The number of these representations differ from one system to another. We have seen some of those supported by this GIS: lines, symbolized points, names, display labels, and contours.

An example of a data structure for line features is shown below:

Feature code, Feature key, Source ID, Theme number, Length, Other information ... ,
Number of coordinates per point, Number of points in line, point 1, point 2, point 3, ...
last point.

A contour, being a line with an elevation, has a similar structure. There is no need to show the number of coordinates per point because they all consist of two coordinates. The elevation of the line is recorded instead:

Feature code, Feature key, Source ID, Theme number, Length, Other information ... ,
Elevation, Number of points in line, point 1, point 2, point 3, ... last point.

A number of symbolized features can be grouped together under one structure:

Feature code, Feature key, Source ID, Theme number, Other information ... , Number of coordinates, Number of symbols in group, symbol1, symbol2, symbol3, ... last symbol.

For each symbol, the following information is stored:

Position, magnification, orientation.

A name string, as well as a display label, could be organized in this structure. Being cartographic features, they require only two coordinates:

Feature code, Feature key, Source ID, Theme number, Font, Size, Other information ... , Number of characters in name, character1, character 2, character 3, ..., last character.

For each character, the following information is stored:

The character, its position, orientation.

Such a structure allows the characters in the name to be positioned and oriented independent of each other, thus allowing the name to follow a curve. It does not allow each character to be associated with a different font or size because such information is stored only once for the entire string. ♦

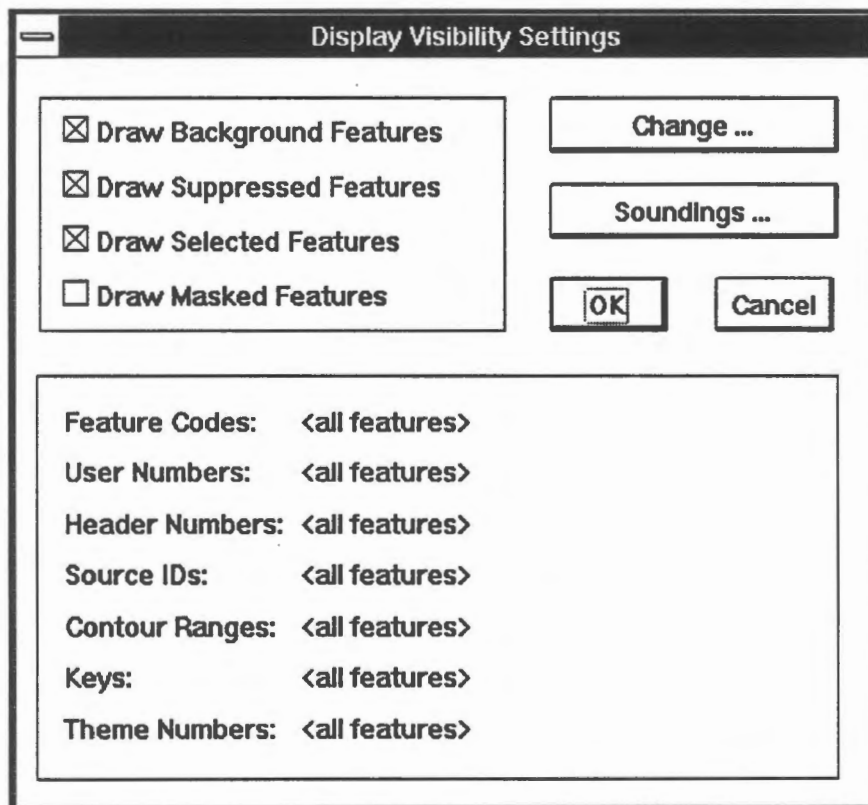
Display Selected Features

The grouping of features according to their feature codes, theme numbers, and other parameters as described in the previous chapter allows you to customize the display by selecting only features of interest. These parameters, collectively known as the *visibility parameters*, control the contents of the display.

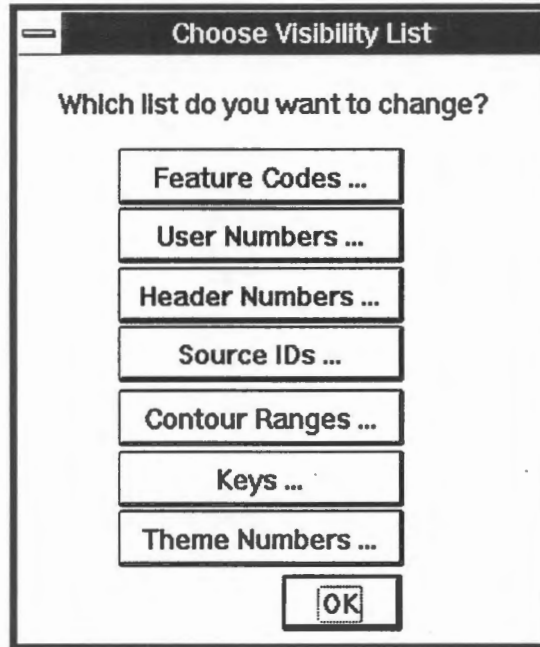
 Display features belonging to the same feature code.

The feature codes identify the type of feature. For example, all trees are associated with the feature code TREE. The following procedure describes how to display only features assigned the same feature code.

1. Select the Options > Display > Visibility Parameters ... menu item.
2. Click on the Change ... button in the Display Visibility Settings panel.

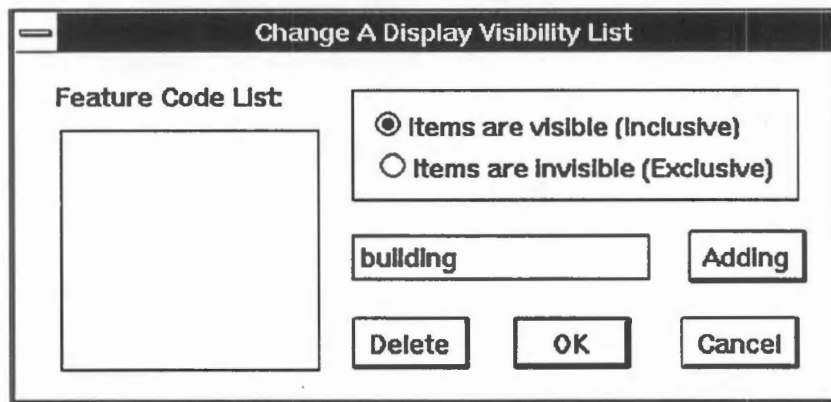


3. Click on the Feature Codes ... button in the Choose Visibility List panel.

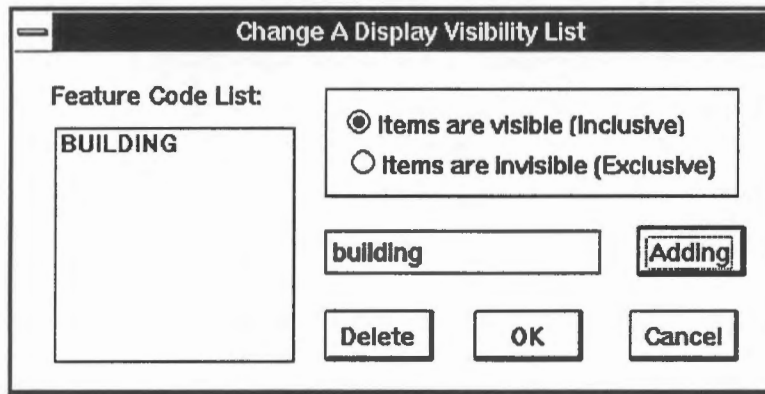


4. Enter the code "building" and click on the Adding button. You can add several codes, but we will use only one in this example.

NOTE Feature codes are not case sensitive. This means that you can use either upper or lower cases letters.



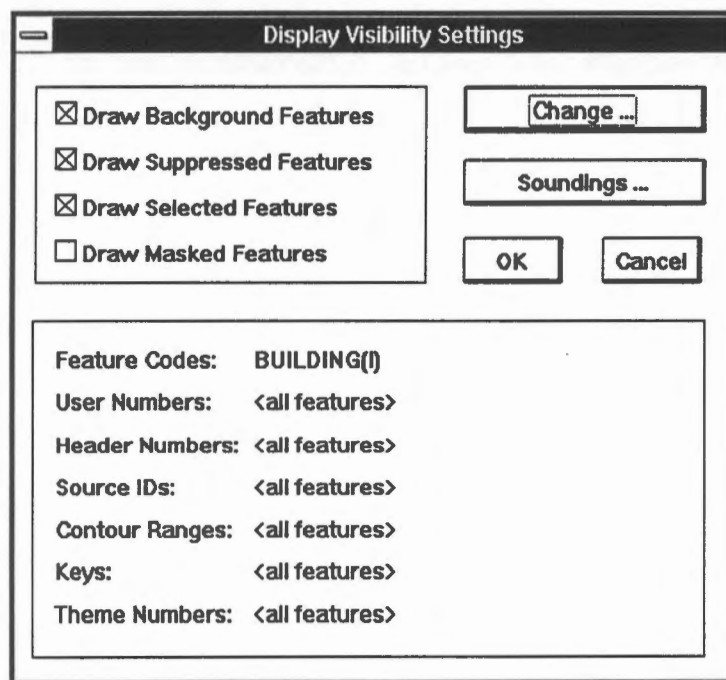
- The feature code will be added to the Feature Code List. Note that all features assigned with these feature codes will be *included* in the next display. Click on the OK button.



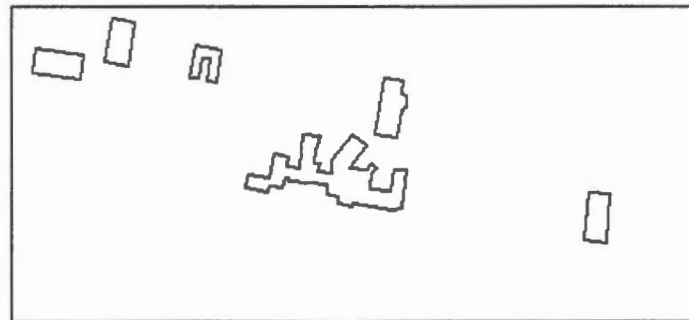
- You will return to the Choose Visibility List panel so that you can include another parameter to control the display. We are using only feature codes in this example. Click on the OK button.

That returns you to the Display Visibility Settings panel. It shows all the parameters you have chosen so far for display control. In this case, only features assigned with a feature code of BUILDING will be shown. Note that the letter (I) behind BUILDING indicates that it will be included in the display.

Click on the OK button.



7. Select the Draw > Overview menu item or the Overview icon on the CARIS Panel to refresh the screen.
8. The display will show only outlines of buildings.



(288200, 740514), 1:2000



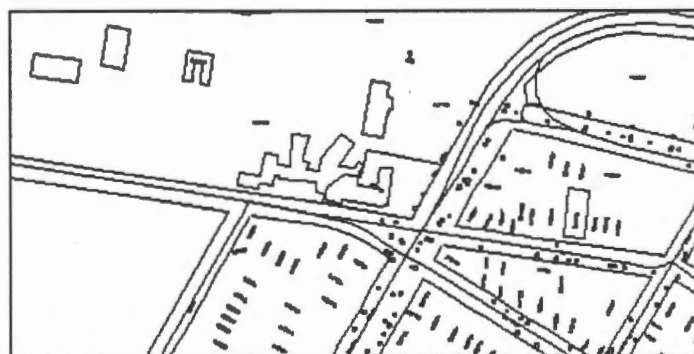
The demonstration Viewing layers of data: Displaying features according to their feature codes shows how to select features of the same feature code for display.

◆ Display all features.

Use the same menu item, Options > Display > Visibility Parameters ..., to include all features for the display. You do this by deleting all feature codes from the Feature Code List in the Change A Display Visibility List panel. ◆

◆ Display all features but parcel boundaries.

Use the visibility list to *exclude* parcel boundaries (feature code PARCEL) from the display to produce a display similar to the one below.

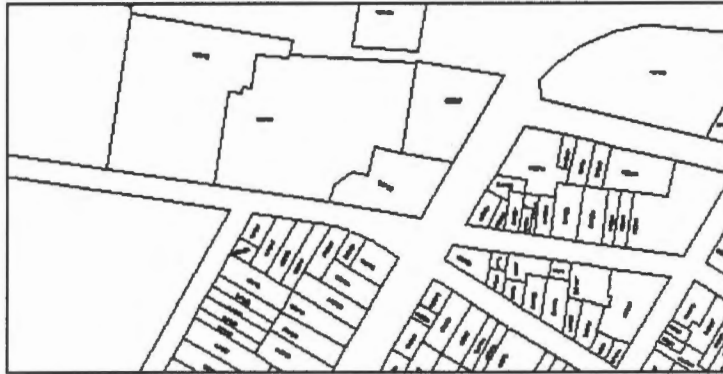


(288200, 740514), 1:2000



◆ ? Display features belonging to theme 50.

Features belonging to theme 50 are topologically related, and they form a layer for land parcels. Display all features belonging to theme 50. You should see a display similar to the one below.



◆ ? The demonstration Viewing layers of data: Displaying features of the same theme shows how to select features of the same theme for display.

◆ ? Use part of a feature code to control display.

You do not have to enter the entire feature code. A number of wild-card characters are available for you to enter partial codes. The character * matches with any string of any length while the character % matches with any letter. For example, TR* matches any code starting with TR and TR% matches any three-character code starting with TR. ◆

◆ ? Does it make sense to include a feature code and exclude another one ?

In the previous exercises, we have either excluded or included features for display based on feature codes. Does it make sense to, for example, include all trees AND exclude all utility poles ? What would happen if you actually try to do this ? ◆

◆ ? Are there utility poles not belonging to theme 30 ?

We have seen in a previous chapter a utility pole that belongs to theme 30. How would you find out using the Options > Display > Visibility Parameters ... menu item whether there are utility poles NOT belonging to theme 30? ◆

Chapter 18

Feature Attributes

The characteristics of geographic features are described by their attributes. A piece of property could have an identification code, one or more owners, an address, an assessment value, and so on. It is also represented in the database as a geometrical object, which in this case is a many-sided polygon. Being a polygon, it has a list of coordinates describing its outline, an area, a perimeter, and other attributes. A highway has a name, a surface type, and so on. If represented by a line, it has a list of coordinates, a length, etc.

Some of these attributes are common to many features, while others are different from one feature type to another. For instance, attributes related to geometrical representations are common to many. Other attributes, such as the owner of parcels and the highway surface type, are quite specific to certain types of features.

In this GIS, we distinguish a set of *system-defined attributes* that are common to many features, and a set of *user-defined attributes* that are specific to feature types. For system-defined attributes, the GIS has already allocated spaces for their storage in the database, and the meaning of them has been defined by the system as well. Feature code, feature key, theme number, and others that we have discussed while using the Information icon on the CARIS Panel are some of the system-defined attributes.

Take the feature code as an example. The users have the freedom to design the codes in whichever fashion they wish as long as they satisfy certain restrictions: they must be within 12 characters long in this GIS and reflect the feature types. Once the codes have been designed, the GIS provides the mechanism to enforce their use. Moreover, the GIS uses the codes to control the colour and symbology of the features on a display.

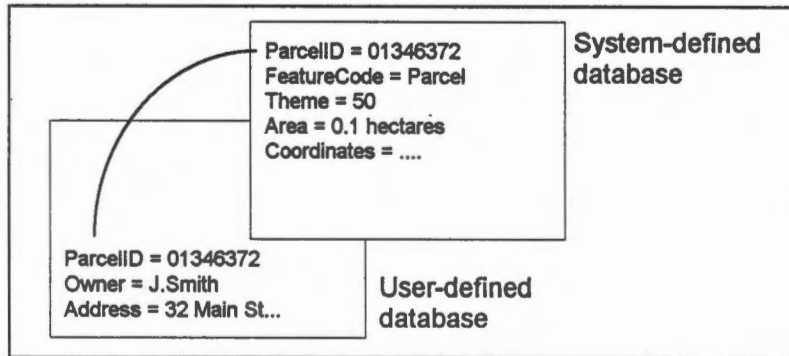
Software modules in the GIS make extensive use of the system-defined attributes in its operation. Any deviation from the expected usage of these variables could cause nasty surprises.

The user-defined attributes vary so much from one feature type to another that it would be difficult for the system to exercise complete control over them. Even for a single feature type, such as a highway, the number of attributes associated with them can vary from one application to another. Hence the number and the meaning of these attributes are left mainly to the user. The GIS only allocates spaces for them at the request of the user and provides limited tools to retrieve and manipulate them.


Some GIS store both system-defined and user-defined attributes in the same database, but others store them separately. In our GIS, as well in several others on the market, such as Arc/Info and Intergraph, two databases are used.

The contents of the system-defined database are controlled by the GIS. The user-defined database consists of a number of relational tables that can be used in conjunction with the system-defined database or separately as an independent relational database.

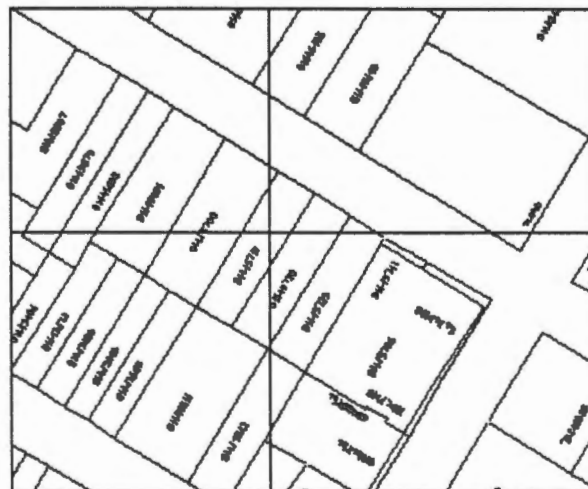
A link is established between the two databases to form a collective set of attributes. In our GIS, we use the feature key, which is a unique identifier for the feature, as the link. The example here shows that the parcel identification code is used to link the attributes from the two databases.



Some features are without keys, indicating that they are not associated with user-defined attributes.

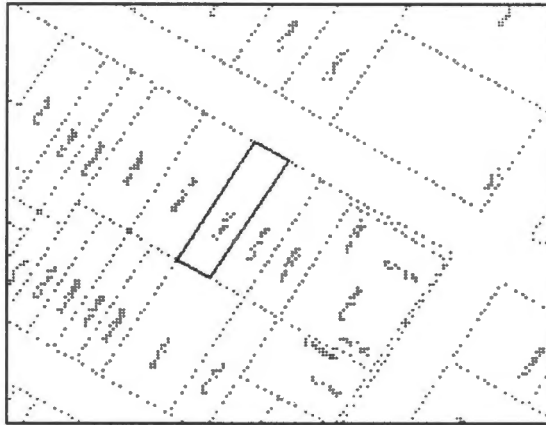
 Display user-defined attributes of a feature.

1. Use the Options > Display > Visibility Parameters ... menu item to display only data belonging to theme 50, the layer of land parcels.
2. Use the MapQuery > By Pointing menu item to select the parcel (01445717) shown in the diagram. You can point at anywhere within the parcel to select it.

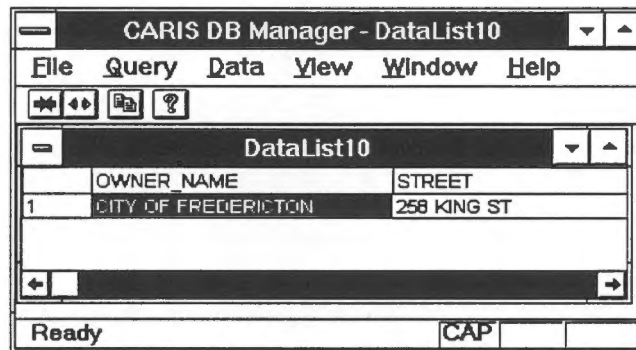


(288577, 740339), 1: 800

- The outline of the parcel will be highlighted. Click on the left mouse button to confirm that this is the parcel you want.



- The crosshair will reappear for you to form a larger collection of features by selecting others. We only want to select one in this example, so click on the right mouse button to exit.
- The database manager, which manages the user-defined database, retrieves the attributes of the selected parcel. You can enlarge this window or use the scroll bar to see the other attributes.




The attributes returned contain the following items:

OWNER_NAME	= CITY OF FREDERICTON
STREET	= 258 KING ST
POSCOD	= E3B 1E2
PROP_DESC	= PARKING LOT
PROP_TYPE	= 220
ASS_SEQ_N	= B049
PID1	= 01445717
NBHD	= 11
TAX_CLASS	= municipal
CURR_ASS	= 77700
TAX_CRED	= 7

The meanings of these attributes and their values are assigned by the user. The following is a description of them:

OWNER_NAME	Name of property owner.
STREET	Street address of property.
POSCOD	Postal code of property.
PROP_DESC	A brief description of the property.
PROP_TYPE	A three-digit code describing the use of the property. Code 220 meaning that it is occupied by a salvage dealer.
ASS_SEQ_N	A sequential number describing the order in which assessors assess properties.
PID1	Parcel identification code.
NBHD	A code assigned to properties in a neighbourhood assessed by one assessor.
TAX_CLASS	Taxation class.
CURR_ASS	The current assessment value. A value of zero means that no assessment value is available.
TAX_CREDIT	Tax credit code. A code of 7 meaning that there are no residential units.

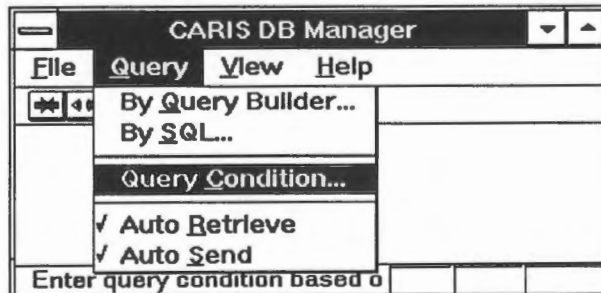
This table is designed by the user, and it can contain more columns if additional attributes are required for the properties. The design of the database tables, sometimes called *database design*, is an important step because a poorly designed database not only wastes storage space but can also cause problems during updating.

 The demonstrations Working with non-spatial attributes: Pointing at a polygon to get its attributes shows how to retrieving attributes of a polygon by pointing at it.

 Retrieve features satisfying certain attributes.

In the previous exercise, we selected a geographic feature and retrieved its user-defined attributes. We can also retrieve geographic features satisfying certain attributes. Since the DB Manager is the module that handles retrieval from the user-defined database, we will invoke this type of search from the DB Manager window.

1. Select the Query > Query Condition ... menu item.



Notice that the Auto Retrieve and Auto Send switches have been turned on, allowing the automatic passing of information between the two databases. By turning off these switches, the two databases can operate independently.

2. A window appears for you to enter the search condition. It allows you to select the relational table to be used. The column name, the operator, and the value together will form the search condition. Several conditions can be linked together either by the AND or OR logical connector.

We will be searching for all properties owned by the CITY OF FREDERICTON. Enter this in the Value field and make sure that the Column field shows OWNER_NAME and that the Operator field shows the equal sign.

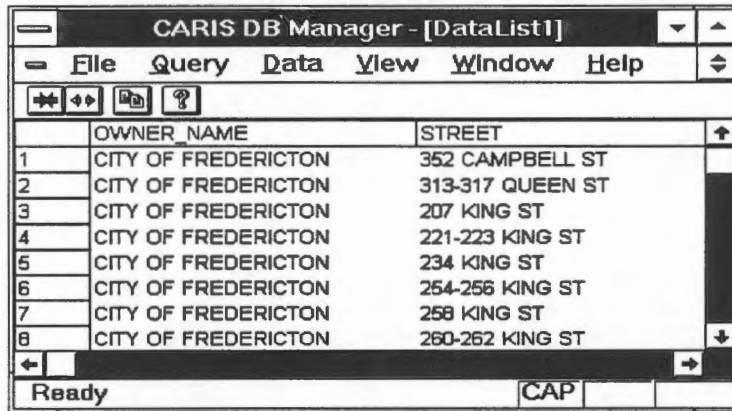
NOTE This interface is case sensitive. Make sure you use capital letters in this example.

The screenshot shows a dialog box titled "Query Condition". It has a "Name:" field with a dropdown menu showing "PROPERTY". Below this are two radio buttons: "AND" (selected) and "OR". To the right of the radio buttons are three fields: "Column:" with a dropdown showing "OWNER_NAME", "Operator:" with a dropdown showing "=", and "Value:" with a dropdown showing "CITY OF FREDERICTON". On the right side of the dialog, there are three buttons: "OK", "Cancel", and "Add". Below the input fields is a large empty rectangular area. At the bottom right, there are three more buttons: "Add", "Delete", and "Clear".

3. Click on Add to include this as a condition, then click on OK to finish entering the conditions.

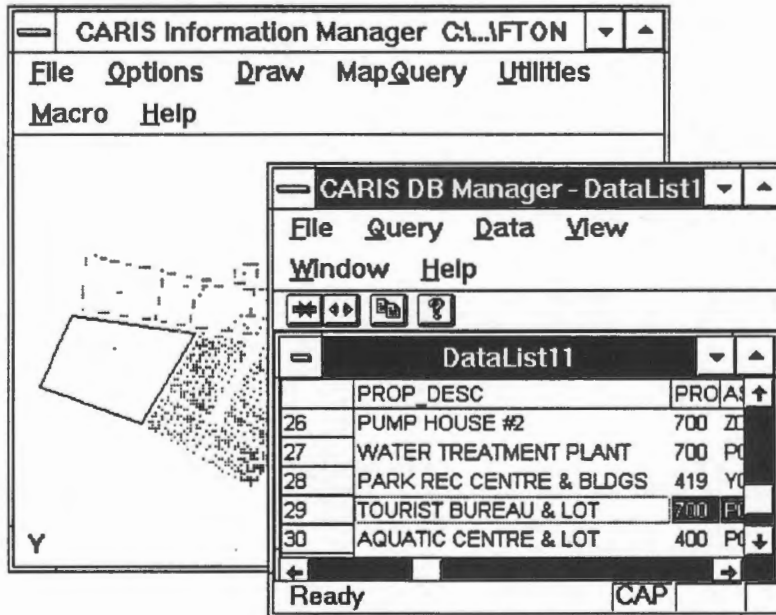
The screenshot shows the same "Query Condition" dialog box. The "Value:" field is now empty. The large rectangular area below the input fields now contains the text "OWNER_NAME = 'CITY OF FREDERICTON'". The "Add" button at the bottom right is highlighted, indicating it has been clicked.


4. The database records satisfying the condition will be returned.




5. If you double click on an entry in the data list, such as entry number 29 in this example, its parcel outline will be highlight in the map window.

NOTE Before you select another item from the data list to highlight, you must stop the highlighting on the map window by moving the cursor over it and clicking on the left mouse button.



 The demonstration Working with non-spatial attributes: displaying polygons satisfying certain attributes shows how to retrieve polygons satisfying certain attributes.

 The demonstration Spatial analysis: finding all vacant lots shows how to retrieve parcels satisfying certain conditions.

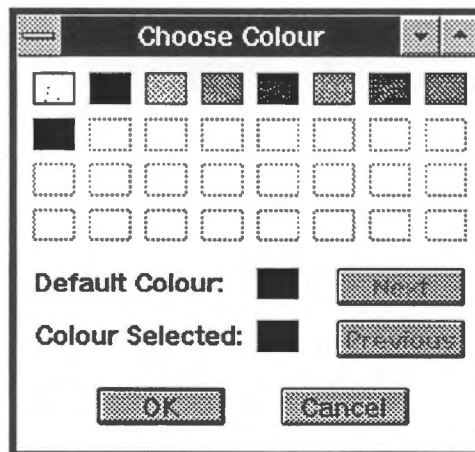
 The demonstration Spatial analysis: finding all vacant lots below \$100 000 in value shows how to retrieve polygons using a complex condition.

 Colour fill the parcels retrieved.

Instead of highlighting the parcels retrieved one by one, you can fill all of them with a colour of your choice.

1. Select the Draw > Fill Polygon > Current Selections ... menu item or the Fill Polygons icon from the CARIS Panel.


A panel will appear for you to choose a colour. Click on the OK button when you are done.



2. The selected parcels will be filled with the chosen colour.

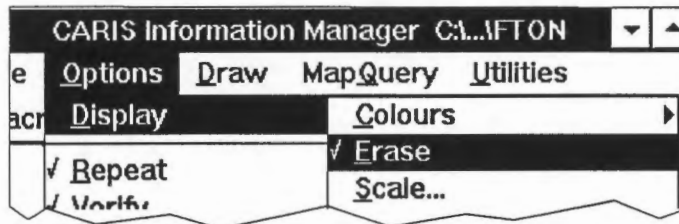
NOTE In order to emphasize the selected parcels in this image, we have both filled and highlighted the selected parcels. To highlight them, use the Draw > Highlight > Current Selections ... menu item or the Highlight Features icon.



 Colour fill polygons and superimpose with line map.

After you have done a colour fill, the line map will be hidden underneath the filled polygons. If you refresh the screen, the colour fill will be gone because the Options > Display > Erase switch was turned on. Turn this switch off if you want to show the line map on top of the filled polygons.

1. Select the Options > Display menu item. If you see a tick mark against it, the screen will be erased before refreshing. In that case, select Erase to turn off the tick mark.

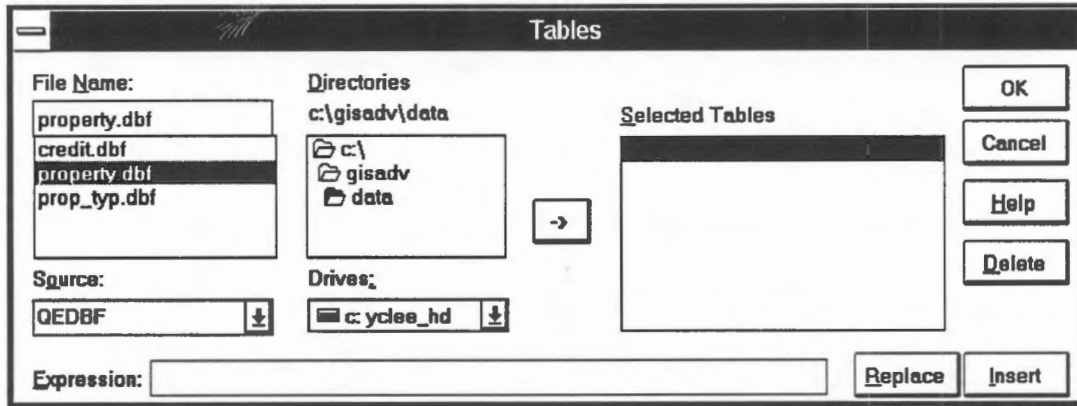



2. Refresh the display. ♦

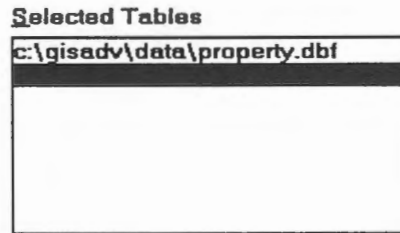
 Query by using the query builder.

This is a more powerful, but a bit more complicated, interface which allows you to enter more sophisticated queries on the relational tables. For instance, it allows us to sort the items returned and include functions in the query. We will use the query builder to find all properties with assessment values larger than or equal to 500 000 dollars and sort them in ascending order. The last item on the list is therefore the most expensive property.

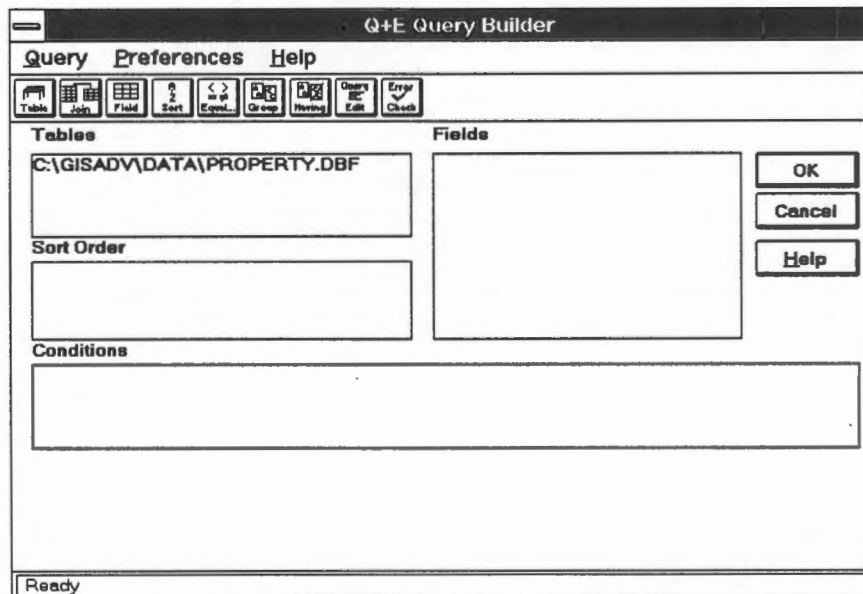
1. Pick the **Query > By Query Builder ...** menu item from the DB Manager. A Tables panel will appear. Select the **property.dbf** database file.



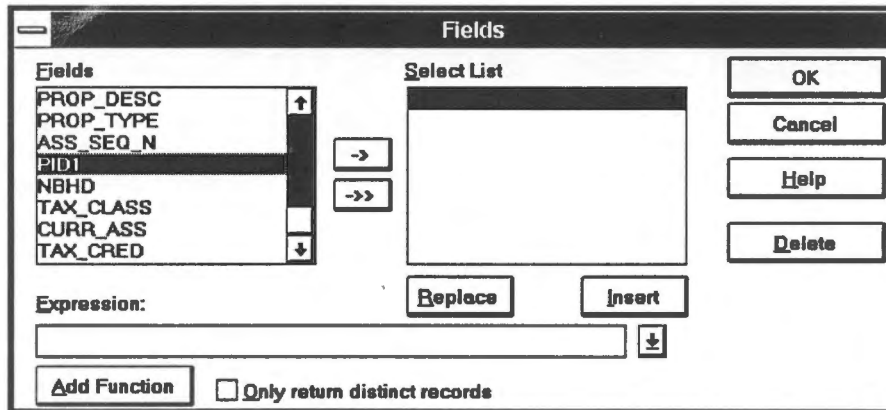
2. Click on the  button.
3. This will enter the file in the list of selected tables.




4. Click on the OK button. This opens the Q+E Query Builder panel.

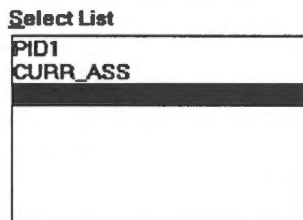



- Click on the  button to open the Fields panel.

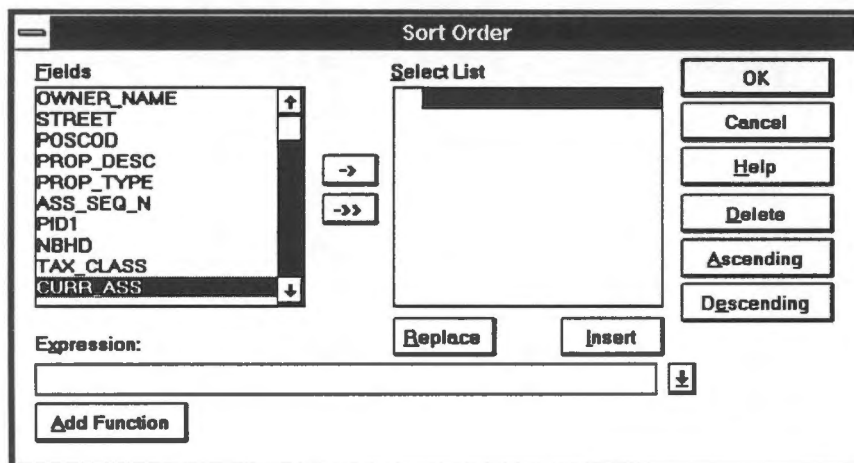


- Select the PID1 field and click on the  button. Do the same for the CURR_ASS field. This will enter the two fields in the selected list. This indicates that we want only these two columns to show in the report.

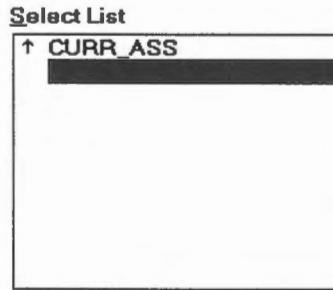
NOTE It is important to include PID1, the key field, in the list.




- Click on the OK button to return to the Q+E Query Builder panel.
- Select the  button to open the Sort Order panel. Select CURR_ASS and click on the arrow key to include it in the selected list.

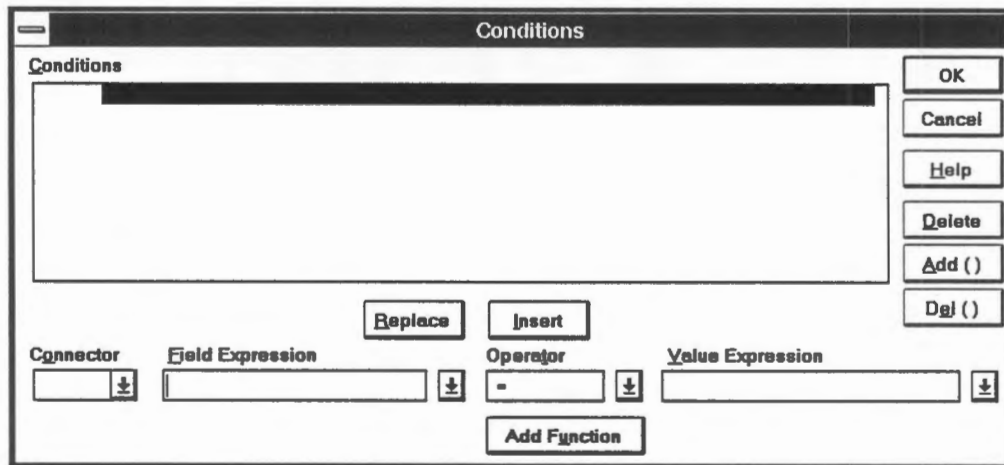


- The up-arrow next to CURR_ASS indicates that the sorting will be in ascending order. To change the order, select CURR_ASS and click the Descending button.

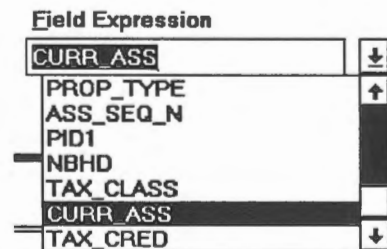


- Click on the OK button to return to the Q+E Query Builder panel.

- Select the  button to open the Conditions panel.



- Select the CURR_ASS field from Field Expression.

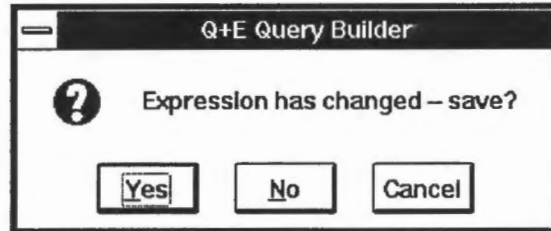


- Select >= from Operator, and enter 500000 in Value Expression.



- Click on OK.

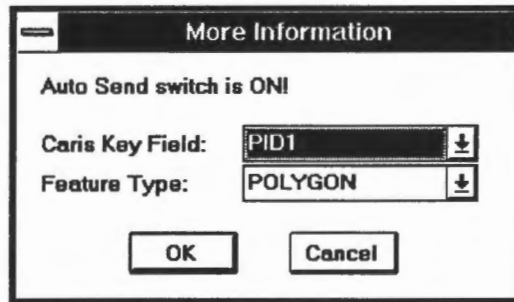
15. Click on Yes to save the expression.



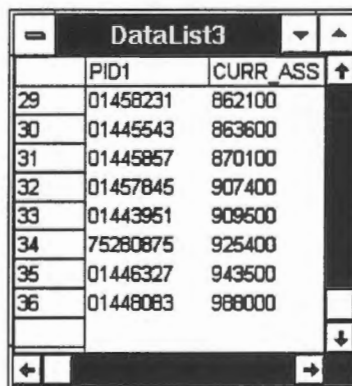
16. When the Q+E Query Builder panel returns, click on the OK button.

17. Make sure that the PID1 field is entered in the CARIS Key Field and POLYGON is entered in the Feature Type field. The key, as we have mentioned, is the link between features in the system-defined database and their attributes in the user-defined database. In this database, we use the parcel identification code as the link.

Click on the OK button.



18. A data list will be returned for all properties with current assessment value greater or equal to 500 000 dollars. The most expensive property is the last one on the list.



	PID1	CURR_ASS
29	01458231	862100
30	01445543	863600
31	01445857	870100
32	01457845	907400
33	01443951	909500
34	75280875	925400
35	01446327	943500
36	01448083	988000


◆ Find properties with assessed values between 200 000 and 500 000 dollars. ◆

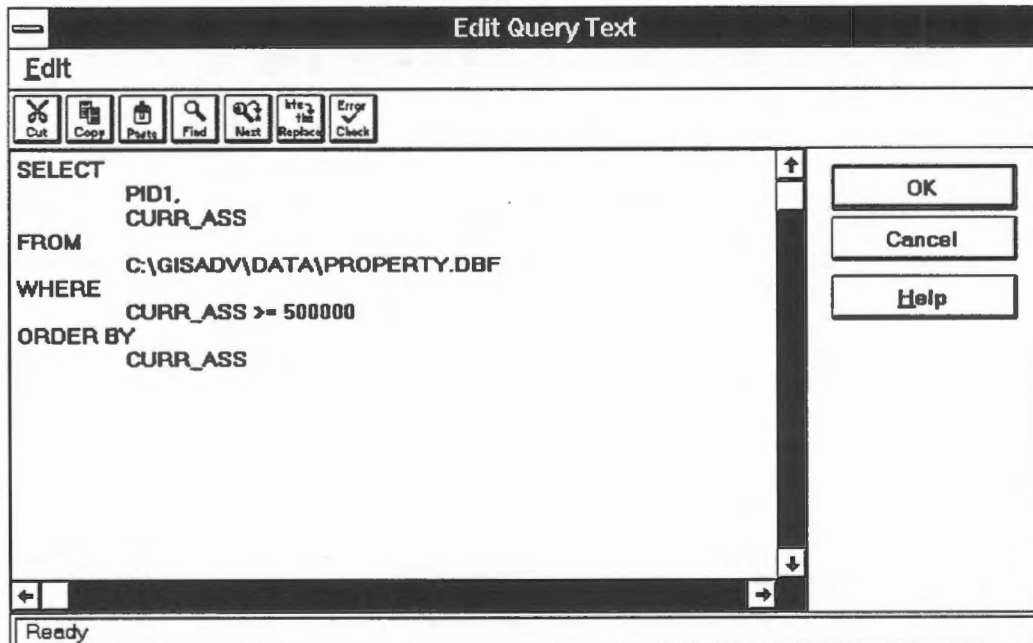
 Passing information between the two databases.

The GIS has control over both databases. When you use the crosshair to select a land parcel, the GIS queries the database of system-defined attributes including coordinates. It uses the spatial indexing system described previously and geometrical computations to determine the parcel containing the centre of the crosshair.

Then the key of the parcel, which is the parcel identification code, is read and passed to another module of the GIS that handles the user-defined database. That module is called the DB Manager in this GIS, which takes the feature key from the system-defined database and retrieves the attributes from the user-defined database. ♦

 SQL statement.

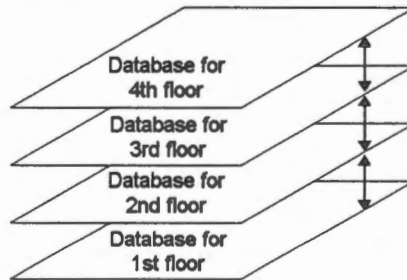
If you select the  button after you have finished entering the query condition, you will notice that the condition has been translated into an SQL (Structured Query Language) statement. You can edit the statement in the panel.



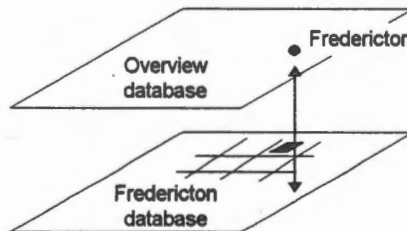
Chapter 19

Vertically Linked Databases

We have so far treated geographic features as being flat. A point can be defined with a third coordinate, thus showing its elevation or depth. But even with three-dimensional points, it is still difficult to show and manipulate solid objects such as multistory buildings. A solution to this is to allow layers of data, such as floors of a building, to be linked so that the user can traverse up and down the chain of databases.

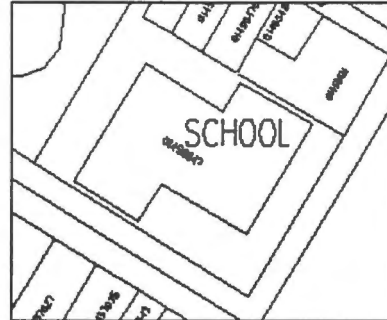


Another reason for chaining databases together is to let a smaller scale database lead to a larger-scale database showing more detail. For instance, a dot representing a city in an overview database could be chained to a large scale database of the city.



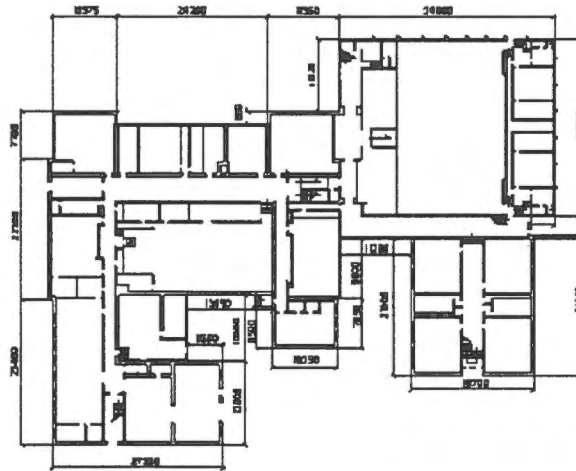
We will show in this example how a school building in the municipality database is linked to its floor plan.

1. Select the Draw > Vertical Database > Draw Down One Level menu item and point at the text SCHOOL.



(289034, 739862), 1:800

2. When the name is highlighted, click the left mouse button to confirm.
3. The floor plan for one of the floors in the school will be shown.



It is a complete database by itself except that we have not created user-defined attributes for it. We also have not built topology for it. Hence you cannot use the DB Manager to query this database. You could, however, perform measurements such as length and area on it.

4. You can return to the municipality database by selecting the Draw > Vertical Database > Draw Up One Level menu item.



The demonstration Viewing vertical layers of data: Vertical databases shows how to view vertical databases.

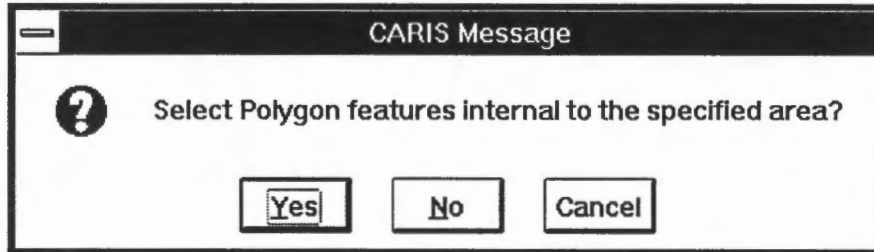
Chapter 20

Select Features Within an Area

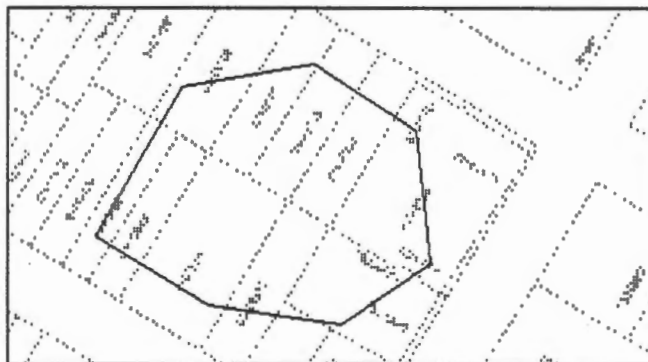
Instead of pointing at the polygons or specifying a search condition based on the user-defined attributes, we can also select polygons within a specified area.

 Select polygons within an area.

1. Pick the MapQuery > By User Defined Area ... menu item.
2. Click on Yes to indicate that you want to select the features wholly or partially overlapping the area you are going to define.



3. Use the crosshair to draw the outline of the area. Click on the right mouse button to close. The polygon will be highlighted for your confirmation. Click on the left mouse button to confirm.

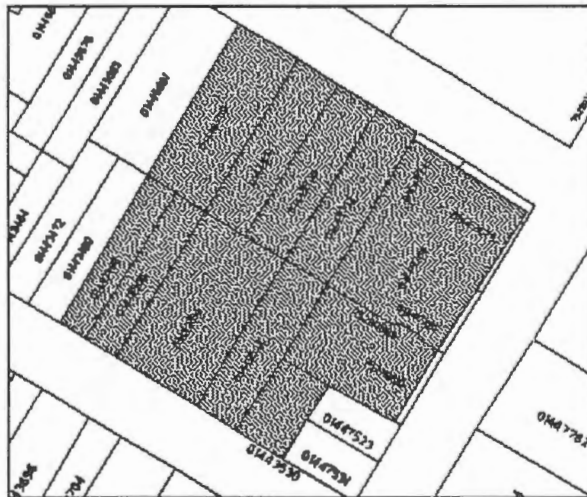


(288577,740339), 1:800

4. The parcels within the polygon will be retrieved from the database.

DataList2		
	OWNER_NAME	ST
1	CITY OF FREDERICTON	25
2	CITY OF FREDERICTON	25
3	CITY OF FREDERICTON	26
4	CITY OF FREDERICTON	27
5	CITY OF FREDERICTON	24
6	CITY OF FREDERICTON	25
7	ALPEN ERIC	25
8	CITY OF FREDERICTON	27
9	G & C VENTURES LTD	12
10	CITY OF FREDERICTON	27
11	KILEEL DEVELOPMENTS LTD	29

5. Colour fill the parcels selected.



Search for polygons overlapping an irregular area.

In the database, we keep a record for each polygon. The structure of each record is rather involved, but in essence it provides information about the theme to which the polygon belongs, the arcs forming its outline, its area, its perimeter, its key, and so on.

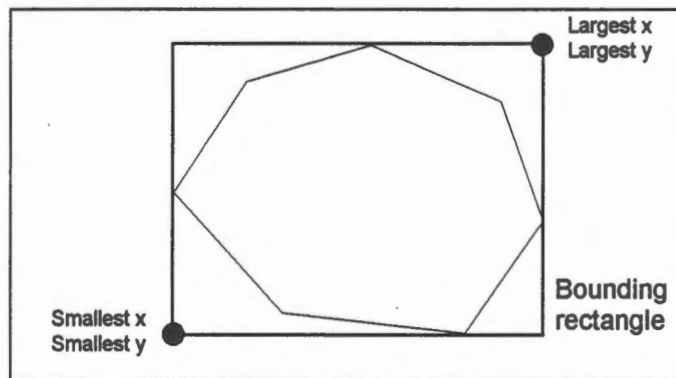
A method of searching for polygons overlapping a user-defined area, called a *window*, is as follows.

1. For each polygon in the database retrieve its outline.

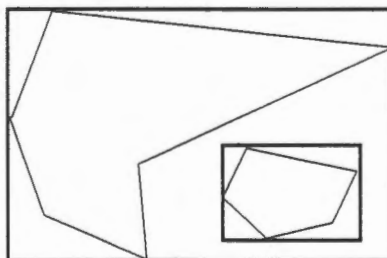
2. If the polygon is entirely within the window, then they overlap.
3. If any side of the polygon intersects any side of the window, then they also overlap.

This method involves a considerable number of geometrical tests on intersections — the vectors on the outline of each polygon will be checked against the vectors on the outline of the window. Although the mathematics involved is not very complicated, the number of intersection tests to be performed could be prohibitively large.

One way of reducing the number of intersection tests is to first check whether the bounding rectangle of the polygon overlaps with that of the window. A bounding rectangle is one bounding a feature, and it can be defined by two diagonal points, one from the smallest x and y coordinates, and the other from the largest x and y coordinates of the feature.



If the bounding rectangle of a polygon is outside that of the window, then they certainly do not overlap. If the two bounding rectangles cross each other, then we need to invoke the intersection test to ascertain whether they overlap. If one bounding rectangle is entirely within another, then we still need to invoke the intersection test because it is possible that they do not overlap at all, as shown in the following diagram.



Each bounding rectangle is composed of just four vectors. This greatly reduces the number of intersection tests required to find out whether the bounding rectangles are overlapping or disjoint.

The bounding rectangles are so useful for solving this and similar problems that they are computed by the GIS and stored as system-defined attributes for the features. ♦

Chapter 21

Thematic Mapping

Topographic and thematic maps are the two basic map types. The main difference between them is their contents. *Topographic maps*, as the name implies, contain mainly topographic information such as water bodies, contours, and vegetation covers. To these, we often add man-made features such as settlements, transportation networks, and major structures. They are general-purpose maps in that their contents are useful to a larger groups of users and they often form the base of thematic maps. For this reason, they are mostly produced by national or regional mapping organizations according to well-defined specifications.

Thematic maps concentrate on the portrayal of a small number of geographical phenomena, and they are more specialized than topographic maps. Common types are soil maps, geological maps, population distribution maps, and land use distribution maps. They are not just produced by professional map makers but also by a much broader group of amateurs who need thematic maps for research, marketing, presentation, or other reasons. Whenever there is a need to include topographic features, thematic cartographers often base them on the official topographic maps.

In this chapter, we will create a land use distribution map using the parcels as the base. To produce such a map, we will first identify the classes of land use to show and the colours to show them with. In this example, we will show the distribution of office complexes and retail stores. The method involved is quite simple — retrieve one type of land use from the database at a time and colour fill the selected parcels.


 Produce a land use distribution map.

1. Pick the *Query > Query Condition ...* menu item from the DB Manager window.
2. Retrieve office complexes (PROP_TYPE is 211).
3. Pick the *Draw > Fill Polygon > Current Selections...* menu item or the Fill Polygons icon from the CARIS Panel to fill the selected parcels with a colour of your choice.



4. Repeat from step 1 to retrieve the retail stores (PROP_TYPE is 210), and fill them with a different colour. The map below shows both types of land use with the retail stores highlighted.

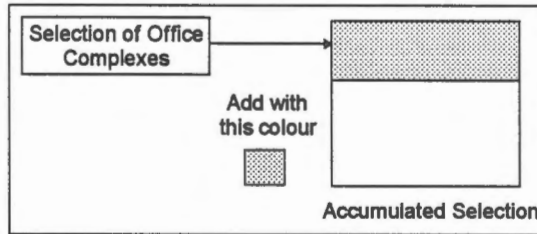


 Save the selections.

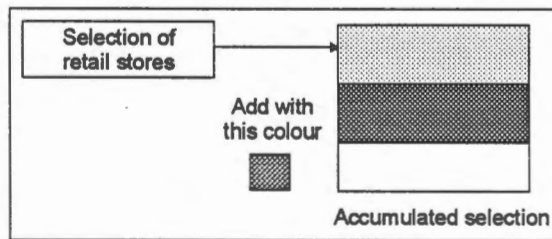
If you refresh the display, all the colour fills will disappear. Using the fill polygons command again will only colour the last selection consisting of retail stores. To avoid losing a selection, you can add it to an accumulation before starting a new one.

1. Select the office complexes as described earlier.
2. Pick the MapQuery > Accumulate Selections > Clear menu item to clear the accumulation. This is just to ensure that you start with a clean slate.

- Pick the MapQuery > Accumulate Selections > Add ... menu item to place this selection in the accumulation. You will be asked to select a colour.



- Repeat from step 1 to select the retail stores and add them to the accumulated selection with another colour. The accumulated selection will now contain two selections.



- To display the two selections, pick the Draw > Fill Polygons > Accumulated Selections menu item. ♦

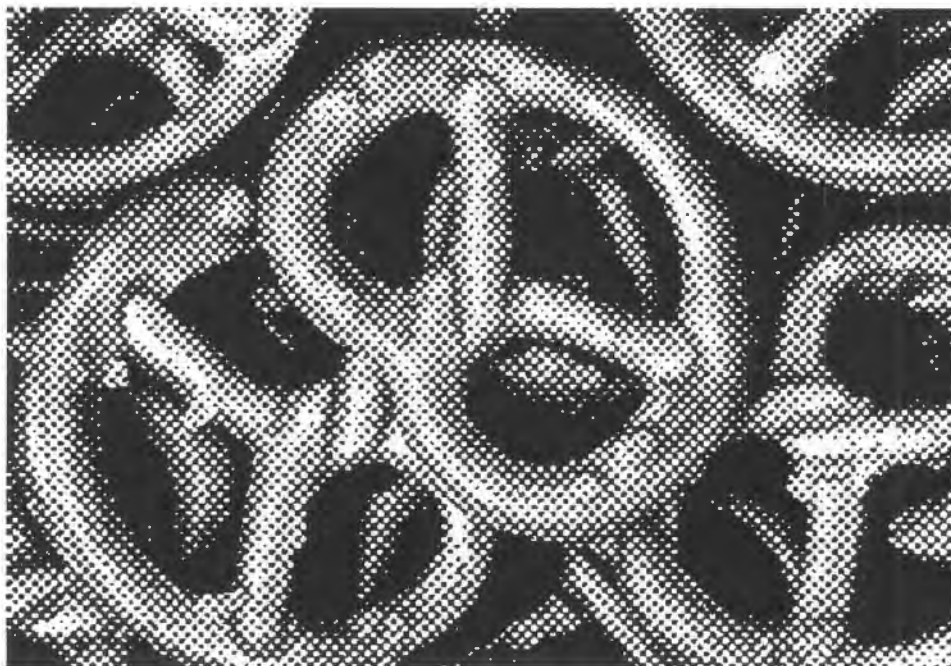
❓ Produce a map showing property values.

Produce a thematic map with parcels showing the following ranges of assessment values. Choose your own colours for the ranges.

Description	CURR_ASS
Not assessed	0
Low value	< 100 000
Medium value	>= 100 000 and < 1 000 000
High value	>= 1 000 000

♦

Part 5
Topology



Topological Relationships

If we say that a school is two kilometres from a hospital, the two features are related by a distance. If we say that a school is located on an island, the two are related by an enclosure relationship. There are many ways features can be related to each other. In a GIS, we are particularly interested in geometrical and topological relationships.

Geometrical relationships change with the coordinate system. Distance is one of these relationships. We have seen earlier that using a flat map and using an ellipsoidal earth to measure the distance between two points could yield different results.

Topological relationships are not affected by the coordinate system. The fact that a school is on an island would not change whether you are looking at a flat map or an ellipsoidal earth.

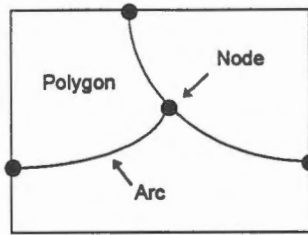
In a GIS, there are two ways of handling relationships. One way is to derive them on-the-fly when requested by the user. This is the approach used when we invoke the distance calculation command to compute the distance between two arbitrary points.

Another way to handle a relationship is to derive the relationships ahead of time, store them in the database, and retrieve them when required. This approach is suitable for relationships that take considerable time to derive and are relatively static. Topological relationships are ideal candidates because their derivations require fairly intensive calculations, and once derived would remain valid until the next update to the geometric data. It is common practice, therefore, to pre-derive the basic topological relationships and store them in the database. In essence, this approach trades storage for response time.

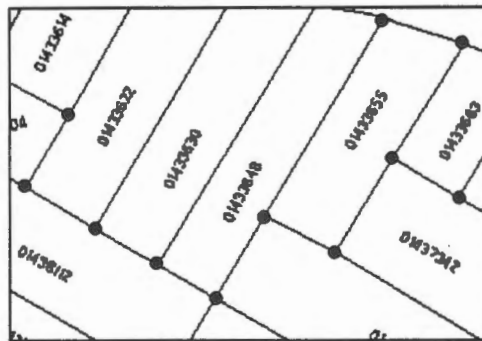
Three basic topological relationships are usually stored: connectivity, adjacency, and enclosure. *Connectivity* describes how lines are connected to each other to form a network. *Adjacency* describes whether two areas are next to each other, and *enclosure* describes whether two areas are nested.

To help derive topological relationships, we impose a restriction on lines which is to not cross each other. We sometimes call lines that observe this restriction and that are topologically related *arcs*. This is not to be confused with circular arcs.

Arcs terminate at *nodes*, and therefore each arc joins two nodes. The arcs in turn form the outlines of *polygons*. The nodes, arcs, and polygons form a hierarchy of topological elements upon which topological structures are built.



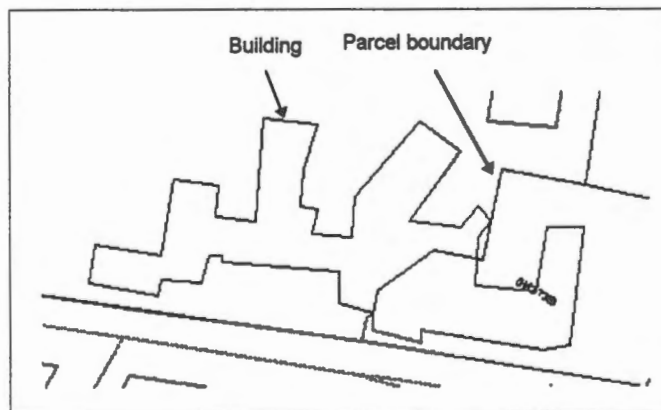
In the map below, the parcel boundaries are arcs and they meet at the nodes. The boundaries in turn form closed outlines of the parcels.



Features that are broken down into arcs, nodes, and polygons that are related to each other topologically in the database to form a *topological layer*. A database could contain one or more topological layers.

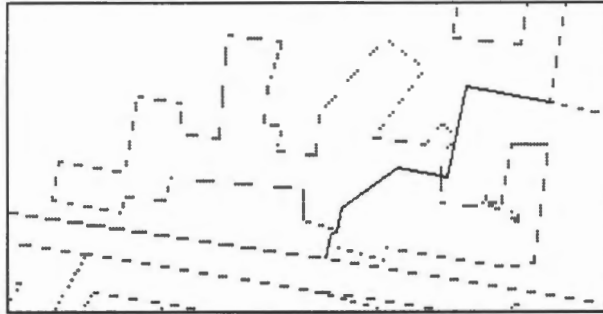
Which lines are arcs and which features make up a topological layer are decisions made by the project manager or the database designer. In this database parcel boundaries and their parcel identification codes are grouped into one topological layer. Hence the parcel boundaries are arcs that stop at intersections.

The buildings and the parcels, however, do not form a topological layer.

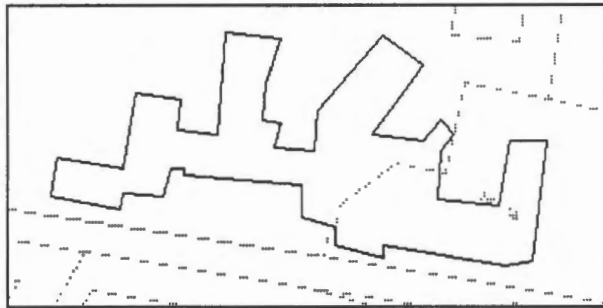


(288172, 740511), 1:800

Hence the boundaries do not stop at the outlines of the building.



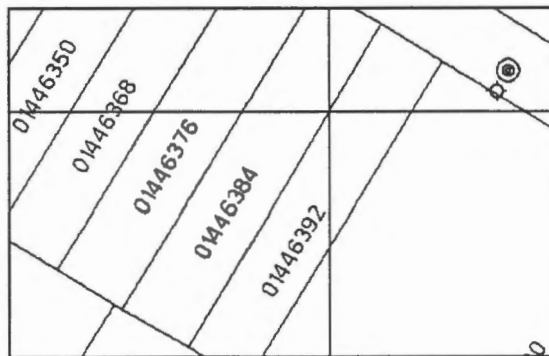
For the same reason, the building's outlines do not stop at the parcel boundaries.



Because the two kinds of features are not in the same topological layer, the database does not store their topological relationships. The GIS therefore does not have information about which buildings are contained in which parcel.

 Display information on arcs.

We have done this in a previous chapter. Use the Information icon on the CARIS panel to display system-defined information for this line.



(288658, 740417), 1:400

The information for this line shows that it belongs to theme 50, it is an arc, and it borders on two parcels with identification codes 01446392 and 01446384 respectively. From this, we can tell which two parcels are neighbours, and with some processing, we can also determine all neighbours of a given parcel.

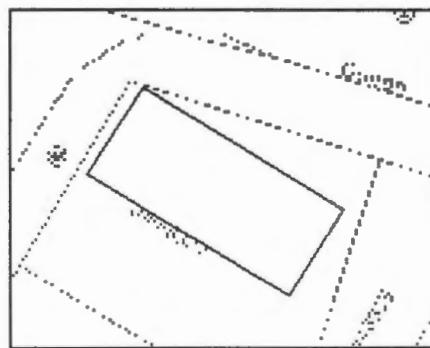
Line Information			
Feature Code:	PARCEL	User #:	50
Description:		Type:	LINE
Source ID:	13L31SE	Theme #:	50
Line Weight:		Scale Indicator:	
Left Polygon:	01446392	Key:	
Right Polygon:	01446384	Length:	50.799 m
Elevation:		Elevation (Y/N)?:	N
Arc (Y/N)?:	Y	Header No.:	1
Compilation Status:	Background		

OK

The identification code of the polygons on two sides of an arc are some of the explicit topological information stored in this database. This information is derived automatically by the GIS. The process of deriving topology is called *topology building* and it is performed by a special module in the GIS not included in this customized version. However, topological relationships were derived and stored in this database to support our analysis. ♦

 Display information on ordinary lines.

1. Use the Information icon to display information for another line, such as the one highlighted in the map below.



(288560, 740486), 1: 300

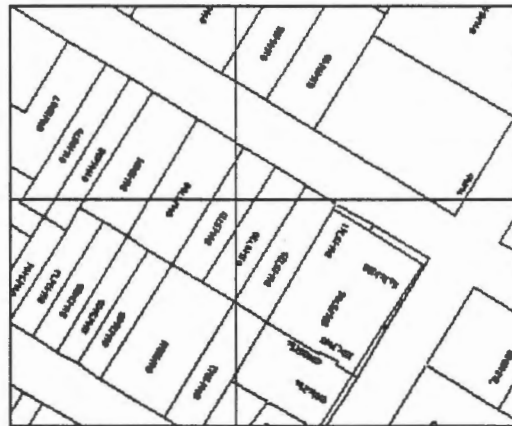
2. Notice that it is not an arc, and therefore its left and right polygons are not known.

Line Information			
Feature Code:	BUILDING	User #:	10
Description:	BUILDING	Type:	LINE
Source ID:	SMKG5761	Theme #:	10
Line Weight:		Scale Indicator:	
Left Polygon:		Key:	
Right Polygon:		Length:	110.462 m
Elevation:		Elevation (Y/N)?:	N
Arc (Y/N)?:	N	Header No.:	1
Compilation Status:	Background		

 Find neighbours of a polygon.

To illustrate the use of explicit topological information in the database, we will retrieve all neighbours of a parcel we have seen before.

1. Use the MapQuery > By Pointing menu item to select the parcel pointed to by the crosshair.



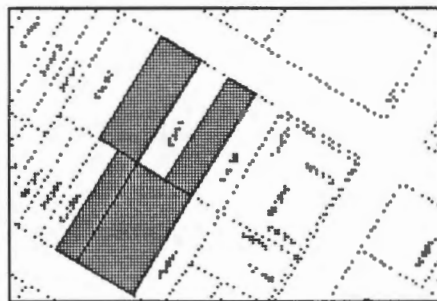
(288577, 740339), 1: 800

2. The attribute list of this parcel will be displayed.
3. Select the MapQuery > Polygon Neighbours > Current Selections ... menu item. This will produce a list of attribute records for the neighbours.

CARIS DB Manager - DataList1		
File Query Data View		
Window Help		
DataList13		
	OWNER_NAME	STREE
1	CITY OF FREDERICTON	254-258
2	CITY OF FREDERICTON	260-262
3	CITY OF FREDERICTON	255 BR
4	ALPEN ERIC	259 BR

Ready CAP

- Use the Draw > Fill Polygon > Current Selections ... menu item or the Fill Polygons icon on the CARIS Panel to colour fill the neighbours.



NOTE Parcels sharing common boundaries as well as common points are regarded as neighbours.



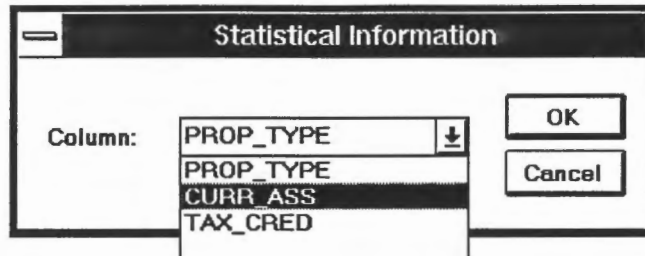
 The demonstration Spatial analysis: Finding the neighbours of a polygon shows how to retrieve the neighbours of a polygon.

 Statistical analysis.

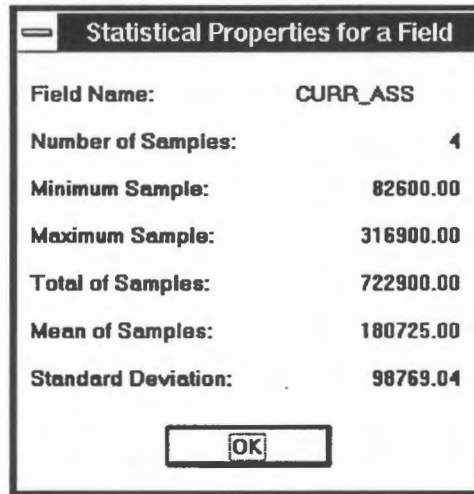
Some simple statistical analysis can be performed on the parcel records returned.

- Select the Data > Statistics ... menu item on the DB Manager window. A panel will appear for you to select the desired calculations.

You can select any column that contains numeric values for analysis. In this example, select the current assessed value of the properties.



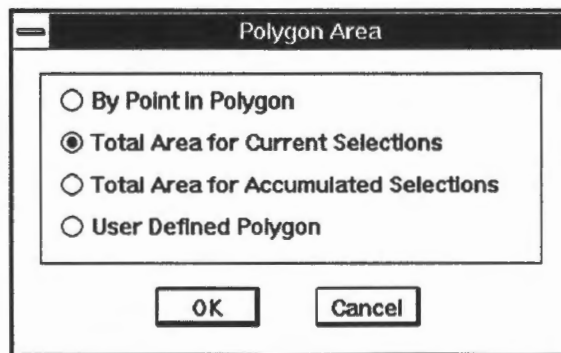
2. The results of the analysis are shown in tabular form.



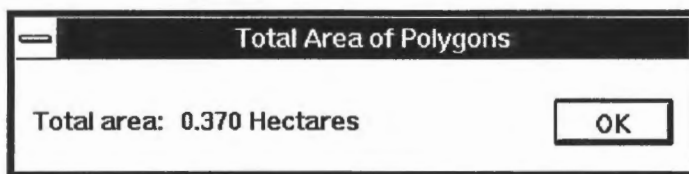
 Calculate the total area of the selected parcels.

The DB Manager handles only user-defined attributes, and areas of parcels are not among them in this database. Instead, they are part of the system-defined attributes. To find the total area of the selected parcels, use the MapQuery menu item.

1. Select the MapQuery > Polygon Area ... menu item.
2. Turn on the Total Area for Current Selections switch.

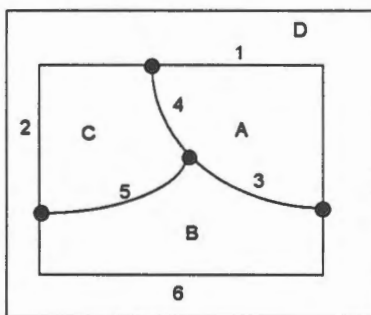


3. The total area in hectares will be displayed.

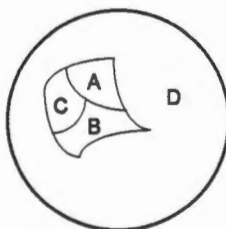


 The method of finding all neighbours of a polygon.

If we know the left and right polygons for each arc, then it is a simple matter to find all its neighbours sharing common boundaries. The following map shows a collection of nodes and arcs forming four polygons. Note that we have also regarded the region outside of the map as a polygon (labeled D).



This is indeed necessary because these parcels exist on a round earth, and the region outside does form a closed area.



The following is a table created to capture the information about polygons sharing common arcs.

Arc	Left Polygon	Right Polygon
1	D	A
2	D	C
3	A	B
4	A	C
5	C	B
6	B	D

To find the neighbours of, say, polygon A, select all arcs with A as either its left or right polygon. This produces arcs 1, 3, and 4 from which we can deduce that polygons B, C, and D are neighbours of A.

This is a highly simplified description of what is actually done by the GIS when you use the MapQuery > Polygon Neighbours command. Note that this method cannot find neighbours sharing a common point. This GIS uses a much more complicated data structure and algorithm to quickly find the neighbours, whether or not they share a common boundary or a common point. We will not go into details on how this is done, but some of the methods explained here are being used. A description of a more powerful structure for organizing topology is given in a later chapter on Topology and Editing. ♦



The demonstration Concepts: Topological structures explains the need for topological information in a GIS.



The demonstration Concepts: Topology building explains the process of building topology.

Use Topology to Identify Polygons by Pointing

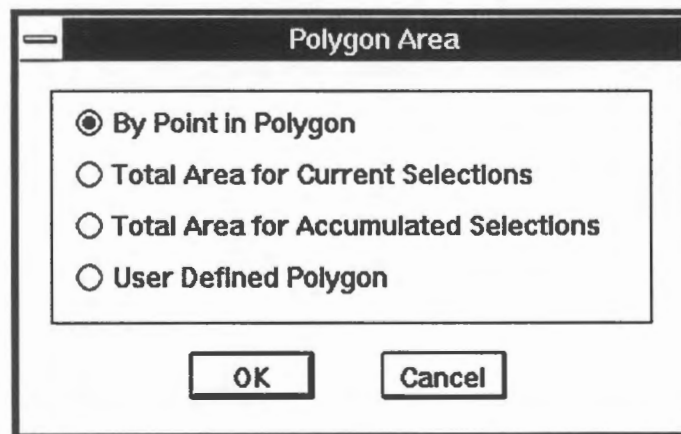
In an earlier chapter, we measured the area of a stored polygon by entering on the screen the points forming its outline. The snap option allows you to select the stored points for the calculation thus obtaining the maximum precision possible. If a large number of points are used to measure highly irregular polygons, this process would be very tedious.

It is not really necessary to hand pick these points if the file has been topologically structured because the system would have known which points make up which arcs, and which arcs make up which polygons. It would then be possible to simply identify the polygon of interest and let the system retrieve the points forming its outline to complete the area calculation.

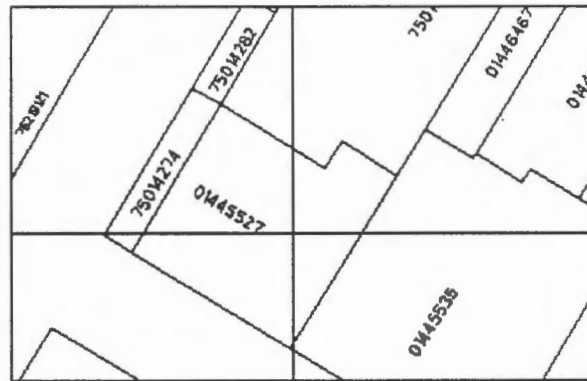
In fact, the relating of arcs to polygons and the calculation of their areas has already been performed by the topology building module called the *topology builder*. In this GIS, the areas computed are stored as one of the system-defined attributes. Hence an area calculation on a topological database is simply a retrieval operation. Unlike the manual method we have used before that involved a lot of pointing on the screen, retrieving the area of a polygon from a topological database involves only pointing at the inside of the polygons.

 Retrieve area of a polygon.

1. Pick the MapQuery > Polygon Area ... menu item.
2. Turn on the By Point in Polygon switch. This allows you to retrieve a polygon by pointing at its inside.

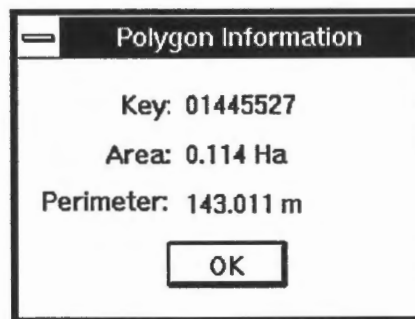


3. Point at the polygon for which you want to retrieve the area. The polygon will be highlighted for your confirmation.



(288706, 740328), 1:500

4. The pre-computed area (and perimeter) of the polygon will be retrieved.



◆ Is the area and perimeter measured on a flat map?

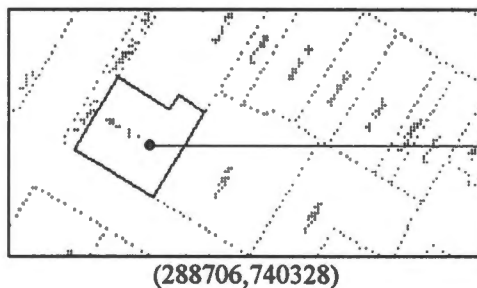
Use the MapQuery command to determine whether the area and the perimeter are computed on a flat or an ellipsoidal surface. ◆

📖 Search for the polygon at which the user is pointing.

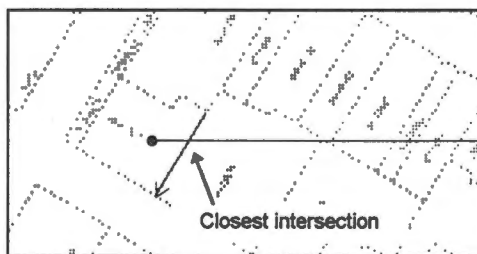
On a line map there are points, lines, and areas. The areas are enclosed by lines. When we point at a polygon, we normally point at its inside. Pointing at its outline is ambiguous because two polygons share the same boundary.

The problem is: given the coordinates of the crosshair, determine which polygon in the database encloses that point. This operation, called the *point-in-polygon operation*, is one of the most fundamental in a GIS.

Here is one solution. Extend an arbitrary line from the point to the edge of the database. The line can be extended in any direction, but one which is parallel to the axes is easier for the computations. Then we perform intersection tests on this line with all the arcs in the database.



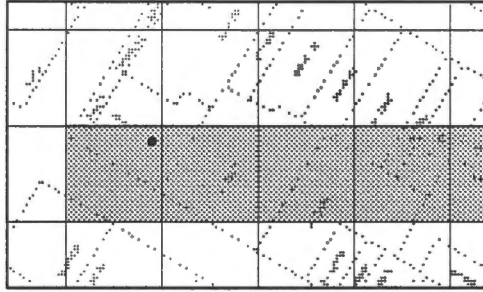
Of all the intersections found, determine which one is closest to the point. For example, the arc highlighted below contains the closest intersection. That arc has a direction as indicated by the arrow.



In the topological database, the polygon to the left of this arc has already been identified as 01445535 and the one to the right as 01445527. You can verify this by displaying information on this arc. The final step in the solution is to determine if the point is to the left or to the right of this arc. We will not elaborate on the method here except to point out that the calculations involved are not complicated.

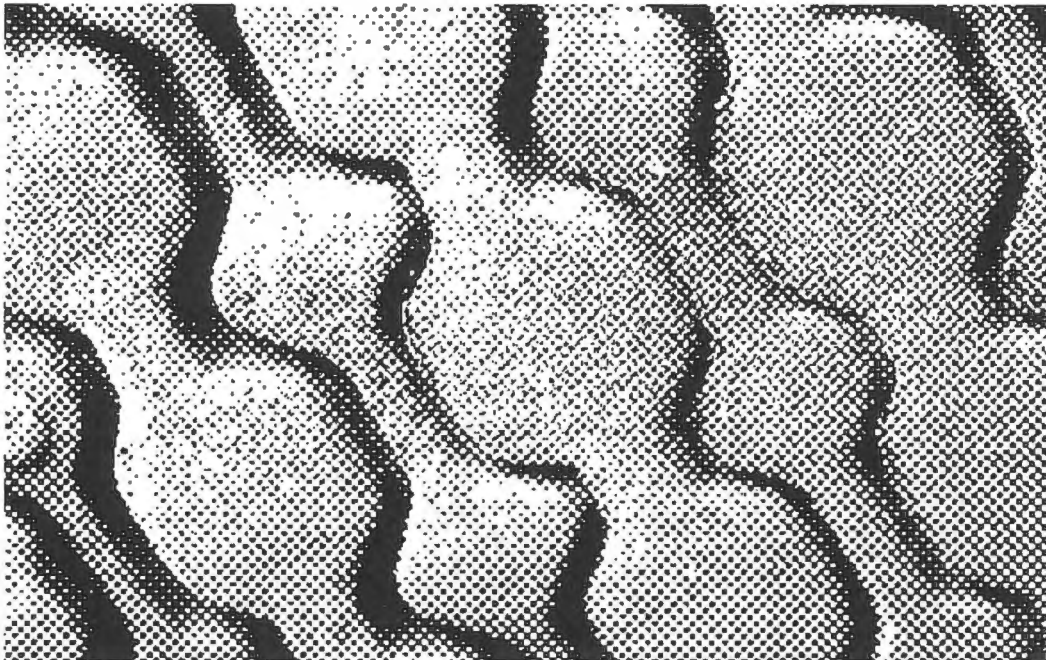
It is a tedious process to perform intersection tests on all arcs in the file to find the closest intersection. But this cannot be avoided unless a spatial indexing system is in place to help minimize the number of such tests.

We have described earlier, in the chapter on Coordinate System and Distance Measurement, a spatial indexing system that logically divides the database into regular cells. This structure, if used on a database of arcs, can tell which arcs fall inside which cells. In the intersection test, we first locate the cells containing the point and the extended line, which are those shaded in the following diagram.



All arcs falling outside the shaded area can be immediately excluded from the intersection test, thus reducing the number of calculations required. ♦

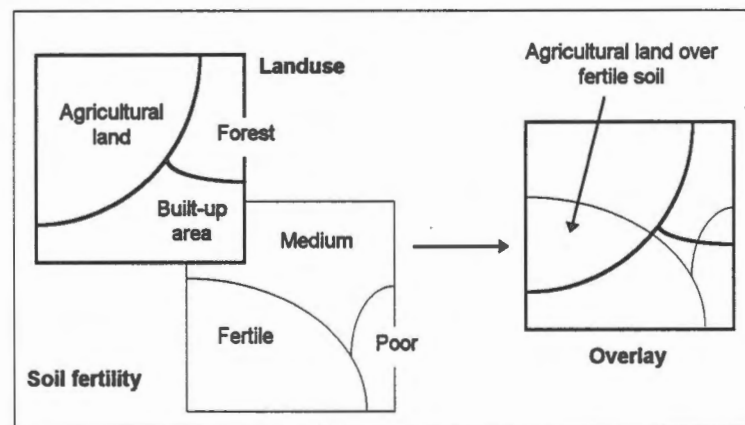
Part 6
Analysis and Modelling



Chapter 24

Overlay Analysis

Overlay analysis is an age-old technique of deriving new information from two or more layers of data covering the same area. For instance, if we have a layer of land use and a layer of soil fertility, we can derive from the two layers the percentage of agricultural land developed over fertile soil.

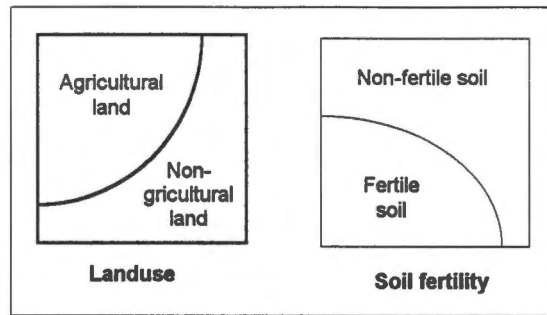


Traditionally, a land use map and a soil map would be placed over each other on a light table, and we would visually inspect the composite to look for areas where agricultural lands overlapped fertile soils. This manual process is tedious and, when many layers are involved, could become extremely difficult for the human eyes to see through them. In a GIS, overlaying is done digitally, and many of these restrictions do not apply.

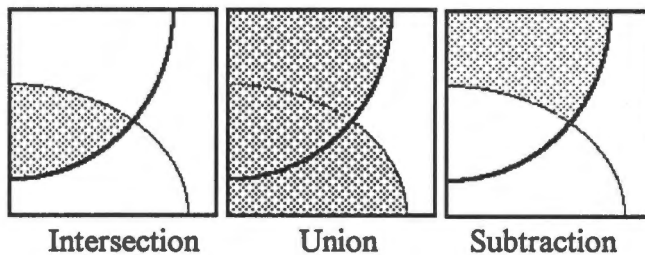
Putting two data layers together to form a composite is only a first step in the overlay analysis. After we have created new areas out of the overlapping regions, we would normally select only those areas of interest to us. For instance, we might be interested only in areas where agricultural land is developed over fertile soil. In another analysis, we might be interested only in areas where agricultural land is developed over poor soil.

A map layer involves different kinds of features. When performing an overlay analysis, we often find it convenient to group features into two classes: those satisfying certain conditions, and those not satisfying the same conditions.

In the land use layer of our example, we are interested in the agricultural lands. So we group all other land use types into non-agricultural lands. Similarly, we identify fertile and non-fertile soils in the soil fertility layer.



After we have reduced each layer into two classifications, we can perform intersection, union, subtraction, and other logical operations on them. The following shows the results of some of these operations. The intersection shows agricultural land over fertile soil, union shows agricultural land and fertile soil combined, and subtraction shows agricultural land over non-fertile soil.

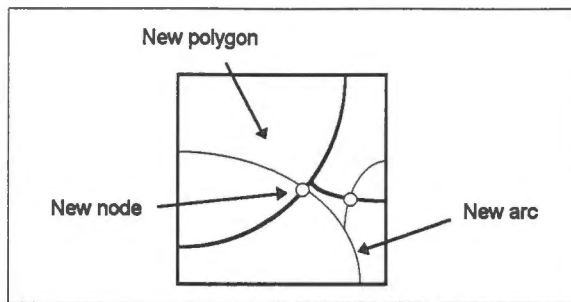


Overlaying does not involve *only* areas. We sometimes overlay areas upon lines to find out, for instance, how many countries a river flows through. We could also overlay areas upon points. An example would be to determine how many trees there are within a park. In theory, overlaying of points upon points, points upon lines, and lines upon lines is possible. They are less common, and many GIS do not support these options.

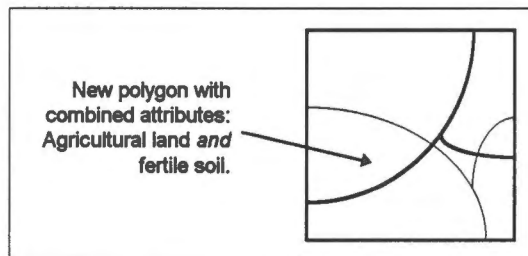
In this chapter, we will address the overlay of areas upon areas using data of both raster and vector formats. Other kinds of overlay analysis will be described in the following chapters.

Vector Overlay

In vector overlay, the input layers and the output composite layer are topologically structured. Hence an overlay operation generates new nodes, arcs, and polygons.



The new polygons will inherit the attributes from the two original layers. For instance, one of the new polygons in our example will have properties inherited from both land use and soil fertility layers. Even an old polygon which was not split during the overlay process will have additional attributes associated with it because it is contained within another polygon from another theme.

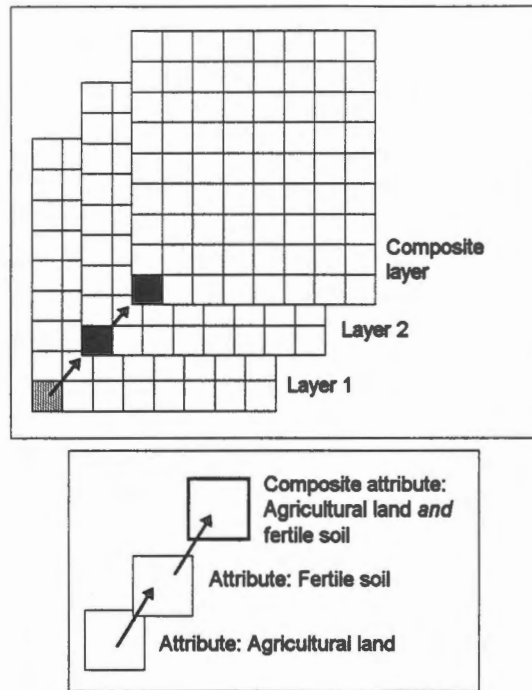


Overlay analysis using vector data involves quite a few calculations — cut arcs from the two layers to form new ones, form new polygons out of the new arcs, and so on. When many arcs and polygons are involved, such calculations are too slow for an interactive environment. In this chapter, we will introduce a simplified version of this that is more suitable for interactive analysis.

Raster Overlay

A faster way of performing an overlay is to use raster data. If both raster layers are of the same resolution and their pixels are aligned, then an overlay operation is merely matching corresponding pixels from both layers and combining their attributes to form a new pixel in the composite layer.

This involves no geometrical calculations and no creation of new polygons. Unless the layers are extremely large, overlay analysis involving raster layers can be done interactively.

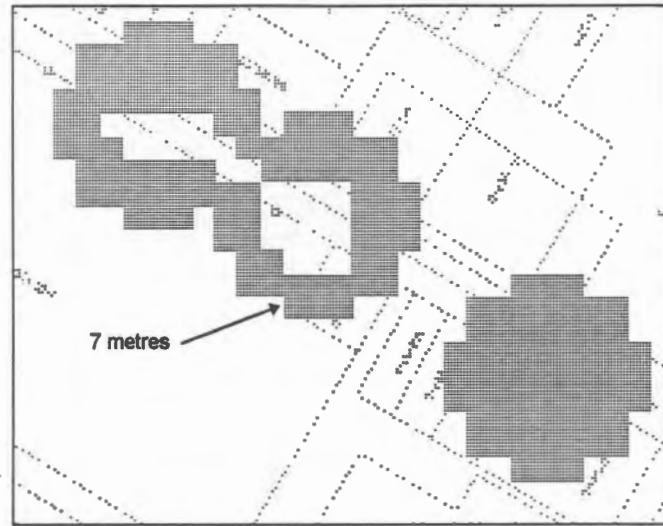


 The demonstration Concepts: Overlay analysis explains the concepts of vector and raster overlay analysis.

Zone Overlay

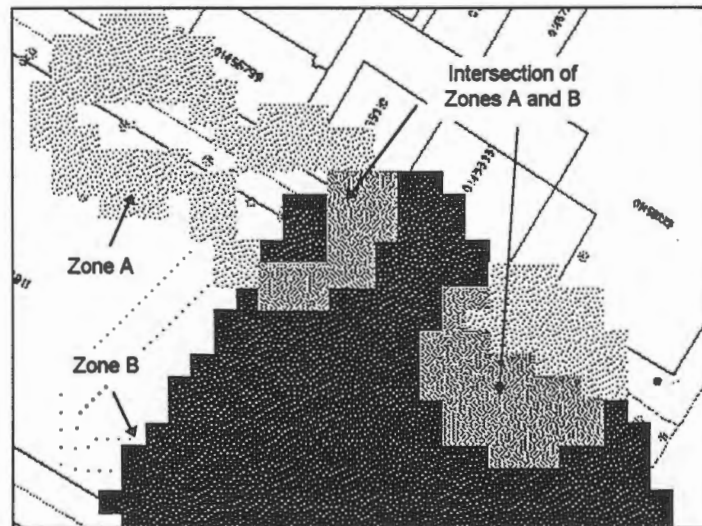
A *zone* is a collection of pixels having the same attributes. In a conventional raster representation, pixels exist independent of each other, and each pixel is associated with a list of attributes. For example, the attribute “agriculture land” is associated with each and every pixel that falls on this type of land use. In a zone representation, all pixels belonging to agricultural land will be grouped together in the file and given the proper attributes only once. This saves us some storage space. Moreover, zones delineate geographical objects such as land use regions.

Zones can be composed of disjoint components, and they can have holes. The zone in this example is composed of two pieces, and one of them has a hole. The resolution for this raster zone is 7.



(289000, 739963), 1:700

After we have created zones, we can perform overlay operations between them. The following example shows the intersection of two zones.



Zone and Vector Overlay

We can also perform a quick overlay analysis of a raster zone over features in vector format. In this kind of analysis, no new polygons will be created, and any polygon overlapping the zone will be retrieved in its entirety.



Chapter 25

Raster Conversion

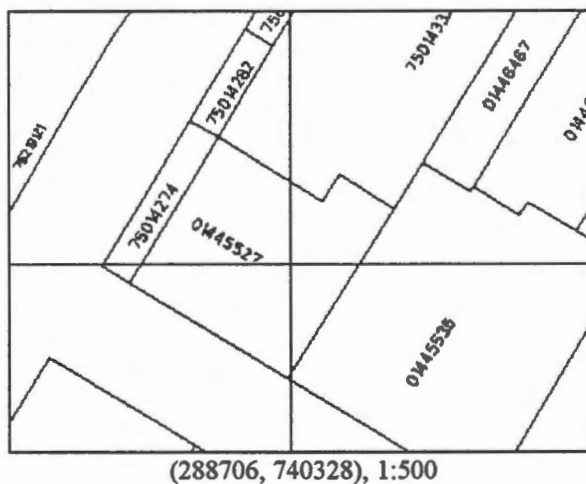
In theory, either the vector or the raster format can be used to represent features in a GIS. In practice, however, one format is more suitable for certain operations. For example, it is easier to compute the length of a vector line, but it is easier to determine whether two areas overlap when they are in raster form.

In order to make the best use of the two formats, we might want to convert data from vector format to raster. There is also a need for the reverse conversion from raster to vector, although this conversion is much more complicated.

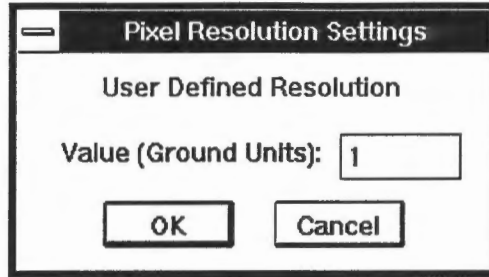
We will describe the conversion of vector polygons to raster zones. As explained in the chapter on Overlay Analysis, zones are composed of pixels with similar attributes and could be used in overlay analysis. They are different from raster images, which can only be used as backdrops.

 Make a zone out of a polygon.

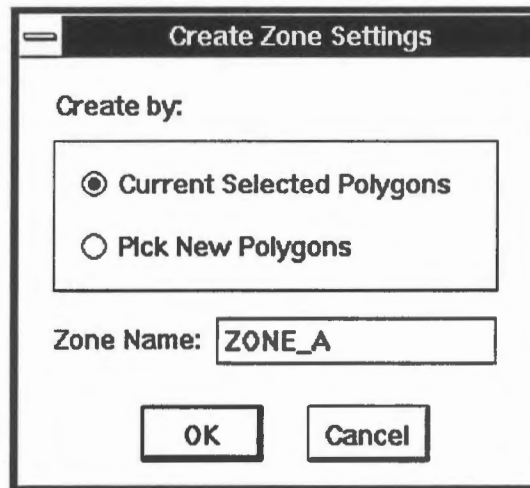
1. Use the MapQuery > By Pointing menu item to select and retrieve attributes for this parcel. It is a gas station owned by Imperial Oil Limited.



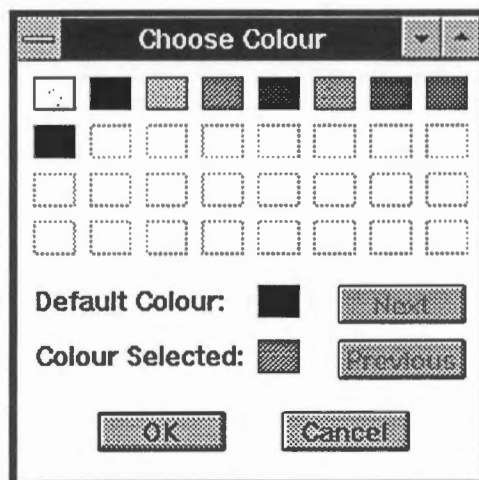
2. Pick the Utilities > Zone > Pixel Resolution ... menu item. A panel appears to allow you to enter the resolution of the raster zone to be generated. Enter a ground resolution of one metre, and click on OK.



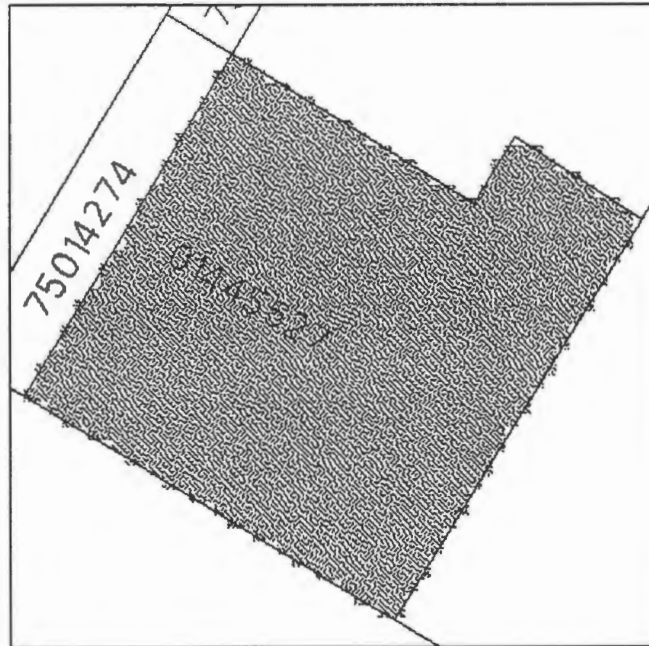
3. Pick the *Utilities > Zone > Create > By Polygons ...* menu item. A panel appears showing that a name, `ZONE_A`, has been generated automatically for the zone. You could assign a different name if you wish. Turn on the *Current Selected Polygons* switch since we have already formed a selection of polygons in a previous step.



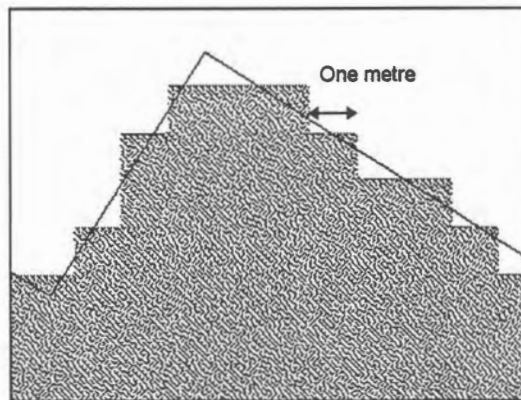
4. Choose a colour to draw the zone.



5. The raster version of the polygon will be shown on top of the vector version.



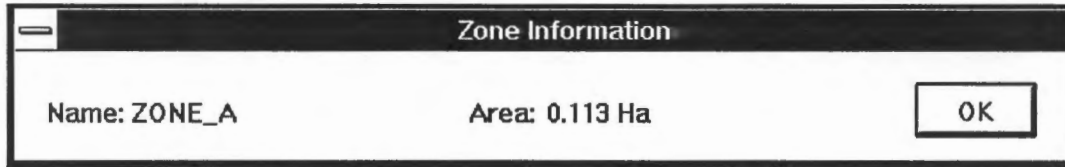
6. An enlarged part of it shows the pixels clearly. You can measure the size of them to verify that they are indeed one metre by one metre on the ground.



📏 Calculate the area of the zone.

1. Pick the Utilities > Zone > Information ... menu item, and point at the zone you have just created.
2. The area of the zone in hectares will be shown.

NOTE Your result might be slightly different



For a raster representation, the area is computed simply by counting the number of pixels and multiplying the result with the area of a pixel, which is one square metre in this case. ♦

❓ Effect of resolution on area.

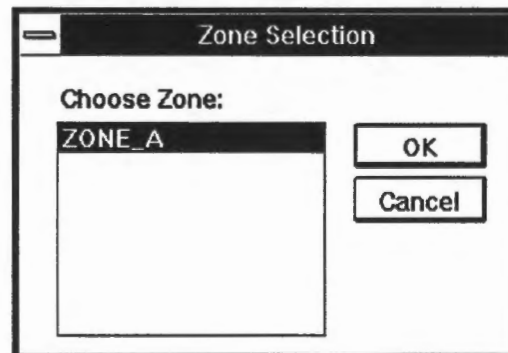
We have retrieved the pre-computed area of this polygon before, and it is 0.114 hectares as computed from its outline. The area computed from the vector representation is more precise, and this difference of 0.001 hectares (10 square metres) is due to the resolution used. Recall that the ground resolution for the vector representation is 0.05 metre, whereas that for the raster representation here is one metre, a difference of 20 times.

Generate zones for the same polygon but use different resolutions and compare the differences in area. You will notice that differences in area are not necessarily proportional to differences in resolution. Why? ♦

🖱️ Draw the zones.

If you refresh the screen, the raster zone for the polygon will disappear. However, zones are always remembered, and you can recall them for display later.

1. Pick the Draw > Zone ... menu item. A panel appears for you to select the zone you want. Choose the zone and click on OK.



2. Choose a colour for the zone.
3. The zone will be drawn on top of the line map. ♦

Select Features within a User-Defined Zone

In a previous chapter, we have used the MapQuery > By User Defined Area ... menu item to select polygons overlapping a user-defined area. You can perform the same operation by using raster zones. In this example, we will find the parcels overlapping a user-defined zone.

This method and the previous one produce identical results. However, the use of zones provides more flexibility — you can create zones with holes, and the zones can be stored away for future use.

In this example, we will use three steps to create a zone with a hole.

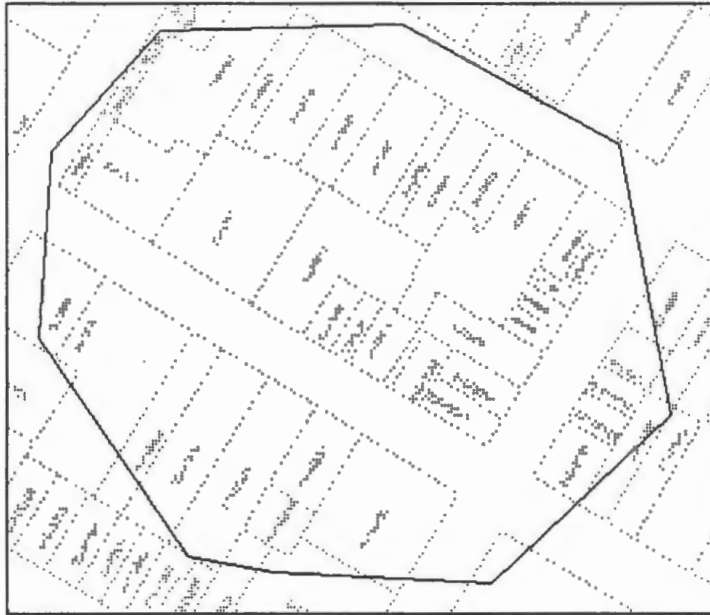
1. Create the outer zone.
2. Create the inner zone.
3. Subtract the inner zone from the outer zone to make a ring.



Create the outer zone.

1. Pick the Utilities > Zone > Pixel Resolution... menu item to define a ground resolution of one metre.
2. Pick the Utilities > Zone > Create > By User Defined Area... menu item.
3. It will be called ZONE_A automatically unless you have previously created a zone with this name.
4. Choose a colour for the zone.

5. Use the crosshair to enter the points for the outer zone.



(288788, 740274), 1:1000

6. A raster zone is created and drawn.



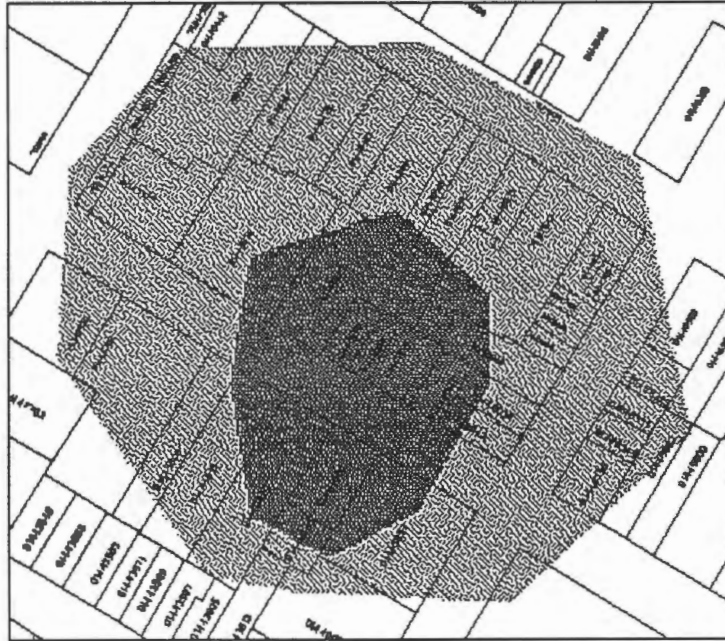
NOTE The zone will cover the parcels underneath. You would have to turn off the Options > Display > Erase switch and refresh the screen again to show them.




 Create the inner zone.

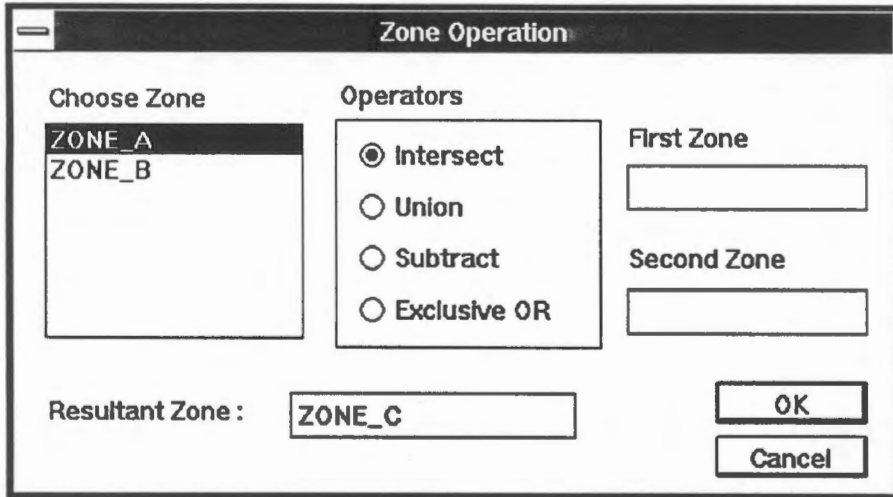
Follow the same procedure as in the previous exercise to create the inner zone. It will be called ZONE_B automatically.

NOTE When you are asked to define the outline of the inner zone, the polygon you have used previously will be highlighted. Enter N on the keyboard to reject it.

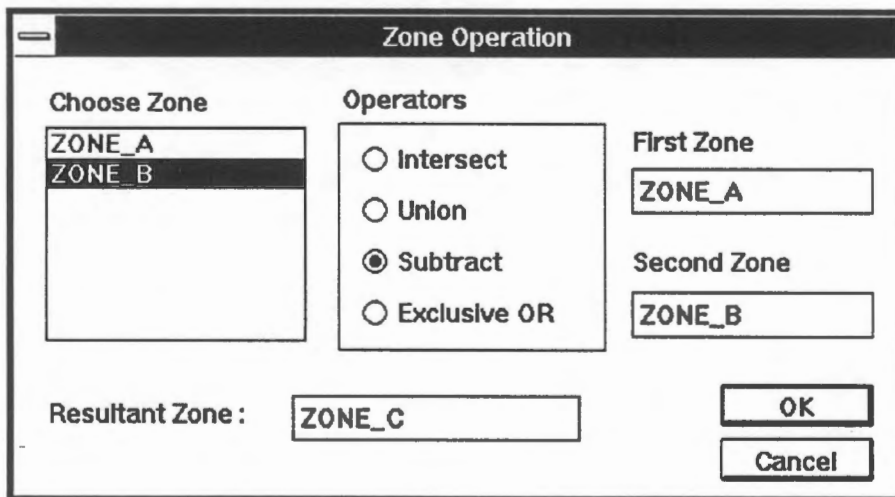


 Subtract the inner zone from the outer zone.

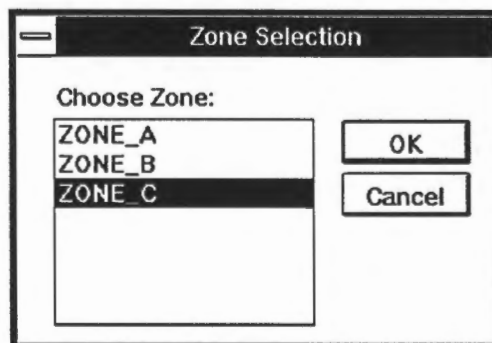
1. Refresh the screen.
2. Pick the Utilities > Zone > Operation ... menu item.
3. The Zone Operation panel shows that two zones, ZONE_A and ZONE_B, have already been created.



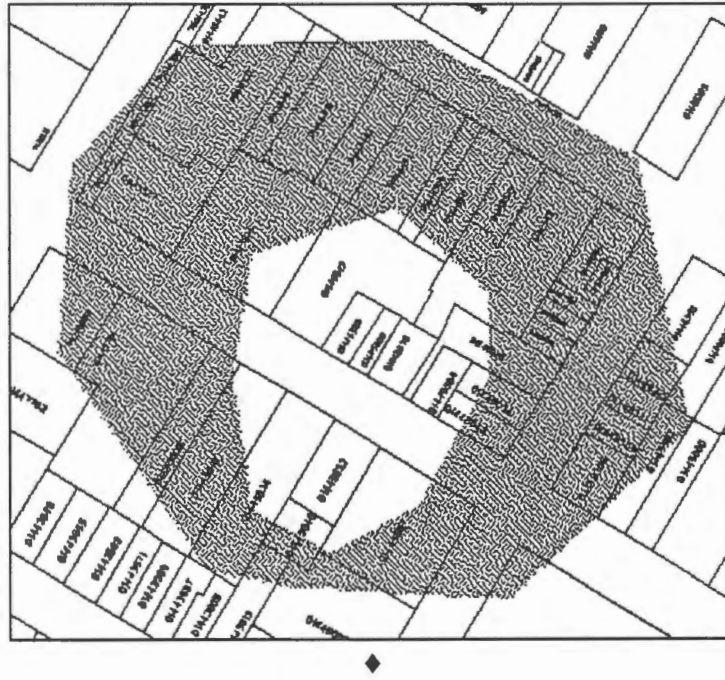
- Click on ZONE_A to select it as the first zone, and then click on ZONE_B to select it as the second zone. Turn on the Subtract operator. This will generate ZONE_C by subtracting ZONE_B from ZONE_A.



- Pick the Draw > Zone ... menu item. It shows that we have now three zones. Select ZONE_C and click on OK.

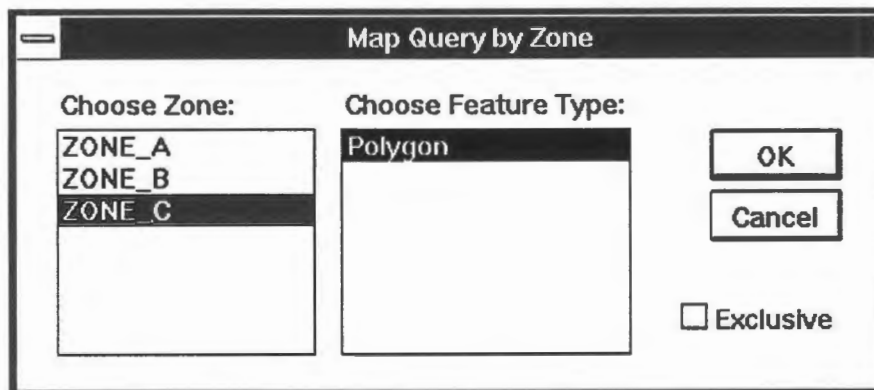


6. Choose a colour.
7. The ring is drawn on the screen.

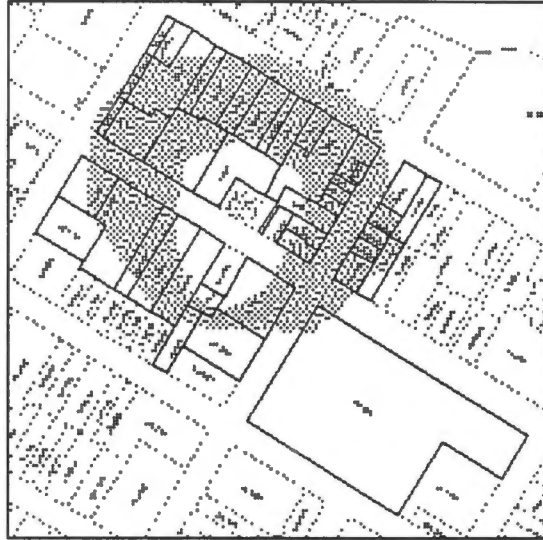


 Retrieve polygons overlapping the ring.

1. Pick the MapQuery > By Zone ... menu item.
2. Select ZONE_C. By leaving the Exclusive switch *off*, we will select polygons inside the zone.



3. The attribute records will be retrieved from the database and a selection formed. Highlight or colour fill the selection.



◆ How to create a zone that has two disjoint parts?

Use one of the zone operation options to create a zone with two disjoint parts as shown in the following diagram.

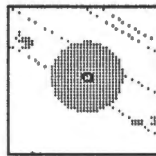


Retrieval Using Buffer Zones

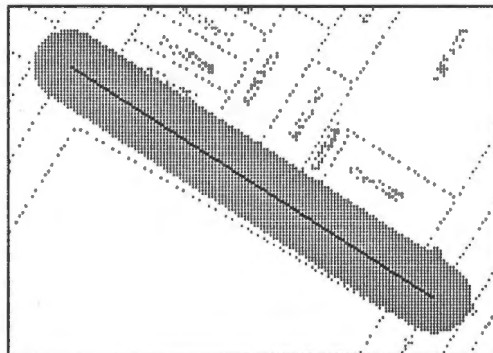
Many analyses make use of regions around features. For instance, the area illuminated by a street light is a circular region around a point. To analyse whether a street is well illuminated, we could generate circles around street lights in the database. The gaps between these circles are the under-illuminated regions.

You could draw circles around the points by hand, but GIS provide functions that could automatically generate circles around points given a radius.

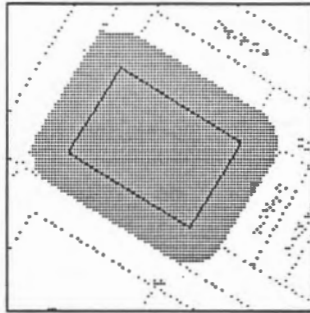
An area surrounding a feature is called a *buffer zone* or a *corridor*. Most of the GIS buffer zone generation commands will generate one that extends to a uniform width around the feature. The circle in the diagram is a buffer zone 10 metres from a point.



Buffer zones around lines form bands with rounded ends like the one in the diagram below. For scenic reasons in certain areas, logging of trees is prohibited along a band of a specific width on both sides of a highway. These bands can be represented by buffer zones around the highway.



A buffer zone around a polygon is an expanded polygon. We will use such a buffer zone in this chapter to illustrate an application involving gasoline leakage from an underground storage tank.



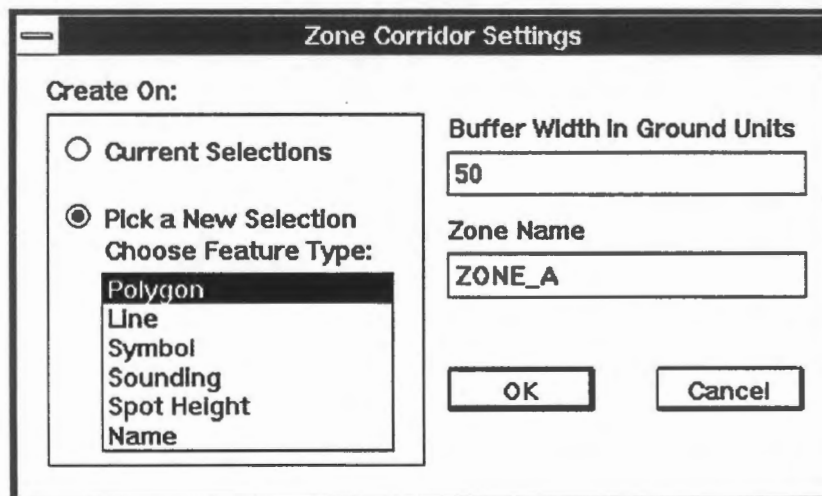
Buffer zones are important tools in modelling, a technique of simulating reality using mathematical models. In this example, we will use a very simple model for gasoline tank leakage, which assumes that the leak affects an area 500 metres around the leaking tank. A more complicated model could involve the underlying geology and other factors. We are concerned with the total value of properties affected if such a leak does occur.

In the database, we do not know the exact location of the gasoline tank, but we do know the location of the gas station containing it. So instead of generating a 50-metre buffer zones around a gasoline tank, we will generate a 50-metre buffer around the property.

The next step is to retrieve the properties overlapping the zone and calculate the total property value from the attributes returned. Because we do not want to include the property value of the gas station in this analysis, the zone without the gas station is a ring.

 Generate the outer buffer zone.

1. Pick a resolution of one metre on the ground.
2. Select the Utilities > Zone > Create > By Corridor ... menu item.
3. Turn on the Pick a New Selection switch, and enter a buffer width of 50 metres on the ground. This will create ZONE_A.



Zone Corridor Settings

Create On:

Current Selections

Pick a New Selection

Choose Feature Type:

- Polygon
- Line
- Symbol
- Sounding
- Spot Height
- Name

Buffer Width In Ground Units

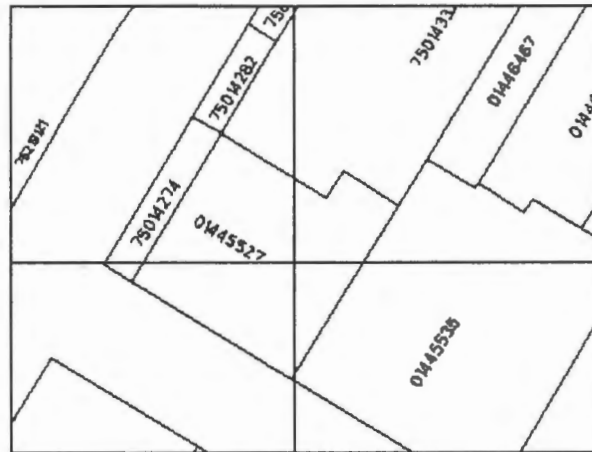
50

Zone Name

ZONE_A

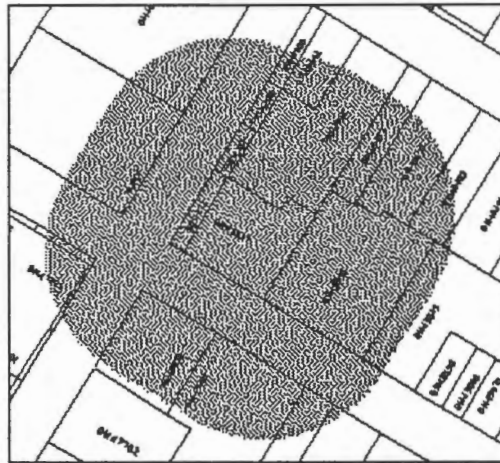
OK Cancel

4. Choose a colour for the zone.
5. Point at the gas station property.



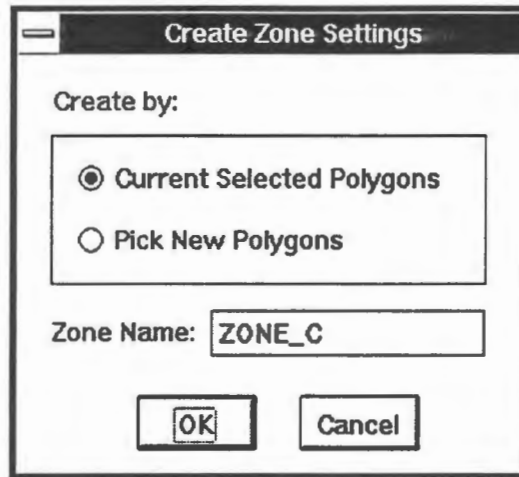
(288706, 740328), 1:500

6. A buffer will be generated around the gas station.

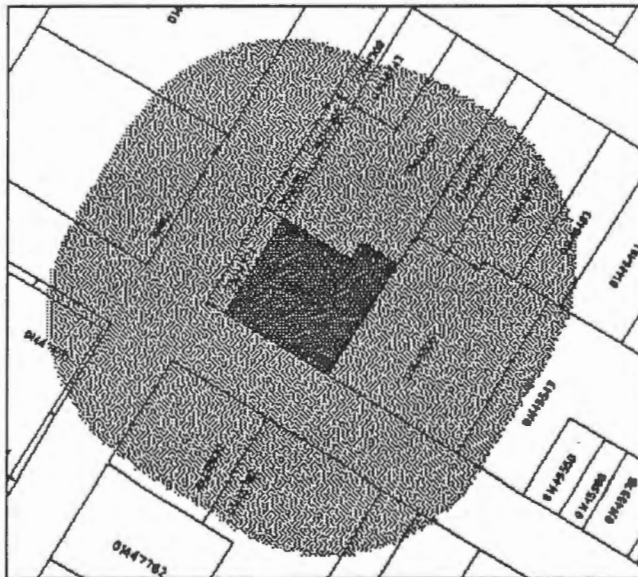


 Make a hole to exclude the gas station from the zone.

1. Make a zone called ZONE_B out of the gas station property. Pick the Utilities > Zone > Create > By Polygons ... menu item to create a zone out of the polygon selected in the previous exercise, which is the gas station in question.



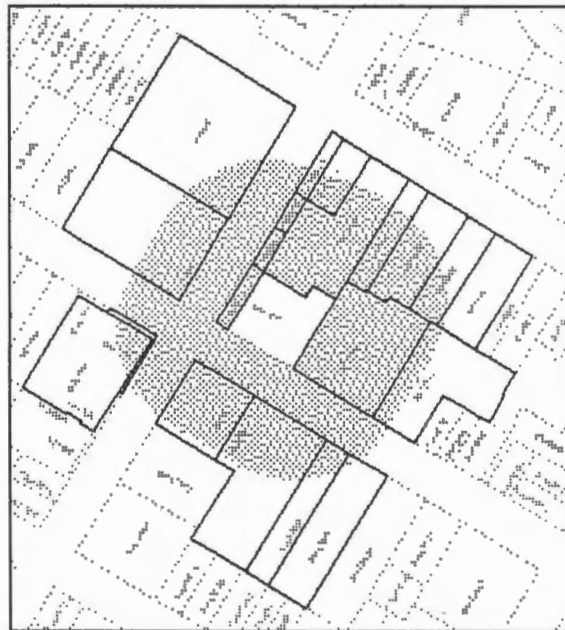
2. Choose a colour and the new zone will be drawn.



3. Select the Utilities > Zone > Operation ... menu item to subtract ZONE_B from ZONE_A creating ZONE_C.



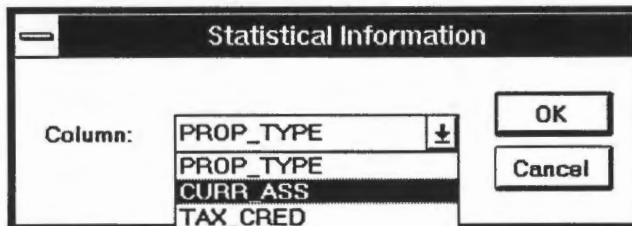
4. Use the MapQuery > By Zone ... menu item to retrieve polygons overlapping the zone.



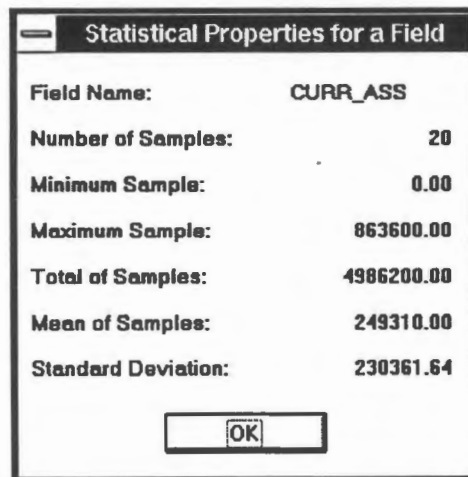
5. Another way to show the selected properties is to colour fill them. It shows that the gas station has been excluded.



- Use the Data > Statistics ... menu item on the DB Manager to generate statistics on current assessment value.
Scroll down the columns to select CURR_ASS and click on OK.

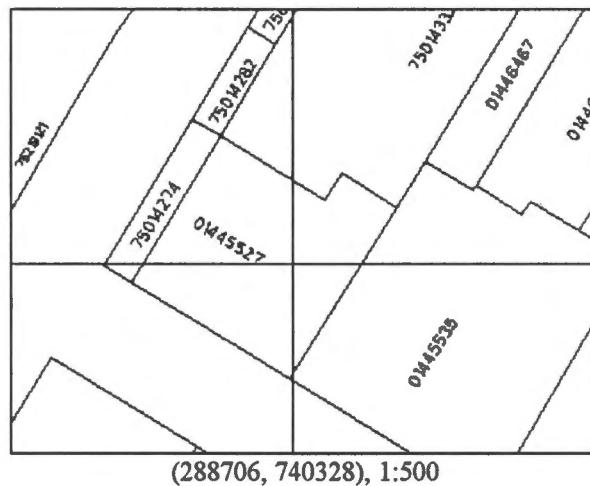


- The total value of properties affected is shown together with a number of other statistical calculations. Note that some of the parcels in the selection have no assessment value, hence a value of 0 for the minimum.



Reduce a Selection

After you have formed a selection, it can be reduced by specifying a further query condition. In the previous chapter, we selected all parcels that could be affected by leakage from the underground gasoline storage tank at the gas station shown below. We also computed the total and average values of these properties but found that some of them have no assessment value. In this example, we will remove from the previous selection those properties without assessment values to obtain more meaningful statistics.



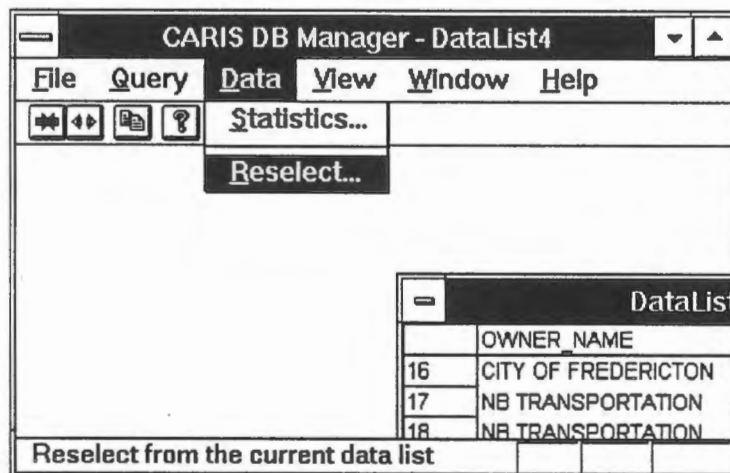
 Form a selection of parcels.

1. As explained in the previous chapter, create a buffer zone 50 metres from the gas station but excluding the gas station itself. Form a selection of parcels overlapping this zone.

NOTE If you have just finished the previous exercise, you would have formed the proper selection for this exercise.



2. Pick Data > Reselect ... from the DB Manager.



3. Select from the current selection those properties with current assessment values *not* equal to zero. The operator used is != for “not equal to”. A new selection will form and a report will be generated.

NOTE In the reselect command, retrieval is performed on the selected data list and *not* on the entire database.

Query Condition

ITY

Column:	Operator:	Value:
CURR_ASS <input type="button" value="↓"/>	!= <input type="button" value="↓"/>	0

4. Pick the Data > Statistics ... menu item to generate statistics for the selected parcels. You will notice that there are now only 15 parcels selected and the average value is higher than last time since we removed those with no assessment value.

Statistical Properties for a Field

Field Name:	CURR_ASS
Number of Samples:	15
Minimum Sample:	82500.00
Maximum Sample:	863600.00
Total of Samples:	4986200.00
Mean of Samples:	332413.33
Standard Deviation:	205963.48

5. Highlight the selected parcels.



Underground Gas Tank Leakage: Two Variations

In this chapter, we will use zone analysis to determine the effect of gas tank leakage. The first analysis is similar to the one we did before, except that we will involve *all* gas stations this time. This is equivalent to finding affected parcels if the gasoline tanks in all gas stations leak at the same time. In the second analysis, we will find the parcels affected by a leakage from *any* of them. We will also change the model and assume that a leakage will affect an area of 200 metres around the contamination source.

The first analysis involves the following steps.

1. Retrieve from the database all gas stations.
2. Generate buffer zones around them to simulate the area affected by a leakage.
3. Overlay the zones and the parcels to find the affected parcels.

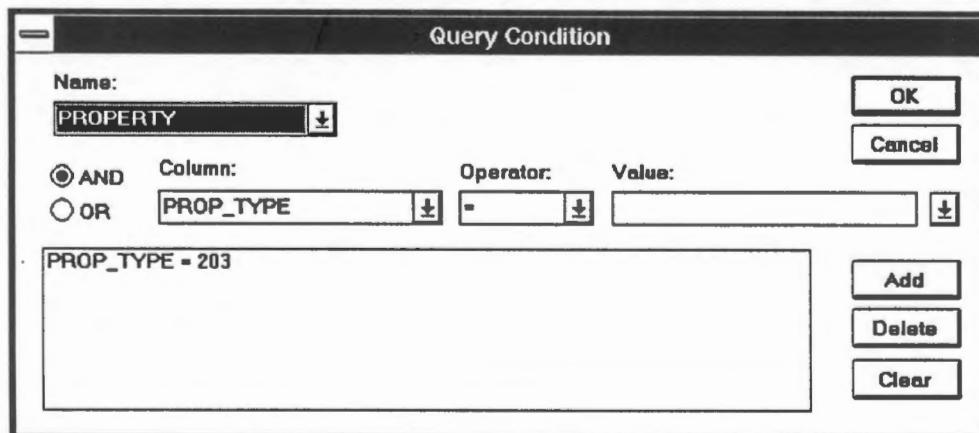
The second analysis involves the following steps.

1. Generate a buffer zone around each of the gas stations.
2. Find the area at the intersection of the buffers. This represents the area with the highest risk of being affected.
3. Overlay the intersection zone over the parcels.

The First Analysis

 Find parcels within 200 metres of any gas station.

1. Select the Query > Query Condition ... menu item from the DB Manager window.
2. Retrieve all properties with type equal to 203, which is gas stations.



Query Condition

Name: PROPERTY

AND OR

Column: PROP_TYPE Operator: = Value: 203

PROP_TYPE = 203

OK Cancel Add Delete Clear

3. There are three gas stations, and their locations are highlighted in the map below.

DataList1		
	OWNER_NAME	ST
1	IRVING OIL COMPANY LIMITED	18
2	CITY OF FREDERICTON	20
3	IMPERIAL OIL LIMITED	31



4. Select a pixel resolution of five metres.

Pixel Resolution Settings

User Defined Resolution

Value (Ground Units):

5. Select the Utilities > Zone > Create > By Corridor... menu item to create a 200-metre buffer zone, called ZONE_A, around the selected properties.

Zone Corridor Settings

Create On:

Current Selections

Pick a New Selection

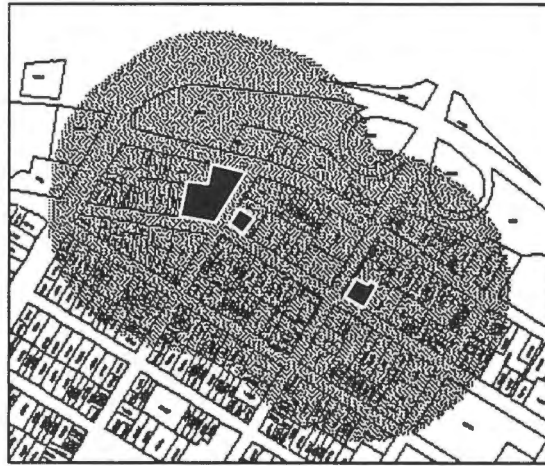
Choose Feature Type:

- Line
- Symbol
- Sounding
- Spot Height
- Name

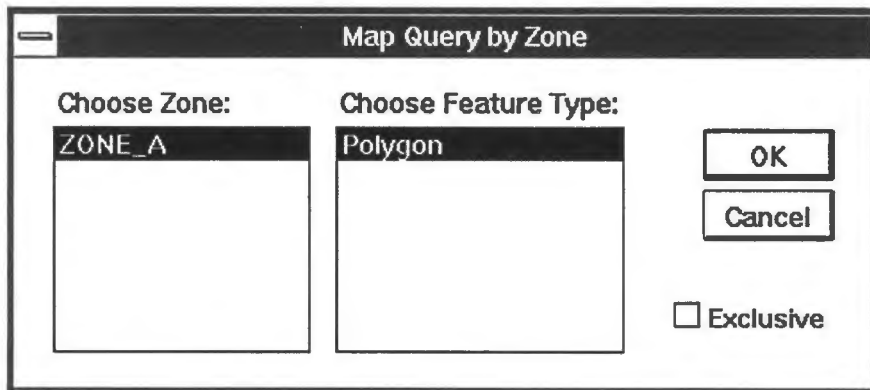
Buffer Width In Ground Units

Zone Name

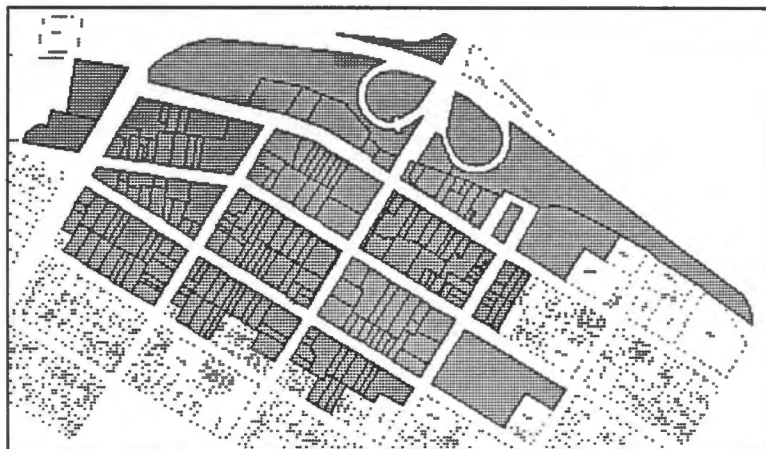
6. The following map shows the zone and the three gas stations.



7. Select the MapQuery > By Zone ... menu item to overlay ZONE_A upon the polygons.



8. The parcels within 200 metres of any gas station are shaded.





The demonstration Spatial analysis: Finding the lots affected by underground gas tank leakage shows how to perform a similar analysis.



What is the average assessment value of these properties.

Compute the average assessment value of these properties, excluding those with assessment values of zero dollars. ♦

The Second Analysis



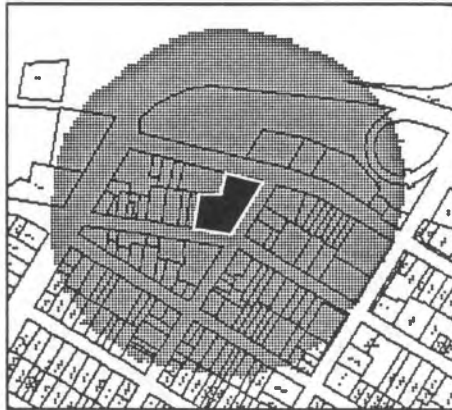
Retrieve the parcels affected by any leakage.

The previous exercise showed that there were three gas stations, and their parcel identification numbers are:

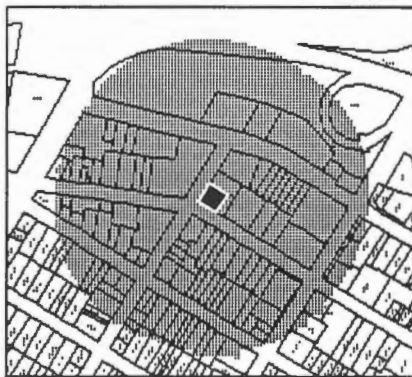
01435700, 01445436, 01445527

We will generate a 200-metre buffer zone around each of them.

1. Generate a buffer zone around parcel 01435700. Call it STATION A.



2. Generate a buffer zone around parcel 01445436. Call it STATION B.



3. Generate a buffer zone around parcel 01445527. Call it STATION C.

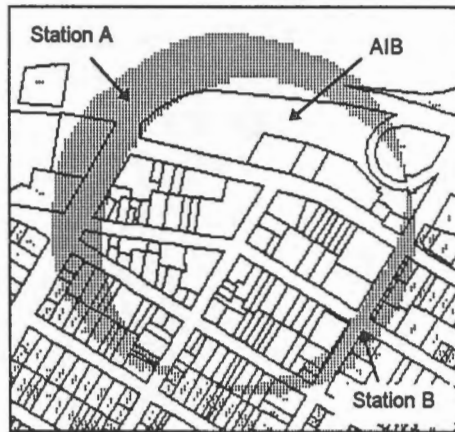


4. Select the Utilities > Zone > Operation ... menu item to generate a zone, called AIB, at the intersection of STATION A and STATION B.

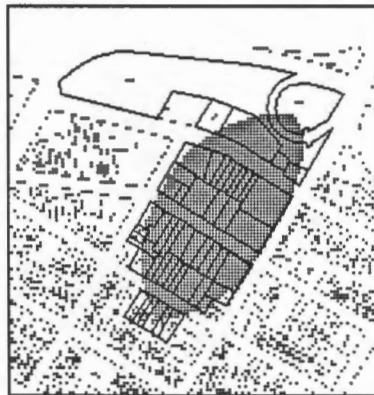
Zone Operation

Choose Zone ZONE_A ZONE_B STATION A STATION B STATION C	Operators <input checked="" type="radio"/> Intersect <input type="radio"/> Union <input type="radio"/> Subtract <input type="radio"/> Exclusive OR	First Zone STATION A Second Zone STATION B
Resultant Zone :	AIB	OK Cancel

5. The diagram below shows the intersection between STATION A and STATION B.



6. Select the Utilities > Zone > Operation ... menu item to generate a zone, called AIBIC, at the intersection of AIB and STATION C.
7. Select the MapQuery > By Zone ... menu item to overlay AIBIC upon the polygons. The diagram below shows the zone AIBIC and the polygons overlapping it.



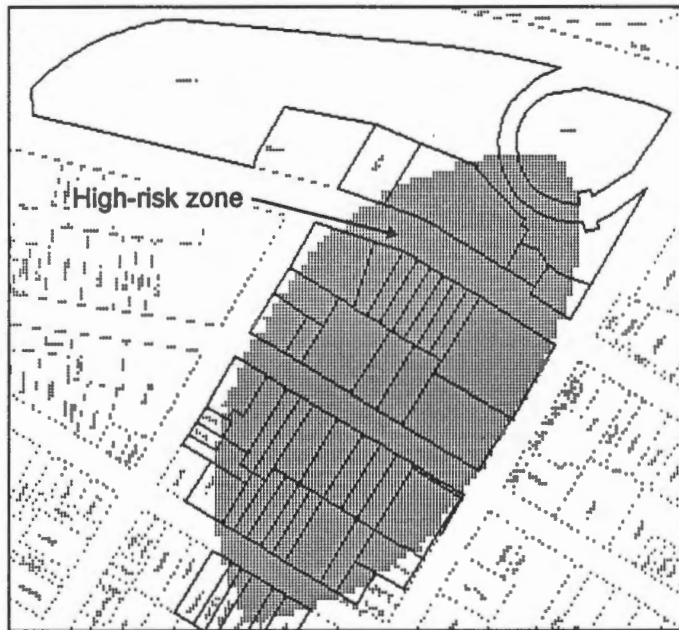
Chapter 30

Area Affected by Underground Gas Tank Leakage

The diagram below shows the result of the previous chapter. The zone is the high-risk area common to the 200-metre buffers around the three gas stations, and the highlighted parcels are those that could be affected by a leakage from any of the gas stations.

In this chapter, we will compute the area of the affected parcels, and this is done in two ways. In the first analysis, the area is computed from the parcels in vector format, which is the area sum of the parcels highlighted here. This is because the zone-over-vector polygon overlay analysis does not cut up the polygons into parts inside and outside of the zone.

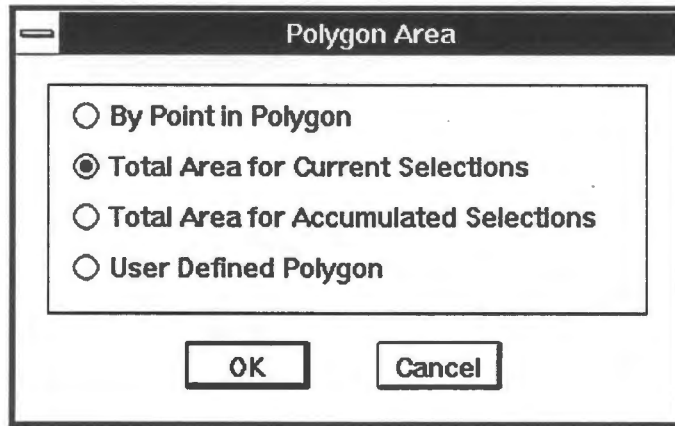
In the second analysis, a more rigorous method is used to exclude those parts of the parcels not in the high-risk zone. This is done through a zone-over-zone analysis.



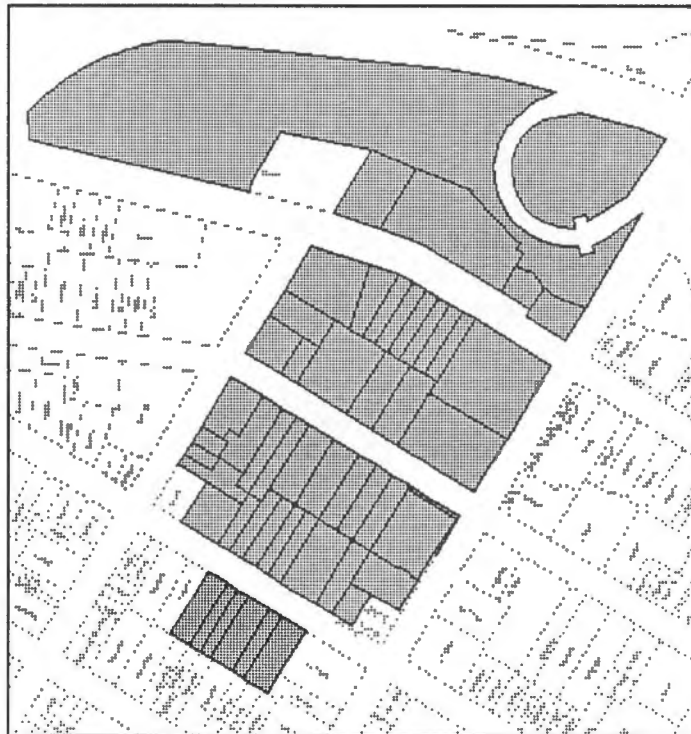
The First Analysis

☞ Compute the area of the affected parcels in vector form.

1. Select the MapQuery > Polygon Area ... menu item to compute the total area for the current selection.



2. The display shows that the total area is 7.806 hectares. It is computed from the shaded area in the following diagram.

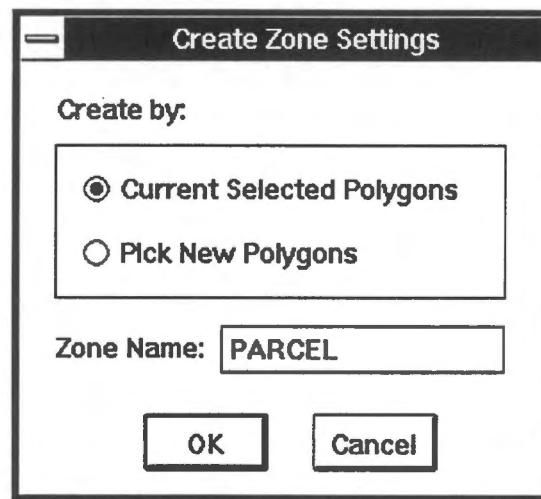


The Second Analysis

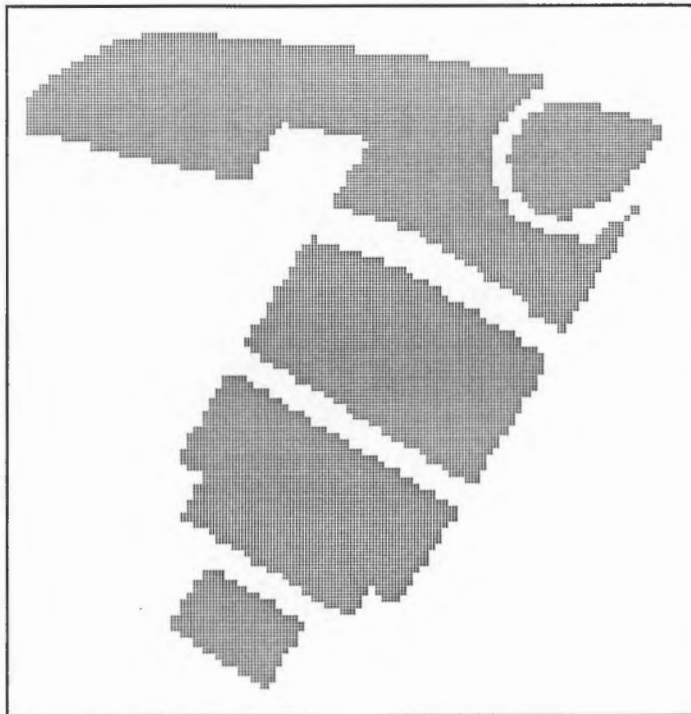
If we want to be more exact and wish to exclude those part of the parcels not within the high-risk zone from the area calculation, we would have to use a zone-over-zone overlay. This can be done by first converting these parcels into a zone and creating an intersection between this and the high risk zone.

 Compute the area of the parcels within the high risk zone.

1. Select the Utilities > Zone > Create > By Polygons ... menu item to convert the current selection into raster zones. Call this zone PARCEL.



2. Use the Draw > Zone ... menu item to draw the zone.



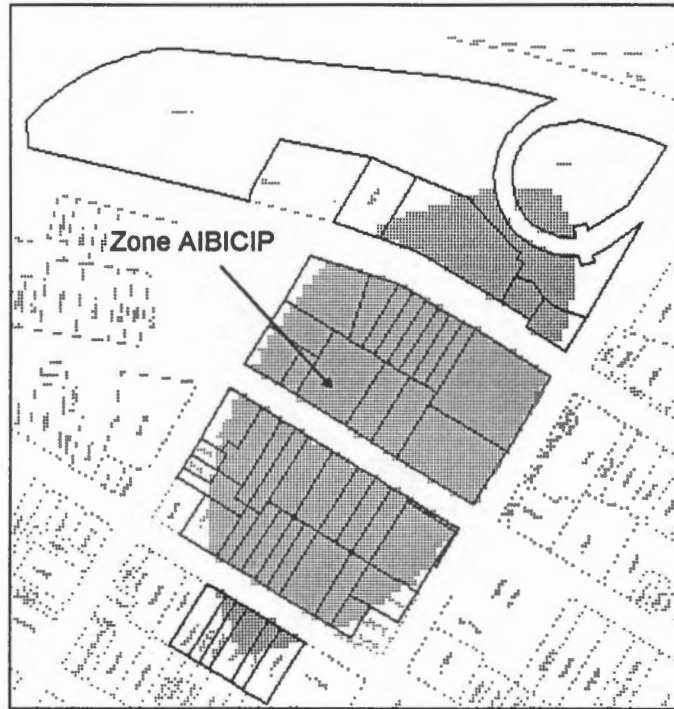
3. Select the Utilities > Zone > Information ... menu item to compute the area of the PARCEL zone.

Point at the zone from which you need to obtain information.

NOTE We have created several zones over the same area. The first one selected might not be PARCEL. Keep clicking on OK until you have reached the proper zone.

Notice that the area is smaller than 7.806 hectares as computed previously using the vector polygons. This is due to the coarse resolution of five metres used for this zone.

4. Use the Utilities > Zones > Operation ... menu item to find the intersection between the high-risk zone AIBIC and PARCEL to produce zone AIBICIP. The relation between the resulting zone and the parcels is shown below.



5. Use the *Utilities > Zone > Information ...* menu item to compute the area of this zone. You will find that the area of the parcels within the high risk zone is 4.088 hectares. ♦

◆ ? What is the difference between zones AIBICIP and AIBIC ? ◆

Overlay Analysis Involving Points

We will show in this chapter how to perform an overlay analysis of a zone over point features. The objective is to find out how many trees are so close to the utility poles that the branches could be in the way of the poles. A simple model will be used, which is to assume that branches extend four metres from the trunk. We will not consider here the height differences between the branches and the poles. The analysis will be performed using three steps.

1. Select all the trees in the database.
2. Build a buffer zone of four metres around each tree.
3. Retrieve the utility poles within this buffer zone. They are the poles too close to the trees.

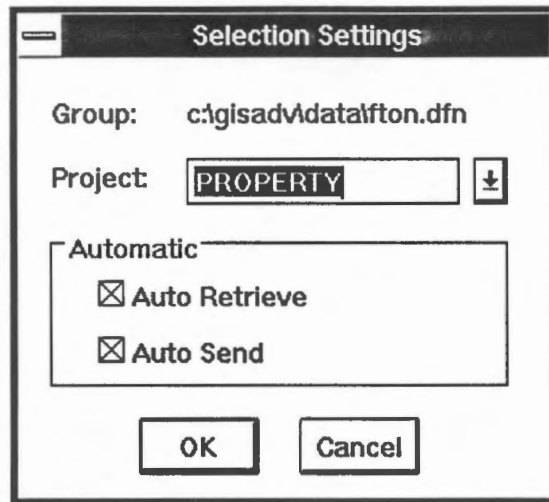
This analysis is different from the previous one is that no user-defined attributes will be used. That is to say, all the analysis will be performed using only system-defined attributes. We could have created databases for the trees and utility poles as well.

 Disconnect from the DB Manager.

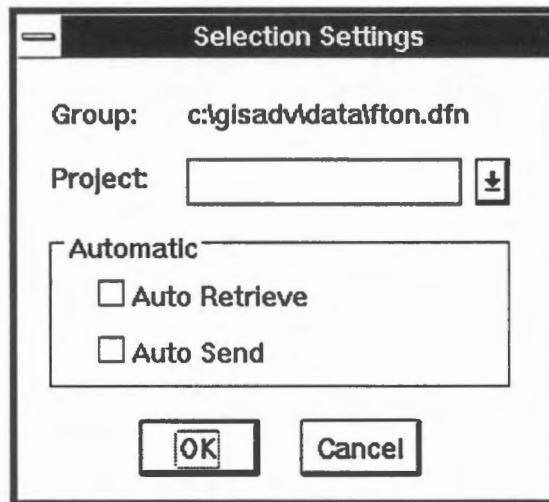
In this exercise, we will be working with tree features without any user-defined attributes. To prevent the GIS from searching the user-defined database, the DB Manager will be disconnected from the Information Manager.

1. Select the Options > Display > Projects ... menu item.
2. A Selection Settings panel will appear showing that the Information Manager is currently connected to the PROPERTY database tables.

Use the cursor to select "PROPERTY" in the Project field and press the Delete key on the keyboard to delete it. Also turn off Auto Retrieve and Auto Send to disconnect the two databases.



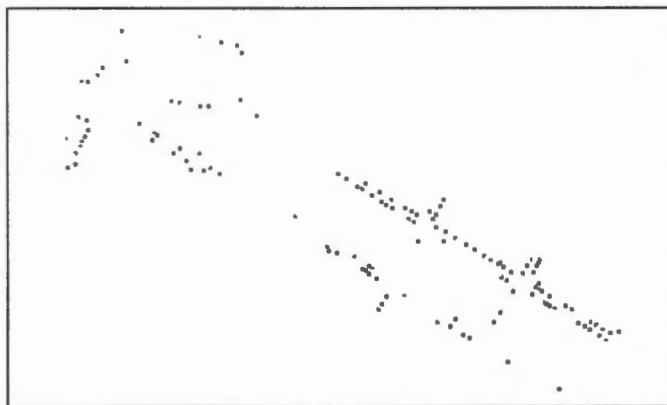
2. Click on the OK button.



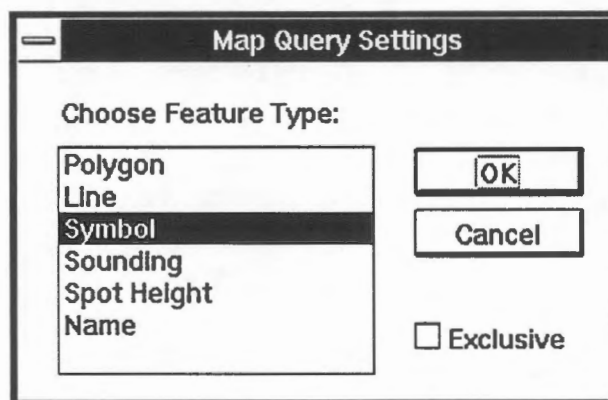
NOTE To connect with the PROPERTY database again, simply return the Selection Settings panel to its previous state.

 Generate buffer zones around all trees.

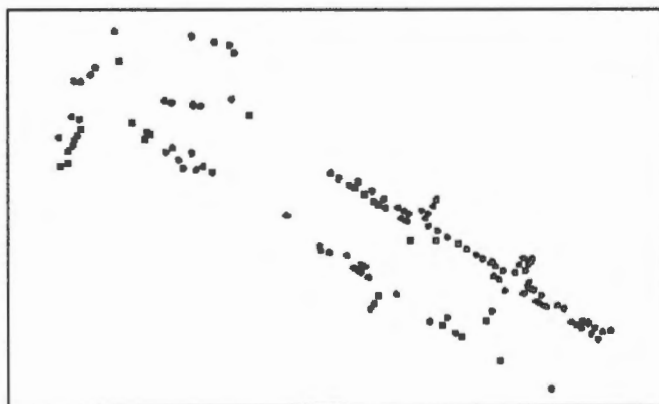
1. Make sure that the DB Manager has been disconnected.
2. Use the Options > Display > Visibility Parameters ... menu item to show trees (feature code is TREE).



3. Select the MapQuery > By User Defined Area ... menu item.
4. A Map Query Settings panel appears. Select Symbol to indicate that you will be selecting symbol features within the area.

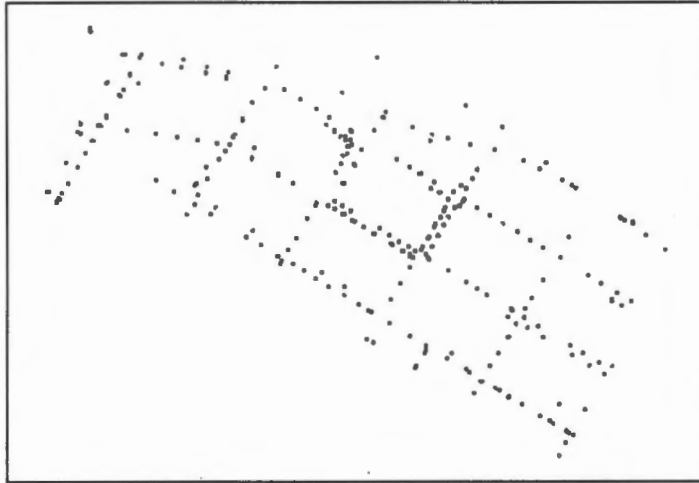


5. Draw a polygon around the trees to select all of them.
6. You might want to highlight the selection to make sure that all trees have been selected.
7. Generate a four-metre buffer zone around the current selection. Call the zone TREES. Use a resolution of one metre.

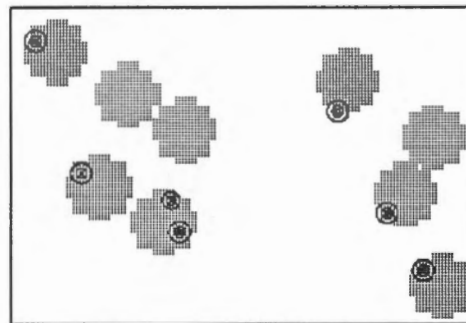


🔍 Retrieve utility poles within the zone.

1. Display only utility poles (feature code is UTILITY_POLE).



2. Select the MapQuery > By Zone ... menu item to overlay the zone TREES over the utility poles, which are point symbols.
3. A message at the bottom of the window indicates that there are 30 utility poles within four metres of any tree.
4. This is a display of some of the utility poles that are too close to trees.



(288824, 740226), 1:300



❓ How many trees are too close to the utility poles ?

We have found out that there are 30 utility poles too close to the trees. How many trees are too close to the poles ? Will there be 30 of them as well ? ◆

 Overlay of zone over points.


Out of overlays of zone over zone, over line, and over point, overlay of zone over point is the simplest. The zone is stored in raster format, and the pixels belonging to the zone are known. To find out if a point feature is overlapping a zone, we need only to check if a pixel in the zone already exists at that location. If the zone is a vector polygon, many more computations would have to be performed to determine whether a point is inside it or not. For an explanation, please consult the technical note in the chapter on Use Topology to Identify Polygons by Pointing. ♦

Generate Buffer Zones Around Lines

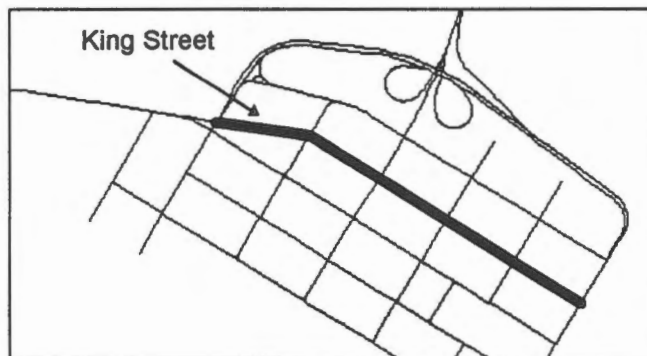
Here we will use an example related to addresses to show the generation of buffer zones around lines. The problem is to find those properties along KING STREET but with addresses other than KING STREET. This example also demonstrates that a line representation of a feature is more convenient than an area representation in this kind of application.

The procedure is as follows:

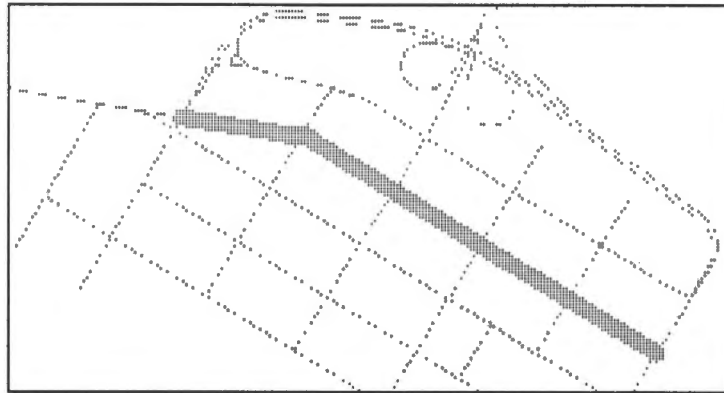
1. Select the centre line of KING STREET.
2. Generate a buffer zone of 15 metres around it. This will overlap any parcel on either side of the street.
3. Overlay this zone upon the parcels.
4. From the records retrieved, select those with street names other than KING STREET.

 Generate a buffer zone around the centre line of KING STREET.

1. Disconnect from the DB Manager and display only street centre lines (theme is 70).
2. Pick the MapQuery > By Pointing ... menu item to select the centre line of KING STREET.

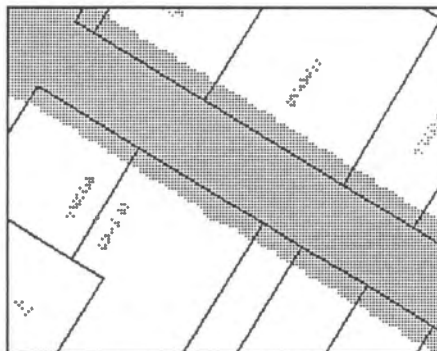


3. Generate a 15-metre buffer zone around KING STREET.

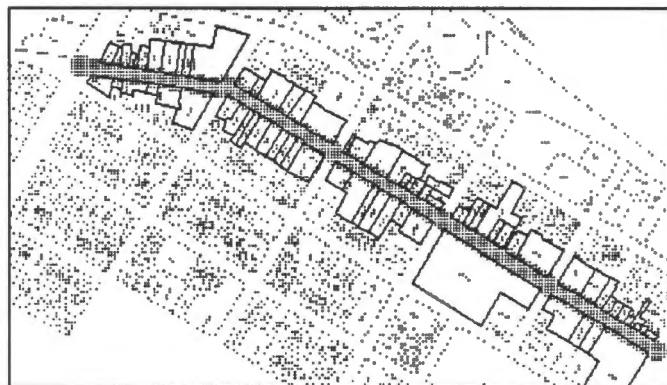


 Retrieve parcels along KING STREET.

The 15-metre buffer zone will overlap the parcels on both sides of KING STREET. Overlaying this upon the parcels will retrieve all those along KING STREET. Before you perform this overlay, connect the PROPERTY database again using the Options > Display > Project ... menu item and change the visibility parameters to display the land parcels (theme 50).



Use the MapQuery > By Zone ... menu item to select the overlapping parcels.



◇ Which of these parcels have street addresses other than KING STREET?

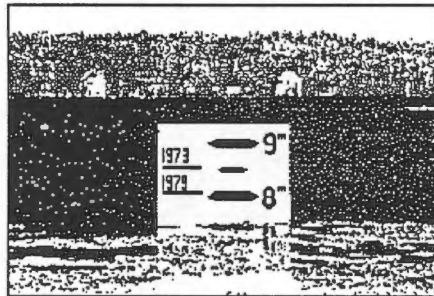
Use the DB Manager to select those parcels along KING STREET but with street addresses other than KING STREET. These parcels are shown below.



Manipulate Two Topological Layers

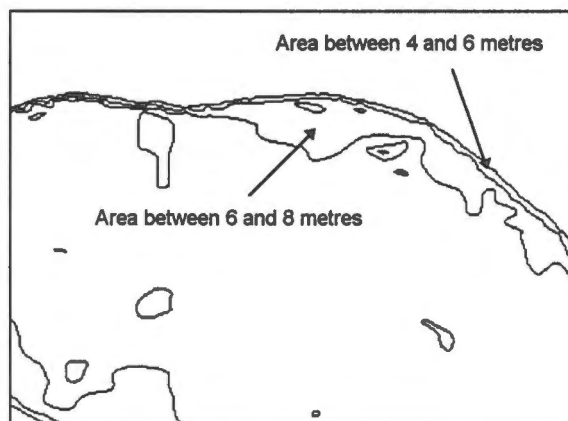
In this database, we have created two topological layers stored under the same file. We have used the layer (theme 50) on land parcels for most of our analysis, and the other layer (theme 90) on contours.

Contour lines form closed polygons which could be nested within each other. By building a topological layer for the contours, we can delineate areas within a certain elevation range. We will use the two layers to find the properties that could be affected if the water from the river rises to 8 metres. In 1973, a record-high flood in Fredericton reached a level of 8.6 metres while in 1979, the second-highest flood reached 8 metres. The following picture showing the water level markers was taken during the 1994 flood.



 Generate a buffer zone of areas at or lower than 8 metres.

1. Disconnect the DB Manager since we do not have user-defined attributes for contours.
2. Display theme 90 for contours.
3. Select these two polygons covering the area at or under 8 metres.

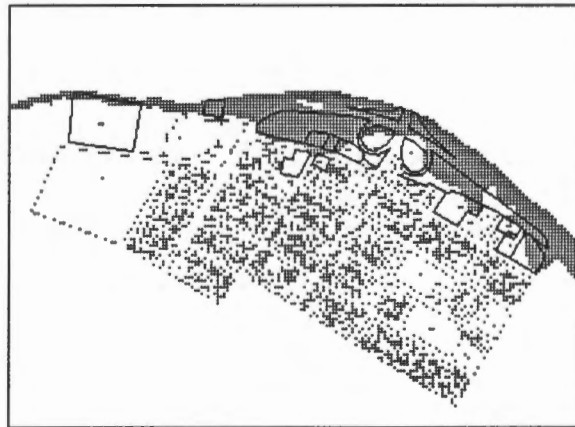


4. Select the *Utilities > Zone > Create > By Polygons ...* menu item to create a zone from the selected polygons. Use a resolution of one metre and call the zone FLOOD.



 Overlay with the properties.

1. Connect with the PROPERTY database.
2. Display only land parcels (theme 50).
3. Select the *MapQuery > By Zone ...* menu item.
4. The following is the properties affected by the flood.



 Compute the total area of the properties under water during this flood.

In this computation, exclude those parts of the properties not under water. 

More for You to Do

There are many more applications you can create from the municipality database. Here are two more for you try.

◆ How many trees are there on KING STREET ? ◆

◆ Is there a house for sale for a young family ?

Find a house for sale (see the entry of HOUSE FOR SALE in the property description) within 150 metres of schools and parks.



Run demonstration Spatial analysis: Finding a house for sale which is close to schools and parks if you want the answer. ◆

Chapter 35

Network Analysis

Network analysis is based on a topologically structured network. We will describe two types in this chapter: shortest path analysis, and network zone analysis.

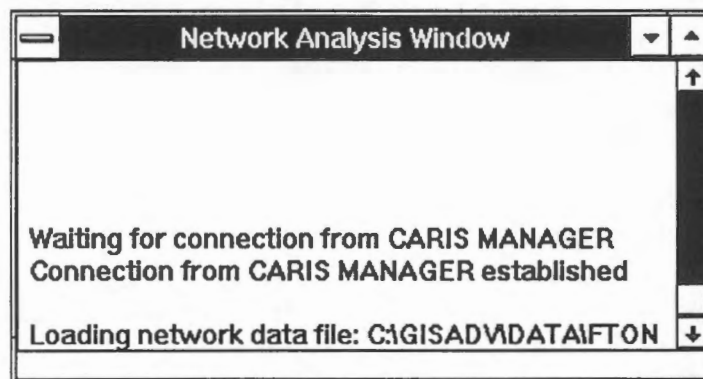
Shortest path analysis is used to find a route of minimum cost from one point to another along the network. If the cost is distance, then a path of shortest distance will be calculated. This kind of analysis is particularly useful for dispatching emergency vehicles such as fire engines. Other than distance, you can assign other measurements such as length of travel time as a cost.

A network zone extends from one point to reach branches of the network within a specified cost. For instance, we can identify all the streets that could be reached within 10 minutes driving time from a police station.

The network analysis software is a separate module that has to be loaded before you can use it.

☞ Load the network module.

1. Select the Utilities > Network > Load Network Module menu item.
2. A Network Analysis Window appears.



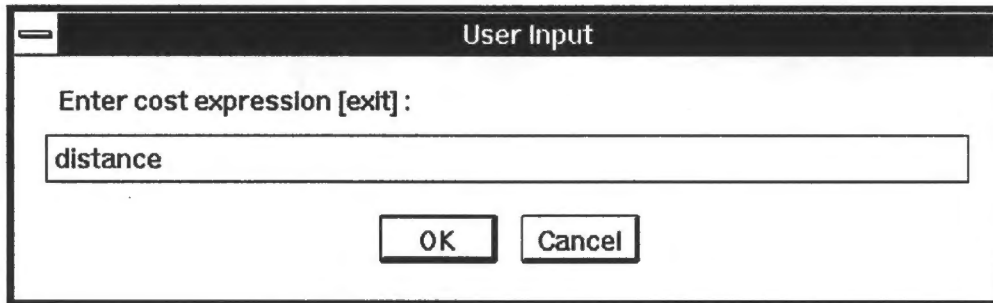
3. Close the DB Manager window to save memory. ♦

☞ Display street network.

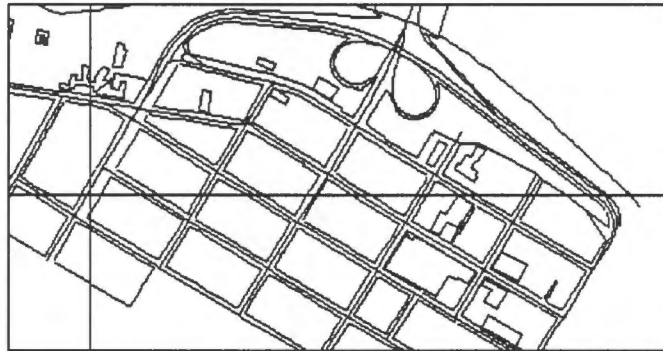
Display only themes 60 and 70 for the analysis. ♦

 Find the shortest path between two points.

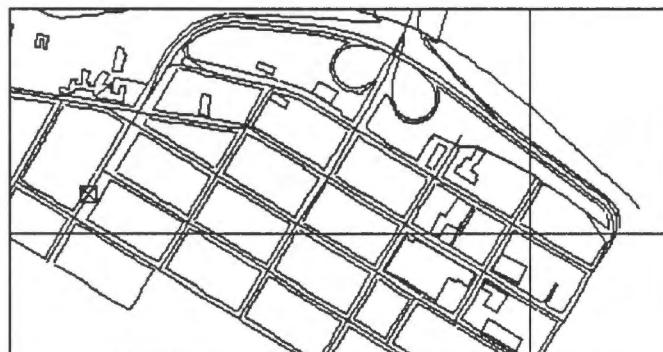
1. Select the Utilities > Network > Cost Model > Define menu item to define the cost model.
2. Enter distance as the cost model.



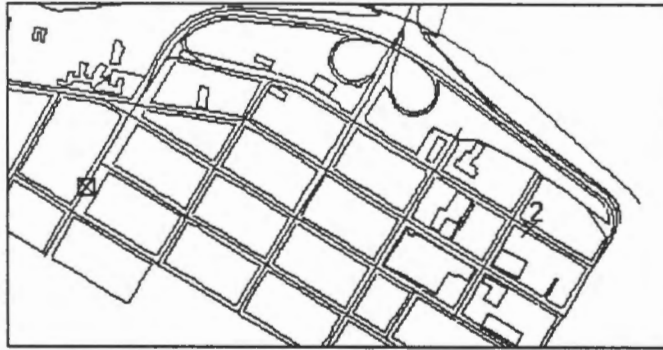
3. Select the Utilities > Network > Route > Calculate Shortest menu item.
4. Point to the origin.



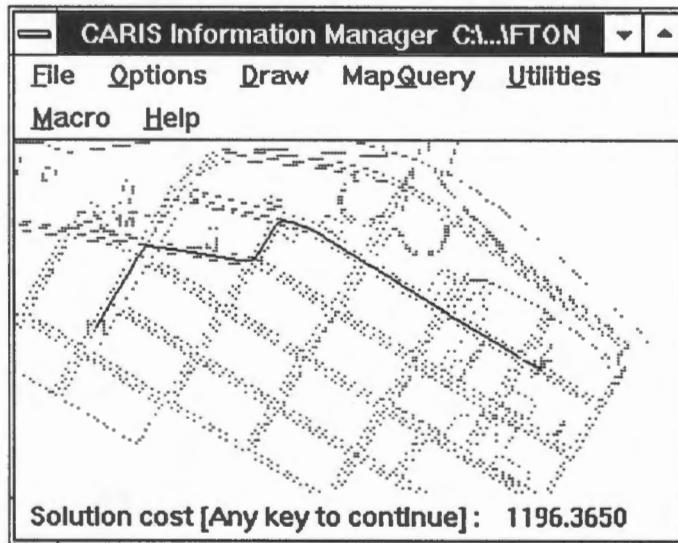
5. Point to the destination.



6. The second point will be marked with a 2.




7. The crosshair will appear again, enter G to proceed with the calculation of the shortest path between the origin and the destination.
8. The shortest path will be highlighted and its cost (distance in metres in this case) will be shown.

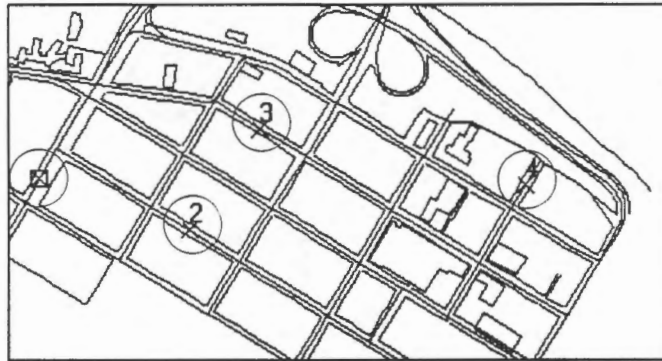


NOTE You might not obtain the exact same cost because of the way you select the origin and destination.

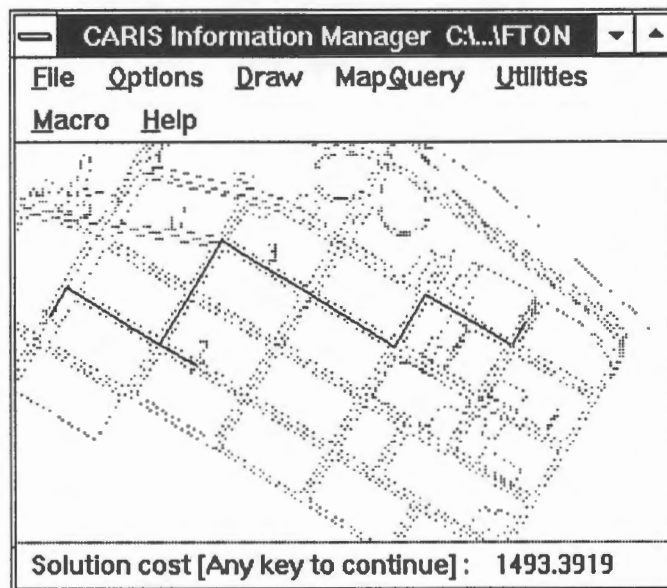


 Find the shortest path between two points while passing through other points.

1. Enter four points as show below. They will be marked sequentially.

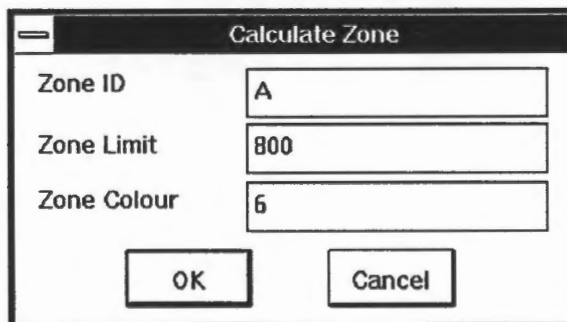


2. The path will be highlighted with its cost (in metres) shown.

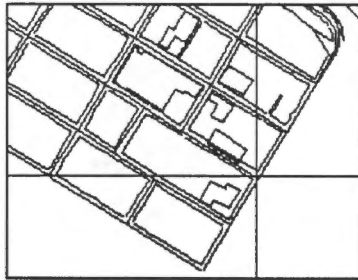


 Find the streets within 800 metres of a point.

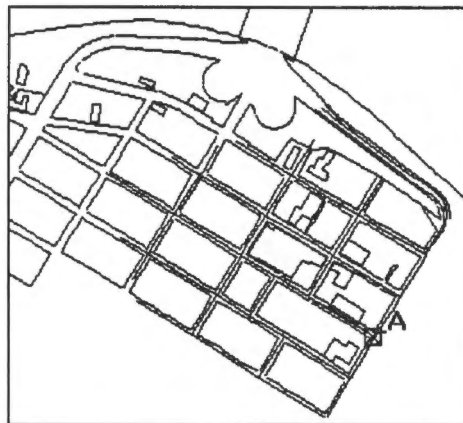
1. Select the Utilities > Network > Network Zone > Calculate menu item.
2. Enter A as the zone ID, 800 metres as the limit, and colour 6 to draw the zone.



3. Enter the point from which the zone will be computed.

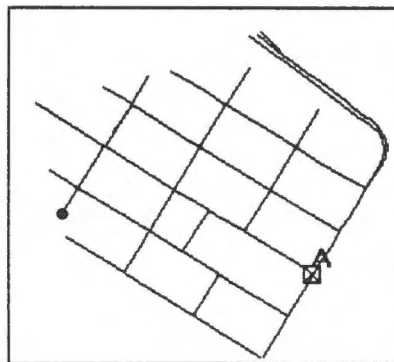


4. The zone will be computed and drawn.



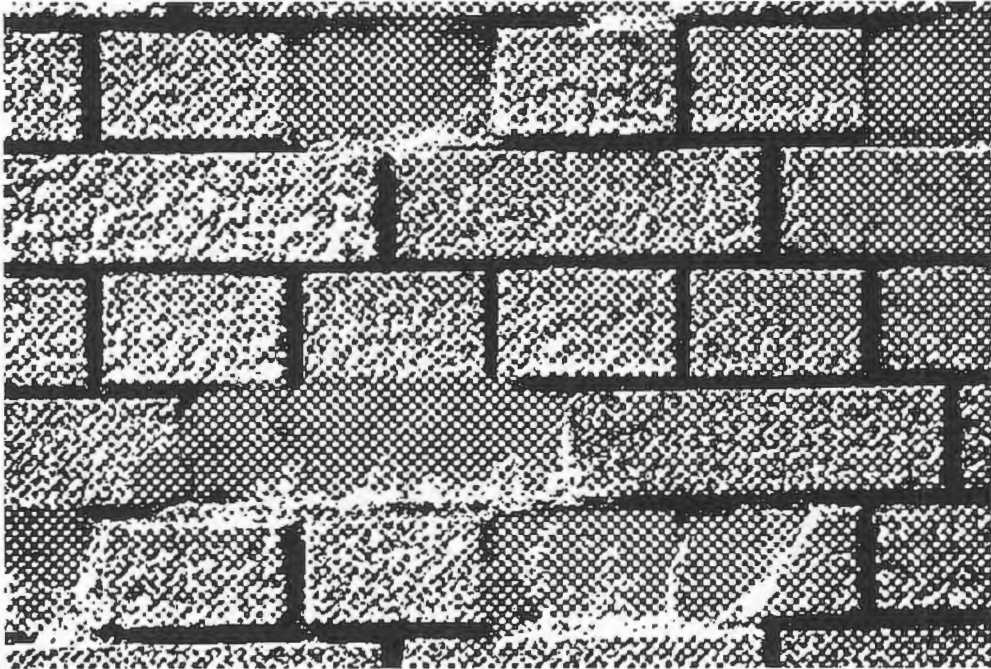
◇ Verify zone calculation.

The shortest route between A and any one of the end points, such as the one marked, in the network zone should be 800 metres long. Verify this.



Part 7

Data Collection and Editing



Chapter 36

Data and Its Collection

The Cost of Data

Data is one of the most important components of a GIS. It is also one of the most expensive, in many cases accounting for over 70% of the total system cost.

Data is costly because of the following reasons:

Geographic databases are voluminous: Geographic databases often cover very large areas, contain many features, and have many attributes associated with each feature. The relatively sparse municipality database that we have been using contains about 1.6 megabytes of vector data covering an area of roughly one square kilometre. The aerial photograph in raster format, covering an area of 4 square kilometres, takes up about 2.7 megabytes of storage. The entire database therefore is about 4.3 megabytes large.

 Determine the size of the database.

To compute the size of the database, use the File Manager in Windows to add up the size of all files FTON.* and the file PROPERTY.DBF.

NOTE Be careful not to delete these files.



Accurate data requires even more storage because of the large number of significant digits required for each coordinate. For raster data, the number of pixels grows significantly with only a slight increase in resolution. For example, a four square kilometre region needs one million pixels at a resolution of four square metres. If the resolution is increased to one square metre, the number of pixels rises to four million.

We need powerful hardware and software to handle large data volumes efficiently, and we also need reliable measures to protect them. All of these add to the system cost.

Data collection and conversion are expensive. Most data collection methods are labour intensive and error-prone, requiring considerable editing afterwards to clean the data. It is sometimes possible to convert existing data from another system. If the other system uses very different data organization, the conversion process can be complicated and in some cases impossible.

Data maintenance is a never-ending task. Data has to be kept up-to-date because geographic features are dynamic and will change over time. Unfortunately, features do not change at the same rate. Man-made ones, such as buildings and roads, tend to change faster than natural ones, such as rivers and mountains. This makes the detection of changes a challenge.

Our knowledge about the real world will change even though the features themselves do not change. For example, a rock formation does not change but our knowledge about its composition could change after more careful investigation.

User requirements for data will evolve and data collected using a different specification would have to be upgraded to meet new and often more stringent user demands. For instance, the accuracy requirements could increase as more large-scale applications are found for the data.

It is simply impossible to completely remove errors in a large database. On one hand, it is difficult to find all the errors. On the other hand, data cleaning, being a manual process, is a source of error itself.

Methods of Surveying the Earth

The most direct method of collecting information about geographic features is to conduct measurements on the ground. The two common methods are ground surveying and GPS (Global Positioning System) surveying.

Ground surveying using high precision instruments such as total stations is the most accurate method of data collection, providing a typical precision of about 5 mm in one kilometre. A *total station* is an instrument that can measure and record angles and distances. This method is also the most labour intensive involving usually more than one person to conduct the survey.

The Global Positioning System (GPS) allows surveying to be conducted by one person on the ground. It can provide centimetre or better accuracy if the proper instrument and procedure are used.

An indirect but faster method is to conduct surveying on an image of the earth taken from high above the ground. This reduces the need to conduct ground surveys which are often hindered by the inaccessibility of some areas and the lengthy time it takes to conduct such a survey. Two common methods of this types are photogrammetric compilation and remote sensing.

Photogrammetric compilation from aerial photographs is the standard method for making topographic maps. The accuracy of the better topographic maps is about 0.5 mm at

map scale for 90% of the well-defined points such as house corners and road intersections. For a 1:1000 map, this is equivalent to an accuracy of 0.5 metres on the ground.

◇ ? Is resolution of this database high enough for storing topographic maps ?

This database has a scale of 1:1000 and we determined its resolution in a previous exercise. Is its resolution high enough for the expected accuracy ? ◆

Remote sensing satellites produce imageries that are like aerial photographs but with three major differences. First, they are taken from a much higher altitude and therefore are much less accurate, not better than metres. Second, they are sensed using wavelengths beyond the visual spectrum, thus providing information not easily visible from aerial photographs. Third, they are more up-to-date because of the regular coverage of the earth by these satellites. As a result, they are used more for collecting thematic data than for topographic mapping.

Data collected on the ground or from space will be processed and converted to a form compatible with a GIS. In some cases, the data collecting system is linked directly with a GIS.

Data Collection and Control Points

Geographic data collection involves the steps of measurement, transformation, and adjustment.

Measurements are done either directly in the field or on an image of the earth such as an aerial photograph or a satellite picture. This often produces raw data which are not directly tied to the geographic or ground coordinate systems and which contain systematic and random errors.

Transformations and *adjustments* are performed on the raw data in order to fit them to a chosen coordinate system and to correct the systematic errors. To help us carry out these operations, we need to know the correct coordinates of some of the surveyed points in the chosen coordinate system.

Points with known coordinates are called *control points*. National mapping authorities create and maintain a network of highly accurate control points with published coordinates. If available for the project area, they are often included as one of the surveyed points to help transform and adjust the rest of the data.

These top quality controls are sparse and far apart. If additional control points are required for a particular project, ground or GPS surveys can be carried out from these

established controls to produce less accurate ones. In less demanding situations, control points and their coordinates can be obtained from existing maps.

A data collection project usually starts with the identification of control points in the project area. In general, the larger the number of control points, the more accurate will be the survey.

Manual Digitizing

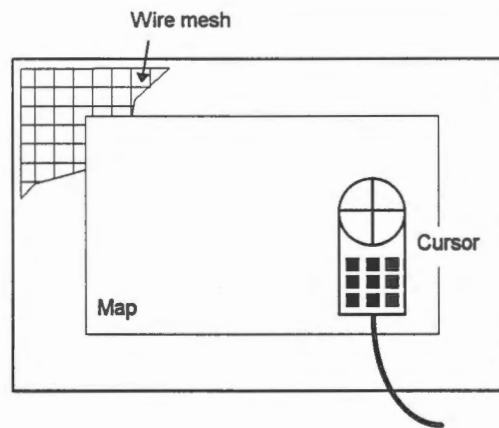
It is more cost effective to base new maps on existing ones. This would not always produce the most accurate and up-to-date maps, but in many cases, lower quality maps are far more desirable than no map at all.

In conventional mapping, new maps are often produced by hand tracing or photographing old ones. If there is a need to change projection, additional calculations and manual drafting would have to be carried out.

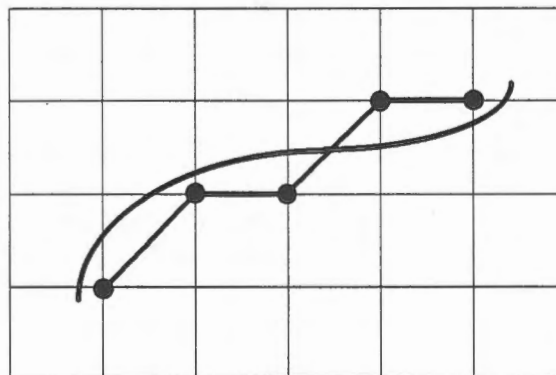
In a digital environment, the need to quickly produce geographic databases from existing maps reaches a new peak because of the urgent need to obtain data in a computer-compatible format — when there is a tool that can handle large volumes of data, we will find good reasons for creating lots of data for it. Because other methods of data capture are too slow and expensive, converting paper maps by manually tracing the lines becomes a very popular method.

The process of converting continuous lines into discrete points so that they can be stored in a computer is called *digitizing*. It involves the manual tracing of lines on a map, resulting in a vector representation of it. There are two modes of digitizing: point and stream mode. In *point mode* digitizing, the user places the cursor over a point and then presses a button on the keypad to send the coordinates to the software. In *stream mode* digitizing, the operator traces a line with the cursor, and the sampling mechanism of the digitizing table picks up the location of the cursor at regular time intervals.

Digitizing is done on a *digitizing table* consisting of a flat surface embedded with a fine wire mesh through which electric currents can pass. A hand-held cursor equipped with a magnetic coil senses its position over the wire mesh, and the coordinates will be sent to the GIS for processing. A keypad on the cursor can be used to enter commands and numbers.



The wire mesh is similar to the resolution grid of a digital database, and it controls the resolution of the digitizing table. In fact, the mesh could be coarser than the resolution, and the table is able to interpolate coordinates in between. In any case, the fixed resolution of a digitizing table results in an imaginary resolution grid from which coordinates are read. A typical resolution for mapping purposes is 0.025 mm (0.001 inch). The diagram below shows the points recorded as the operator traces along the curved line.



The accuracy of a digitizing table (about 0.075 mm or 0.003 inch) is roughly three times worse than its resolution. Factors that contribute to digitizing table inaccuracies are:

- **Instability:** There is a random error associated with the coordinates generated for a fixed point. Hence the coordinates read from a stationary cursor could drift.
- **Non-repeatability:** After a cursor is moved from a spot and placed again exactly at its previous location, the reading of coordinates at the point could differ.
- **Skewness:** The wire mesh does not form square grids.

The accuracy of line tracing on a digitizing table is further affected by the following factors.

- Line thickness: There is a higher chance of drifting the crosshair inside a thicker line. The thinner lines on a printed map are about 0.1 mm wide.
- Crosshair thickness: A thicker crosshair tends to obscure thin lines underneath, making it difficult to trace them with high precision.
- Physical limitations of the human operator: The skill, dedication, and stamina of the person affects the accuracy of pointing and tracing.

As a result, the line tracing accuracy of an experienced operator is about 0.2 mm. The pointing accuracy would be better than that, about 0.1 mm. They are both appropriate for topographic mapping, which expects an accuracy of 0.5 mm at map scale.



The demonstration Concepts: Digitizing explains point-mode and stream-mode digitizing.

Digitizing and Control Points

Control points are used to transform the raw survey data to the desired coordinate system. When you read the coordinates of a point in the digitizing table, the coordinate system used is that of the table. It has an origin selected by the user and the axis of this coordinate system is parallel to the embedded wire mesh. This system is not related to the geographic or ground system at all.

If we know the coordinates of a number of points on the table coordinate system as well as on the ground coordinate system, we can compute the relationship between these two systems. The process of computing this relationship is called *registration*. In most map digitizing, the ground coordinate system is used rather than the geographic coordinate system.

A digitizing project starts with the identification of points on the map with known ground coordinates to serve as control points. We have mentioned the use of control points in data collection, and they serve a similar function here. Again, how we obtain their ground coordinates depends on the accuracy required. If we can find high accuracy national control points on the map, we can request their coordinates from the authority. A quick but less accurate method is to read their coordinates from the map. The grid lines on a map are well-defined points already labelled with coordinates. They make excellent control points for map digitizing.

We can also register raster images to the ground coordinate system using control points. This allows us to superimpose an aerial photograph or a satellite image on an existing line map in the database.

The Procedure of Digitizing

The following steps are normally followed in digitizing a new map.

1. Identify control points on the map and enter their coordinates into the GIS.
2. Register the map using the control points.
3. Identify feature to be digitized.
4. Enter the attributes, such as classification code and key, for the feature.
5. Digitize the feature.
6. Repeat from step 3 until done.

Raster Scanning

A faster method of converting graphics into digital form is to scan them automatically. This produces a raster representation instead of a vector one.

Part of the time spent in manual digitizing is to enter attributes, such as classification code and name, for the features. In raster scanning, the image will be converted to a raster representation automatically, but the attributes of the features on the image cannot always be picked up so easily.

It is possible to separate the original map into layers according to their attributes, for example separate the highways from the houses, before scanning them. This would at least allow us to quickly assign the same attributes to features scanned from the same layer. However, some manual pre-processing would be required to separate the map into layers in the first place.

Conventional cartography often produces colour separations that can be scanned separately. For instance, scanning the blue layer would capture all water feature, and scanning the brown one would capture all the contours. Some manual processing would again be required if a project needs finer classification of features under the same colour.

Unlike manual digitizing, during which the operator performs a selective process to eliminate noises and features not relevant to the database, raster scanning picks up everything on the document, including dirt and imperfections. A post-processing step is then required to clean them.

In short, raster scanning rarely produces automatically a clean digital database with features properly associated with attributes. When there is a need to operate the data in vector form, the process of converting from raster to vector is not trivial.

Chapter 37

Accuracy and Data Editing

Data editing is necessary for three main reasons:

- Correct input errors.
- Modify data to reflect changes in the real world.
- Enhance the map cartographically.

In the first case, the operator has made the mistakes. In the second case, the data is too old. They both result in databases that are wrong in one or more of the following ways. We will address the third case in the following chapter.

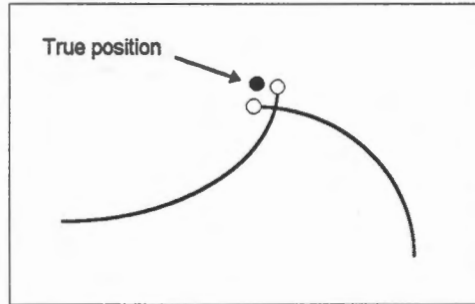
Data with incorrect attributes. Attributes are characteristics of a feature, and wrong attributes would result in the incorrect representation of features. For instance, a piece of property given the wrong owner, the wrong address, or the wrong assessment value is no longer the feature it is supposed to represent.

Data with incorrect position. Spatial data are never one hundred percent accurate because both data collection devices and their operators introduce certain degrees of error. This also causes errors in the coordinate systems because they are established from measurements on the shape of the earth.

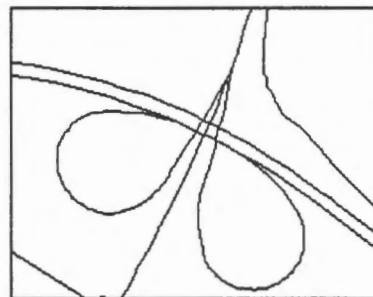
Missing data. A highly accurate map may be incorrect because it does not contain all the features intended for the database. On the one hand, some of the highways, houses, and lakes may be missing from a topographic database. On the other hand, the database may contain all the features but miss some of their attributes. For example, it might not contain the assessment value for some of the properties. This might be due to human errors during input, or because such data were not available at the time of input.

Inconsistent data. Logical inconsistency is caused by incorrect relationships among the data items. A common problem is that the positions of geographic features are correct but the relationships among the features are not, causing logical inconsistencies.

In any spatial database, we can expect its data to contain a certain amount of error. For a high quality topographic map, we can only expect 90% of its well-defined features to be within 0.5 mm of their correct positions at the map scale. But an accurate database might not be logically consistent. In the diagram below, the end-points of the two lines are within 0.5 mm of their true positions at map scale and are therefore “accurate” according to the specification for topographic maps. However, the two end-points do not meet as they are supposed to. This inconsistency causes a topological error.

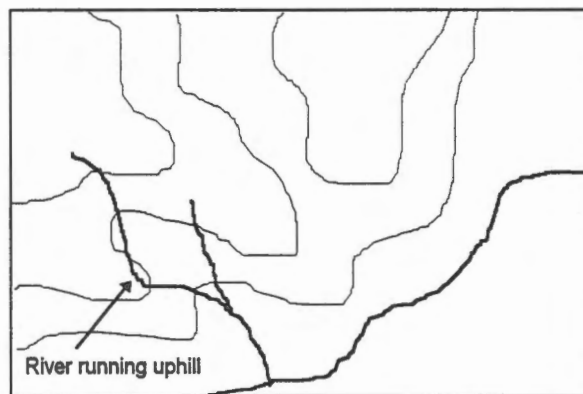


Another example of logical inconsistency is shown in the highway network below. Although the position of the lines are within the accuracy specification, they are not smooth as highways should be. In other words, the position of the points along the line are correct within the specification but are not correct relative to each other.



(288822, 740560), 1:2700

Another common type of logical inconsistency is related to the terrain. In the diagram below, part of the river crosses the same contour three times, giving the impression that it is flowing uphill along part of its course.



Cartographic Enhancements and Data Editing

The third reason for editing is to make a map more legible. A conventional map, as well as a display on the screen, is a communication medium that graphically conveys to users geographic information. Because of the limited space provided by a map and the large amount of information that is normally shown on it, cartographers use various techniques to enhance the graphic message for easier reading. Cartographic enhancements address the following issues:

Name placement. Text annotations are added to the map to depict some of the feature attributes. For example, we add text to show the names of roads, rivers, buildings, and so on. Names for linear and area features are often curved, and names in dense areas of the map have to be arranged so as to avoid clustering. Contour labels follow special cartographic rules to ensure that they can be read without rotating the map.

Generalization. This is a technique of making use of the limited space available on the map to effectively depict geographic information. At a small scale, features that are too thin or small to be visible have to be exaggerated in size. This expansion could interfere with the surrounding features, and the affected features have to be moved to maintain topological consistency.

Symbolization. Annotation text alone is not enough to represent all the attributes of features. In fact, too many annotations could create clustering. A very effective way of visually representing the attributes is to symbolize the point, line, and area features. For instance, a dashed line can be used to indicate that a highway is under construction, and a double line can be used to indicate that the highway has multiple lanes.

Graticule. A network of grids on a map labelled with coordinates form a graticule, which helps the reading of positions. A map is not just a visual depiction of the geographic features, it is also a database. In order to facilitate geographic feature retrieval and spatial analysis, maps are superimposed with a graticule labelled with coordinates. This allows reading of the coordinates of any point on the map, and when given the coordinates of a feature, the graticule helps the quick location of it on the map. These coordinates can then be used to compute distance, area, angle, and other measurements.

Margin information. Supplementary information about the map, such as its scale, map projection, reference ellipsoid, map identification number, and year of the survey are placed around the margin of the map.

Cartographic enhancements are largely manual operations, and a GIS provides interactive editors for this purpose. There has been considerable research conducted on automating these operations, and some important advances have been made.

When the map is not too dense, name placement can be automated to a large extent. It becomes difficult in a crowded place. An interesting problem regarding name placement is the automatic generation of a label at the visual centre of an area. There are many good techniques for this, and most GIS would provide such a function.

Symbolizing point features and line features can be largely automated. Symbolization of area features mainly consists of colour and pattern filling and are quite easy to automate.

Much GIS software provides the function of generating a graticule automatically. A potential problem here is interference of the grids with names on the map.

Margin information is often quite uniform for maps in a series. They have the same scale, map projection, and so on. Hence margin information is often created as templates which are then superimposed on the maps.

Chapter 39

The GIS Editor

In the following chapters, we will use the GIS Editor to learn methods of editing geographic data. To open the Editor, click on the CARIS Editor icon on the CARIS Suite



and select the file called FTON.DES

The Editor is different from the Information Manager that we have used so far in the following ways.

- The Editor allows you to change the database while the Information Manager does not.

NOTE In the version of CARIS customized for this guide, you cannot save the changes. If you are using a full version of CARIS, we do not expect you to save the changes before you exit from the Editor.

- The Editor does not work with the DB Manager and therefore only operates on system-defined attributes.
- The arrangement and the label of some of the menu items are different. For example, MapQuery > Point Location in the Information Manager is Query > Point Location in the Editor. Although they are labelled differently, their operations are identical.

Other than those differences, the operation of the Editor is similar to that of the Information Manager, and many of the concepts we have learned before will be applicable here.

NOTE The screen captures used in some of the examples might look different due to changes you might have made to the database.

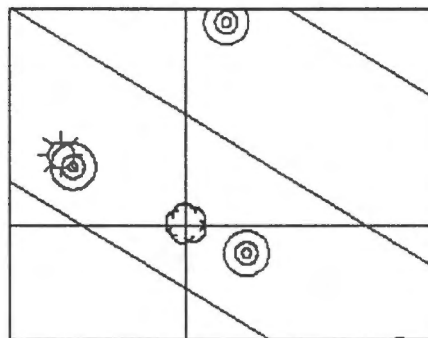
Move Features

Geometrical transformations change the location, shape, or size of features. Move, or shift, is the simplest of the transformations. We will describe the shifting of point features and names in this chapter.

To move a feature, we have to identify it and then point to its new location. We will illustrate this by moving a point feature.

 Move a tree.

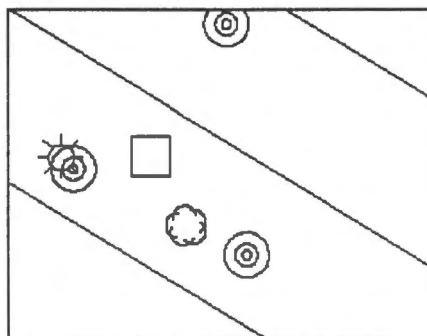
1. Select the Edit > Symbol > Move menu item and point at the tree.



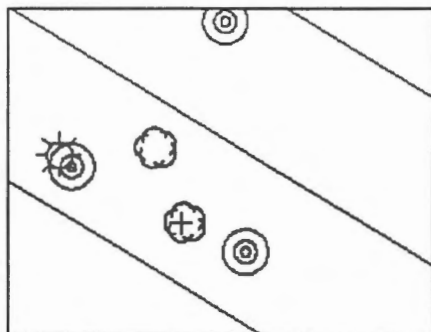
(288684, 740312), 1:200

2. The tree will be highlighted; click on the left mouse button to confirm.
3. A box will appear, and you can move it to the new location for the tree and click on the left mouse button.

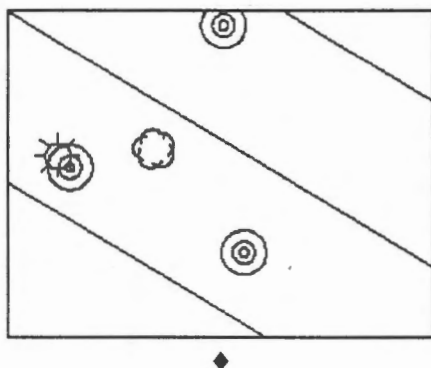
NOTE The box is called a *minimum bounding rectangle* because it is the smallest upright rectangle that bounds the symbol.



4. A tree will be drawn at the new location and a cross will be placed at where the old symbol was. Although the image of the old symbol remains, it has been moved in the database.



5. Refresh the screen to update the display.



The demonstration *Edit graphic data: Moving a symbol* shows how to move a symbol.

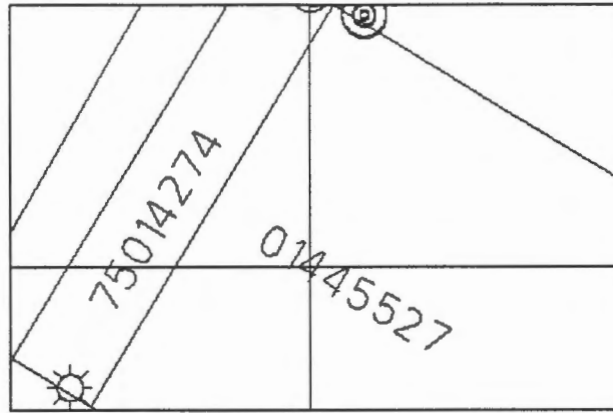


Move a name.

We will move a display label to a new location. To the Editor, a display label is just a name.

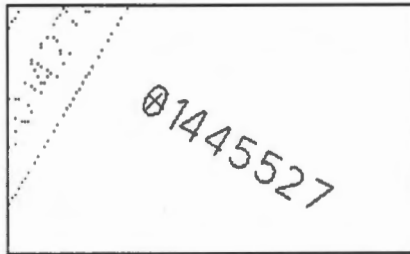
1. Select the *Edit > Text > Move > Whole Name* menu item and point at the name.

NOTE You select a name by pointing at any of its characters.



(288700, 740335), 1:200

2. The name is highlighted for confirmation.



3. After you have confirmed that this is the name to move, a minimum bounding rectangle of the name appears. Move it to the desired location and click on the left mouse button.
4. A cross will be placed over the old name.
5. Refresh the screen. ♦



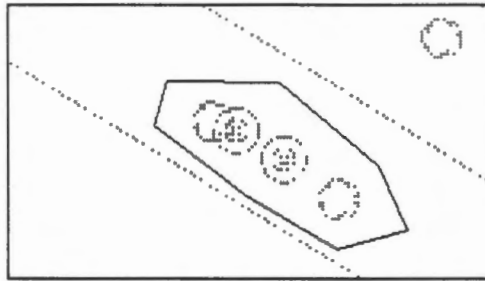
The demonstration Edit graphic data: Adding a name shows how to add a name.



Move a group of symbols.

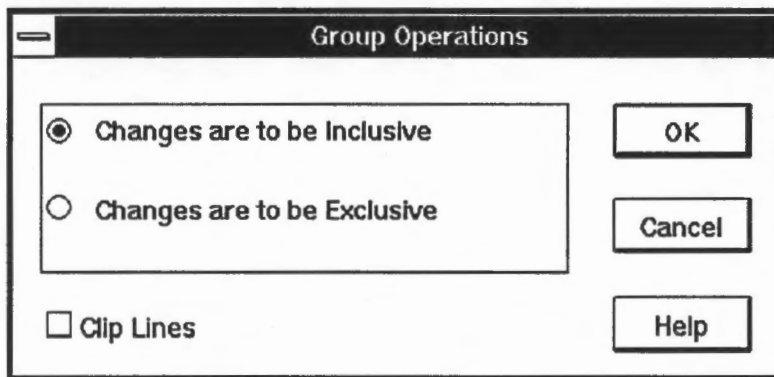
To move a group of features, such as symbols, you first put them into a group. Then you use a group move command to move them.

1. Select the Group > Define Group menu item.
2. The crosshair will appear for you to draw a polygon around the symbols you want to form into a group. After you have entered the last point, click on the right mouse button to close the polygon.

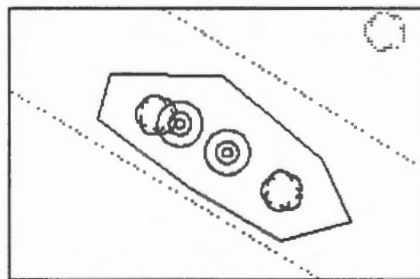


(288715, 740288), 1:200

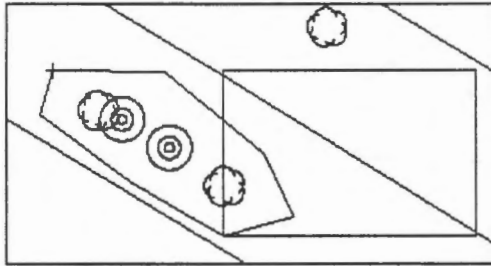
3. Click on the left mouse button to confirm.
4. Select the Group > Feature > Move menu item.
5. Turn on the button to indicate that you want to move symbols inside the window. Click on the OK button.



6. The features inside the polygon will be highlighted for confirmation. Click on the left mouse button to confirm.



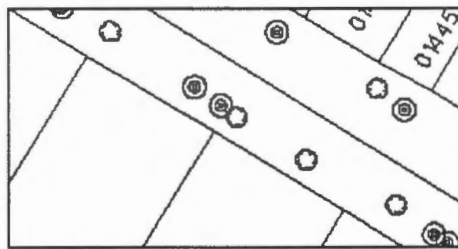
7. The bounding rectangle for the polygon appears. Move it to the desired location and click on the left mouse button.



8. Refresh the screen. ♦

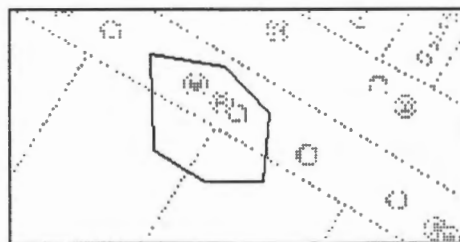
 Move different types of features together.

You can move points, lines, and other features together by enclosing them in the polygon. In this example, we will move a few symbols and lines in one operation.



(288757, 740260), 1: 400

1. Use the Group > Define Group menu item to group these features together.



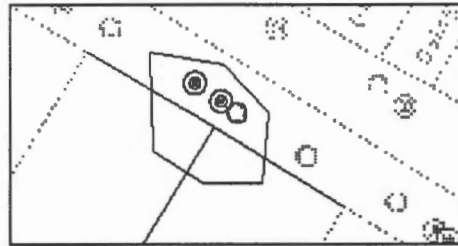
2. Use the Group > Feature > Move menu item to move them. Turn off the Clip Lines button to not clip the lines.

Changes are to be Inclusive

Changes are to be Exclusive

Clip Lines

All features selected for the move will be highlighted. Since lines are not to be clipped, any line inside or overlapping the polygon will be moved as a whole.



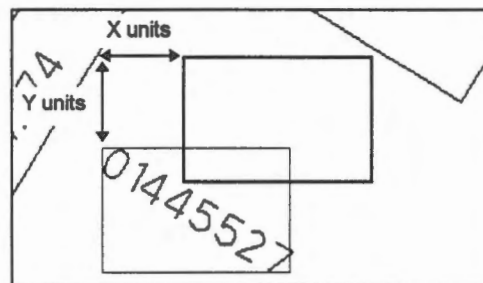
◆ Can you move part of a line?

Try to do a similar move as in the previous example, but this time turn on the Clip Lines switch. ◆

📖 Performing a shift in the database.

The shifting of a single point feature is simple — its old coordinates are replaced by the new coordinates of its new location.

The shifting of a name involves a bit more work. The amount of shifts in X and Y are computed from the old and new bounding rectangles, and they will be applied to each of the characters in the name.

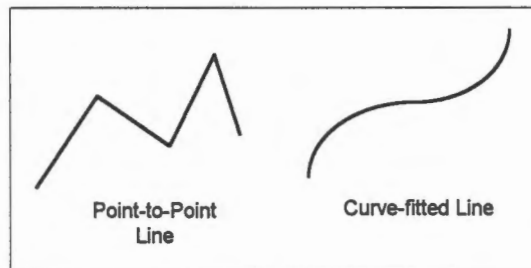


The same method is applied to the shifting of a group of symbols. ◆

Chapter 41


Modify Shape of a Line

This GIS recognizes two basic kinds of lines: point-to-point, and curve-fitted. Basically, a *point-to-point* line is angular suitable for building outlines, while a *curve-fitted line* is smooth and suitable for contours.

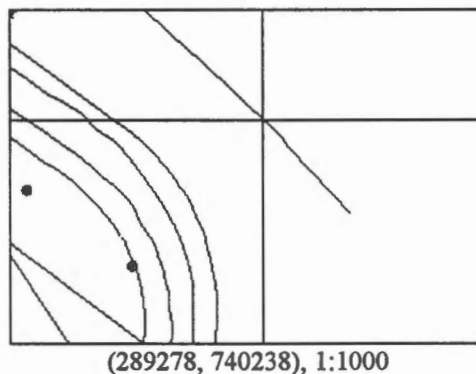


The shape of a line can be changed by one of the following editing operations:

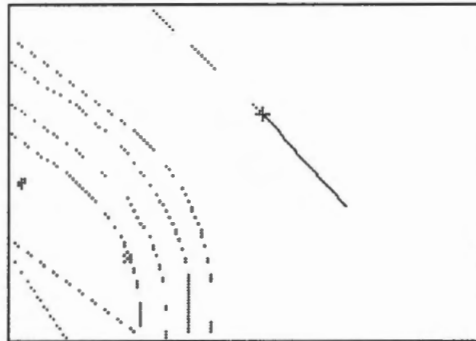
- Clip the end of a line.
- Move individual points on a line. This can be performed only on point-to-point lines.
- Extend a line.
- Change an interior segment of a line.

 Clip the end of a line.

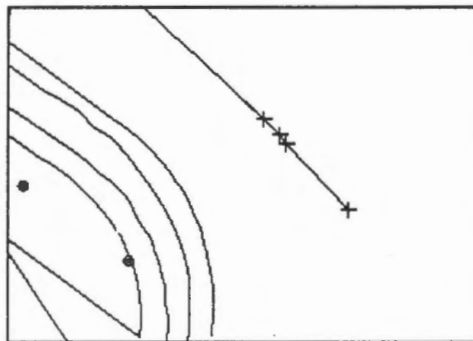
1. Exclude contours (theme 90) from the display.
2. Select the Edit > Line > Change > Clip menu item.
3. Point at the spot where you want the clip to occur (289299, 740302).



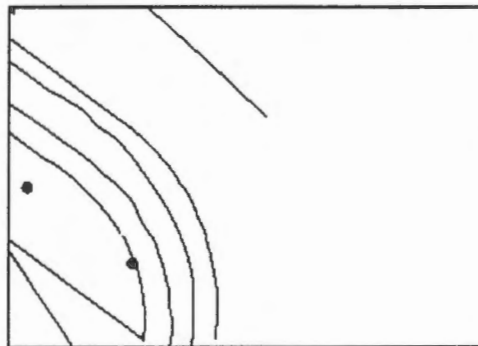
- The shorter end of the line will be highlighted. Click on the left mouse button to clip it, or enter X on the keyboard to clip the other end instead.



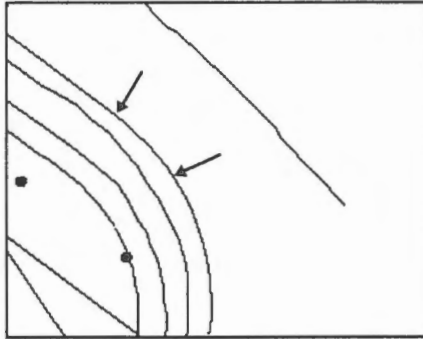
- Crosses will be drawn over the deleted end. When the crosshair appears, click on the right mouse button to quit this command.



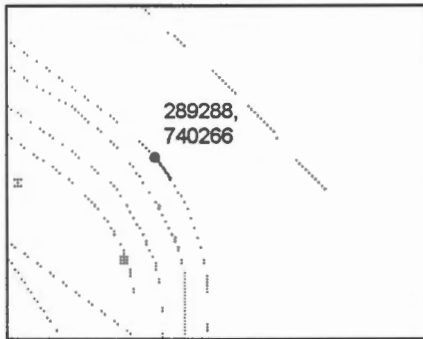
- Refresh the screen to show the updated map.



A visually continuous line could be composed of several disconnected pieces. For example, the line shown here breaks at the two points indicated by the arrows.

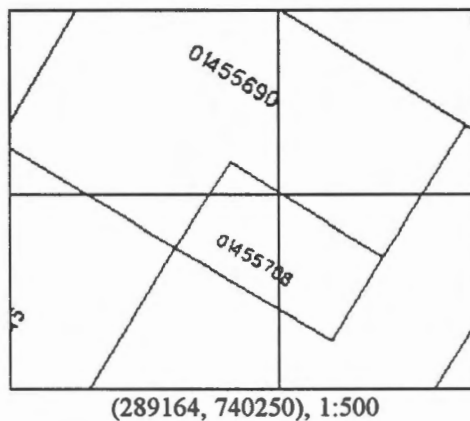


So if you enter a cut point at (289288, 740266), the highlighted portion would appear to cover only part of the line.

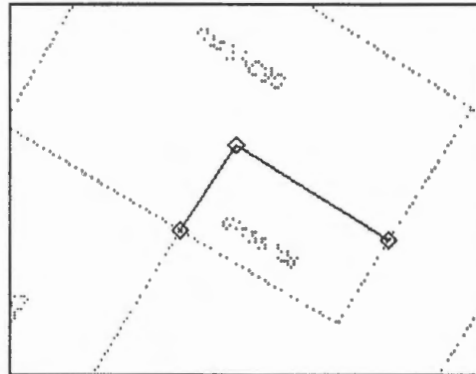


 Move individual points on a line.

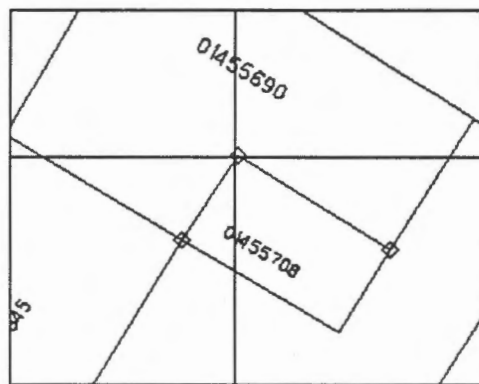
1. Select the Edit > Line > Move Points in Line menu item and point at the line to be changed.



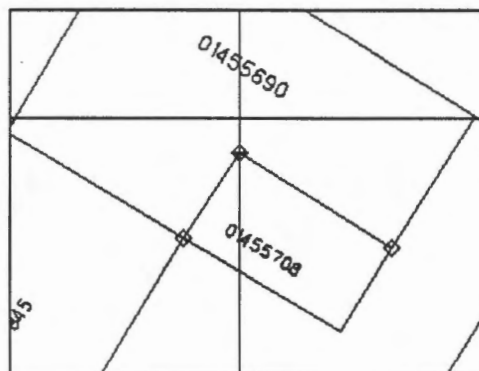
2. All points on the line will be highlighted. Click on the left mouse button to confirm.



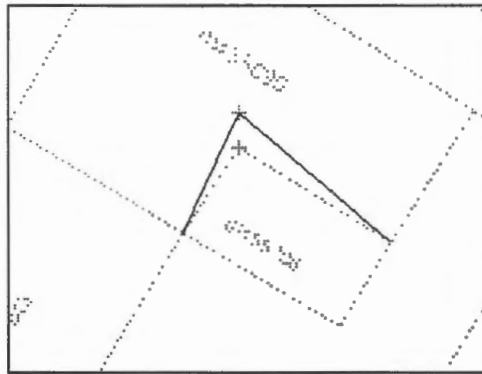
3. Select the point to move.



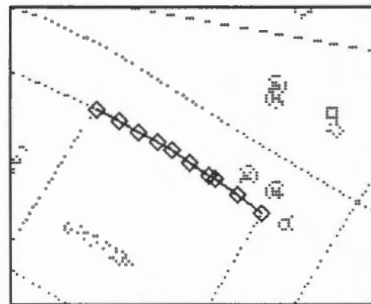
4. Point to its new location.



5. You can move more points if necessary. We will only move one point in this case, so click on the right mouse button to finish moving points.
6. The new line will be highlighted for confirmation.



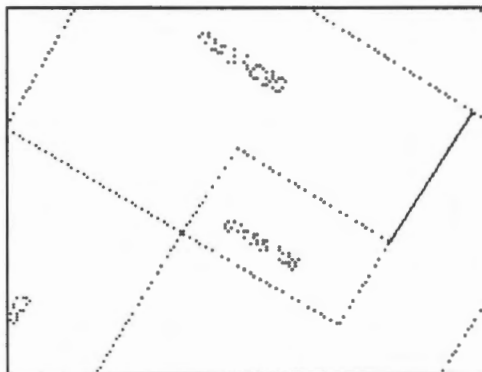
In this GIS, you cannot move individual points in a curved line, such as the one highlighted here. You can see that a curved line is also composed of individual points, except that they are spaced closer together.



(288210, 740430), 1:500

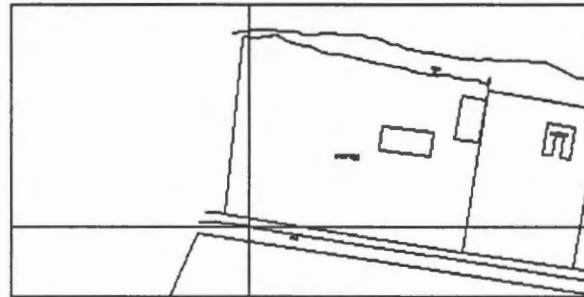
❓ Clip a straight line.

You can clip a straight line using the move point command. Try it on this line.



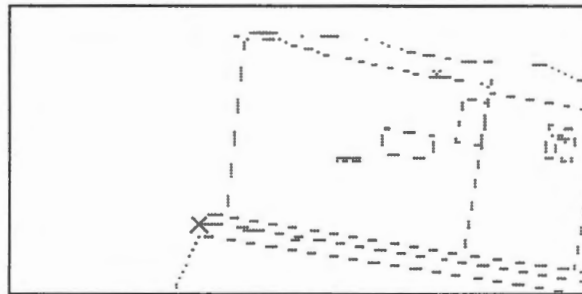
 Extend a point-to-point line.

1. Select the **Edit > Line > Change > Extend Line** menu item and point at the line to be extended.

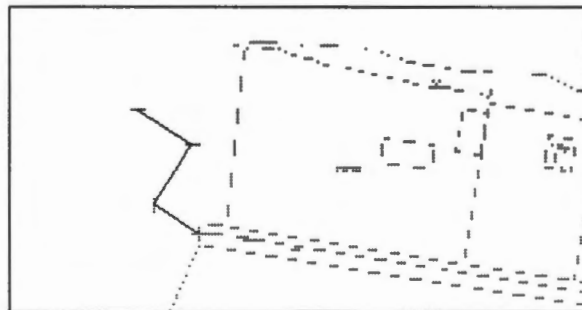


(287814, 740607), 1:2500

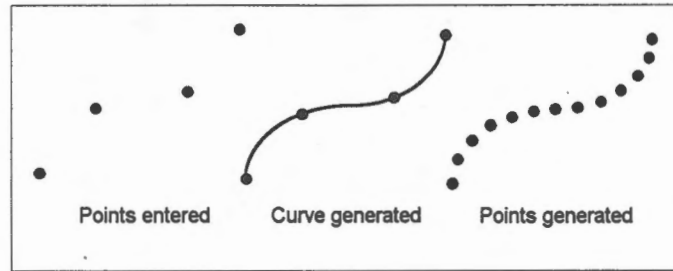
2. The line will be highlight for confirmation. Click on the left mouse button.
3. The closer end will be highlight for confirmation. Click on the left mouse button. If you enter X on the keyboard, the other end will be selected instead.



4. The crosshair will appear. Enter the points to extend the line and after adding the last point click on the right mouse button. The extension will be highlighted for confirmation.

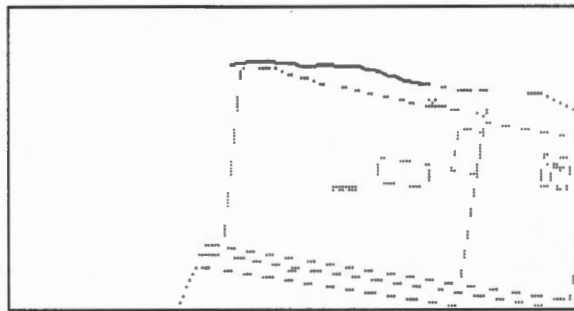


In the previous example, a point-to-point line is extended by entering every point in the extension. For curved lines, we need to enter only a few points, and the rest of the points making up a smooth extension will be generated automatically.

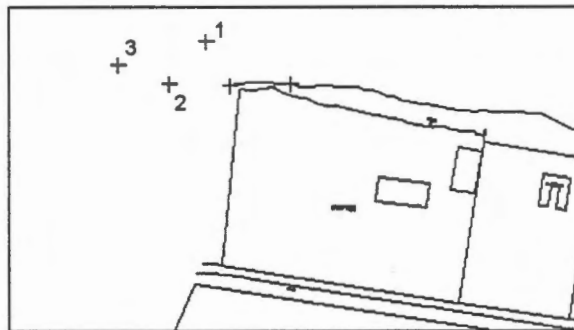


 Extend a curved line.

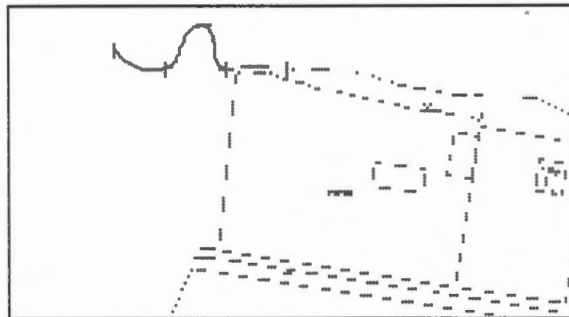
1. Select the Edit > Line > Change > Extend Line menu item to extend this line.



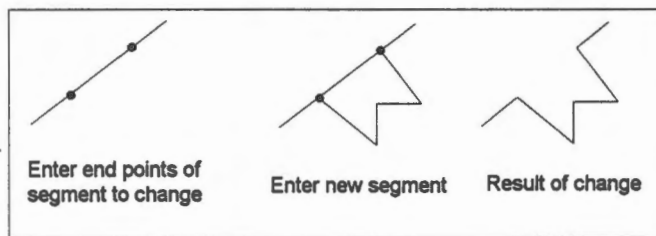
2. Enter the three points in the order shown.



3. The extension is a smooth line.



To change an interior line segment, we first identify the segment by its two end points. We then enter the new line segment.

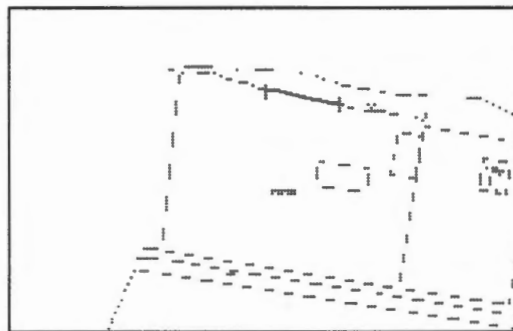


 Change an interior segment of a point-to-point line.

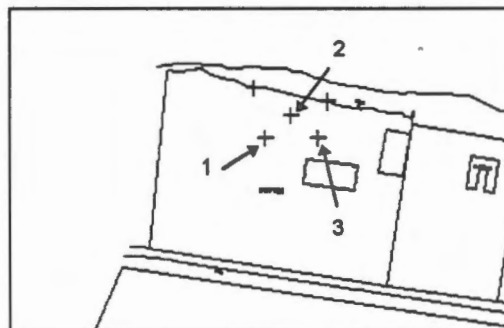
1. Select the Edit > Line > Change > Part of a Line menu item to identify the part of the line to be changed.

NOTE You should select the line by pointing at one end of the segment to be changed.

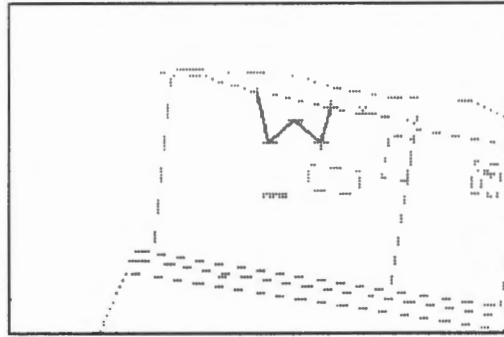
2. Confirm the line, point at the other end of the segment, and confirm the segment.



3. Enter the points for the new segment in the following order.



4. Confirm the new segment.



◆ Change the interior segment of a smooth line.

Find a smooth line and change a part of its interior. You can tell if a line is a smooth one by the closeness of its points. Use the Edit > Line > Move Points menu item to highlight the points on the line. ◆

📖 Curved lines and data compression.

A curved line is composed of many points that are spaced very close to each other. In this database, a point is composed of two coordinates with six digits each. Hence a line composed of ten points would need 120 digits to store.

If the points are spaced very close together, the difference in X and Y coordinates between adjacent points would be small. For instance, this line

(288170, 740500), (288177, 750504), (288180, 750530), (288190, 750541)

of four points is composed of 48 digits. The difference between the first two points in X and Y are (7,4), that between the second and the third point are (3, 26), and that between the last pair are (10, 11). Using the differences in coordinates instead of their absolute values, a more compressed form of representation using only 21 digits can be achieved:


(288170, 740500), (7,4), (3,26), (10,11).

In the first representation, the *absolute coordinates* of the points are used while in the second representation, their *relative coordinates* are used except for the first point. It is easy to compute the absolute coordinates of the points from their relative coordinates. ◆

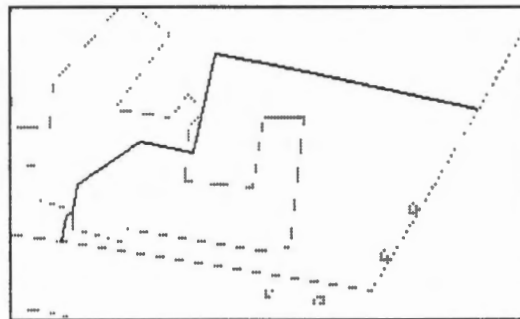
Chapter 42

Line Splitting and Resolution

Line splitting is a simple editing operation, but it has some interesting side effects caused by resolution. We will discuss a few of them in this chapter.

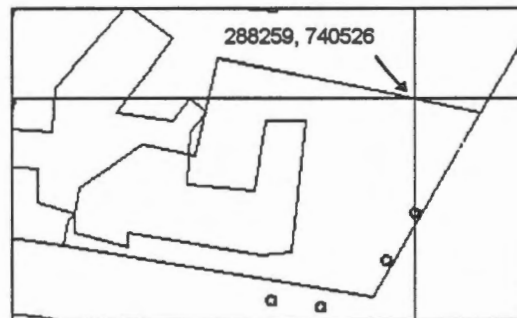
 Split a line into two.

We will split this line, which is 150.466 metres long on the ground, into two parts.

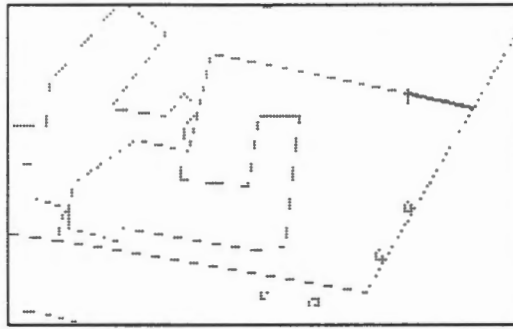


(288173, 740531), 1:800

1. Display only theme 60.
2. Select the **Edit > Line > Change > Split** menu item and place the crosshair at the cutting point (288259, 740526).

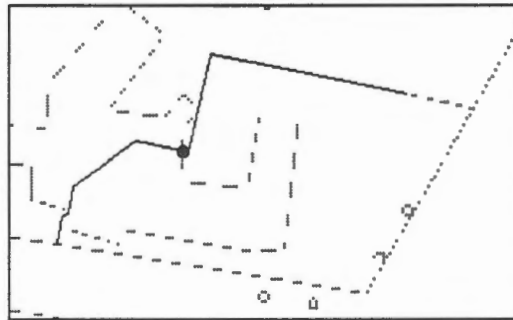


3. Use the Information icon on the CARIS Panel to show that the line has actually been split. In this example, the shorter piece has a length of 18.170 metres and the longer one has a length of 132.295 metres. They add up to 150.465 metres, 0.001 metre shorter than the original line. This difference is due to the resolution of the database.

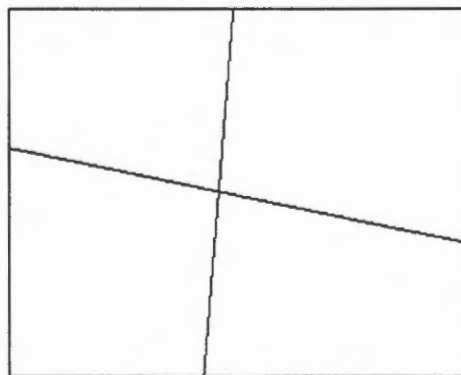


 Split a line at its intersection with another line.

We will split the city block where it meets the building.



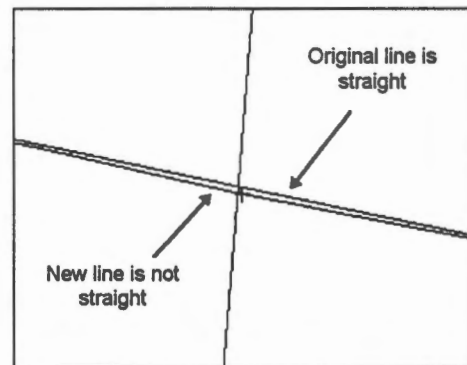
1. Enlarge that part of the database since it is difficult to point precisely at their intersection. The following is a display at 1:5. If the Options > Display > Scale ... menu item does not allow you to enter such a large scale, enter the largest possible, say 1:10, and then use the Zoom In icon to magnify further.



(288199, 740511), 1:5

2. After the split, you will notice that the cutting point does not occur at the location intended. This is because the closest resolution point is not at the intersection of the two lines, but at the location of the cross.

As a result, the shape of that part of the city block has been changed from a straight line to a bent one.



3. If you have zoomed in too much, the zoom out icon would not work. You could, however, use the Draw > Previous > Most Recent menu item to return to the previous display.
4. The city block has now been split into three lines. In this example, the three pieces have lengths of 49.822, 82.572, and 18.072 metres. They do add up to 150.466 metres indicating that the errors due to resolution could be canceled out. ♦

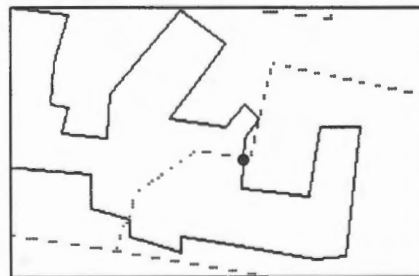


The demonstration *Concepts: Effects of Resolution* explains this and other effects of resolution on editing.

◆ Split the building where it meets the city block.

We have split the city block at the point where it meets the building. The building is still one continuous line. Split it where the city block was cut so that four lines meet at the same point.

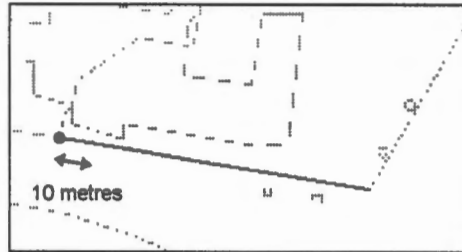
There is a method of performing this precise split without having to draw the database at a large scale.



Would that part of the building close to the cut point change shape as well? ♦

 Split a line at a specified distance from a point.

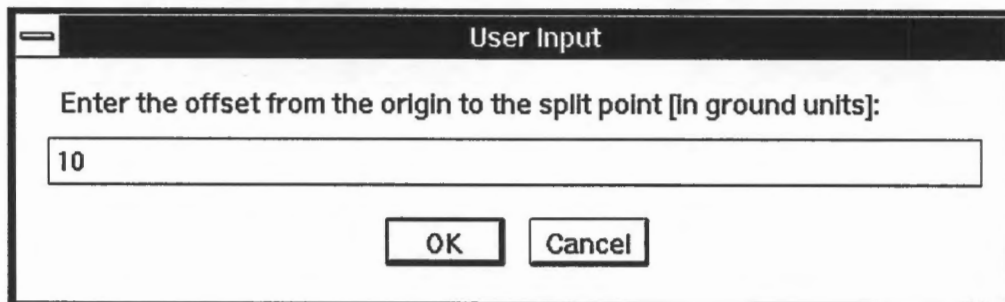
This part of the city block is 84.312 metres long. We are going to split it at a point 10 metres from the left. You will see that because of the resolution, the piece created will not be exactly 10 metres long.



1. Select the **Edit > Line > Change > Split With Offset** menu item and point close to the left end of the line.

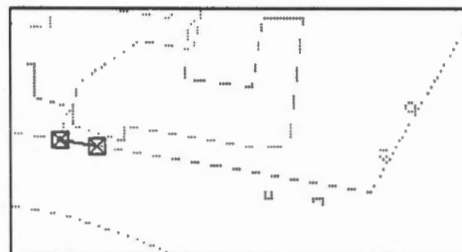


2. Enter an offset of 10 metres for the split.



A dialog box titled "User Input" with a close button in the top-left corner. The text inside reads "Enter the offset from the origin to the split point [In ground units]:". Below this text is a text input field containing the number "10". At the bottom of the dialog are two buttons: "OK" and "Cancel".

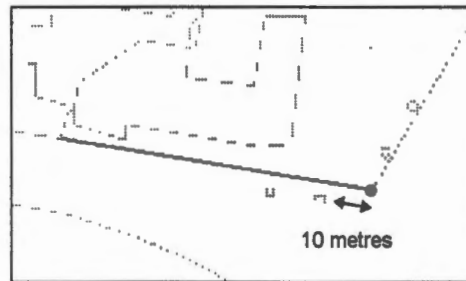
3. The piece will be highlighted for confirmation.



4. Measure the length of the two pieces. They are 10.028 and 74.284 metres, adding up to 84.312 metres which is identical to the original length. ♦

❓ Split with the offset from the other end.

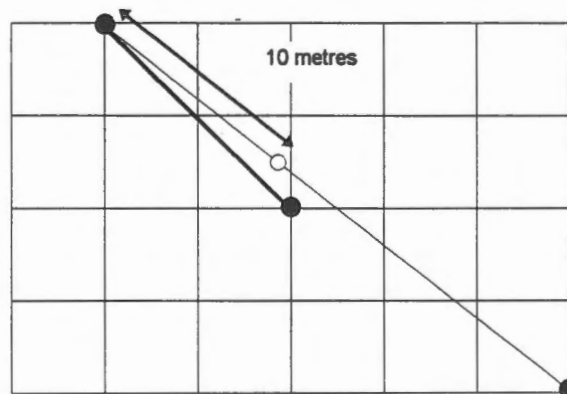
Split the city block at a point also 10 metres from the other end. Check its length and see whether it is off from 10 metres by the same amount as in the previous exercise.



📖 Why is it not always possible to create lines with specific lengths?

We have seen in the previous examples that it is not always possible to create a line with a specific length such as 10 metres. This is again due to the resolution of the database that controls the geometrical accuracy of the features.

The diagram below shows the reason. Before the line was split, a length of 10 metres ends at the open dot. After the line was split, its new end must be located at a grid intersection, thus causing the 10-metre long line to shift from the open dot to the black dot resulting in a line longer than 10 metres. Under another situation, a shorter line or even one with the correct length could be created.



Chapter 43

Modify Text

Text strings (also called names) are cartographic features placed around, along, and inside features for annotation purposes. Their position need not be accurate, except that they should be related unambiguously to the features they annotate. Consequently, we have quite a bit of freedom in their placement. The common changes we make to text, other than moving it as explained in the previous chapter, are:

- Edit text.
- Rotate and stretch text.
- Bind text along a curve.

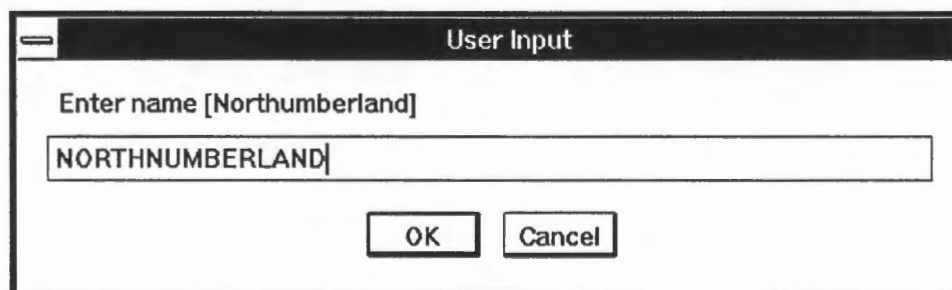
 Edit text.

1. Display all features. Select the **Edit > Text > Change** menu item and point at this text, which is part of a street name.



(288449, 740355), 1: 500

2. You will be asked to enter a number of attributes for this text, including its feature code, size, line weight, font, and so on. Choose **OK** for all of them to keep the original values except for the name. We will change it to **NORTHNUMBERLAND** in upper case.

A dialog box titled 'User Input' with a text input field containing 'NORTHNUMBERLAND' and 'OK' and 'Cancel' buttons.

User Input

Enter name [Northumberland]

NORTHNUMBERLAND

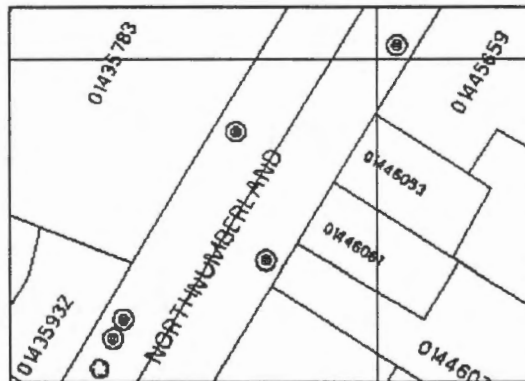
OK Cancel

3. After the text has been updated on the display, refresh the screen.

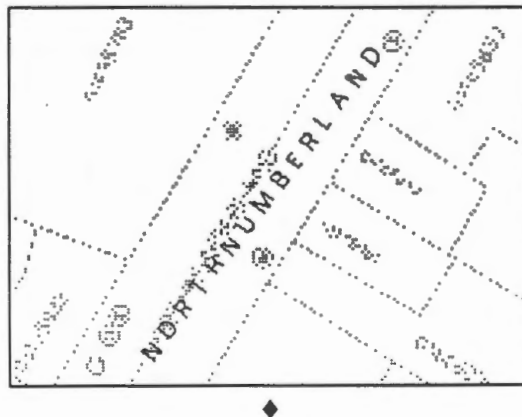



 Rotate and stretch text.

1. Select the Edit > Text > Rotate/Stretch menu item and point at the text we have just changed.
2. Point at the new position for its last character.



3. The text will fill the space between its first character and its new end.



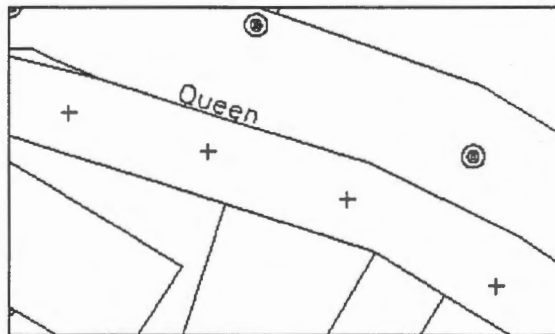
 Bind text along a curve.

1. Select the Edit > Text > Move > Fit to New Curve menu item and point to this text.

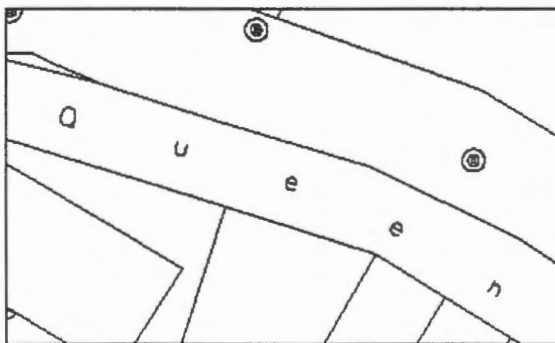


(288593, 740505), 1:400

2. Use the crosshair to draw crosses over the curve you want the name to follow.



3. The characters of the text will be spread evenly over the curve you have just defined. Notice that each character has been rotated so that it is at a right angle to the curve.



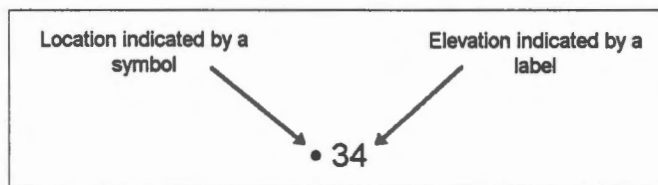
The demonstration Edit graphic data : Adding a Name Along a Curve shows how to add a name along a curve.

Chapter 44

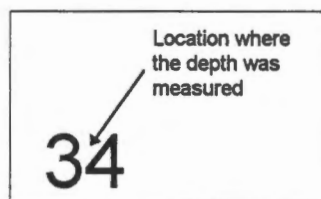
Add Point Features

We will describe how to add three kinds of point features in this chapter. They are symbols, spot heights, and soundings. Symbols, as we have explained before, depict point features. Spot heights are elevations above the ground, and soundings are depths below the water's surface.

As indicators of elevation, spot heights and soundings have positions as well as numeric values. In the case of spot heights, their positions are indicated by symbols.



For a sounding, its location is not indicated by a symbol. Instead, the *visual centre* of the sounding is placed over the spot where its depth was taken. The term visual centre (or optical centre) means what an operator would estimate to be the centre of the sounding label. This could be different from the mathematically calculated centre of gravity.

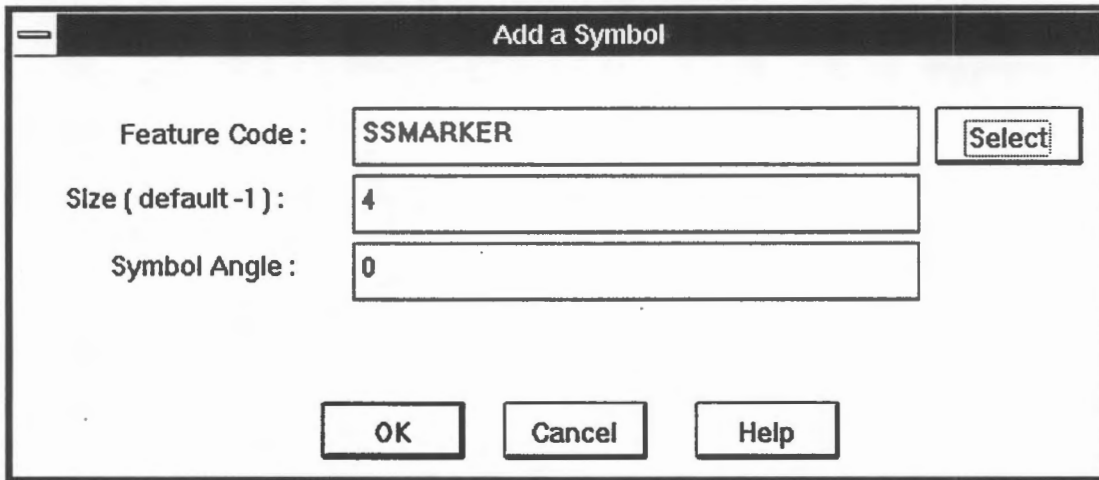


There are good cartographic reasons for this difference in presentation between a spot height and a sounding. Measurements on the ground are usually more accurate than those under water. Hence the use of a more precise point to indicate the position of a spot height is appropriate. The other reason is that there are a lot more soundings on a hydrographic chart than spot heights on a topographic map because of their importance to navigational safety. The already high density of soundings on a chart simply does not leave enough room for additional symbols.

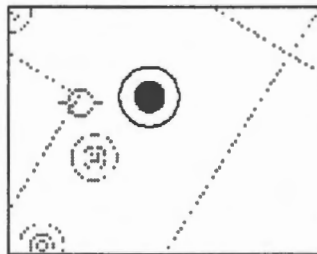
 Add a symbol.

1. Select the Edit > Symbol > Add menu item.
2. Enter the feature code, size, and angle for the symbol.

NOTE Make sure the display parameter includes the feature code SSMARKER, otherwise you will not be able to see the symbol added.



3. Place the symbol at the desired location..



NOTE If you have entered a wrong feature code, and the symbol is undefined, you will see a question mark.

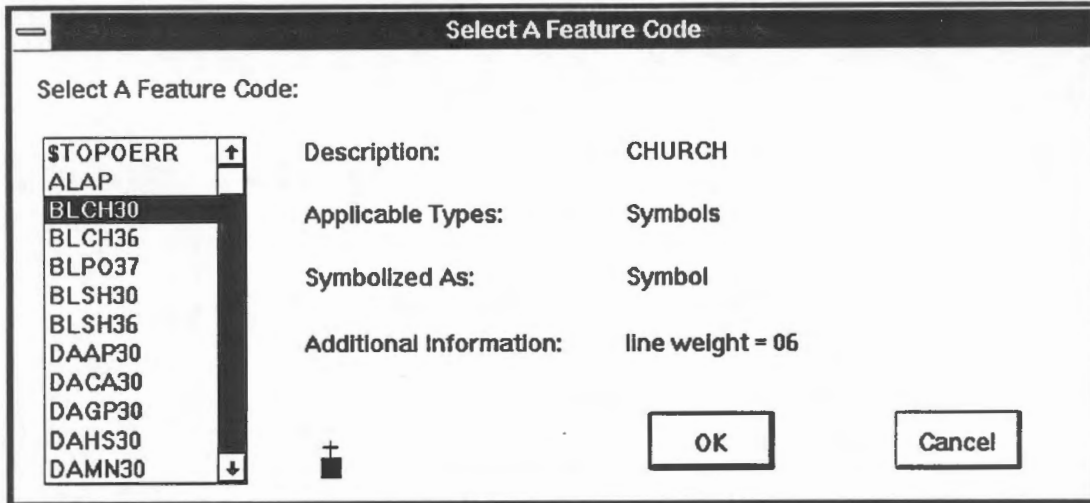


The demonstration *Edit graphic data : Adding a symbol* shows how to add a symbol.



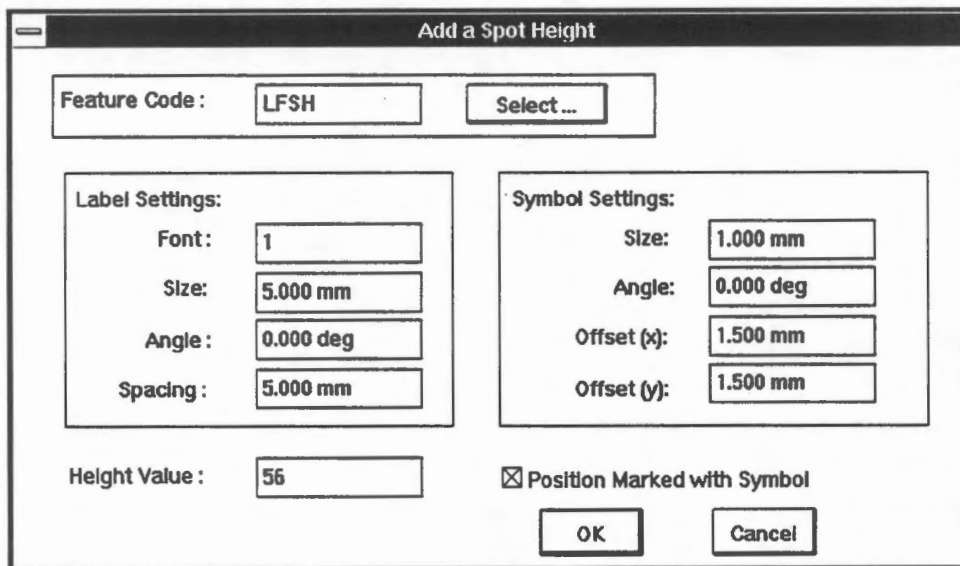
Try other symbols.

The symbols for the point features are controlled by their feature codes. Try other feature codes by clicking on the SELECT button in the Add a Symbol panel. If you click on a feature code, a description and a picture of it will appear.

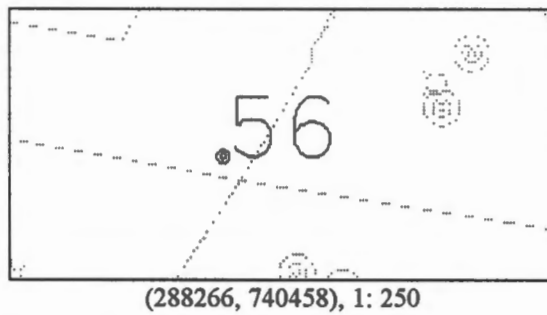



 Add a spot height.

1. Select the *Edit > Spot Height > Add* menu item.
2. Select a feature code of LFSH, 5.0 mm for the size and spacing for the label, 1.0 mm for the size of the symbol, and 1.5 mm for offsets in x and y. The offset indicates the spacing between the label and the symbol. Click on the OK button to continue.



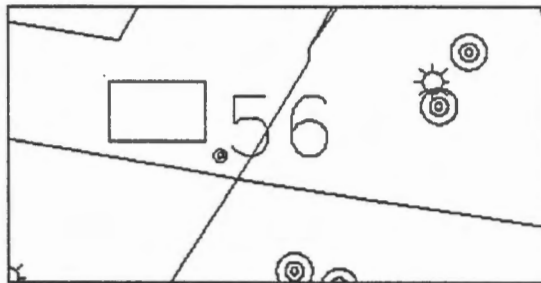
3. The crosshair appears for you to enter the position of the spot height. The symbol will be placed at the location you have selected.



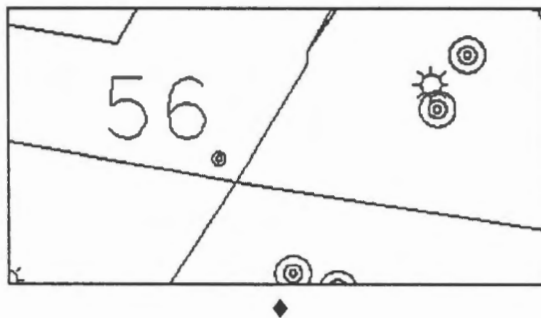
 Move the spot height label around.

For a spot height, the symbol indicates the position, whereas the label indicates its height. If the label interferes with other graphics on the map, we would have to move it around. As in our example, the label is placed on top of the street centre line. We will move the label to a less crowded spot leaving the symbol untouched.

1. Select the Edit > Spot Height > Move Label menu item and point at the symbol of the spot height you want to edit.
2. A bounding rectangle for the label appears. Move it to the new location for the label.





3. When the crosshair appears again, press the right mouse button to quit. Refresh the screen.



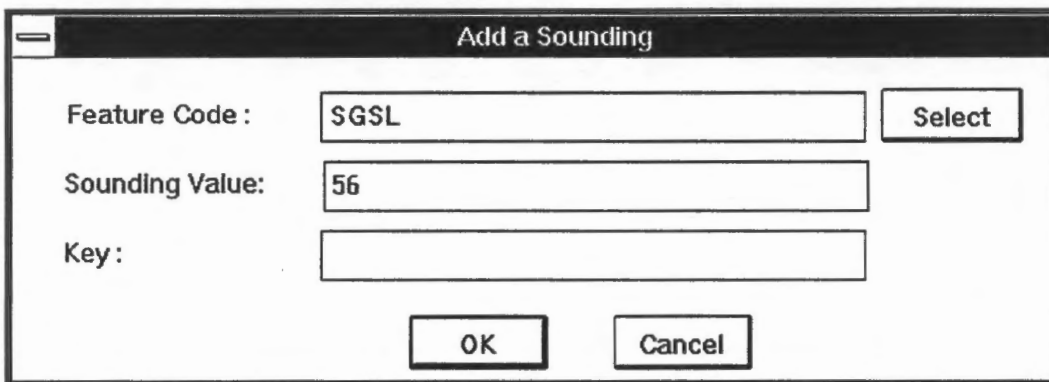
- ◆ Move the symbol and the label of the spot height together.

The symbol and the label of a spot height are tied together as a unit although you can move the label without moving the symbol. Try the Edit > Spot Height > Move menu item. ◆

 The demonstration Edit graphic data: Adding a spot height shows how to add a spot height.

 Add a sounding.

1. Select the Edit > Sounding > Add menu item.
2. Enter the feature code and the sounding value.

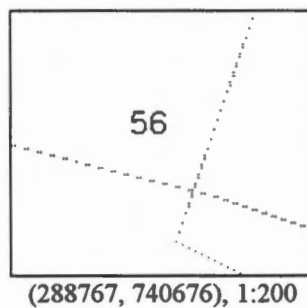


Feature Code :

Sounding Value:

Key :

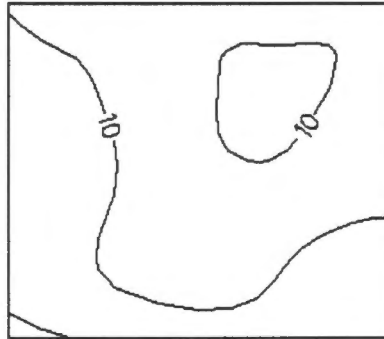
3. Place the sounding at the desired location.



Chapter 45

Add Contour Labels

Contour labels annotate the height of contours and are cartographic features. They are text strings placed along contours and oriented so that they can be read without turning the map.



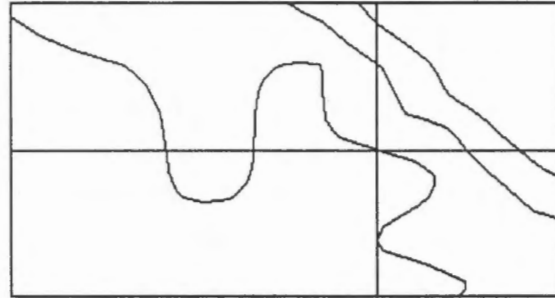
We could add contour labels as names through the following steps:

1. Retrieve the height of the contour to be labelled.
2. Identify on the contour the appropriate location for the label.
3. Add the label as a text string and orient it along the contour.
4. Open a gap on the contour around the label.

This method of adding contour labels is not only tedious but also leaves a gap on the contour, thereby affecting the completeness of the database. We will show in this chapter a better way of adding contour labels that does not leave permanent gaps on the contour. To make this possible, we need the help of specially designed data structures and editing operations.

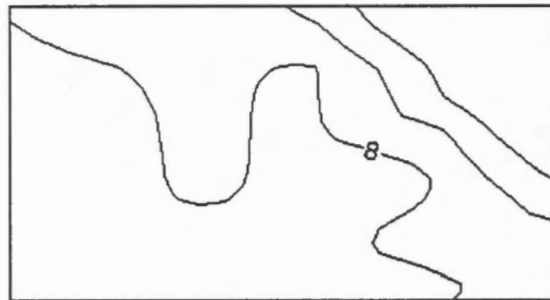
 Label a contour.

1. Display features with feature code CONTOUR.
2. Select the *Edit > Line > Change > Add Contour Label* menu item.
3. Place the crosshair at the spot you want the label to appear and click on the left mouse button.



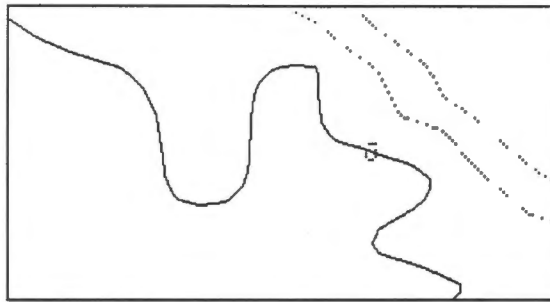
(289089, 740312), 1:1500

4. Confirm the contour when it is highlighted.
5. Enter a feature code of CONTOUR for its feature code.
6. When you are asked to enter its style/point size, just click on the OK button to select the default.
7. Enter a size of 10 mm for the contour labels.
8. Use the default line weight.
9. Enter 1 for the font.
10. Use the default margin around the contour labels.
11. Enter 1 for the contour interval. This will produce a label without decimal places. Entering 0.5 means that you want one decimal place for the label.
12. The contour label will be highlighted. Confirm it.
13. Refresh the display and you will see the label placed at the chosen location and oriented properly.

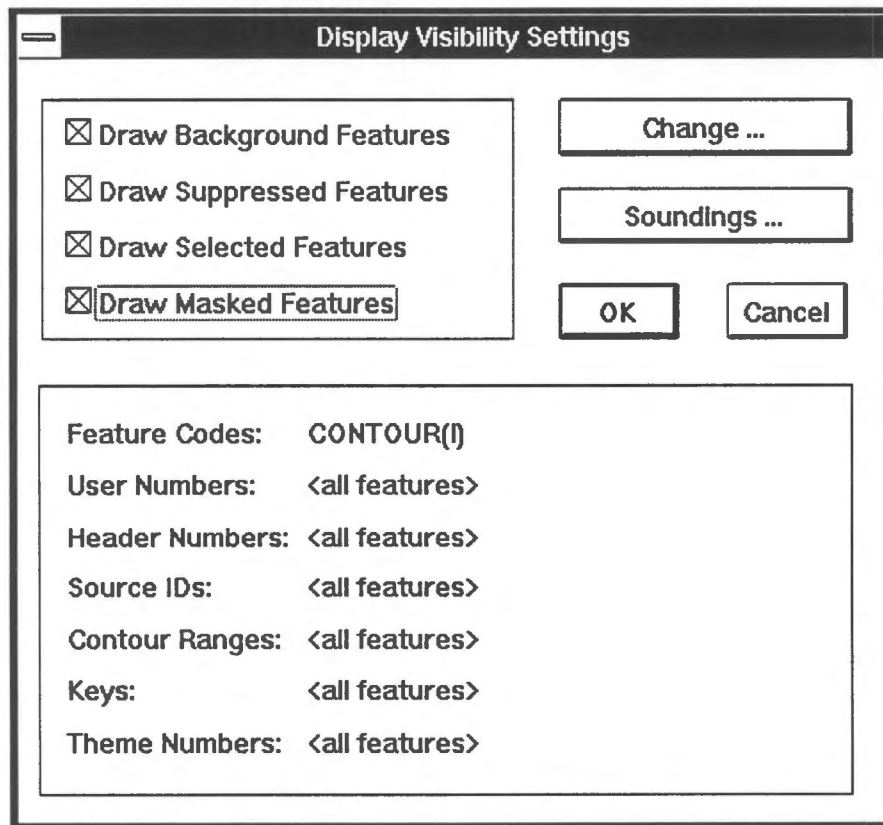


The previous exercise showed that you only need to point to the spot for the contour label, and the rest of the editing is performed automatically by the GIS. The system knows the elevation for the contour because it is stored as part of the contour data structure. The shape of the line around the chosen spot helps the system to determine the proper orientation for the label.

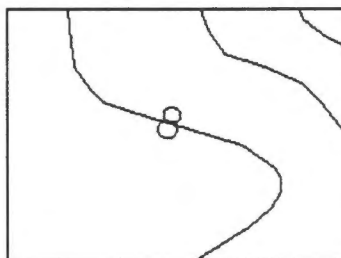
This GIS does not leave a permanent gap around the label. To confirm this, use the Information icon on the CARIS Panel to point at the label. You will see that it is still a continuous line. The gap was created by a temporary mask placed over that part of the contour.



Use the Options > Display > Visibility Parameters ... menu item to turn on the Draw Masked Features switch.



Refresh the display and you will see that the contour is drawn uninterrupted over the label.



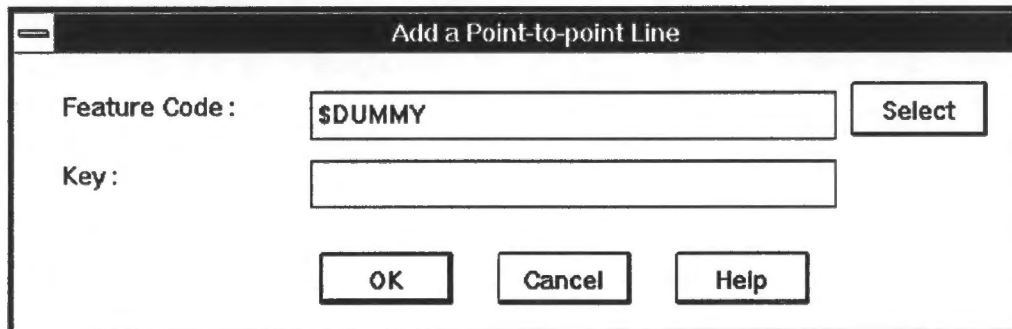
Chapter 46

Add Line Features

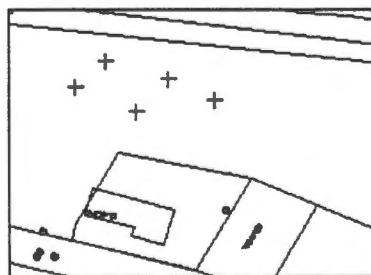
In this chapter, we will enter curve-fitted lines, point-to-point lines, parallel lines, and squared lines. Curve-fitted lines are defined at input by a small number of points, and the GIS will generate a smooth curve over these points to create the eventual line. They are suitable for smooth lines such as highways. Point-to-point lines are angular lines such as parcel boundaries. The operator must enter every one of its points. Parallel lines are generated at a distance from existing lines. Squared lines turn at 90° angles and are suitable for building outlines.

 Add a curve-fitted line.

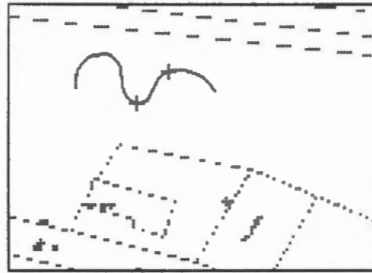
1. Display all features. Select the Edit > Line > Add > Fitted Curve menu item.
2. Enter the feature code. We will use \$DUMMY to show that we are not entering anything real.



3. Use the crosshair to enter five points on the curve-fitted line.



4. A smooth curve will be fitted over the five points. Click on the left mouse button to confirm.

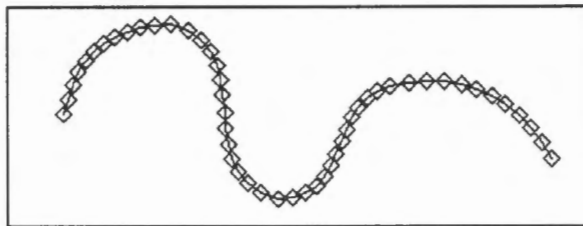


5. Refresh the screen afterwards to show the updated map. ♦

 The demonstration *Edit graphic data: adding a three-point curve* shows how to add a curve using three points.

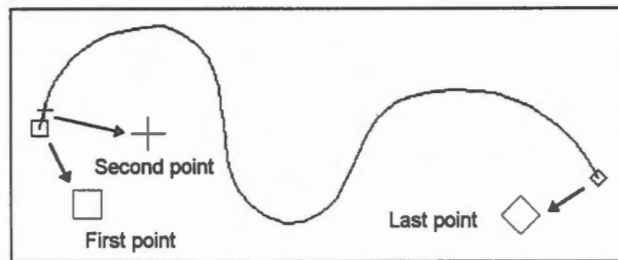
❓ Show the points on the curve-fitted line.

Use a command we have introduced before to show the points making up the curved line we have just created.



❓ Show direction of a line.

1. Select the *Edit > Line > Highlight Points* menu item and point at the line we have created.
2. Different symbols will be drawn over three points on the line to show the direction of the line.



 Add a point-to-point line.

Use the Edit > Line > Add > Point-to-Point menu item to add an angular line using the following three points:

	X	Y
Point 1	288511	740600
Point 2	288531	740600
Point 3	288562	740600

and highlight the points on the line. ♦



The demonstration Edit Graphic Data: Adding a Point-to-Point Line shows how to add a point-to-point line.


 Add a straight line as a curve-fitted line.

Use the add curve-fitted line command to add a straight line defined by the following three points:

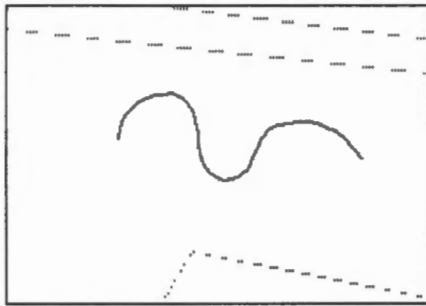
	X	Y
Point 1	288511	740590
Point 2	288531	740590
Point 3	288562	740590

Show the points on the line and see how different it is from the point-to-point line created in the previous question. Do you think this is a good method of generating straight lines ?

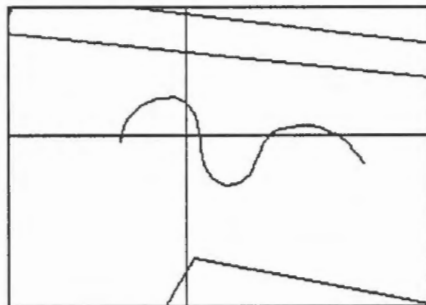
♦

 Add a parallel line.

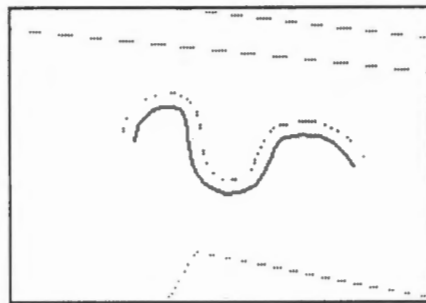
1. Select the Edit > Line > Add > Parallel menu item.
2. Enter a feature code of \$DUMMY.
3. Select the line we have previously added.



4. Enter a point on the parallel line.

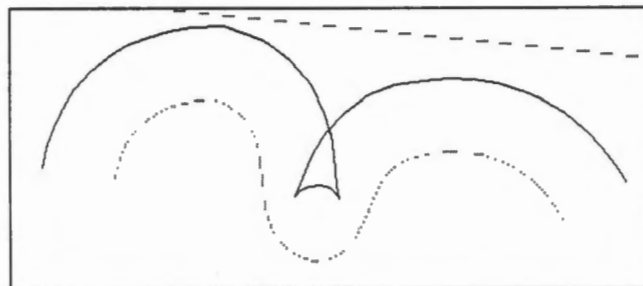


5. Confirm the parallel line.

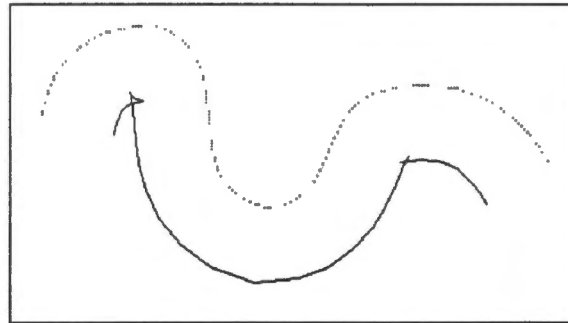


6. Refresh the screen. ♦

Wide parallel lines are not easy to generate automatically. We will first generate a wide parallel line above the existing line.



We will then generate a parallel line below it.

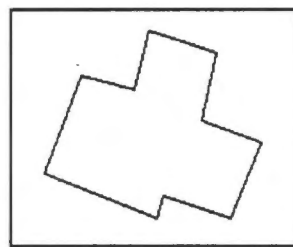


These examples show the difficulty of using vector representation to generate parallel lines. It is possible to implement a better parallel line generating program, but the calculations would be quite involved.

Buffer zone generation around a line is similar to parallel line generation. We have seen the creation of buffer zones in raster format, and they do not show these problems. It is easier to generate parallel lines in raster than in vector format. Because of this, some systems use a raster approach to generate parallel lines and then convert the raster result back to vector format.

◇ Generate parallel lines from point-to-point lines.

Generate parallel lines of various widths from this house and notice the problems you may encounter.



(287870, 740268), 1:400

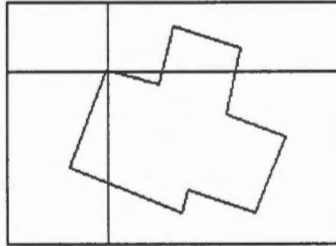


☞ Digitize a line with 90° turns.

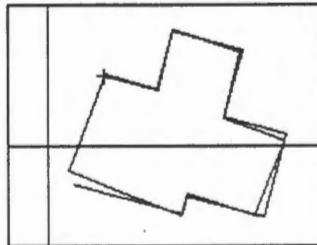
The outline of a house consists mostly of 90° turns. If the angles are not 90° as they are supposed to be, the outline is not logically consistent although it might be positionally accurate. We can correct this problem by forcing the turns to be 90° when digitizing the house.

In this example, we will digitize another house on top of the old one but with 90° turns.

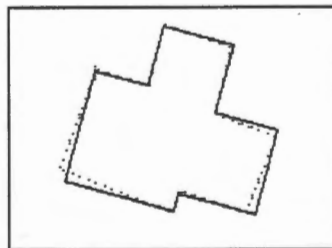
1. Select the *Edit > Line > Add > Squared* menu item.
2. Enter a feature code of *\$DUMMY*. We could have used the proper feature for a house, but we have picked a different one here so that the new outline draws with a different colour.
3. Digitize the first point over an existing corner. Use the search option to select the exact corner.



4. After you have entered the last point, press the *C* key on the keyboard to automatically close the house.



5. Confirm the new house outline.



❓ What is the effect of logical consistency on positional accuracy ?

In order to make positionally accurate data logically consistent, we have moved some of the points, thus affecting their accuracy. Measure the amount of shift in the corner points as a result of corner squaring. Do you think the amount of shift has caused some of the corners to be located more than 0.5 mm from their true locations at map scale ? We assume that the points on the original outline are accurate to within 0.5 mm. ♦

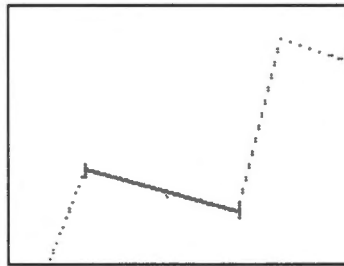
❓ Why does the choice of the first corner affect the shape of the squared house ?

Create other squared versions of the same house by starting with another corner. You will find out that the resulting houses are not identical in shape. Why ? ♦

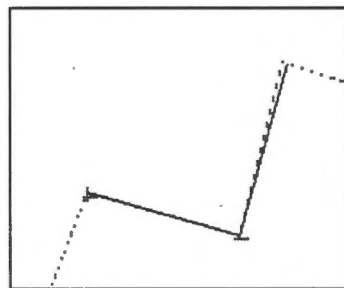


The method of squaring closed lines.

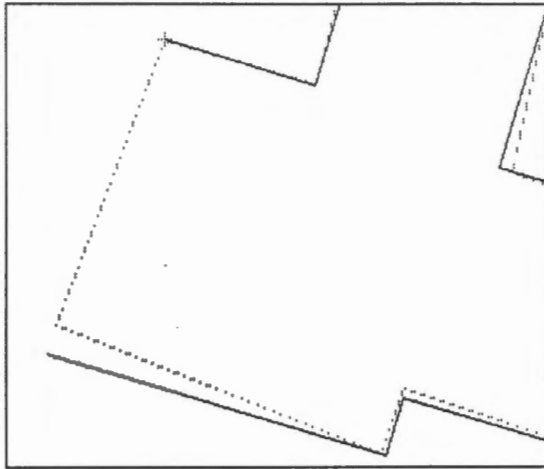
The process of squaring closed lines is not as simple as it may appear. After you have entered the first and second points, they are used to form the first segment.



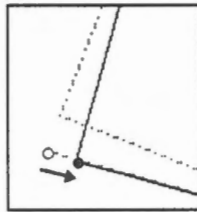
After you have entered the third point, the third corner will be adjusted to maintain a 90° turn from the first segment. Hence the new corner and the old one might not coincide. This shift will most likely happen over all subsequent points.



An interesting thing happens at the last point. If you simply join the last point to the first point, the two resulting turns might not be 90° .



Therefore, an adjustment is made to the last point to ensure the correct angles.



Chapter 47

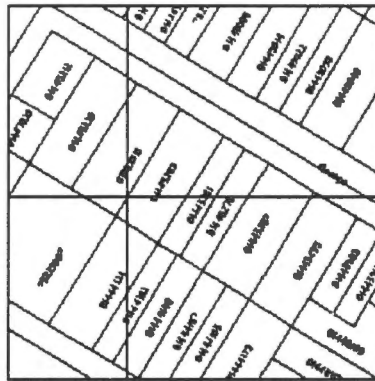
Topology and Editing

Editing affects topology when it changes the relationship among the geographic features. An arc that used to meet another arc could be clipped so that they no longer meet, a new arc could be added to split a polygon in two, and an arc could be deleted to merge two polygons into one. We will investigate the effect of editing on topology in this chapter.

 Highlight arcs around a node.

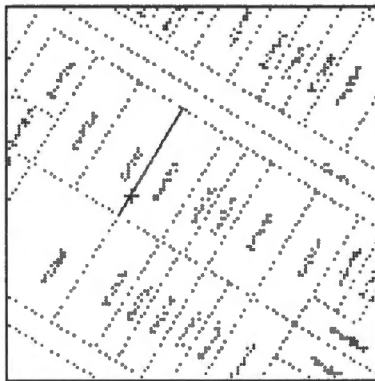
This is to show that the GIS has linked arcs around a node for quick retrieval.

1. Select the Edit > Topology > Cycle > Arcs Around Nodes menu item and point at this arc.



(288569, 740026), 1:1200

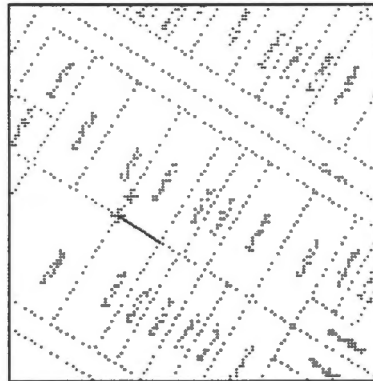
2. The arc will be highlighted. Click on the left mouse button to confirm.



3. The closer node will be highlighted. Click on the left mouse button to confirm.



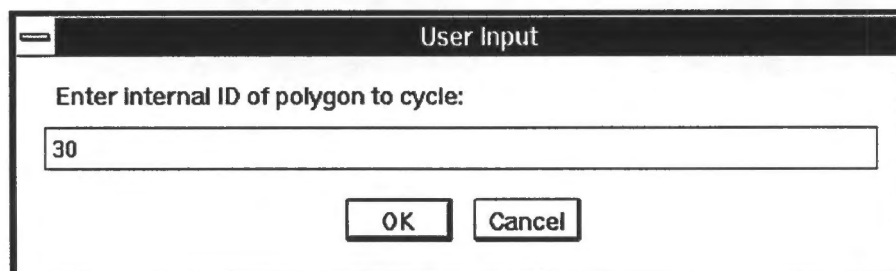
4. The next arc linked to the same node will be highlighted.



5. Clicking the left mouse button will highlight each of the other arcs linked to this node. Click on the right mouse button to quit from this command. ♦

 Highlight arcs around a polygon.

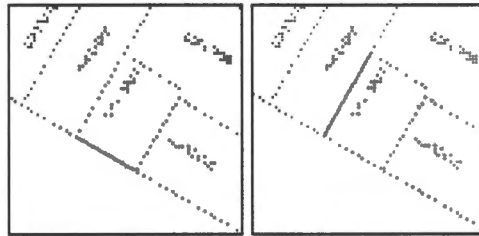
1. Select the Edit > Topology > Cycle > Arcs Around Polygon menu item.
2. Enter the internal identification (not the polygon key) of the polygon to cycle. The internal polygon identifications were assigned by the system sequentially when the polygons were created. The polygon key was assigned by the user instead. Although they are both used to identify polygons, the key is more meaningful to the user.



A dialog box titled "User Input" with a standard window control bar. The text inside reads "Enter internal ID of polygon to cycle:". Below this text is a text input field containing the number "30". At the bottom of the dialog are two buttons: "OK" and "Cancel".

NOTE The internal polygon identifications are not normally revealed to the user.

3. The first arc of the polygon will be highlighted. Clicking on the left mouse button will highlight in turn each of the arcs forming the polygon.

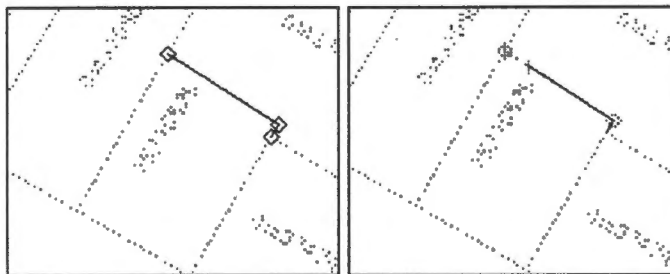


(288026, 740172), 1:800

4. Click on the right mouse button to quit. ♦

 Change the geometry and see its effect on the topology.

1. Use the Edit > Line > Move Points in Line menu item to move an end point of this arc away from its original position.



2. Use the Edit > Topology > Cycle > Arcs Around Nodes menu item to cycle around this changed arc.

You will notice that this arc has been topologically detached from the other arcs since its geometry has changed. The other unchanged arcs are still topologically connected.

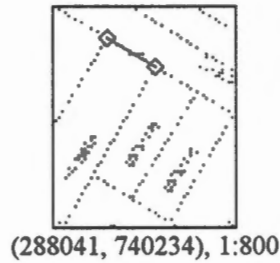
♦

In general, changes in geometry have potential effects on topology. There are two approaches to correcting topological problems after editing. One is for the editor to make these corrections in real-time after each change. The other is to let a separate module make the corrections in batch mode after you have done all the changes. This GIS uses the second approach, and it keeps a record of the topological changes made to the database. This information will be passed to the topology builder to reconstruct the topology only for the affected features.

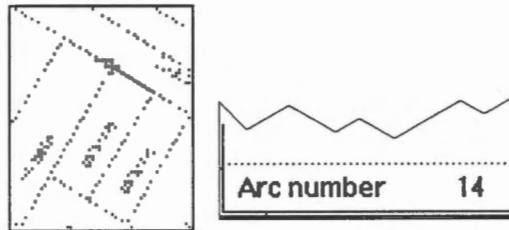
 Topological data structure for this GIS.

We have described a simple topological data structure in the chapter on Topological Relationships. That structure shows whether two polygons are neighbours, but it does not show if arcs connect. We will show here how that structure can be extended to include connectivity.

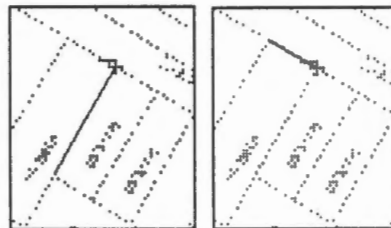
Cycle around the right-hand node of this arc.



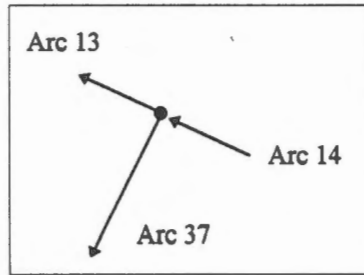
When the *next* arc is highlighted, its internal number 14 will be shown on the message bar at the bottom of the window.



The next arcs after that have internal numbers -37 and -13 respectively. A negative number meaning that the arc is directed away from the common node.



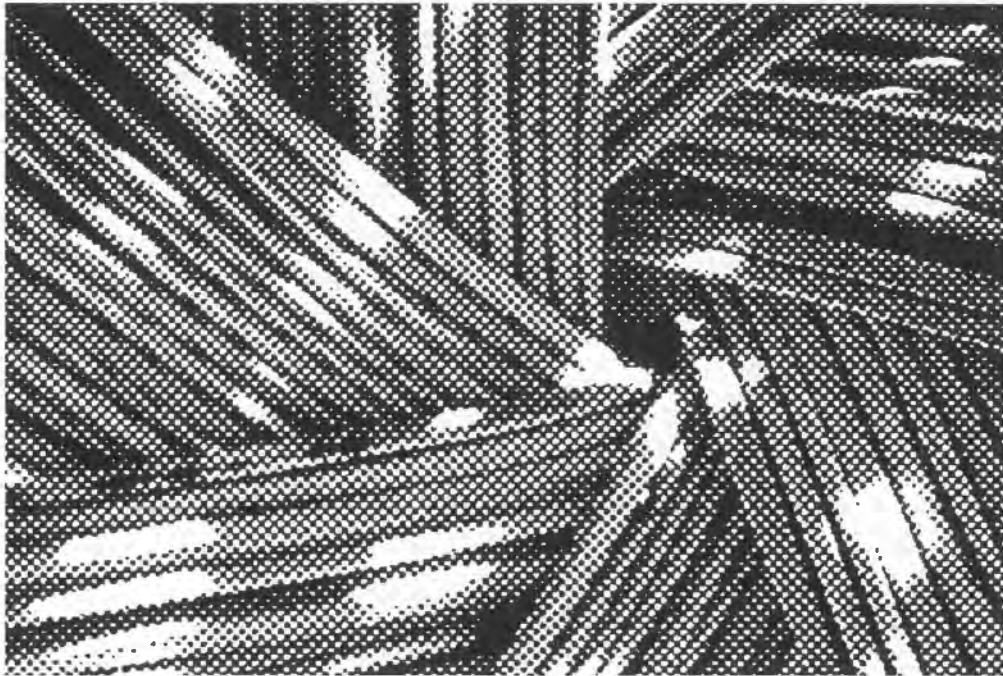
The three arcs are related to each other in the following way.



In order to facilitate the retrieval of all arcs sharing a common node, this GIS links each arc to two others in the clockwise sense. In this example, there is one node with two arcs directed away from it and one towards it. This topological relationship is captured in the following table forming part of the topological structure.

Arc	Next arc at beginning	Next arc at end	Left polygon	Right polygon
13	14	--	--	--
14	--	-37	--	--
...				
37	-13	--	--	--

Part 8
Data Sharing



Data Conversion and Integration

The task of integrating data from different sources is both important and difficult. It is important because the increasing cost of data collection accentuates the need to share data among users. Even in an organization that does not rely heavily on external data, internal data specifications and organizations could change, making it necessary to integrate new data with the old.

One of the advantages of a GIS is to allow integration of data into a single database. However, there are several factors that could make this a challenging task. In this chapter, we will address several of these factors including differences in data structure, projection, resolution, accuracy, graphic quality, interpretation, currency, and attributes.

 Prepare the FTON file for the exercises.

To ensure that you see the proper features on the FTON file in these exercises, do the following.

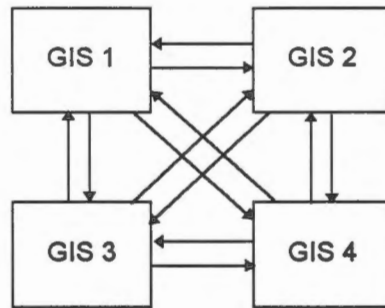
1. Open the FTON file and use the Options > Display > Visibility Parameters ... menu item to exclude contours from the display.
2. Exit from the Editor. ♦

Difference in Data Structure

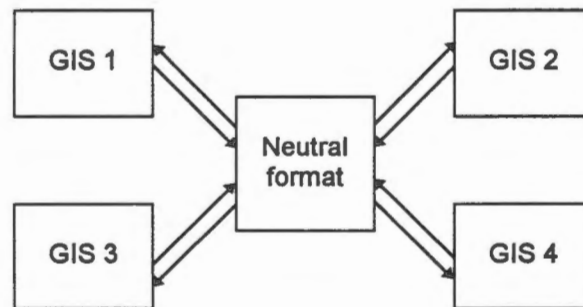
Different GIS could use different formats to store the same geographic feature. For example, one of them might use a vector structure and the other a raster structure. Converting data from vector to raster is simple, but the reverse conversion is complicated.

Converting between different raster formats is not as difficult as converting between different vector formats. This is because vector formats are more complicated, contain more information, and hence have more variations. Depending on the nature and magnitude of the differences, the conversion of data from one vector structure to another is not always easy.

There are two approaches to converting data between different GIS. One is to use a direct translator that converts one GIS structure to another. When there are n different systems involved, we need $n(n-1)$ one-way translators to cover all possible data conversions among them. In this example involving four different GIS, we need 12 such translators.



The other approach is to let all translations go through a neutral format, requiring only $2n$ one-way translators.



The first approach uses direct translators but there are more of them. This is usually not a problem until one of the GIS structures changes, affecting $2(n-1)$ translators. The second approach involves fewer translators but changing one of the structures affects only two translators. However, all translators are indirect because they must go through a neutral format. Hence the design of the neutral format has significant effects on data conversion.

In theory, the neutral format has to be very flexible and accommodating. A restrictive format, such as one not allowing points to have three coordinates, would cause data degradation during the conversion. But a format that is too flexible tends to be complicated and difficult to use. In reality, several of these neutral formats have been developed, each catering to a group of users with similar requirements.

Most GIS would provide translators to and from a number of these neutral formats. When there is a definite need to convert large volumes of data efficiently from one particular GIS to another, a direct translator between them will be written.

In this chapter, we use a digital topographic map produced by Geomatics Canada that was imported to this GIS through an interchange format called DXF. DXF was originally designed to exchange Computer-Aided Design (CAD) data between AutoCAD applications. As the use of AutoCAD for geographic data capture and editing increases, DXF becomes one of the standards of exchanging data between GIS.

Geomatics Canada used a translator to export data from the National Topographic Database (NTDB) to DXF, and we have used another translator to import data from DXF to the CARIS format. Because DXF was not specifically designed for geographic data, there is some loss of information during the translation process. For instance, DXF does not recognize topology and therefore no topological information is passed to us.

Difference in Scale

The NTDB file was digitized from 1:50 000 topographic maps, while the database we have used so far is at a scale of 1:1000. Note that the scale of a GIS database is mainly used as a general guide to its accuracy and levels of detail, and was derived from the scales of its input documents.

Difference in Resolution

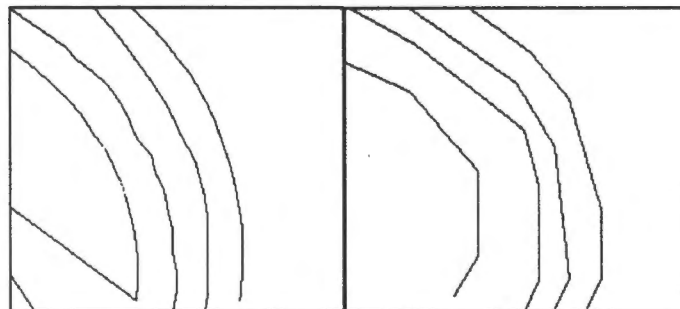
This NTDB file was created for a much smaller scale mapping with less accurate positional data. Consequently, the file is based on a coarser resolution.

❖ Determine the resolution of this file.

Select the CARIS Editor icon and open the file NTDB_UTM. Use a method described earlier to verify that this file has a resolution of 2.5 metres on the ground. ♦

The resolution has a direct effect on the cartographic quality of the lines as a coarser resolution produces coarser lines. In this example, the left diagram is from FT0N, the database of finer resolution we have been using so far. The right diagram is from a corresponding area from NTDB_UTM where the lines are obviously more angular in appearance.

NOTE The coordinates are different because of the difference in map projections.



FT0N (289285, 740236)

NTDB_UTM (682977, 5092557), 1:1000

When we copy the features from NTDB_UTM to FTON, the features are moved to a file of finer resolution. This increase in resolution will not improve the cartographic quality of the lines unless we apply a smoothing operation on them after the move. It is important to note that cartographic smoothing of lines does not increase their positional accuracy, although a smooth line could convey a false impression of higher positional accuracy.

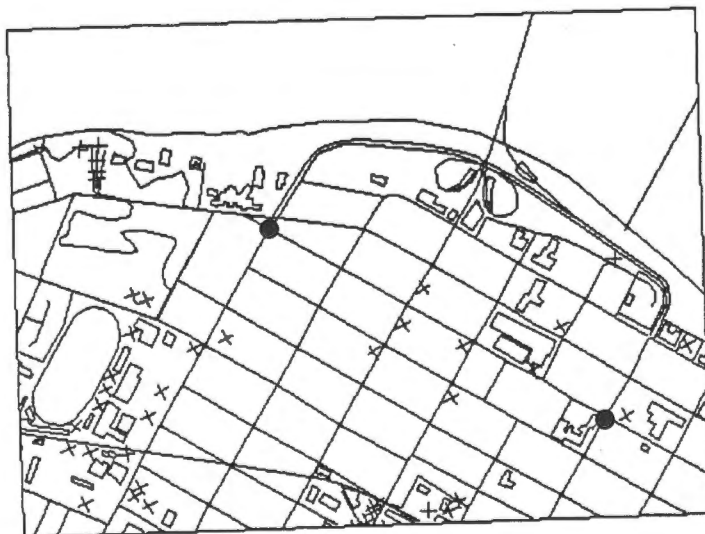
Difference in Map Projection

The NTDB files are based on the Universal Transverse Mercator (UTM) projection, and it distorts the map in a somewhat different way than the stereographic projection used for the FTON database. We will illustrate the following differences.

- The ground coordinates are different from one projection to another while the geographic coordinates are similar. Although geographic coordinates are better for large databases and are more accurate, they are less popular for GIS. This is probably because the early GIS rely heavily on paper maps for data. Consequently, integrating data from another GIS database in ground coordinates requires map projection changes.
- Different projections distort distances in different ways. Because of the small area covered in this file, such differences are small. Nevertheless, even small differences can cause misfits of data that must be corrected later.

 Measure distance between two points.

1. Click on the CARIS Editor icon and open the file NTDB_UTM.
2. Select the Query > Distance between 2 Points menu item to measure the distance between the two points marked below. Use the search option to point exactly at the intersections.



3. The following are the coordinates of the two points.

Coordinates of a Point			
X: 272781	XG: 681952.50	LAT: 45-57-50.432N	<input type="button" value="OK"/>
Y: 2037092	YG: 5092730.00	LONG: 66-39-06.028W	
	ZG:		

Coordinates of a Point			
X: 273140	XG: 682850.00	LAT: 45-57-33.224N	<input type="button" value="OK"/>
Y: 2036890	YG: 5092225.00	LONG: 66-38-25.058W	
	ZG:		

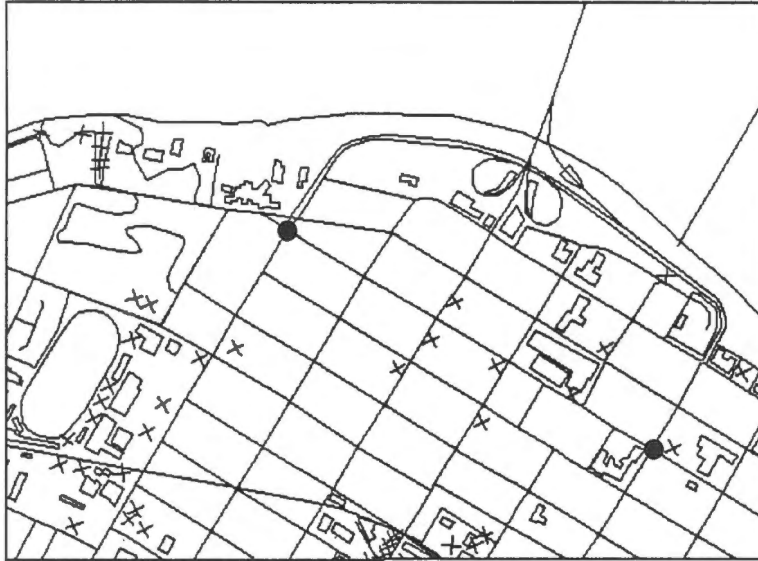
You will notice that the ground coordinates are within a very different range from what we have seen before. However, the geographic coordinates are similar as we will show later.

4. The distances between the two points are 1029.821 metres on the flat map and 1029.812 metres on the ellipsoid. ♦

We have converted the NTDB file from the UTM projection to the stereographic projection so that it agrees with the one we have used.

 Measure distance between the same two points on a different projection.

1. Select the File > Open ... menu item to close the current file and open the file NTDB_STE.
2. Measure the distance between the same two points. Use the search option to point exactly at the intersections.



3. The following are the coordinates of the two points.

Coordinates of a Point			
X: 5764879	XG: 288243.95	LAT: 45-57-50.433N	OK
Y: 14808751	YG: 740437.55	LONG: 66-39-06.029W	
	ZG:		

Coordinates of a Point			
X: 5782502	XG: 289125.10	LAT: 45-57-33.225N	OK
Y: 14798093	YG: 739904.65	LONG: 66-38-25.059W	
	ZG:		

4. The ground coordinates are familiar to what we have seen before. In theory, the geographic coordinates should be identical to those obtained from the UTM version of this file. The minor differences are due to computational errors.
5. The distances computed are 1029.761 metres on a flat map and 1029.828 metres on the ellipsoid. Notice that this reduction in ground distance is a characteristic of this map projection as we have observed before. ♦

The following summarizes the differences.

Left point	Latitude	Longitude
NTDB_UTM	45-57-50.432	66-39-06.028
NTDB_STE	45-57-50.433	66-39-06.029

Right point	Latitude	Longitude
NTDB_UTM	45-57-33.224	66-38-25.058
NTDB_STE	45-57-33.225	66-38-25.059

	Ground distance	Geodetic distance
NTDB_UTM	1029.821m	1029.812m
NTDB_STE	1029.761m	1029.828m

The UTM and stereographic versions of the NTDB file are identical except for the map projection. Their geometric differences are due mainly to projection distortions and are minor.

❓ Determine resolution of this file.

When converting from the UTM to the stereographic projection, we have also changed the resolution of the file to conform to that of FTON. Verify that the resolution of this file is 0.05 metre on the ground. ♦

Difference in Accuracy

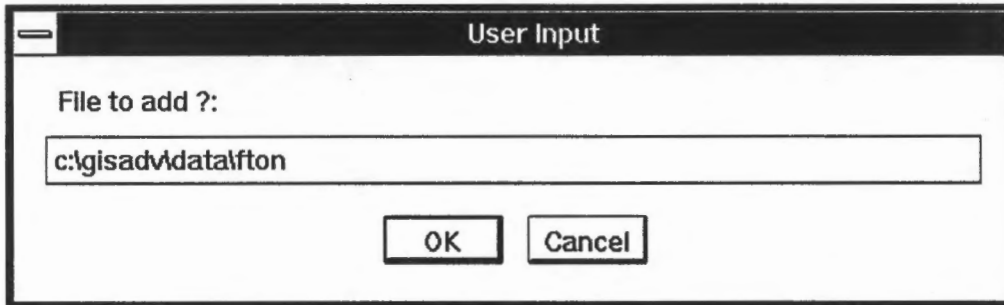
The scale of a database is related to the level of detail shown and the accuracy of the coordinates. In this exercise, we will first open the NTDB_STE file and superimpose the FTON file over it to illustrate the differences.

After incorporating data from another source, it is often necessary to decide which contains more accurate data. The factors we can consider are: currency of the survey, accuracy of the surveying method, scale of the source document, and resolution of the database. In case of doubt, a ground survey can be conducted to verify the data.

📏 Measure positional accuracy on street intersections.

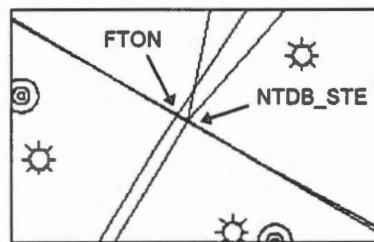
1. If the NTDB_STE file is not yet open, open it.
2. Select the Macro > Load menu item to load the file FTON. This will show both files on the screen, but you are allowed to access features only on NTDB_STE.

NOTE The location of your FTON file could be different from what is shown here.

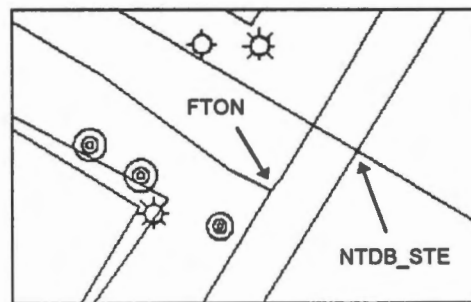


3. Refresh the screen, and you will see both maps on the screen.
4. Centre at the left point of the street we have seen before. You will notice that there is a gap of about 1.3 metres between the same point from the two files.

NOTE You cannot search or edit features on the FTON file.



5. A wider gap of about 8.9 metres occurs between the right points from the two files.



◇ Measure length of street.

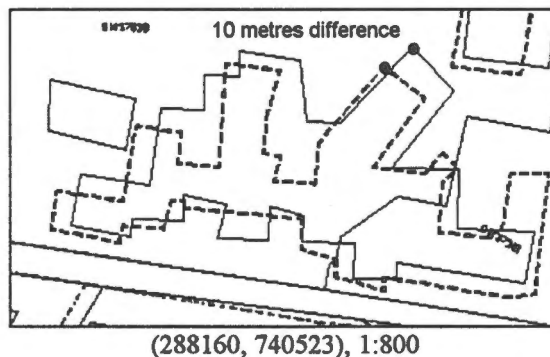
We have measured the length of the street from the left to the right point in a previous exercise. Measure the length of this street on FTON. Their coordinates are:

left point (288243, 740438)
right point (289117, 739901). ◆

📏 Measure positional accuracy on house corners.

1. A display of these buildings from the two files shows that their corners do not match, and a difference of 10 metres is noted at one of the corners.

NOTE We have shown the buildings from the FTON file in dashed lines to distinguish them from those of the NTDB file. On the screen, the two lines are of different colours.

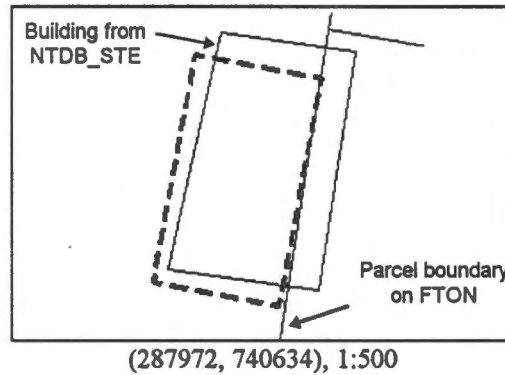


The expected positional accuracy of well-defined features on a 1:50 000 map is 25 metres on the ground (0.5 mm at map scale). Disregarding other differences due to interpretation and up-to-dateness, the house corners from the NTDB file are within the accuracy specification.

◇ Measure positional accuracy on other features.

Although the topographic map positional accuracy specifications do not cover the less definite features, such as coastlines, it is still useful to determine the accuracy of these features. ◆

Features rarely exist in isolation. When a feature of lower accuracy is copied to another database, a topological error could happen. In this example, a building from NTDB_STE crosses the parcel boundary on FTON. The correct location of that building, as shown on the FTON database, is situated entirely on one parcel.

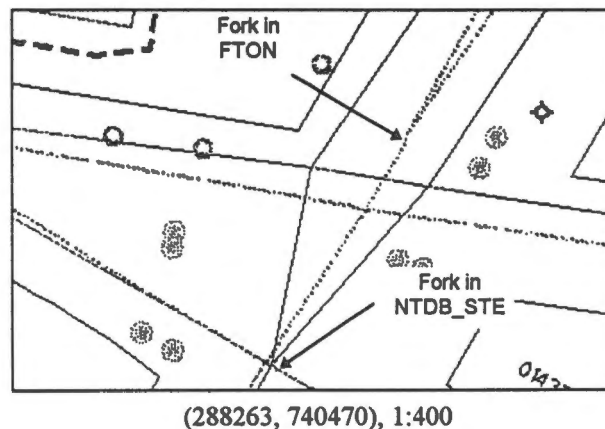


Difference in Data Collection Convention

Both the NTDB and the FTON databases cover buildings, but the shape and number of houses are different. The same applies to the centre lines of the streets. Other than positional accuracy, the following are additional factors that contribute to the differences.

Different criteria for feature selection: Features are included or omitted according to the specifications designed for the mapping project. For example, a building that appears in a larger scale database could be omitted in a smaller scale database because its size is too small to show.

Different interpretation of features: The outlines of the features are traced by operators when collecting data. Different operators could interpret features in different ways, thus affecting their shapes. In the following example, an undivided road changes to a divided one, causing a single line to fork. Different interpretations on where the change occurs places the fork at different locations.

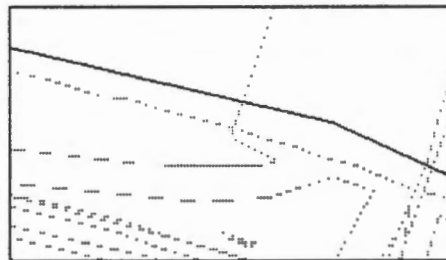


Difference in Data Currency

Geographic features change with time. Take buildings as an example. New buildings might be added, old buildings might be demolished, and the shapes of existing buildings might be altered. Hence two databases of the same features might not agree.

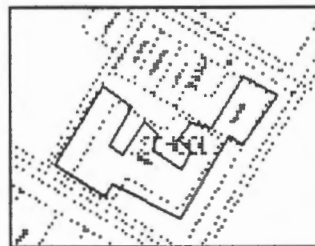
Difference in Attributes

Even when the same features are captured in the databases, the attributes associated with them might not be identical, and one may contain more attributes. For instance, the FTON database uses alphanumeric feature codes and the NTDB_STE uses numeric ones. Hence the river bank in the FTON database has a feature code of RIVER, whereas its feature code in the NTDB_STE file is 1452-1011 meaning "other waterbodies".



(288774, 740680), 1:1200

The classification of features in one database might be different in the other. This building has a feature code of BUILDING in the FTON database, but it is given a feature code of 160-1001 in the NTDB_STE indicating that it is a school building. Note, however, the fact that this is a school building is carried in the user-defined attributes of the FTON database.



(289032, 739865), 1:1800

Another example of a difference in feature classification is given by the highways. The highway network captured in the FTON database is classified either as STREET or ROAD. That in the NTDB_STE is classified much finer. The segment highlighted has a feature code of 1003-1001 indicating that it is an all season road with hard top, undivided, with more than two lanes, and so on.



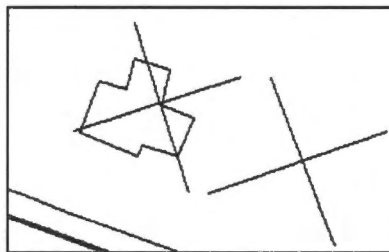
(289067, 740010), 1:1800

◆ Investigate other differences in feature classification.

Refer to the list of feature codes in Appendix B to investigate other differences in feature classification. ◆

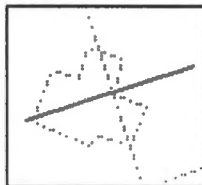
Difference in Geometric Representation

The same feature could be represented differently in two databases. This example shows that a building represented by a closed loop in FTON is represented by a symbolized point in NTDB_STE. The “symbol” has been given a feature code of 147-1003 meaning “other buildings”.



(287870, 740270), 1:800

Although the crosses appear like symbolized point features, they are not really stored as such. If you display information on them, you will notice that they are actually lines.



What has happened is that these crosses were symbols in the originating database, but the data translation program has decomposed each of the symbols into a series of lines.

Chapter 49

Transformations

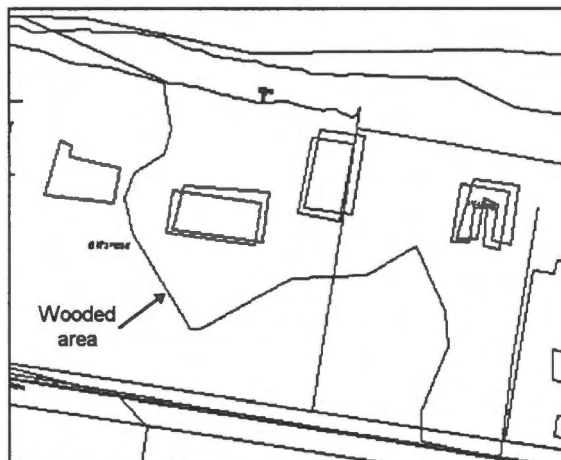
Other than the move (or shift) that we have seen in the previous chapter, geometrical transformations also involve scaling, rotation, and other changes.

One reason for transforming features is to correct geometrical distortions. If a house has been entered into the database at a wrong location, we can shift it to the proper position. If the scale and orientation are also wrong, then more involved transformations will have to be applied. This kind of geometrical correction is particularly useful when integrating data from other sources of different quality.

In performing any transformation, we need control points. We have explained earlier that a control point is used to relate two coordinate systems, and it has known positions in both systems. Control points can also be used to transform features from one position or shape to another in the same coordinate system. In either case, the idea is the same — if we know the before and after positions of some critical points of a feature, then the changes in shape or position can be determined.

We will use geometrical transformations to show how to fit features from one database to another. Let us suppose we would like to copy the outline of the wooded area from the NTDB file to FTON. The FTON database, being of larger scale and higher resolution, can be regarded as more accurate.

We notice the difference in location of the same building in the two databases. We will assume that all other features, including the wooded area outline, are distorted in a similar manner. The correct building corners (from FTON) being more accurate and the distorted ones (from NTDB) can serve as control points to compute the type and magnitude of distortion. We can then apply the same transformation to the wooded area outline to make it conform better to the FTON database.



We will do the transformation in several ways, using a different number of control points (house corners) each time. In the first case, only one control point is used, allowing us to shift the features relative to each other. In the second case, two control points are used, allowing us to correct for differences in orientation and scale as well. In the third case, three control points are used. This will correct for shift, orientation, and different scale changes in the x and y directions.

☞ Prepare the FTON file for the exercises.

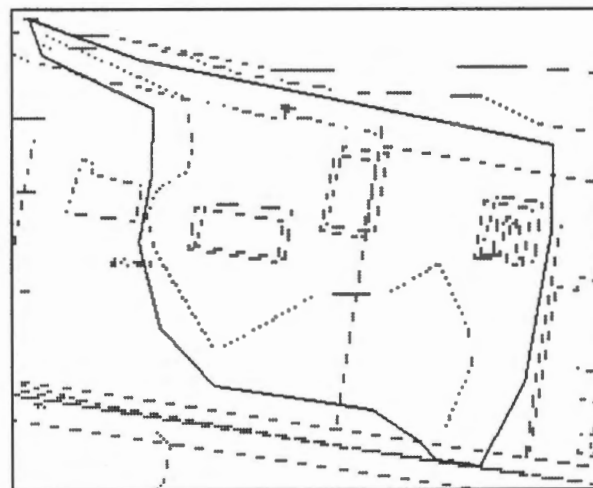
To ensure that you see the proper features on the FTON file in these exercises, do the following.

1. Open the FTON file and use the Options > Display > Visibility Parameters ... menu item to exclude contours from the display.
2. Exit from the editor. ♦

☞ Use one control point to transform.

We will use the geometrical transformation command to shift the features on NTDB_STE to make them fit better with those on FTON.

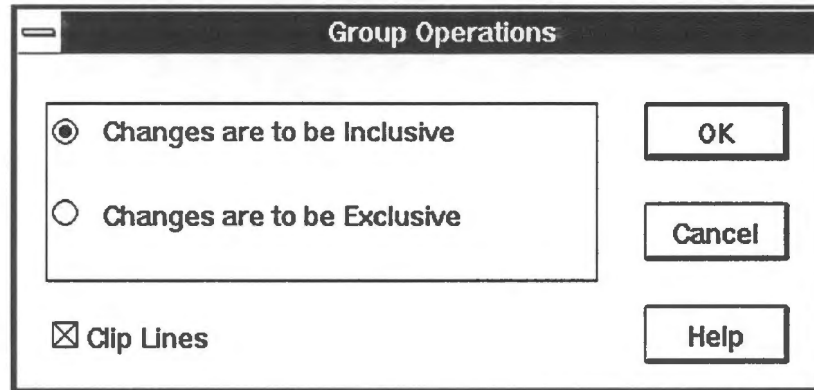
1. Open the file NTDB_STE.
2. Use the Macro > Load menu item to load the file FTON.
3. Use the Group > Define Group menu item to form a group containing buildings and the wooded area outline we want to copy.



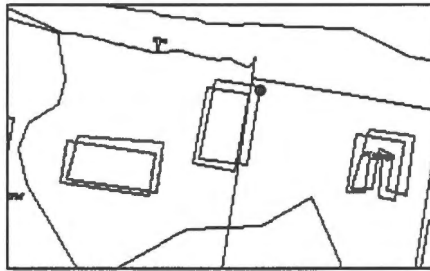
(287980, 740600), 1:1500

4. Use the Group > Geometric Trans menu item to perform a geometric transformation.

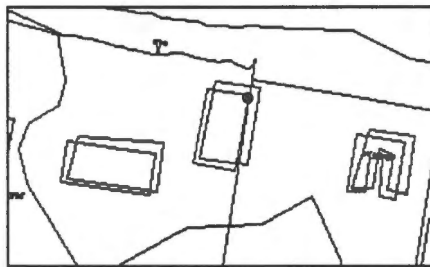
5. Select the inside features to be transformed, turn on Clip Line, and click on OK.



6. The features to be transformed will be highlighted. Click on the left mouse button to confirm.
7. Select one of the corners as a control. You might want to use the search option to point precisely at the corner.

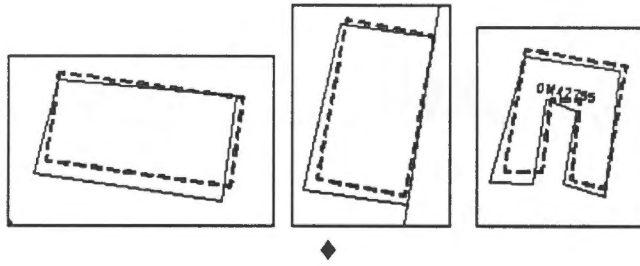


8. Point to the correct location of this corner.



9. You can enter more than one control points. In this example, we are using only one. Click on the right mouse button to finish entering the control points.
10. The transformed data will be highlighted. Press the left mouse button to confirm.

The differences in position of the three buildings affected give an indication of the effectiveness of the transformation. Notice that the middle building which previously crossed the parcel boundary has now been moved back to the proper parcel.



◆ Use other control points to correct for the shift.

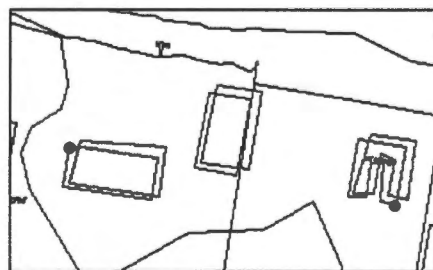
The use of a different control point would produce different amounts of shift. Try this using the other control points. Before you do this, exit from the current edit session and open the files again to return to the original file.

NOTE If you are using a production version of CARIS, do not save the edit when closing the editor.

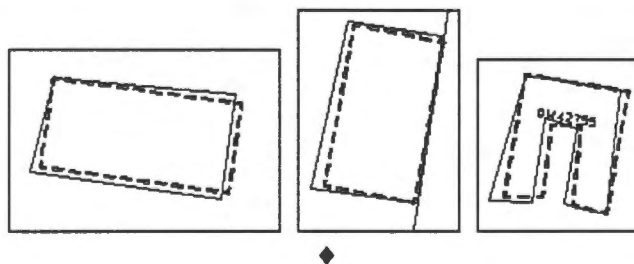


☞ Use two control points to transform.

1. Open the NTDB_STE file and load the FTON file.
2. Use the following two control points to transform.

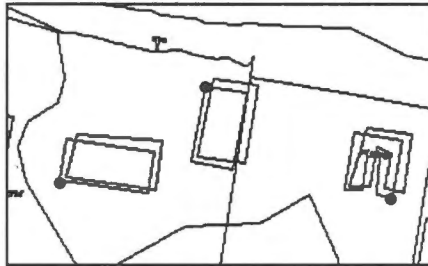


3. Comparing the building corners shows that they conform better to each other.

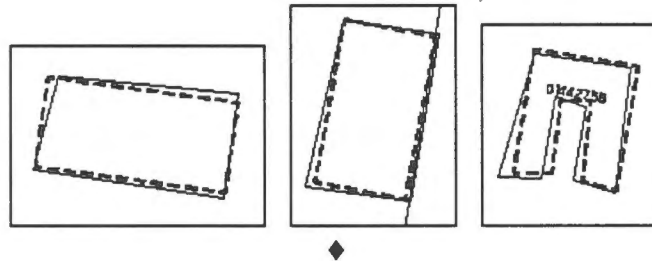


 Use three control points to transform.

1. Open the NTDB_STE file and load the FTON file.
2. Use the following three control points to transform.



3. The building corners are in even better agreement after the transformation..




NOTE The distribution of the control points also affects the transformation. Choose points that are distributed evenly over the area to be transformed, and avoid using control points that fall in a straight line.

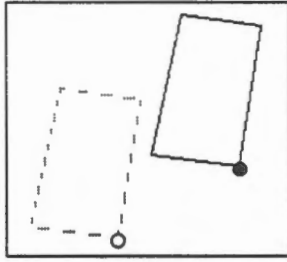
 Use more than three control points to transform.

If we use more than three control points, we cannot ensure that all points match after the transformation. However, some points are guaranteed to match. What are these points?

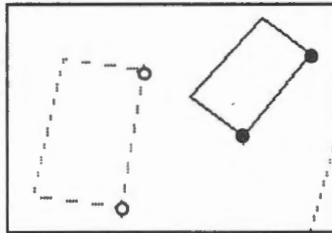


 The shift, similarity, and affine transformations.

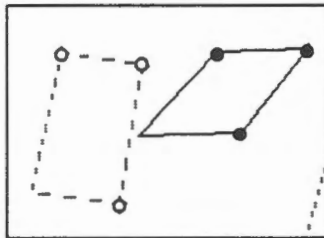
The use of one, two, or three control points to transform produces three different types of geometrical transformations. The first type, called *shift transformation*, uses one control point. If we know the correct position of a control point and its shifted position, we can compute from them the amount of shift. This amount of shift can then be applied to correct the other features.



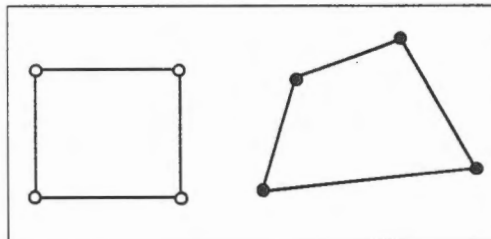
The second type, called a *similarity transformation*, uses two control points. This allows you to control the position, orientation, and size of the transformed result. Note, however, that the shape of the transformed object is similar to that of the original one, hence the name, similarity transformation.



The third type, called *affine transformation*, uses three control points. In addition to the shift, rotation, and scaling mentioned earlier, this transformation allows a different scale in x and y directions. Consequently, the shape of the transformed object has been modified. However, parallel lines will be preserved as parallel lines.



The affine transformation is more general than the similarity transformation. The similarity transformation in turn is more general than the shift transformation. There are even more general transformations than the affine. One of them is the *projective transformation* that uses four control points. It can be used to perform transformations like this.



Part 9
Wrapping Up



Chapter 50

Create a Geographic Database

In this chapter, we will review the process of creating a geographic database. We will base our discussions on the municipality database and the GIS software we have been using.

Define Data Requirements

A geographic database covers a certain area and serves a specific purpose. Some are used to manage forest stands in a country, some to aid geological explorations within a region, and some to help highway construction for a city. The municipality database we have been using supports land use planning, property records, and network analysis for part of the City of Fredericton.

In order to better define the contents of the database, we first consult our potential users and then consolidate their requirements for data into a single design often called a *database schema*.

The process of developing a database schema for a certain application is called *database design*, and it is carried out at several levels of detail. A high-level design for our municipality database reveals that we need the following types of geographic features:

- Properties each corresponding to a land parcel.
- Street centre lines for network analysis.
- Trees, street lights, utility poles, and fire hydrants.
- Contours for flood analysis.

As for attributes, we decided that they are required only for land parcels. Again, input from the potential users will be sought to create the attribute list as shown in Appendix C.

Such a design has to be implemented with a GIS. But before we can do this, we have to identify the sources of data because the scale, projection, and other characteristics could affect the implementation plan.

Identify Sources of Data

Data for a geographic database can be obtained in various ways. You can gather data from ground surveys, digitize them from existing maps, or transfer them from other systems. A review of the data collection methods is given in the chapter on Data and Its Collection.

For the municipality database, we transferred most of the data from a provincial properties database. The contour lines were interpolated from a digital elevation model. The floor plan of the school was converted from a file produced by AutoCAD, a Computer-Aided Design (CAD) software. We have also obtained digital data from the National Topographic Data Base (NTDB) maintained by Geomatics Canada.

Implement Database Using a GIS

Next, we will determine how to implement a geographic database using a particular GIS. Issues related to the graphics are map projection, resolution, and symbology.

For a municipality database where the geographic extent of the project area is small, and the requirement for accuracy is moderate, a wide selection of map projections is available. A reasonable choice is one that is used by the local mapping authority because more maps would be available in that form.

If you obtain data from other sources or digitize them from existing maps, you would have been given the projection, resolution, and other geometric characteristics of data. You could retain the original state of the imported data unless there is a need to convert them for conformance with existing data.

Resolution should be chosen to accommodate the most accurate data required for the project. Symbology for most of the regional and national series of maps has been standardized. It is possible either to acquire a digital version of the symbology or to create one from the paper maps.

A good organization of the database groups features into layers using one or more parameters in which feature code is one of the most popular. A *feature code*, being a classification code for the feature, is a compact representation of the feature attributes. Several standards on feature coding have been developed nationally and internationally. They should be adopted as much as possible.

The decision on what groups of features form a topological layer depends on the immediate and future needs. Features belonging to a topological layer have to be decomposed into nodes, arcs, and polygons. Putting too many features into one layer provides more topological information but creates significant storage overhead. The tendency is to create as many layers as possible, and let the user perform an overlay analysis to create cross-layer topology when needed.

Import Data From Other Sources

After you have identified the source of data, the next step is to convert the data from their format to yours. In the fortunate case that you are both using the same GIS software, this might involve just the transfer of files. When different computer hardware or operating systems are involved, binary files might not be compatible, and you must first convert them to ASCII.

In most cases, the formats are different and a translation program would have to be used. Unless a direct translator already exists between the two GIS, you would have to choose a neutral data interchange format.

The choice of the neutral format depends on the translators available on both systems and the information you need to preserve through the translation. Some of the neutral formats do not support topology, some of them do not recognize symbology, and some of them do not support curve-binding text. If you are not sure about the behaviour of a certain translator, test it out on a small sample of data first.

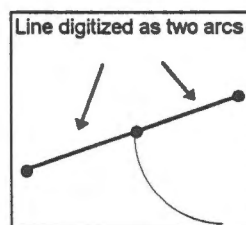
The municipality data are stored in a similar GIS system and hence the translation was trivial. Property attributes were imported, in the form of text files, from the regional land records office. The DXF data interchange format was used to convert the NTDB data from Geomatics Canada and the floor plan from the School Board.

Digitize Your Own Data

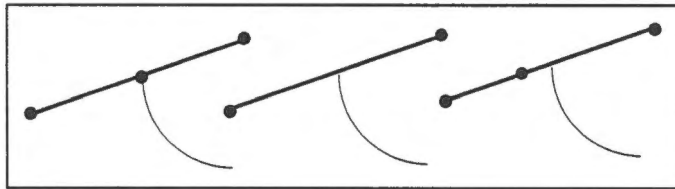
If the data you want are not available from other sources in digital form, you would have to collect them yourself. If they are available on existing maps, you can digitize them manually or scan them automatically. Automatic scanning would not be appropriate in our case because of the amount of processing required to convert them to the vector format expected by the GIS.

In our case, no additional digitizing was required. But the parcel boundaries we imported were originally digitized from existing maps.

There are two methods of digitizing them. The method of *arc digitizing* starts and stops lines at intersections. Hence the line in the following diagram will be digitized in two pieces.



The other method is called *spaghetti digitizing* that does not require lines to stop at intersections. Any of the following ways of digitizing the line (and other variations) are acceptable using this method.



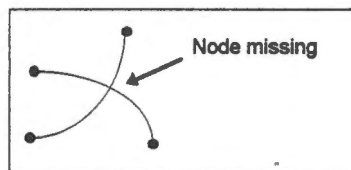
At the same time the parcel boundaries are digitized, the parcel identification numbers are entered as display labels inside the polygon. The labels and the boundaries are given the same theme number to indicate that they are topologically related.

Prepare Data For Topology Building

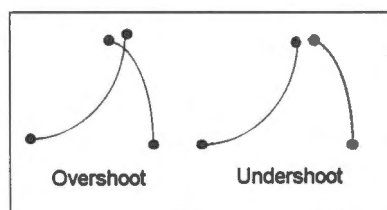
We need topological information to support most analysis. Topology can always be implied from the geometric representation of features, but explicit topological information saves processing time. Digitized or imported data generally do not contain such information, and special software, called the topology builder, has to be used to derive topology.

The task of the *topology builder* is to take features in vector form, derive topological relationships among them, and store these relationships in the database. In order for the topology builder to work properly, the input data must be *topology ready*, meaning that their geometry is clean enough to avoid topological ambiguities. The following are some of the common problems that affect the topology readiness of a file.

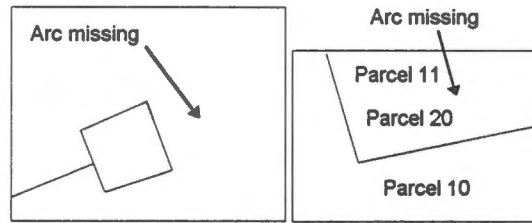
- The lines are not in arc form in that lines do not start and stop at intersections.



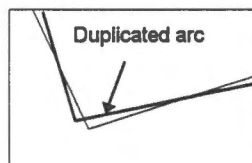
- The file is not free of undershoots and overshoots.



- There are labels missing from polygons.
- There are arcs missing.



- There are duplicate arcs.



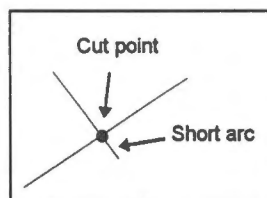
Newly digitized data are normally not topology ready because of errors. Imported data might or might not be topology ready depending on the extent of editing being applied to them at source.

When we import data from another system, or digitize them from a map of different projection and quality, the geometrical differences could cause topological conflicts. We have seen some of them in the chapter on Data Conversion and Integration.

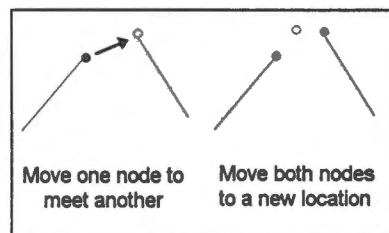
Some automated and manual processings have to be carried out in order to prepare data for topology building. Converting lines into arcs can be done automatically because it merely involves the cutting of lines at their intersections.

The removal of overshoots and undershoots are more complicated and cannot be fully automated. To remove overshoots, two steps are involved. Firstly, we cut lines at their intersections. This can be done automatically with the overshooting parts ending up as very short arcs. After this, arcs shorter than a certain length can be deleted, thus removing the undershoots.

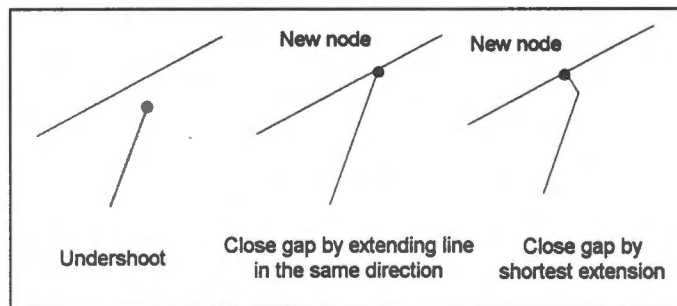
The difficulty here is to decide on the appropriate threshold on arc length because one that is too loose would remove legitimate arcs as well and one that is too tight would not be effective at all. For this reason, the complete removal of overshoots requires some operator intervention.



Undershoots are corrected by extending arcs to close the gaps. Unlike overshoots, there are several cases that require different strategies to handle. In the first case, the gap occurs around the ends of arcs. If the position of one node is known to be more accurate than the others and is identified as such in the database, we can move the other nodes to meet it. Otherwise all nodes will move to meet at the correct spot.



In the second case, a gap occurs between the end of one arc and the middle of another. Such a gap can be closed by extending the undershooting arc along its direction or across the shortest distance. A new node would have to be created at where the extensive meets the other line.



Other problems of topological ambiguities, missing data, duplicated data, and other problems probably have to be detected and edited manually.

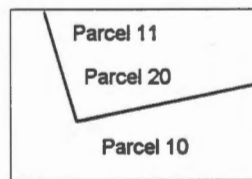
Build topology

After we have made the data topology ready, we can invoke the topology builder to derive topological information and store them in the database. Topology building involves considerable geometrical processing to determine if and where arcs meet, which arcs form closed loops to make polygons, which islands are enclosed by which polygon, and which display label belongs to which polygon.

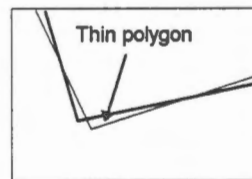
The reason for associating display labels with polygons is to assign attributes to polygons automatically. In this case, the display label contains the parcel identification code, and it can be used as the polygon key in the database.

The topology builder can help to detect a number of topological inconsistencies that can be traced back to geometrical errors. It is not possible for the topology builder to catch all errors, but it could look for revealing signs and mark them for user verification. A short list of these signs is given here.

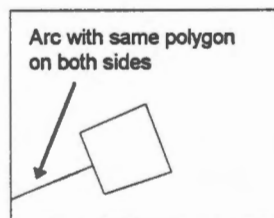
- Missing polygon label. This can be a digitizing error or the label could have been omitted intentionally to indicate a polygon without user-defined attributes. For example, the polygon forming the highway of the FTON database does not contain a display label.
- Duplicated polygon labels. This can be caused by the entry of an additional label or by a missing arc.



- Very thin polygons. This can be caused by a duplicated arc.



- An arc with the same polygon on both sides. This can be caused by a missing arc.



Topology building is an iterative operation. After an initial processing, topological errors are flagged and corrected. Alternating rounds of topology building with editing will be performed until we have eliminated all problems. At that point, we would have created a geometrically clean and topologically correct database available for spatial analysis and other GIS operations.

Chapter 51

A Survey of Concepts

This guide covers both GIS applications and concepts. The approach is operation oriented, making it necessary to present the same concept from different perspectives when the various GIS operations related to the concept are being introduced. In this chapter, we consolidate and supplement the concepts we have covered.

Coordinate System

A unique property of geographic features is that they are positioned in a coordinate system on which geometrical calculations are based. The *dimensionality* of a coordinate system indicates the number of coordinates we use to uniquely locate a point. A two-dimensional system uses two coordinates.

Geographic coordinates in degrees of latitude and longitude are based on ellipsoidal reference surfaces, and there are a number of these. Different ellipsoids are used to map different parts of the earth, and they are modified as we acquire more accurate measurements of the earth's shape.

Curved surfaces are not very convenient for display and for making measurements. The grid of latitudes and longitudes on an ellipsoidal surface form a network of curvilinear lines, making measurements, such as distance, angle, and area, more difficult than corresponding measurements on a flat surface. Hence the need to transform the earth's surface onto a flat map.

Please also refer to chapters on Geographic and Ground Coordinates, Data Conversion and Integration, Transformations.

Map Projection

A *map projection* transforms an ellipsoidal earth to a flat map, always causing some geometric distortions. Different map projections were developed to minimize one kind of distortion or another.

After being transformed onto a flat surface, it is then possible to use a rectangular grid to locate points and perform other geometrical measurements on the earth's surface. The units of measurement on such a grid are usually based on metres or feet on the ground, thus the term *ground coordinate system*.

Maps produced by different projections would not agree with each other in two aspects. First, they are distorted differently. Second, the ground coordinates of the same point from two different projections seldom agree because of the different rectangular grid being used.

Ground coordinates from any map projection could be converted back to geographic coordinates, and in theory the geographic coordinates of the same point from different map projections should agree. Discrepancies could be caused by computational errors and the different reference surfaces being used.

Please also refer to chapters on Geographic and Ground Coordinates, Coordinate Systems and Distance Measurement, Coordinate System and Area Measurement, Data Conversion and Integration.

Scale

A GIS database, unlike a conventional map, does not have a scale in the usual sense. Scale is an important characteristic of maps because they are rarely as large as the area they cover. In a digital database only coordinates are stored and there is no physical paper involved. Consequently the units of measurement for the coordinates can be made to be one-to-one, or in other words, scaleless. However, geographic data are often collected from documents such as maps and aerial photographs that *do* have scales. A scale indicator carried in the database helps to convey a sense of accuracy and level of detail.

Please also refer to chapters on Data Conversion and Integration.

Resolution

A more important characteristic for digital databases, which is related to scale, is *resolution* of the coordinate system. It indicates the shortest line or smallest gap that could be represented. Its most visible effect is the smoothness of the lines.

Conceptually, resolution implies a grid over the database. The spacing between the grid lines, if measured in ground units, is called the *ground resolution*. Only grid intersections are assigned coordinates.

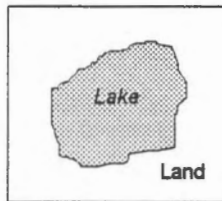
Resolution is related to the number of significant digits we assign to the coordinates, and it affects the accuracy of databases. Naturally, one needs a higher resolution to support a more accurate database.

Please also refer to chapters on Resolution, Vector Representation, Raster Representation, Line Splitting and Resolution, Data Conversion and Integration.

Two Organizations of Space

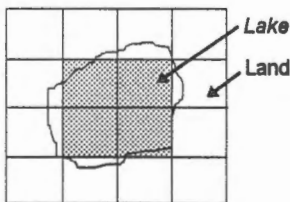
A geographic database deals with space, and there are two ways of organizing the space covering the project area — by features or by arbitrary cells.

In the first case, the features are identified, their outlines are delineated, and then a computer representation is used to store the outlines. We call this a *feature-oriented* partition.



In the second case, the project area is subdivided arbitrarily into cells of space. Although cells of any shape could be used, it is more common to use regular ones. We call this a *space-oriented* partition.

Our real interest lies with the features contained in these cells. Since each cell can overlap several features, such as lake and land in this example, a sampling process has to be carried out. This normally associates each cell with the dominating feature.

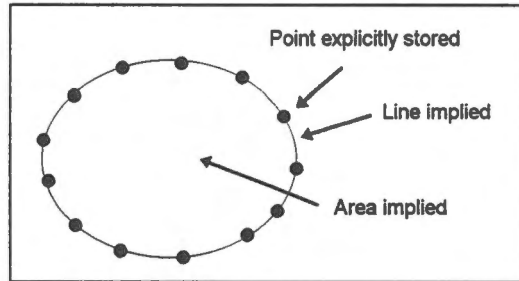


Vector Representation

In a feature-oriented partition, the feature outlines are important geometric primitives. It is not possible to store smooth lines in a computer, and the lines must be decomposed into a series of points. When the lines are displayed, straight line segments (or *vectors*) are used to join adjacent points to form continuous lines. The closer the points, the shorter the vectors, and the smoother the lines would appear. Because of the use of vectors to join points, this is called the *vector representation*.

A problem with the vector representation is that we only record points at discrete locations and selected spots along lines. That leaves the space between them empty. As shown in the diagram, the points are the only geometric primitives stored explicitly. A line is in fact implied because the vectors joining adjacent points are only mathematical

abstractions. As a result, a certain number of geometrical calculations have to be carried out to determine whether an arbitrary point lies on a line. For the same reason, we have to determine computationally whether a point is inside or outside of an area.



Please also refer to chapters on Vector Representation, Retrieve Information on Features, Overlay Analysis, Data Conversion and Integration.

Raster Representation

A space-oriented partition using regular cells produces a *raster representation* where the cells are called *pixels*. Each pixel is stored with its attributes that describe the feature overlapping the cell. When the pixel size is large, each pixel can contain more than one feature. It is possible, although not very common, to assign each pixel to multiple features and record the percentage of area covered by each.

Please also refer to chapters on Raster Representation, Overlay Analysis.

Vector Versus Raster Representations

The two representations complement each other in that vector representation focuses on features, whereas the raster one focuses on space. As a result the vector representation is more suitable for manipulating features and the relationships among them. It is, for example, easier to edit the features and store topological relationships among them.

We mentioned earlier that one of the problems with vector representation is that the space within feature outlines is implied, making it difficult to manipulate space. The raster representation, on the other hand, is more suitable for manipulating space than manipulating objects.

Using raster representation, every location in space (a pixel) has a representation in the database. To determine on what feature a given point falls, we merely have to look up the corresponding pixel in the database and retrieve the feature associated with it. Raster is

also the representation of choice when generating buffer zones and performing overlay analysis, operations that involve space.

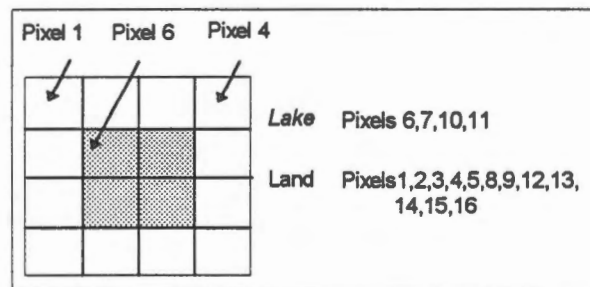
In a raster representation, the extent of the features is implied because the pixels related to a geographic feature are not packaged together in the database. This makes it awkward for manipulating objects and difficult to store topological relationships among them. The only explicit topological relationships stored for a raster representation are those between one pixel and its neighbours.

There are an increasing number of GISs that support both vector and raster representations. Depending on the application, the user could choose the appropriate representation. This requires the conversion from one representation to another and perhaps the storage of the same feature in two different forms. The conversion from vector to raster is simpler, while the reverse conversion is quite complicated.

Please also refer to chapters on Vector Representation, Raster Representation, Raster Conversion.

Harmonizing Vector and Raster Representations

It is possible to collect all pixels belonging to a feature and package them as a unit. In this example, we associate a lake with all its pixels. In essence, we have produced a feature-oriented partition but use a raster representation for each feature. In this GIS, we call them *zones*.



Please also refer to chapters on Overlay Analysis, Raster Conversion.

Geographic Features

Each geographic feature in a database is associated with one or more geometric representations. The choice between vector and raster representations, as we have pointed out earlier, depends on the application in mind. When vector representations are used, we can choose a point, a line, or a polygon to represent a feature. The choice partly

depends on the operations to be performed on the features and partly on the amount of detail we want to record about them. For instance, a house would become a point if we do not know or wish to represent its dimensions. A two-sided street would become a single line if we do not need to show its width. In fact, a single-line street network is easier and more suitable for network analysis.

Other than the geometrical representations, geographic features are also associated with other attributes describing their characteristics. Two common attributes are the feature code and the feature key. A *feature code* describes the kind of feature, while a *feature key* identifies which feature. In other words, a feature code is assigned to a group of features of the same property, but a feature key is assigned to individual features.

There are many ways of designing feature codes and feature keys (or feature names). Several national and international organizations have designed feature codes intended to be used as standards. These coding systems differ in the length of the code and the level of feature classification.

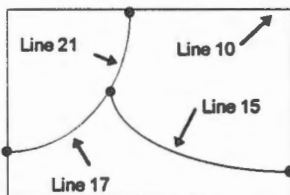
Some geographic features, such as highways, have already been numbered. Most cadastral offices have a scheme for naming parcels and properties. These unique identification codes can serve as feature keys. However, there are many features, such as buildings, without standard names. Users might have to rely on the GIS to assign arbitrary names to them or develop their own naming conventions.

Please also refer to chapters on Different Perspectives of Reality, Vector Representation, Feature Attributes, Retrieve Information on Features, Data Conversion and Integration, Creating a Geographic Database.

Topology and its Representation in a GIS

Topology describes spatial relationships among geographic features and is independent of the coordinate system. That is to say, the topological relationships, such as connectivity, adjacency, and enclosure, are not affected by changes in the coordinate system. It is important to handle topology in a GIS because users ask topological questions such as: Are there properties facing a recreational lake for sale?

In a vector representation of geographic features, the outlines of the features have already been delineated. From this, we can deduce the topological relationships. As in this example, it is obvious to us that lines 10, 15, and 21 form a closed polygon and that lines 15, 17, and 21 meet at a point. To a GIS, however, these relationships are not obvious but have to be derived by geometrical calculations which could be quite time consuming.



In order to avoid the computations every time we need to know about topological relationships, it is common practice to pre-derive and store them in the database. The process of deriving topological relationships is called *topology building*.

The topology builder, a piece of software to derive topological relationships, makes several assumptions about the graphics. For example, it expects that lines will not cross each other but start and stop at intersections, which are called *nodes*. Lines that terminate at nodes are called *arcs*.

Please also refer to chapters on Topological Relationships, Topology and Editing, Creating a Geographic Database.

Data Digitizing and Editing

Digitizing existing maps is one of the most popular methods of geographic data capture. Although data could be obtained from ground surveys, photogrammetric compilation, and other techniques, converting data from existing maps is more cost effective.

There are three methods of converting map data into digital form. The first method is to manually trace the data on a digitizing table. This produces a vector representation of the features.

The second method is to scan them automatically, producing files in raster format. If necessary, the raster file can be converted to vector format using raster-to-vector conversion software. These conversions are not trivial and cannot be fully automated for complicated maps.

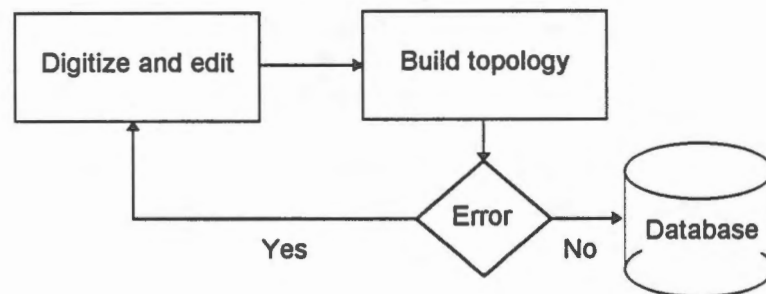
The third method is to first scan the map, and then use an operator-assisted technique to trace over the raster image to produce lines in vector format. With the help of an operator, some of the tricky problems of converting from raster to vector are handled manually.

When digitizing a map for cartographic output, we only need to ensure that the geometry is visually clean. Minor imperfections that are visually insignificant can be ignored. But when digitizing a map for generating a topological database, we need to avoid any imperfections that could cause topological inconsistencies. Undershoots and overshoots of lines have to be cleaned up, and lines have to be in arc form to help the topology builder perform the task easier.

Digitizing lines in arc form is quite tedious as the operator must start and stop at intersections. A more popular method is called *spaghetti digitizing* by which the operators ignore the intersections and terminate lines wherever they wish. After digitizing, the spaghetti lines can be converted to arcs by cutting them at intersections. The removal of overshoots and undershoots is then performed to make the data topology ready. Most of these can be done automatically.

When digitizing a point or a line, their attributes such as feature code and feature key are entered at the same time, normally through the keyboard. The same technique cannot be used for polygons because when we are digitizing their outlines, polygons are not formed yet. A common method is to digitize some of their attributes, particularly their keys, as annotation labels placed inside their respective polygon outlines. It is then the topology builder's job to determine which label is inside which polygon and thereby picking up these attributes automatically.

Editing and topology building are iterative processes. The first attempt at topology building might not produce the desired result because the input data are not clean enough. After correcting these problems with the Editor, we pass the data through the topology builder again. The steps of editing and topology building are repeated until a clean topological database is built.



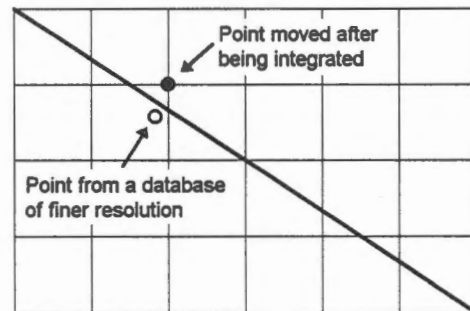
Please also refer to chapters on Data and Its Collection, Accuracy and Data Editing, Cartographic Enhancements and Data Editing, Topology and Editing, Creating a Geographic Database.

Data Conversion and Integration

There is an increasing need to share data among GIS users. When there is digital data available from another source, it might be more cost effective to acquire them instead of collecting them from scratch. But data sharing could be complicated by differences in coordinate systems, data accuracies, feature representations, and feature definitions.

Converting data from one projection to another can be done automatically provided that we know all the necessary map projection parameters. You can expect, however, that such conversions could introduce some computational errors.

Differences in resolutions could cause topological problems. In the following example, a point from a database of finer resolution is being integrated into one with a coarser resolution. Because the closest resolution grid is on the other side of the line, the point will be moved relative to the line, causing a topological error.



Differences in data accuracies will also cause topological problems. They are not easy to detect automatically and almost always have to be corrected manually.

Differences in feature representations could be difficult to reconcile. For instance, it is difficult to convert raster data from one system to vector data in another system. Moreover, houses represented as point features from a small-scale database cannot easily be changed into area features expected by a large-scale database.

Feature definition differences could cause many problems. When we define a geographic feature, such as a highway, we define its geographic extent as well as its attributes. A highway, for example, could include the edge of the pavement, the pavement, the shoulder, or even the easements on both sides. Even when the same definition is being used, the attributes collected for them could be different from one organization to another — some might include the number of lanes and some might not. Consequently, organizations using different definitions of features cannot share data easily.

We can write a direct translator between two GIS databases for data interchange. A more common method is to go through a neutral format called the *interchange format*. This indirect method uses two translators to do the job of one. Nevertheless, the translators to

and from the neutral format are easier to write because these formats are more readily available and are more stable than the proprietary formats used by the GISs.

A problem of using neutral formats for data interchange is the number of such formats available and none of them has been universally accepted. They are different in their flexibility and capability to handle geographic features and their relationships. A wrong choice of interchange format could cause degradation or even damage to data.

Please also refer to chapters on Data Conversion and Integration, Transformation, Transformations, Creating a Geographic Database.

Data Retrieval

GIS databases are large. They cover extensive geographic areas, contain many features, and associate many attributes with each feature. The large volume of data makes the retrieval of data a challenge.

The retrieval problem is compounded by the multi-dimensionality of the data and the complex query conditions. One of the common queries, called the *range search*, finds all geographic features within a user-defined window. This kind of search is used when we zoom in and out of a database.

A brute force method of performing such a search is to compare every geographic feature in a database against the window to decide if they overlap. To determine if two features overlap, geometrical intersection tests are performed on the vectors making up the features. For a large database, this method of solving the range search is too slow.

GIS tackle this search problem from different angles. First of all, the shapes of the geographical features are simplified to *become minimum bounding rectangles*. This allows us to quickly discover features which obviously do not overlap and eliminate them from the more expensive intersection tests using the individual vectors.

Next, a *spatial indexing system* is used to reduce the search to a localized region of the database. There are many different schemes of spatial indexing, but they are all based on some kind of sorting in two or more dimensions.

Data retrieval can also be based on non-spatial conditions, such as finding all properties belonging to a certain owner. Operations like this can be used to control the contents of the display and to retrieve data for analysis.

Please also refer to chapters on Coordinate systems and Distance Measurement, Display Selected Features, Feature Attributes, Retrieve Information on Features, Select Features Within an Area, Use Topology to Identify Polygons by Pointing.

Data Analysis and Modelling

Data retrieval finds existing data within the database. Analysis uses the retrieved data to derive new information, and modelling simulates a real world situation in the computer. These are the functions that make a GIS a powerful decision support system.

There are so many analytical and modelling functions for geographic data that it is impossible for a GIS to support all of them. Consequently, different GISs emphasize different functions, thus giving each GIS its own unique character.

The choice of either raster or vector representation influences the type of analytical and modelling functions supported by a GIS. In general, raster-based systems provide more analytical functions related to areas (such as neighbourhood analysis) and vector-based systems provide more functions related to topology (such as network analysis).

One of the most popular analytical functions related to area is *overlay analysis*, which combines two or more layers of data to reveal new information. For example, a layer of population distribution combined with a layer of supermarkets can reveal whether they are distributed optimally to serve the community.

Overlay analysis is easier for raster data than for vector data. But in both cases, attributes from the layers are combined to form composite attributes. Although overlay analysis can involve many layers, it is basically done two layers at a time.

For raster data, the unit of processing is a pixel and not a feature. Corresponding pixels from the two layers are matched and their attributes combined for the resultant layer. That means no geometrical calculations are performed, making overlay analysis for raster data quite straightforward.

For vector data, the unit of processing is a feature. Point, line, and area features from the layers are combined so that logical relationships among them, such as union and intersection, can be derived. Polygon overlay is the most common type, and we will use it for illustration here. The concepts, however, apply to point and line features as well.

Different methods of overlay analysis are used for vector data, offering various levels of sophistication. At the highest level of sophistication, polygons from the layers are cut to form new polygons for the resultant layer, and their attributes are combined to become composite attributes for the new polygons. This involves the calculation of intersections, the cutting up of polygon edges, and the formation of new polygons.

At a lower level of sophistication, new polygons are not formed. Polygons from one layer is merely tested against those from the other layers to see whether they overlap or not. Depending on the application, the overlapping or the non-overlapping polygons are

reported. A mixture of raster areas, sometimes called zones, and vector polygons could be used in this type of overlay analysis.

At the lowest level of sophistication, an overlay analysis involves merely the displaying of two layers of data simultaneously on the terminal. A mixture of raster and vector data can be used. It is then left to the user to analyse the result visually.

Modelling generates new scenarios for the analysis. For example, when we wish to determine the growth of a burning forest fire, we need to generate areas representing the extent of the fire at regular time intervals. In order to do this, we need to know the factors (such as forest cover, wind direction, and humidity) and the way they interact to affect the spread of the fire. This often requires expert knowledge and sophisticated mathematical models.

A simple modelling tool is called *buffer zone generation*, which generates areas surrounding seeds of point, line, and area features. The outline of the buffer zones are normally at a constant distance, specified by the user, from the seed features. This is suitable for simple models that assume uniform growths of geographic phenomena from features. For instance, we can assume that the spread of fire is uniform from a source and hence the burned areas at regular intervals form concentric circles. Buffer zones are often used in overlay analysis.

A GIS cannot support all the analytical and modelling functions. If a function is not available, the user would have to program it. There are several ways of extending the functionalities of a GIS. Some GIS provide macro languages that can be used to produce application programs. Some GIS support subroutines that can be called from a user program, and some of them allow the passing of data between the GIS and an external program.

Please also refer to chapters on Overlay Analysis, Retrieval Using Buffer Zones, and those on applications.

Where to Go from Here ?

In this Guide, we have touched on the surface of the vast subject of GIS. We have used a GIS software to illustrate the concepts and to show how one GIS does the job. You will notice that most of what you have learned from this Guide will apply to others because most GIS on the market are based on similar principles. There are, however, some notable differences among them.

User interface. Most of the systems are menu driven, but the window and menu layouts would be quite different from one GIS to another. Moreover, the same operation could be called by different names.

Data types and operations supported. Geographic features are represented by geometric primitives such as points, lines, and polygons in the GIS. Other types that could be found in GIS are text, spot height, sounding, contour, and image. You can tell the number of types available by the commands supported by the software. For example, the presence of a contour-add command normally means that there is a contour type.

A GIS supporting more of these primitives is more flexible in that real world features have direct representations in the database. For instance, if a GIS does not support the contour type, then an ordinary line would have to be used instead. Because a contour is a line associated with an elevation, it contains more information than an ordinary line.

The data types supported by a system are a reflection of its dominant applications. Systems mainly used for map production would have more geometrically-defined primitives such as circles, ellipses, curve-fitting lines, and curve-binding text. Systems used mainly for spatial analysis would support more raster types and topological primitives such as nodes and arcs.

Number of databases used. The GIS we have used, as well as a number of others on the market, employs two databases to store data — one for system-defined data and the other for user-defined data. Some other GIS use only one database.

There are advantages and disadvantages for both schemes. The reason for using two databases is because traditional database management systems (DBMS) are not entirely suitable for managing spatial data. That has forced some GIS vendors to develop their own data management modules for spatial data and incorporate an off-the-shelf relational DBMS to handle non-spatial data. In this way, a two-database system is supposed to give the user the best of both worlds. With data for the geographic features separated into two databases, GIS developers are faced with the challenge of hiding this separation from the users. This is not an easy task, and you probably have noticed the awkwardness this can produce as you used this GIS.

GIS using one database, and hence one data management module for both spatial and non-spatial data, are faced with another challenge, which is to get around the deficiencies of traditional DBMS in handling spatial data.

One of the deficiencies is in the way data are organized. Because relational DBMS are mainly designed for business applications and are not optimized for multi-dimensional data, the use of them for spatial data involves a certain amount of storage and processing overheads. As a result, GIS using one database tend to be slower and consume more storage space. However, developers of traditional DBMS have recognized this problem and some of them are taking steps to correct the situation. Oracle, a major relational DBMS vendor, has recently developed a multi-dimensional version called Oracle-MD.

Learning the operation of one GIS provides you with the experience to make the learning of another easier. The knowledge of one GIS becomes a reference with which to compare other systems. But in order to make sense of the similarities and differences, you must have a good grasp of the concepts.

One way to learn concepts is to read the many books on GIS now available. We have provided a selected list of them here. There are many magazines, professional journals, and newsletters related to GIS and computer. Some of these are free for your asking if you qualify. Attending short courses, workshops, and conferences on GIS will put you in contact with the latest developments. An excellent source of information is the Internet if you have access to it and the time to browse through the large collection of GIS-related sites. A list of some of the useful sites are included at the end of this chapter.

GIS is a dynamic, multi-disciplinary, and exciting field. We hope you will continue to be curious and explore the many possibilities it offers.

Selected Books on GIS

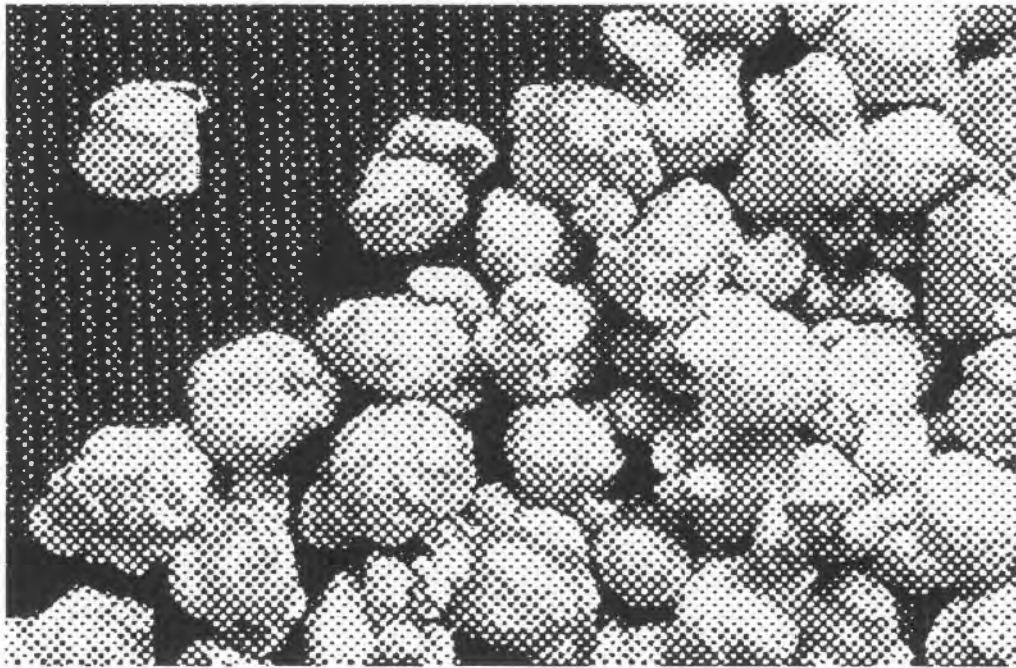
- Antenucci, J.C. et al. (1991). *Geographic Information Systems, A Guide to the Technology*. Van Nostrand Reinhold.
- Aronoff, S. (1989). *Geographic Information Systems: A Management Perspective*. WDL Publications.
- Burrough, P.A. (1986). *Principles of Geographical Information Systems for Land Resources Assessment*. Clarendon Press.
- Laurini, R. and D. Thompson (1992). *Fundamentals of Spatial Information Systems*. Academic Press.
- Maguire, D.J., M. F. Goodchild, and D.W. Rhind (Eds.) (1991). *Geographical Information Systems*. Two Volumes. Longman Scientific and Technical.
- Peuquet, D.J. and D. F. Marble (Eds.) (1990). *Introductory Readings in Geographic Information Systems*. Taylor and Francis.
- Ripple, W.J. (Ed.) (1987). *GIS for Resource Management: A Compendium*, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping.
- Taylor, D.R.F. (Ed.) (1991). *Geographic Information Systems, The Microcomputer and Modern Cartography*. Pergamon Press.
- Tomlin, C.D. (1990). *Geographic Information Systems and Cartographic Modeling*. Prentice Hall.

Internet Sites

There are numerous Internet sites on the World Wide Web (WWW) related to GIS. Web browsing software such as Netscape provide search facilities, and a search using keywords such as GIS would generate useful leads. The following sites are good starting points.

Internet GIS and RS Information Sites	ftp://gis.queensu.ca/pub/gis/docs/gissites.html
Interesting and useful sites on GIS	http://www.seic.okstate.edu/serv.html
GIS WWW resource list	http://www.geo.ed.ac.uk/home/giswww.html
Frequently-asked questions on GIS	ftp://abraxas.adelphi.edu/pub/gis/FAQ

Appendices



Appendix A

Glossary of Terms

Those terms marked with a □ are specific to CARIS, the GIS we use in this Guide.

Arc

A topological entity that starts and ends at a node. Although it is often graphically depicted as a line, an arc only denotes a topological link between two nodes, and the geometrical shape of the link is not important. It is also called a *link*, an *edge*, or a *1-cell*.

Attributes

Characteristics of geographic features. Examples are area, length, owner name, and address.

Attributes, Non-Spatial

Attributes that are not related to the geometry of features. Examples are population of a country and name of a highway.

Attributes, Spatial

Attributes that are directly or indirectly related to the geometry of features. Examples are length of a road and area of a country.

Attributes, System-Defined □

Those attributes which are managed directly by CARIS.

Attributes, User-Defined □

Those attributes which are not managed directly by CARIS but are stored separately in a relational database.

Bounding Rectangle

The smallest upright rectangle enclosing a feature, and it is also called a *minimum bounding rectangle*. It can be defined by two diagonal points, one from the smallest x and y coordinates, and the other from the largest x and y coordinates of the feature. The main use of this rectangle is to approximate the shape of feature so as to reduce the time it takes to retrieve features.

Buffer Zone

A modelled region around a feature. It is usually described as a polygon with its boundary at a constant distance from the given feature. A buffer zone around a point is a circle.

Co-locating lines

Lines that share the same space. For example, a river and an international boundary co-locates if they share the same line.

Control Points

They are points which can be used to relate and then transform geometric objects from one configuration to another. If the two configurations are in different coordinate systems, control points are used to transform the objects from one system (such as a digitizer coordinate system) to another (such as a ground coordinate system). If the two configurations are in the same coordinate system, control points are used to transform from a distorted configuration to a more accurate one.

Coordinate, Absolute

Coordinate with its value measured from a single origin in the coordinate system. See also Coordinate, Relative.

Coordinate, Relative

Coordinate with its value measured from a nearby point instead of from the single origin in the coordinate system. The purpose is to reduce its magnitude so as to save storage space. See also Coordinate, Absolute.

Coordinate System

It is a means of specifying unique position for geographic features. In a two-dimensional coordinate system, two coordinates are used to specify position on a plane. In a three-dimensional coordinate system, an additional third coordinate is used to specify elevation or depth.

Coordinate System, Geographic

A coordinate system defined over a mathematical body, usually an ellipsoid, approximating the earth. The coordinates are in degrees of latitude and longitude.

Coordinate System, Ground

A coordinate system defined over a flat map. The coordinates are usually in metres along the Northerly and Easterly directions.

Coordinate System, Scaled Ground

A coordinate system which is a scaled version of the ground system. The constant scale factor applied to the coordinates helps to change them from decimal numbers to integers, thus reducing computer storage.

Corridor

Same as Buffer Zone.

Data Interchange Format

A neutral format used to convert files from one GIS format to another. The neutral format serves as a hub in the interchange environment and helps to reduce the number of translators.

Database

Related data collected with a purpose and serving specific applications.

Database Schema

A description of the organization of data in a database. The headings of a relational table showing the attributes is a form of database schema.

DBMS

Database Management Systems. They are computer software designed to manage large databases. Many GIS contain a DBMS component to access and manage data.

Digitizing

It is the process of converting line graphics into digital data by breaking continuous lines into a series of discrete points. See also Raster Scanning.

Digitizing, Arc

Lines in arc form are produced from this method of digitizing. In other words, the operator starts and ends every line at an intersection. This is the preferred form for input to topology builders. See also Digitizing, Spaghetti.

Digitizing, Point Mode

Using this mode of digitizing a line, the operator selects every single point on the line to be recorded. See also Digitizing, Stream Mode.

Digitizing, Spaghetti

Lines are digitized in a seemingly unstructured manner using this method. In contrast with Arc Digitizing, lines do not necessarily start and stop at intersections. A software is often used afterwards to cut lines at intersections and turn them into arcs.

Digitizing, Stream Mode

A digitizing method by which the operator only traces along a line to be digitized. A sampling mechanism will automatically record points along the line at fixed time intervals. See also Digitizing, Point Mode.

Digitizing, Table

A device for manual digitizing of maps.

Display Label

An annotation label for a polygon. It is normally digitized by the operator to record attributes of a polygon. The topology builder will pick up the display labels, determine the polygon it belongs to, and assign the attributes to the polygon accordingly.

DXF

Digital Exchange Format developed to exchange data between AutoCAD systems. AutoCAD is a Computer-Aided Design (CAD) system also used to digitize and edit maps.

Ellipsoid

A mathematical object created by rotating an ellipse about one of its axes.

Feature Code

A code that describes the feature class a geographic feature belongs to. The code HW4L denoting a four-lane highway is an example of a feature code.

Feature Key

A code that uniquely identifies a feature. Highway name such as HIGHWAY 10 is an example of a feature key.

Feature Snapping Option

An option for point entry which allows the GIS to select an existing point in the database as input. If you measure the distance between two points and have invoked the feature snapping option, you would have measured the distance between two existing points in the database.

Features, Cartographic

They are not real world features but are used to annotate a map. River name is an example of a cartographic feature.

Features, Geographic

They are features that occupy space in the real world.

Geodetic Length

The length of a line measured along an ellipsoid.

Geographic Analysis

A process of deriving new information from geographic data. Using the coordinates of points on a line stored in a database to compute its length is one type of geographic analysis. Some geographic analysis involve more complex calculations.

Geographic Modelling

A process of simulating geographic phenomena using mathematical models.

GIS

Geographic Information Systems. A computer system that supports input, storage, retrieval, analysis, modelling, and output of geographic data.

GPS

Global Positioning Systems. A system of satellites and control networks developed and maintained by the United States. It allows the determination of one's position on earth using special receivers.

Line, Curve-fitted

A smooth line.

Line, Point-to-Point

An angular line.

Map

A flat representation of the earth's surface, almost always reduced in size.

Map, Thematic

A map with a focused theme such as land use, geology, and population distribution. See also Map, Topographic.

Map, Topographic

A map covering a wide range of features commonly required by a wide range of users. The features on a topographic map typically includes transportation networks, water bodies, contours, vegetation covers, and major man-made features. It is also called a *base map* because it often form the base of thematic maps.

Map Projection

The technique of transforming an ellipsoidal or spherical earth to a flat map. Since all map projections show some kinds of geometrical distortion, different map projections have been developed to minimize certain distortions.

Map Projection, Conformal

A projection that preserves angular measurement over a very small area.

Map Projection, Equidistant

A projection that preserves distance measurement between certain points.

Map Projection, Equivalent

A projection that preserves area measurement over the entire map. This is achieved at the expense of angular and length measurements, which are in such case distorted.

Map Projection, Stereographic

A type of conformal map projection.

Map Projection, UTM

Universal Transverse Mercator Projection. A type of conformal map projection that systematically covers the entire world except for the polar regions.

Name

Annotation text on a map. Same as Text.

Network Analysis

Geographic analysis using networks such as highway and river networks. One of the popular network analytical function is shortest path analysis.

Node

A topological entity that defines the start and end of arcs. It is also called a *vertex* or a *0-cell*.

NTDB

National Topographic Data Base. A digital topographic database maintained by the Department of Natural Resources of Canada.

Overlay Analysis

A type of geographic analysis that uses two or more layers of data and performs logical operations on them to derive new information. For example, overlaying a land use map over a soil fertility map will allow us to isolate agricultural land developed over poor soil.

Overshoot

The end point of a line extending beyond its correct position. See also Undershoot.

Partition, Feature-Oriented

A subdivision of space by geographic features. Using this approach, the outlines of geographic features are delineated and stored in the geographic database together with their attributes. See also Partition, Space-Oriented.

Partition, Space-Oriented

A subdivision of space by arbitrary cells. Using this approach, the project area is subdivided into cells usually square in shape and constant in size. These cells are stored in the geographic database together with an indicate of which and what geographic features they fall on. See also Partition, Feature-Oriented.

Photogrammetry

The technique of measuring three-dimensional position of features from stereoscopic pairs of aerial photographs. It is the standard technique of producing topographic maps.

Pixel

Acronym for Picture Element. It is a cell, usually square in shape, in a raster image.

Pointing Box

An imaginary box used by CARIS to control how close a user must point to a feature to select it. The smaller the pointing box, the more precisely the user must point at the desired feature.

Point-in-Polygon Operation

The mathematical procedure of determining if a given point is inside a given polygon.

Polygon

A topological entity that is composed of arcs forming a closed area. It is also called an *area* or a *2-cell*.

Polygon label

Same as display label.

Range Search

The search for features in a database within a rectangular window.

Raster representation

The representation of features using a matrix of pixels.

Raster Scanning

The automated conversion of a line map into a raster image through the use of a scanner. See also Digitizing.

Reference Ellipsoid

The ellipsoid used in a particular coordinate system.

Registration

The process of measuring control points and undertaking other operations in order to compute the transformation between two coordinate systems. Usually, it is between a digitizing table system and a ground coordinate system.

Remote Sensing

The collecting of information about the earth from a distance, usually by sensors mounted on satellites.

Resolution

It is the smallest gap or the shortest line that can be represented in a database.

Resolution, Grid

An imaginary grid over a database with the spacing of the grid representing the resolution. Points can be recorded only at the grid intersections.

Resolution, Ground

The resolution measured in ground units.

Run-Length Encoding

A method of compressing raster images.

Scanner

An electronic device that optically scans a document and converts the image on it to a raster representation in digital form.

Selection, Accumulated

A permanent collection of features selected. See also Selection, Current.

Selection, Current

Features selected by the latest command. It is temporary and will be replaced after another command that generates a new selection. See also Selection, Current.

Shortest Path Analysis

The determination of the least-cost path from one point of a network to another. A shortest-distance path minimizes distance whereas a shortest-time path minimizes time.

Sounding

A point with measured depth. See also Spot Height.

Spatial Indexing

It is like the index at the end of a book except that the geographic features are referenced. Such an index helps the GIS to retrieve geographic features quickly.

Spot Height

A point with measured elevation. See also Sounding.

SQL

Structured Query Language. A popular language for manipulating relational databases.

Symbol

Same as Symbolized Point.

Symbolized Line

A line displayed with a pattern. For example, a political boundary can be symbolized as a dashed line.

Symbolized Point

A point displayed with a symbol. For example, a church, which is a point in a small-scale map, can be symbolized as a cross.

Text

Same as Name.

Theme

A collection of features with explicit topological relationships among them. They form a topological layer.

Topological Layer

A collection of features for which explicit topological relationships among them are required. See also Theme.

Topology

The study of relationships, not affected by changes in the coordinate system, among geometric objects. Examples of topological relationships include connectivity, neighbourhood, and enclosure.

Topology Builder

Software that derives and stores the topological relationship among features.

Topology Ready

A state of the GIS database indicating that it is geometrically clean enough as to not cause topological ambiguities and conflicts in the topology builder.

Total Station

A land surveying instrument that combines the functions of a theodolite, an electronic distance measuring device, and a data recorder.

Transformation

The shifting, scaling, rotation, and other geometrical changes of geographic features and coordinate systems.

Transformation, Affine

A geometrical change caused by skewing an object.

Transformation, Projective

A geometrical change that transforms a quadrilateral into any other quadrilaterals.

Transformation, Shift

A geometrical change that causes an object to change location.

Transformation, Similarity

A geometrical change that preserves shape.

Undershoot

An end point of a line falling short of its correct position. See also Overshoot.

Vector

A line with direction and magnitude.

Vector Representation

A representation of geographic features by their outlines.

Visibility Parameters

A set of parameters, such as feature codes and theme numbers, that controls the contents of a display.

Zone

A collection of pixels having the same attributes. For example, all pixels forming a house could be defined as a zone.

Zone, Network

A collection of routes covering, for example, portions of the network within 10 minutes of travel time from an origin.

Appendix B

Data Descriptions

This is a description of the four databases used in the exercises: FTON, NTDB_UTM, NTDB_STE, and SCHOOL

FTON

Part of the City of Fredericton, New Brunswick, Canada supplied by the New Brunswick Geographic Information Corporation (NBGIC).



Themes

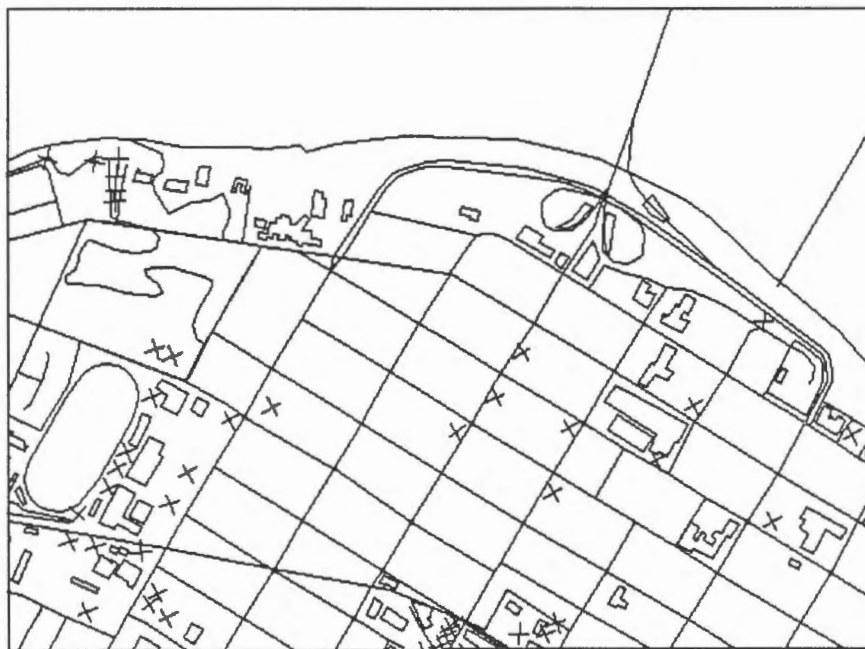
- 10: Links to graphics or other databases.
- 20: Trees and street lights.
- 30: Hydrants and utility poles.
- 50: Parcels.
- 60: City blocks, buildings, and river bank.
- 70: Street network.
- 90: Contours

Feature Codes

BUILDING
CITY_BLOCK
CONTOUR
DRAG
HYDRANT
NEATLINE
PARCEL
PARCEL_ID
POINTER
RIVER
ROAD
STREET
STREET_LIGHT
STREET_NAME
TREE
UTILITY_POLE

NTDB_UTM and NTDB_STE files

Part of the City of Fredericton, New Brunswick, Canada from the National Topographic Data Base (NTDB) supplied by Geomatics Canada, Ottawa, Canada. The stereographic version (NTDB_STE) was converted from the original UTM version (NTDB_UTM).



Themes

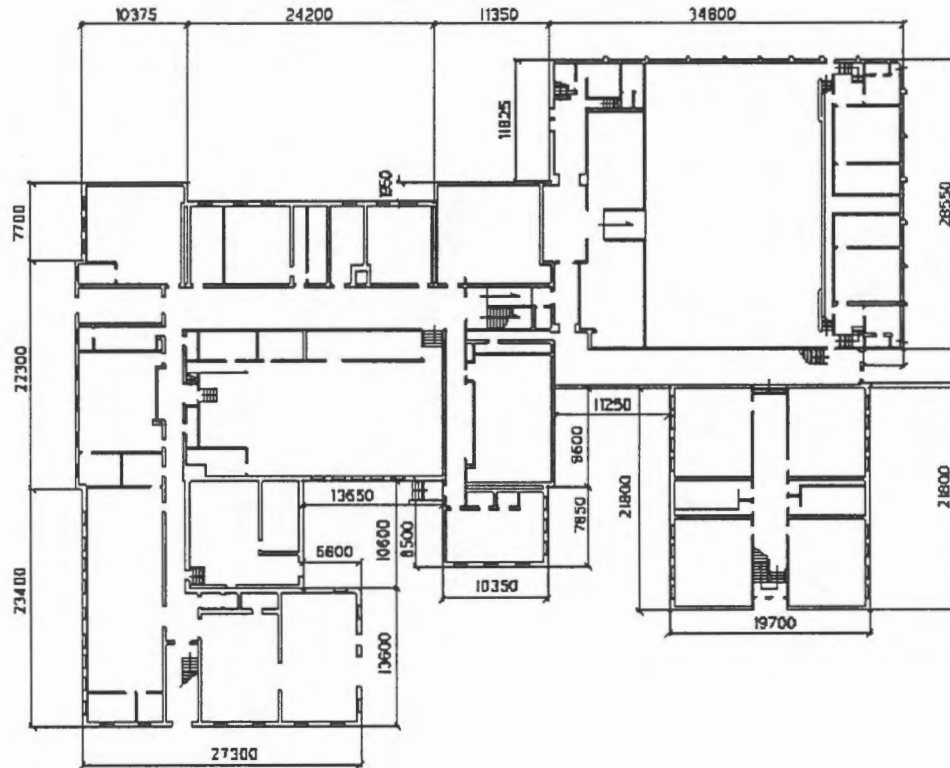
- 100: River bank
- 200: Vegetation.
- 300: Street network
- 400: Other features
- 500: Neatline

Feature Codes:

94-1001	Bridge, other than covered or moveable surface.
106-1001	Building, arena.
113-1001	Building, church.
114-1001	Building, city hall.
120-1001	Building, courthouse.
133-1001	Building, hospital.
143-1001	Building, museum.
147-1001	Building, others.
147-1003	Building, others.
149-1001	Building, parliament building.
150-1001	Building, penitentiary.
153-1001	Building, police station.
160-1001	Building, school.
166-1001	Building, sportsplex.
184-1001	Built-up area.
250-1001	Cemetery.
440-1001	Embankment.
823-1001	Park/sports field.
955-1001	Railway, standard gauge, level with ground, operational, side track.
956-1001	Railway, standard gauge, level with ground, operational, single.
998-1001	Road, less than two lanes, all season, undivided, level with ground, operational, hard surface.
1003-1001	Road, more than two lanes, all season, undivided, operational, level with ground, hard surface.
1005-1001	Road, dry weather, undivided, level with ground, operational, loose surface.
1011-1001	Road, buildings on one or both sides, operational, hard surface.
1016-1001	Road, two lanes, all season, undivided, level with ground, operational, hard surface.
1020-1001	Road, two lanes or more, all season, divided, level with ground, operational, hard surface.
1198-1001	Race Track, other than drag strip racing.
1220-1001	Stadium.
1318-1001	Transformer station, electric.
1412-1001	Vegetation, wooded area.
1452-1011	Waterbody.

SCHOOL

One floor of George Street Junior High School, Fredericton, New Brunswick, Canada supplied by the New Brunswick Board of School District No. 18.



Theme 1: Entire database.

Feature Codes

AUTOCAD0	Triangle
DIM1	Dimension
DIM2	Dimension pointer
DOR	Door
STR	Stair
WAL1	Outer wall
WAL2	Inner wall
WIN	Window

Appendix C
Database Table

This is a listing of the database table PROPERTY. The following are the column headings.

OWNER_NAME	Name of property owner.
STREET	Street address of property.
POSCOD	Postal code of property.
PROP_DESC	A brief description of the property.
PROP_TYPE	A three-digit code describing the use of the property.
ASS_SEQ_N	A sequential number describing the order in which assessors assess properties.
PID1	Parcel identification code.
NBHD	A code assigned to properties in a neighbourhood assessed by one assessor.
TAX_CLASS	Taxation class.
CURR_ASS	The current assessment value. A value of zero means that no assessment value is available.
TAX_CREDIT	Tax credit code.

The property type PRO_TYPE takes one of the following codes, not all of them present in this database.

100	Residential.
101	Residential lots, serviced.
120	Residential improved.
121	Condominium units.
200	Commercial.
201	Commercial land - unimproved.
203	Service station.
207	Convenience stores.
209	Banks & other financial institutions.
210	Retail stores.
211	Office complex.
212	Commercial and residential, combination.
213	Office and residential. combination.
216	Automobile dealerships.
217	Parking facilities.
218	Funeral homes or parlours.
220	Salvage dealers.
299	Other.
300	Industrial.
302	Industrial land - unimproved unserviced.

- 303 Warehouses, including cold storage plants.
- 313 Printing and publishing plants.
- 315 Fabricated metal product plants.
- 319 Utilities.
- 335 Water /sewage treatment plants.
- 400 Institutional.
- 401 Institutional land, unimproved.
- 402 Schools.
- 405 Hospitals.
- 408 Libraries.
- 409 Museums/art galleries and historical sites.
- 412 Post offices.
- 419 Municipal buildings unspecified.
- 421 Police stations.
- 422 Provincial buildings.
- 423 Federal buildings.
- 504 Parks.
- 508 Social clubs/service clubs.
- 510 Arenas/civic centres.
- 576 Aquatic centres.
- 800 Mission hall for salvation army.
- 801 Church.
- 802 Cemetery.
- 900 Miscellaneous land.
- 911 Right of way.

The tax credit code TAX_CRED takes one of the following values.

- 1 Single unit, owner occupied.
- 2 Residential units, one owner occupied.
- 3 Complex, one residential unit, owner occupied.
- 4 Complex, two residential units, one owner occupied.
- 5 More that two residential units, one may be owner occupied.
- 6 One or two residential units, not owner occupied.
- 7 No residential units.
- 8 Two or more residential units, all owner occupied.
- 9 Complex, two or more residential units, two or more owner occupied.

Appendix D

Tips for Running CARIS

- If you encounter problem in selecting a feature, do the following:
 - Refresh the screen and try again.
 - Make sure the display window is large enough to show the feature.
 - Check the “By snapping to a feature” switch in the Options > Pointing Method... menu item.
 - Make sure the visibility parameters include this feature.
- If you encounter problem inputting from the keyboard, make sure the window is active by clicking on it.
- Make sure the drawing has finished before proceeding with the next operation. You can interrupt the drawing by clicking on the display window.

Index

- accuracy 170, 225
 - digitizing table 167
 - line tracing 168
 - map 164
- adjacency 97
- analysis 2, 254
 - network 158
 - overlay 111, 254
 - involving lines 152
 - involving points 147
 - raster 113, 254
 - vector 112, 254
 - zone 114
- statistical 103
- arc 97, 250
 - information 99
- arcs
 - around a node 216
 - around polygon 215
- area measurement 35
- bounding rectangle 92, 175, 253
- buffer zone 127, 255
 - around lines 152
 - around points 148
 - create 128
- CAD 220
- CARIS
 - Information Manager 11
 - panel 12
 - suite 8
- connectivity 97
- contour
 - data structure 68
 - information 67
 - label 172
 - add 203
- control point 165, 231
- coordinates
 - absolute 189
 - enter 22
 - geographic 20, 244
 - ground 20, 222, 244
 - read 19
 - relative 189
- corridor 127
- data
 - analysis *see geographic analysis*
 - conversion 219, 239, 252
 - cost 163
 - integration 219, 252
 - interchange format 220, 239, 252
 - modelling *see geographic modelling*
- database 2, 256
 - create 237
 - design 78, 237
 - PROPERTY 147
 - schema 237
- datum 25
- DB Manager 12, 87
- DBMS 1, 256
- demonstration, starting 7
- digitizing 250
 - arc 239, 251
 - manual 166, 239
 - point mode 166
 - procedure 169
 - spaghetti 240, 251
 - stream mode 166
 - table 166
 - accuracy 167
 - resolution 167
- display label 65, 251
- display window 11
- distance
 - geodetic 27
 - ground 27
 - measurement 26
- DXF 220
- editor 174
- enclosure 97
- erase switch 82
- false origin 25
- feature
 - attributes 75

- system-defined 75
- user-defined 75
- cartographic 41
- close by 63
- code 60, 75, 238, 248, 251
- geographic 41, 248
- key 76, 248, 251
- line 44
 - add 206
 - curve-fitted 181, 187, 206
 - data compression 189
 - data structure 68
 - information 59, 100
 - modify shape 181
 - parallel 208
 - point-to-point 181, 208
 - squared 210, 212
 - symbolization 47
- move 175
- point 46
 - add 198
 - data structure 69
 - symbol information 66
- geographic analysis 1, 2, 254
- geographic modelling 1, 2, 254
- GIS 1
 - starting 8
- GPS 164
- help 13
- Info Manager icon 8
- length measurement 28
- line pattern library 47
- logical consistency 170, 211
- map
 - thematic 93
 - topographic 93
- map projection 21, 244
 - conformal 25, 36
 - equidistant 36
 - equivalent 36
 - stereographic 25
 - UTM 222
- message line 11
- minimum bounding rectangle
 - see bounding rectangle*

- mouse 12
- name 62
 - data structure 69
 - font 63
 - modify 195
 - size 63
- network
 - analysis 158
 - zone 161
- node 97, 250
- NTDB 221
- overshoot 240
- panning
 - by entering coordinates 24
 - by pointing 16
- partition
 - feature-oriented 246
 - space-oriented 246
- photogrammetric compilation 164
- pixel 51, 247
- point symbol library 46, 50
- pointing box 31
- point-in-polygon 108
- polygon 97, 250
 - arcs around 215
 - area 103, 106
 - key 66
 - label *see display label*
 - left 61, 104
 - neighbours 101, 104
 - right 61, 104
- query
 - builder 82
 - condition 79
- range search 253
- raster 51, 247
 - conversion 117
 - data compression 57
 - data structure 56
 - scanning 169, 250
- reference ellipsoid 20, 244
- relationship
 - geometrical 97
 - topological 97
- remote sensing 165

- resolution 37, 244
 - digitizing table 167
 - effect
 - on area 120
 - on data integration 252
 - on length 194
 - on line splitting 190
 - grid 38, 244
 - ground 37, 244
 - raster 54
 - screen 49
- run-length encoding 57
- scale 15, 244
- selection
 - accumulate 95
 - reduce 133
- shortest path 159
- snapping to feature 30, 31
- sorting 85
- sounding
 - add 202
- source ID 60
- spatial indexing 17, 33, 108, 253
- spot height
 - add 200
 - move label 201
- SQL 87
- symbol
 - add 199
 - information 66
- text *see name*
- theme number 61
- topological
 - data structure 105, 217
 - layer 98
 - layers 155
- topology 97, 249
 - building 100, 106, 240, 242, 249, 251
 - editing 214
 - ready 240
- total station 164
- transformation 165, 231
 - affine 236
 - projective 236
 - shift 235
 - similarity 236
- undershoot 240
- vector 44, 51, 246
 - data structure 68
- visibility parameters 70
- zone 114, 248
 - area 119, 145
 - create 117, 121
 - draw 120
 - information 145
 - network 161
 - operation 124
 - overlay 125
- zooming 14

GSC/CGC OTTAWA



OOG 02601458

