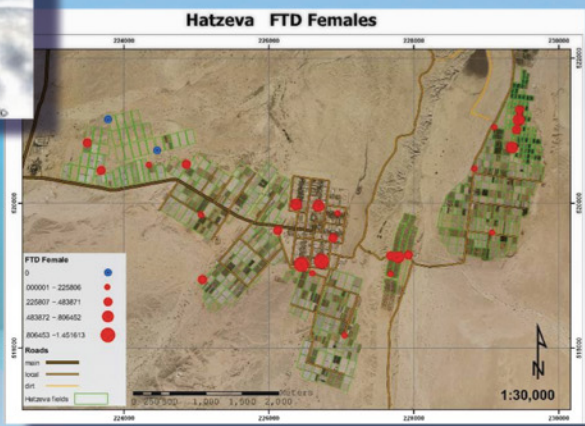
# Designing and Implementing a Geographical Information System

## A Guide for Managers of Area-wide Pest Management Programmes




Sponsored by the Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture

Joint FAO/IAEA Programme
Nuclear Techniques in Food and Agriculture

# DESIGNING AND IMPLEMENTING A GEOGRAPHICAL INFORMATION SYSTEM

## A Guide for Managers of Area-wide Pest Management Programmes

DESIGNING AND IMPLEMENTING A GEOGRAPHICAL INFORMATION SYSTEM
A Guide for Managers of Area-wide Pest Management Programmes

IAEA, VIENNA, 2006
IAEA/FAO-GIS


Printed by the IAEA in Austria
April 2006

# FOREWORD

Over the past two decades, the use of computer software and mapping methods known as geographical information systems (GIS) has been adopted by an ever growing variety of professionals. Every activity that deals with location dependent information can use GIS, and agriculture is no exception. The potential of GIS and remote sensing (RS) to facilitate the planning and implementation of area-wide integrated pest management (AW-IPM) programmes is enormous but unfortunately, these methods are still much underused.

AW-IPM programmes, especially those that integrate the sterile insect technique (SIT) with other surveillance and control methods, would benefit considerably by drawing on GIS/RS. These programmes are often implemented over large areas of even tens of thousands of square kilometres, where surveillance methods are deployed and large data sets are systematically generated on a daily basis. The acquisition of geo-referenced data sets on pest presence/absence, relative abundance, disease prevalence, crop damage, etc., that will allow accurate spatial and temporal analysis is important for proper and timely decision making to efficiently plan and implement any operational pest management programme.

Animal health and plant protection officials and pest control programme managers might be intuitively aware of the importance of employing GIS as an analytical tool. However, they often lack a deeper understanding of its capabilities. Since GIS is a desk exercise using computers, data analysis is often left to the computer staff without proper directives from the programme managers on programmatic needs. This is unfortunate as it will usually NOT bring the desired GIS-processed information to the decision makers.

This manual targets area-wide pest control programme administrators and managers of FAO and IAEA Member States in an attempt to demonstrate the type of data processing and spatial analysis that can be expected of GIS. The manual does not aim to provide guidelines for the detailed operation of software or equipment, but rather tries to give users a grasp of the capabilities and outputs of GIS/RS and how it can help in the planning, implementation and decision making process of operational programmes. The manual should provide programme administrators and managers with sufficient insight to assign to their GIS experts explicitly defined tasks so that management level officers will see a return in useful and constructive feedback.

The IAEA is grateful to M. Silver (GIS Consultant, Arava Development Co. Ltd, Idan, Israel) for preparing this manual.

The officers responsible for this publication were M. Vreysen and W. Enkerlin of the Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture.

**CONTENTS**

# INTRODUCTION

A Geographical Information System (GIS) forges a marriage between unlikely partners. Information systems deal with tables of raw data, often numerical, collated into sums or averages. To the lay observer, a database seems somewhat unconnected to reality. On the other hand, geography offers a very obvious representation of reality. Viewing a map, we recognize roads and rivers, and if the scale is large, we can make out individual buildings, even pinpointing our own house. So how do these two disparate ways of presenting information come together in a GIS? It is in fact their opposite approaches that makes the two halves of the GIS model compliment each other and merge so well together.

A map, ironically, lacks particular information, regardless of how detailed it is. We can follow every twist and turn in a road, but what is the name of the road? How long is it? Paved or unpaved? These kinds of detail require tabular data ‘behind’ the geographical features. In the context of pest management programmes, a map can reveal at a glance the distribution of pests, but without some measure of the level of infestation that glance shows us only half of the truth. On the other hand, we can pour over columns and rows of pest capture data, calculating aggregations such as FTD (flies per trap per day) — however, the numbers alone cannot give us a full picture of pest distribution and relative abundance unless we can visualize where those flies actually are. When the mirror reflection of reality that we see in maps is combined with facts and measurements, then we have a GIS.

# Chapter 1

## ELEMENTS OF EVERY INFORMATION SYSTEM

Before going into geographical aspects, we must first review the basics of every information system (IS).

- Systems analysis always begins at the end. Designing an IS starts by defining what resulting information is to be extracted. The analyst needs to know what questions are being asked, what the final reports should look like, how often reports are to be created and how detailed they must be.
- Based on the end product which is expected from the IS, tables of input data and queries are designed. This stage must cover frequency of data input, level of detail, and how to aggregate the raw data into relevant, understandable information.
- Now a suitable computer software application is chosen to handle the data entry, storage and queries. The family of software which serves this purpose is known as a database management system (DBMS). A DBMS is used to create the tables of data, to construct a user interface (electronic forms for data entry, see Section 2.4.), to design queries and to print out the required summary reports.
- Data must of course be collected and transferred to the computer in a timely fashion, matching the frequency of updates in the initial design requirements. The systems analysis should lay out explicit channels of communications and data flow such that everyone involved in collecting and entering data knows who sends what to whom, in what format, and how often.
- Finally, the crucial but all too often overlooked issue of data backup must be addressed. A backup strategy needs to be decided upon and tested from the start (see Section 4.3). This includes choosing a backup medium, laying down a rigid routine for running backups, and performing test runs periodically to ensure that backup data will be available when they are needed. Not if, but when! Computer systems crash, data can be ‘accidentally’ deleted; thus backups must be an integral part of every IS.

### 1.1. Geographical elements in a GIS

In order to combine tabular data with geographical (or ‘spatial’) data and create a GIS, we need access to additional elements:

Some background map layers must be available to serve as a frame of reference for the geographical data. These might include ortho-rectified aerial photos, scanned topographic maps or satellite images (for small scale mapping). All of these are examples of ‘raster’ layers: graphical images that are geo-referenced so that features in the images appear at their proper location in the map’s coordinate system. Note that any raster layer, to be applied in GIS work, must be ortho-rectified, i.e. geo-referenced and tied to the local coordinate system. Most aerial photography companies offer ortho-rectified products. Satellite imagery, on the other hand, is usually supplied unprojected, that is in a geographical (latitude–longitude) coordinate system. Thus, other local layers will not overlay the satellite image correctly unless it is reprojected to the correct local coordinate system. Reprojection of raster layers is a sophisticated and somewhat expensive procedure, which is unavoidable.

Vector layers — point, line or area — will be obtained for features of interest, and will be used to display tabular data. For example, a point vector layer of trap locations (Fig. 1) might be linked to data of pest captures in the traps. Or an area layer of fields with crop data might be cross-referenced with trap capture data to better understand the distribution of a certain pest. Vector layers can be obtained from survey maps (in CAD or GIS computer file format), or from a government geographical survey department.

*Figure 1. Raster background (aerial photo) with vector layers.*

Often, GPS instruments, barcode readers and other data collectors will be used to map out vector layers and accumulate data in an efficient manner (see Section 1.4).

Needless to say, GIS software will be employed to analyse both spatial and tabular data. Modern GIS applications include tools for querying attributes, querying locations and editing geographical and tabular data, and a rich set of cartographic tools for producing maps. In a GIS, a well thought out, visually pleasing map supplements the conventional print-out from a standard information system.

### 1.2. Interaction between tabular and geospatial elements

In the Introduction, we mentioned merging of tabular data with geographical data. But what exactly does this mean? A sharpened sense for match-making is required to successfully merge database technology with geographical methods. Every geographical layer obviously contains the location details for all features in that layer. So, a layer of traps, for example, will hold the latitude–longitude information for each trap. In addition, every GIS file format also has the capability to maintain a table of information regarding each feature (Table 1).

We might require the trap layer to contain each trap's type, its host tree, the date the trap was first placed, the date it was last checked, etc. The well known GIS 'shapefile' format is actually a collection of computer files: one contains the the geographical data (with file extension *.shp), another contains the tabular data (with file extension *.dbf) and the third keeps the index between the first two (with file extension *.shx). Our layer of traps will be made up of at least three files named traps.shp, traps.shx and traps.dbf. So this popular format for maintaining geographical data handles for us the connection between the geographical location of features and the data connected with those features.

TABLE 1. SHAPEFILE ATTRIBUTE TABLE FOR A TRAP LAYER

Attributes of Traps

| FID | Shape* | ID | HOSTREE | TYPE | GPS_HEIGHT |
|---|---|---|---|---|---|
| 7 | Point | 2062 | fig | Tefri | -146.739 |
| 8 | Point | 2063 | guava | Tefri | -142.143 |
| 9 | Point | 2064 | guava | Tefri | -136.582 |
| 10 | Point | 1041 | citrus | Jackson | -142.701 |
| 11 | Point | 2065 | citrus | Tefri | -137.795 |
| 12 | Point | 2066 | fig | Tefri | -130.891 |
| 13 | Point | 1044 | mango | Jackson | -132.105 |
| 14 | Point | 4014 | mango | Jackson | -134.856 |
| 15 | Point | 2087 | mango | Tefri | -139.231 |
| 16 | Point | 1045 | mango | Jackson | -135.622 |
| 17 | Point | 4015 | mango | Jackson | -141.309 |
| 18 | Point | 2091 | citrus | Tefri | -139.936 |
| 19 | Point | 2092 | citrus | Tefri | -135.252 |
| 20 | Point | 2069 | date | Tefri | -158.225 |
| 21 | Point | 1995 | date | Jackson | -157.671 |
| 22 | Point | 1048 | fig | Jackson | -161.905 |
| 23 | Point | 2071 | date | Tefri | -161.085 |
| 24 | Point | 2070 | date | Tefri | -163.845 |
| 25 | Point | 1046 | citrus | Jackson | -161.614 |
| 26 | Point | 2095 | mango | Tefri | -162.762 |

Record: 12 Show: All Selected Records (0 out of 41 Selected.) Options

However, much of the data in any information system is constantly growing and changing, and this is true also in pest management programmes. Traps are checked on a weekly or semi-weekly basis, and each monitoring cycle means a new set of data. The inherent attribute tables in a shapefile are not designed for handling or accumulating this kind of dynamic data. Furthermore, integrated pest management (IPM) programmes often will consist of field workers in remote locations collecting raw data and sending them to a central office for processing and GIS analysis. So the data collection will be physically separate from the data processing and GIS analysis. For these two reasons:

(1) Rapidly changing data,

(2) The split between data collection and GIS analysis;

we often need to choose a robust database separate from the GIS layers for maintaining tables of trap data. This required separation between the spatial data and the monitoring data means that the geographical layers must always be linked (or ‘joined’, in database terminology) to external data. So, a few obvious requirements for implementing a GIS in the context of an IPM programme come to mind:

(1) A rapid communications channel between the branch offices and the main office;

(2) Everyone must be ‘talking the same language’.

In computer terms, item (2) means a uniform, agreed upon file format for the external data. All GIS software can read and import data files in at least two standard formats: simple ‘flat’ text files (also known as ‘comma-delimited’ or *.csv format (Table 2)), or the well known *.dbf format. Every DBMS can export queries in either of these two formats. So choosing either of these two formats should always lead to a smooth transfer of data.

TABLE 2. A COMMA-DELIMITED TEXT FILE OF CAPTURE DATA

tblCaptures.txt - Notepad

File Edit Format View Help

```
2,1044,2/1/2003 00:00:00,2,4
3,1042,22/1/2003 00:00:00,1,20
4,1043,30/1/2003 00:00:00,1,10
5,1044,3/1/2003 00:00:00,1,10
6,1045,13/1/2003 00:00:00,1,12
7,1047,13/1/2003 00:00:00,2,4
8,1048,23/1/2003 00:00:00,2,15
9,1980,3/1/2003 00:00:00,2,15
10,1996,23/1/2003 00:00:00,2,25
12,1996,2/1/2003 00:00:00,3,20
13,1048,22/1/2003 00:00:00,3,20
14,1042,22/1/2003 00:00:00,2,20
15,1043,30/1/2003 00:00:00,2,20
16,1044,25/1/2003 00:00:00,2,10
17,1048,25/1/2003 00:00:00,2,15
18,1048,25/1/2003 00:00:00,3,25
20,4015,2/1/2003 00:00:00,1,10
21,4015,20/1/2003 00:00:00,1,20
22,4014,2/1/2003 00:00:00,1,20
23,4016,20/1/2003 00:00:00,1,25
25,4012,15/2/2003 00:00:00,1,0
26,1047,25/2/2003 00:00:00,1,5
27,1044,26/2/2003 00:00:00,2,0
28,4012,26/2/2003 00:00:00,3,12
29,4013,26/2/2003 00:00:00,3,22
30,4014,26/2/2003 00:00:00,2,13
31,4015,26/2/2003 00:00:00,2,30
32,4016,12/8/2003 00:00:00,1,10
34,4015,12/8/2003 00:00:00,1,23
35,4014,15/8/2003 00:00:00,3,98
36,4013,15/8/2003 00:00:00,1,1
37,4012,21/8/2003 00:00:00,1,0
38,4016,21/8/2003 00:00:00,2,2
40,4016,25/8/2003 00:00:00,1,9
41,4015,28/8/2003 00:00:00,1,3
42,1041,28/8/2003 00:23:47,3,10
```

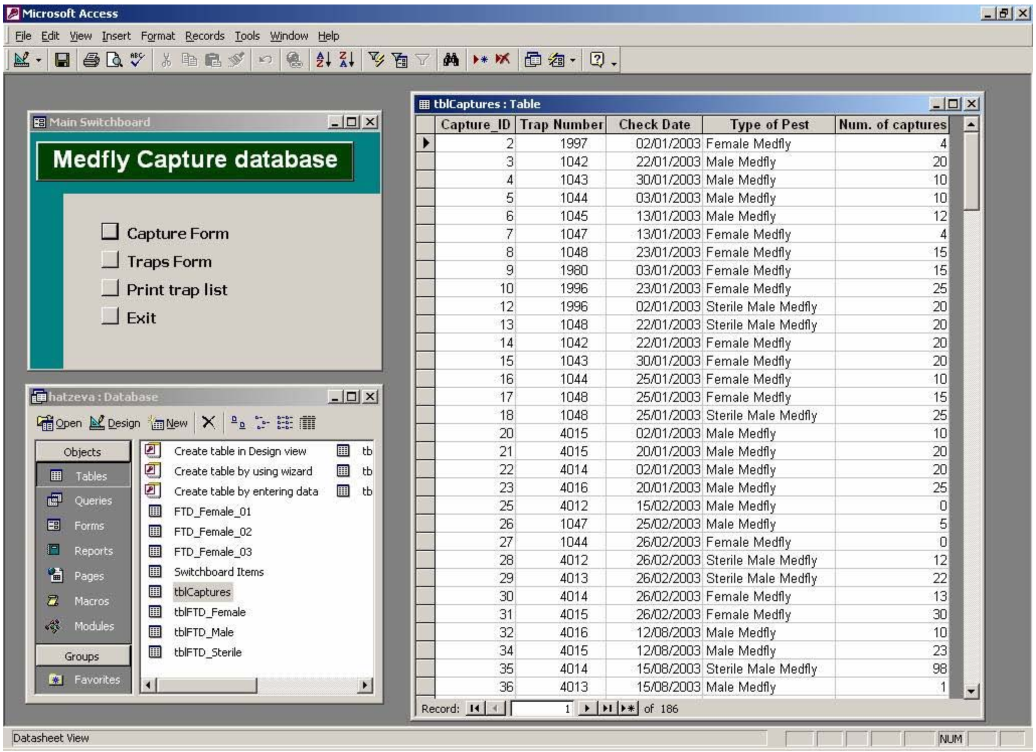
But modern GIS products can open or import a plethora of data file formats, including reading data tables straight from the database itself. This option — a direct, on-line database link — is the preferred approach when the tabular data are maintained in the same computer or on the same network as the spatial data. In every other case, when data collection and GIS analysis are separate, the programme administrator must choose one of the standard formats for transferring data from the field monitors to the GIS operators.

Furthermore, when linking spatial layers to external data tables, another hurdle must be faced. The data and geographical layers must have identical indexes. After all, the purpose of the exercise is to join tabular data with the geographical layers. This join will be possible only if both halves of the GIS data model have some common attribute. For example, pest capture data might be collected at a field office, with each trap identified by its trap ID number in the database. The trap locations, on the other hand, are stored in the layer of traps, identified again by a unique ID number. The exact same ID must appear for each trap in both the geo-referenced layer of traps and in the data table in order to link or join capture data to the trap locations. These dual sets of trap IDs must be kept synchronized. Thus whenever new traps are employed in the field, their geographical location within the GIS must include each trap's unique Trap_ID, and that same Trap_ID must appear in the data tables where capture data will be entered.

In Tables 3 and 4, an external table of capture data ("tblCaptures") has been linked to the geographical layer of traps ("HaTraps") using the synchronized Trap_ID field.

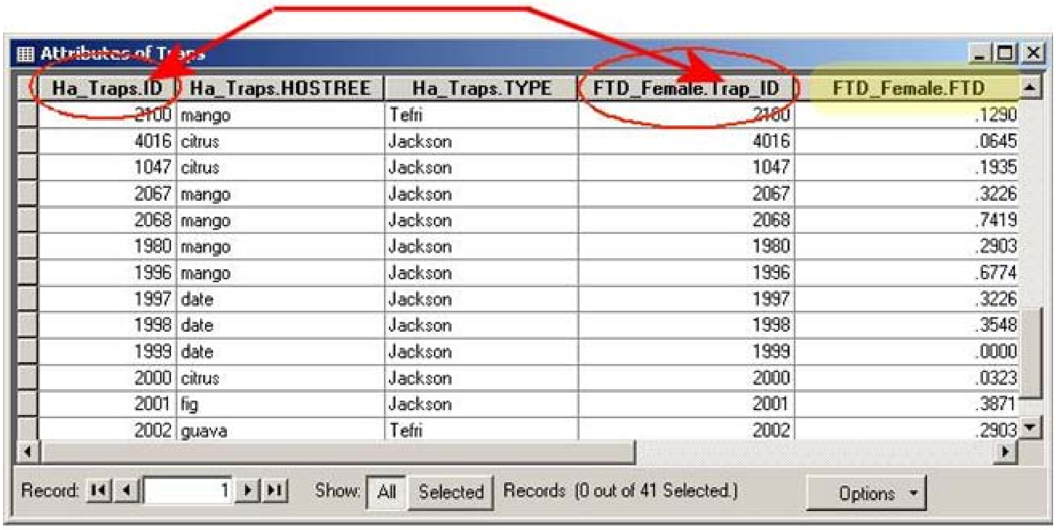
Once we have agreed on a standard file format, arranged for rapid transfer of the data to a GIS operator, and successfully linked the new data to a geographical layer, what next? We now choose from the various cartographic tools in the GIS arsenal to display the information visually with colours

TABLE 3. SAMPLE CAPTURE DATA WITHIN A MICROSOFT ACCESS TABLE




| Capture_ID | Trap Number | Check Date | Type of Pest | Num. of captures |
|---|---|---|---|---|
| 2 | 1997 | 02/01/2003 | Female Medfly | 4 |
| 3 | 1042 | 22/01/2003 | Male Medfly | 20 |
| 4 | 1043 | 30/01/2003 | Male Medfly | 10 |
| 5 | 1044 | 03/01/2003 | Male Medfly | 10 |
| 6 | 1045 | 13/01/2003 | Male Medfly | 12 |
| 7 | 1047 | 13/01/2003 | Female Medfly | 4 |
| 8 | 1048 | 23/01/2003 | Female Medfly | 15 |
| 9 | 1980 | 03/01/2003 | Female Medfly | 15 |
| 10 | 1996 | 23/01/2003 | Female Medfly | 25 |
| 12 | 1996 | 02/01/2003 | Sterile Male Medfly | 20 |
| 13 | 1048 | 22/01/2003 | Sterile Male Medfly | 20 |
| 14 | 1042 | 22/01/2003 | Female Medfly | 20 |
| 15 | 1043 | 30/01/2003 | Female Medfly | 20 |
| 16 | 1044 | 25/01/2003 | Female Medfly | 10 |
| 17 | 1048 | 25/01/2003 | Female Medfly | 15 |
| 18 | 1048 | 25/01/2003 | Sterile Male Medfly | 25 |
| 20 | 4015 | 02/01/2003 | Male Medfly | 10 |
| 21 | 4015 | 20/01/2003 | Male Medfly | 20 |
| 22 | 4014 | 02/01/2003 | Male Medfly | 20 |
| 23 | 4016 | 20/01/2003 | Male Medfly | 25 |
| 25 | 4012 | 15/02/2003 | Male Medfly | 0 |
| 26 | 1047 | 25/02/2003 | Male Medfly | 5 |
| 27 | 1044 | 26/02/2003 | Female Medfly | 0 |
| 28 | 4012 | 26/02/2003 | Sterile Male Medfly | 12 |
| 29 | 4013 | 26/02/2003 | Sterile Male Medfly | 22 |
| 30 | 4014 | 26/02/2003 | Female Medfly | 13 |
| 31 | 4015 | 26/02/2003 | Female Medfly | 30 |
| 32 | 4016 | 12/08/2003 | Male Medfly | 10 |
| 34 | 4015 | 12/08/2003 | Male Medfly | 23 |
| 35 | 4014 | 15/08/2003 | Sterile Male Medfly | 98 |
| 36 | 4013 | 15/08/2003 | Male Medfly | 1 |

TABLE 4. LINKING CAPTURE DATA TO A GEOGRAPHICAL LAYER OF TRAPS IN ARCVIEW




| Ha_Traps.ID | Ha_Traps.HOSTREE | Ha_Traps.TYPE | FTD_Female.Trap_ID | FTD_Female.FTD |
|---|---|---|---|---|
| 2100 | mango | Tefri | 2100 | .1290 |
| 4016 | citrus | Jackson | 4016 | .0645 |
| 1047 | citrus | Jackson | 1047 | .1935 |
| 2067 | mango | Jackson | 2067 | .3226 |
| 2068 | mango | Jackson | 2068 | .7419 |
| 1980 | mango | Jackson | 1980 | .2903 |
| 1996 | mango | Jackson | 1996 | .6774 |
| 1997 | date | Jackson | 1997 | .3226 |
| 1998 | date | Jackson | 1998 | .3548 |
| 1999 | date | Jackson | 1999 | .0000 |
| 2000 | citrus | Jackson | 2000 | .0323 |
| 2001 | fig | Jackson | 2001 | .3871 |
| 2002 | guava | Tefri | 2002 | .2903 |

and symbology (see Section 3.3), to achieve a graphical impact. A GIS layer of cities, connected to a data table of their population, offers one straightforward example: large cities are displayed with a large symbol, and towns with smaller populations with a smaller symbol. If we have a layer of agricultural plots linked to data of a pest infestation, we can display the fields using a colour ramp to show higher infestation with darker colours.

### 1.3. Spatial queries

Let us examine in more detail the possibilities the GIS offers for querying data by location. Using GIS tools we can answer questions such as: What features are near other features? What point features are contained within an area? For example, the owner of a fast food company wants to add a new restaurant to his existing chain of outlets. He might want to find locations that are at least 30 km from existing stores and less than 1 km from a main highway. Then, using population statistics, he would ask how many people live within 20 km of each of the potential locations. Translating these types of questions to pest control terms, a programme manager might need to ask: Which traps are located inside or near fruit orchards? How large an area surrounds each trap (trap density)? Sorting and packing facilities with dumps of spoiled fruit might become hot spots for pest infestation, so a manager might be concerned whether there are enough traps near all the packing houses in that area. A GIS can tell how many traps there are within 1 km of each packing house. When first planning the layout of monitoring traps, a GIS can be used to properly cover an area and ensure the required density of traps. Alternatively, after traps have already been positioned, the manager can request a map indicating how much area each individual trap 'covers' (see Section 3.1).

Furthermore, spatial queries are combined with data queries to extract more information (see Section 3.2). Once the monitoring and capture data are available and linked to the trap locations, we can answer questions such as: Which traps have both FTD (flies per trap per day) above 1.0 and are near residential areas? In SIT projects, after a release of sterile males, a high rate of sterile fly captures indicates a successful release. By cross-referencing traps with a low rate of sterile captures with geographical layers such as elevation, proximity to buildings, etc., the project manager can isolate reasons for the poor distribution and improve the sterile release plan and thus increase efficiency.

### 1.4. How the GPS fits in

The GPS system, an array of 24 satellites rotating around the earth twice a day at an altitude of some 20 000 km, allows anyone with an inexpensive GPS receiver to find their location to within 10–15 m. This extraordinary technology is indispensable to any GIS project in that anyone can 'create' their own geographical layers. For example, a tourist information office might need to produce maps of new trails for hikers, which do not appear on old survey maps. They can send someone out with a GPS, at a much lower cost than for re-surveying the whole region, and obtain very accurate up-to-date GIS layers of the trails. In a similar way, pest control programmes need to map out the locations of their monitoring traps — a trivial mission with a GPS instrument.

We categorize GPS instruments into two groups: navigation units and data loggers. Navigation GPS instruments usually cost between US $200 and 1000. They are designed for recreational use, to allow travellers to find their way in foreign cities or boaters to navigate along a coastline. The more expensive models will have a large colour screen and the ability to upload, store and display road maps or coastline maps. While all navigation instruments have the ability to record 'waypoints', this feature is mostly intended to aid travellers in backtracking their steps, and not as a means for data collection. Cheaper instruments can usually store a maximum of 500 waypoints, and they record only the waypoint number and its coordinates, with no option to add additional details. More expensive units might have memory for 750 or more waypoints, and an option to enter a short text comment with the waypoint number. Transferring a table of waypoints to a computer and converting to GIS layers usually requires some additional software, can be tricky (especially regarding projections into the local coordinate system) and will not contain any information other than the waypoint number.


*Figure 2. Trimble's "GeoXH" Windows CE GPS field computer.*

Data logger GPS instruments (Fig. 2), on the other hand, contain very sophisticated data design and collection abilities. These instruments often cost upwards of US $4000, but they will have a long lasting rechargeable battery, waterproof case, large internal memory and, most importantly, data handling functions. The instrument will interface smoothly with a PC, allowing easy upload and download of GPS data files, as well as with computer software to convert the GPS files directly into GIS layers. The attribute tables inside the GPS are fully user defined, giving total flexibility in collecting a wealth of information — textual, numerical, menu based — together with the geographical location of features. Using a properly configured data logger GPS, the GIS staff can go into the field and collect a variety of features and a full set of attributes for each feature. Then, returning to the office, they connect the GPS to their GIS computer to download the collected data, and within a few minutes they can display all those features with all relevant attributes, correctly located in the local coordinate system within their GIS project.

Which class of instrument should a pest control programme administrator choose? A GPS unit will be used in the early stages of the IPM project to map out fields, crops and trap locations. These kinds of mission require recording of data together with the feature location. The fields might need to include the farmer's name, planting date, type of greenhouse, etc. A trap layer should have the trap ID, type of trap, date first located and host tree. So a data logger GPS is the correct choice. But after the initial chore of mapping out static feature layers, long term monitoring begins, and many programmes expect their monitors to keep an eye on the area and, in addition to routine monitoring, report on irregular events: orchards with fallen fruit, picnic areas with uncollected garbage, untreated weeds infested with pests. For this kind of job, monitors can be equipped with a low cost GPS to be used to quickly record the location of such incidents.

A word about accuracy should be added. Today every GPS unit, whether it costs US $100 or US $10 000, can deliver measurements to a precision of at least 15 m. (Precision can be brought down to less than 1 m, and even better than that, by utilizing a system called differential correction. This system requires that the GPS be 'differential ready', and it will employ a special external antenna which receives correction signals from a base station. Most base stations charge a fee for use of their service.) However, when we apply GPS technology in the context of an IPM programme, where the dimensions of the treated areas are usually tens or hundreds of square kilometres, we have no need for an accuracy better than 10 m. So the question of precision becomes a 'non-issue'. Every GPS, if operated correctly, will give sufficiently close readings.

## 1.5. Remote sensing

### 1.5.1. What is remote sensing?

Remote sensing (RS) refers to the process of collecting details about the earth from a distant vantage point. GIS and RS techniques have developed in parallel over the past few decades, their paths often overlapping, and complementing each other. The complex procedures known as image analysis are common to both GIS and RS. Sources of RS data include aerial photography taken at altitudes of a few kilometres, satellites with sophisticated imaging or radar equipment orbiting the earth at several hundred kilometres, and weather satellites positioned in geostationary orbits (together with communications satellites) at about 38 000 km. Aerial photography, which will be covered more thoroughly in Section 3.5, supplies map makers with a fairly accurate way to create topographic maps and survey maps through the process known as photogrammetry. This procedure uses two overlapping aerial photos in a special stereoscopic viewer to accurately digitize features on the surface. In addition, colour aerial photos serve as an excellent backdrop for other map layers, affording the finished map a very realistic, tangible 'feel'.

High altitude satellite imagery is mostly employed in weather forecasting, and due to the relatively low resolution (the greater part of North America is covered by one such image) has only limited applications for mapping, and only at a very small scale. The chosen altitude of high altitude weather satellites locks them into a geostationary orbit such that they are always positioned over the same spot on the earth's surface. This is ideal for following weather patterns, but not for map making.

### 1.5.2. Low- to mid-altitude satellites

This leaves the third source of RS imagery, which is most often used in GIS applications, from the low- to mid-range altitude satellites. Often these satellites are put into an orbit of a few hundred kilometres, known as sun synchronous polar orbits. They circle the earth more or less north to south, and they return to the same spot of the earth's surface at the same time of day periodically. This allows for easy comparison of images taken over several weeks or months, since each image will have identical lighting. The LANDSAT series, in operation since 1974 (LANDSAT 5 and 7 are still supplying data) orbit at about 700 km, and return to the same spot every few weeks. They cover a swath of 185 km, giving a resolution of 30–100 m. The French SPOT program, first launched in 1986, covers a swath of 60 km, and supplies black and white images at a resolution of 10 m, and colour images at 20 m resolution. SPOT 5 was put into an 830 km high orbit in 2002, and carries instruments which give panchromatic resolution as low as 2.5 m. Each of the SPOT satellites, like the American counterpart, cycles back over the same location on the earth's surface every 26 days, at the same hour of the day. And there are many other RS satellite programs in operation: the Indian Space Research Organization maintains the IRS series of satellites, and in 1999 the highly publicized Ikonos satellite began supplying stunning imagery at a resolution of 1 m in black and white (or 4 m in colour).

TABLE 5. RECOGNIZING SURFACE FEATURES USING MULTI-SPECTRAL SCANNER (MSS) IMAGERY

*(Best MSS bands for identifying surface features (from: http://rst.gsfc.nasa.gov).)*

| Item | Category | Best bands | Salient characteristics |
|---|---|---|---|
| a | Clear water | 7 | Black tone in black and white and colour. |
| b | Silty water | 4, 7 | Dark in 7; bluish in colour. |
| c | Non-forested coastal wetlands | 7 | Dark grey tone between black water and light grey land; blocky pinks, reds, blues, blacks. |
| d | Deciduous forests | 5, 7 | Very dark tone in 5, light in 7; dark red. |
| e | Coniferous forest | 5, 7 | Mottled medium to dark grey in 7, very dark in 5; brownish-red and subdued tone in colour. |
| f | Defoliated forest | 5, 7 | Lighter tone in 5, darker in 7 and greyish to brownish-red in colour, relative to normal vegetation. |
| g | Mixed forest | 4, 7 | Combination of blotchy grey tones; mottled pinks, reds and brownish-red. |
| h | Grasslands (in growth) | 5, 7 | Light tone in black and white; pinkish-red. |
| i | Croplands and pasture | 5, 7 | Medium grey in 5, light in 7, pinkish to moderate red in colour, depending on growth stage. |
| j | Moist ground | 7 | Irregular darker grey tones (broad); darker colours. |
| k | Soils-bare rock – Fallow fields | 4, 5, 7 | Depends on surface composition and extent of vegetative cover. If barren or exposed, may be brighter in 4 and 5 than in 7. Red soils and red rock in shades of yellow; grey soil and rock, dark bluish; rock outcrops associated with large land forms and structure. |
| l | Faults and fractures | 5, 7 | Linear (straight to curved), often discontinuous; interrupts topography; sometimes vegetated. |
| m | Sand and beaches | 4, 5 | Bright in all bands; white, bluish, to light buff. |
| n | Stripped land-pits and quarries | 4, 5 | Similar to beaches — usually not near large water bodies; often mottled, depending on reclamation. |
| o | Urban areas: commercial, industrial | 5, 7 | Usually light toned in 5, dark in 7, mottled bluish-grey with whitish and reddish specks. |
| p | Urban areas: residential | 5, 7 | Mottled grey, with street patterns visible; pinkish to reddish. |
| q | Transportation | 5, 7 | Linear patterns, dirt and concrete roads light, in 5; asphalt dark in 7. |

Common to each of the above programs, the satellite has equipment on board that captures several different wavelengths of light. Most will capture three bands of the visual spectrum, and in addition a few bands of infrared and perhaps also ultraviolet. This equipment is termed a ‘multi-spectral scanner’, or MSS, usually capturing seven bands (Table 5). The most sophisticated equipment uses a technology known as ‘hyperband’ to record up to 250 separate, narrow wavelength bands. Figure 3 shows how different surface features appear in the separated bands of a multi-spectral satellite image. It should be clear why the separation between bands is important. By comparing the strength of the reflected light at different wavelengths, more information can be drawn than from visible light alone.

Using suitable image processing software (IDRISI, for example), the different colours can be chosen to display each band, thus emphasizing certain features. In agricultural applications, for example, vegetation strongly reflects the near infrared (NIR) wavelength. By applying red to the NIR band this 'false colour' rendering will show a cultivated area with healthy crops as bright orange/red, whereas unhealthy crops will appear as light pink. Soil moisture reflects a slightly different wavelength. By examining these two bands, the relation between crop health and irrigation can be tested.

### 1.5.3. How is remote sensing applicable to area-wide pest management?

Remote sensing (RS) imagery can be used effectively and simply as background map layers, just like aerial photos (see Section 3.5). This would pertain especially to the mapping of extensive areas (hundreds of kilometres wide), at a small mapping scale. For example, an administrator needing a country-wide map of pest management efforts might rely on satellite images as a backdrop to the traps and pest infestation layers. Other background layers would be either very expensive or too detailed for the purpose. A LANDSAT image printed at a scale of 1:500 000 (a 1 m-wide printout = 500 km) looks quite impressive.

However, RS technology has a much more elaborate application. During the early stages of planning an IPM programme, only scarce information is available concerning pest infestation, and at times there are not even enough detailed data to map out the cultivated areas and crops. Administrators need to formulate a plan of action despite this high level of uncertainty as to the scope of the problem. In these situations, RS could supply valuable information to fill these gaps. By examining the NIR band, cultivated areas are easily mapped to a relatively high degree of accuracy without the need for expensive, time consuming (and sometimes physically impossible) ground surveys. Using sophisticated image analysis it is possible to determine terrain, topography, and to some extent even soil types and moisture levels. Armed with these geographical layers, an educated guess can be made as to where outbreaks of pest infestation are most likely to occur. These layers might be cross-referenced with climate data to further narrow down the best target regions, thus giving the IPM administrator a toe-hold to begin drawing up a more focused course of action.

Image analysis requires highly specialized software and equipment, and very skilled technicians employing complex mathematical algorithms. These prerequisites are beyond the reach of basic GIS staff members, so IPM administrators wishing to obtain useful RS data would be well advised to seek professional consultants and retain outside specialists to perform the image analysis and supply the requested data layers.

# Chapter 2

## DATABASE DESIGN ISSUES

All too often tables of data are entered (on paper or into easy to use spreadsheets) without serious thought being given to the organization of the tables, to future needs to collate and summarize, or to proficiency in data entry. Somewhere down the line, when the time comes to draw conclusions from the mounds of data, it becomes painfully obvious that the original design of the tables, which might have seemed quite simple at the time, must be totally reorganized in order to summarize and present the desired conclusions. The time for proper database design is before the first row of data is entered.

### 2.1. 'Flat' data tables

We usually speak of two forms of database: a 'flat file' and a relational database. The flat file design can be created on paper, in a computer based text file, or using a spreadsheet program. The principle behind a flat design is simple: each row of data must contain an entry for each of the required attributes in the database. A simple example will clarify the concept: A large corporation is compiling a list of all employees. The list must include the employee's name, age, the department he or she works in, and his/her manager. So it might look something like this:

| Employee | Age | Department | Manager |
|---|---|---|---|
| John | 25 | Sales | Mr. Smith |
| Jim | 42 | Sales | Mr. Smith |
| Jane | 35 | Manufacturing | Ms. Jones |
| Janet | 30 | Manufacturing | Ms. Jones |
| Julie | 45 | Manufacturing | Ms. Jones |
| Jack | 41 | Sales | Mr. Smith |
| Jed | 26 | Manufacturing | Ms. Jones |

The repetitive nature of a flat table is part of its design. Often, in working with such tables, people try to avoid tedious retyping of data and slip into one of two pitfalls. First they might just skip entering the information in some of the rows (i.e. "We know that Ms. Jones is Head of Manufacturing, why write it in every time..."); then later, when the time comes to summarize all employees working under each manager, errors or misinformation will appear. The second pitfall which should be avoided is to sort and accumulate the information into subtables in advance. For example, why not just make a separate table of employees for each department? But then the Vice President asks for a list of all employees over 40, and it becomes necessary to flip through all the department tables to find the information. Both of these pitfalls should be avoided.

Applying the same idea to capture data in the context of a pest control programme, we might see the following:

| Trap No. | Capture date | Trap type | Host tree | Captures |
|---|---|---|---|---|
| 1000 | 10/01/04 | Tefri | Citrus | 3 |
| 1001 | 10/01/04 | Tefri | Citrus | 4 |
| 1002 | 10/01/04 | Tefri | Mango | 2 |
| 1003 | 10/01/04 | Mcphail | Guava | 1 |
| 1000 | 25/01/04 | Tefri | Citrus | 6 |
| 1001 | 25/01/04 | Tefri | Citrus | 5 |
| 1004 | 25/01/04 | Mcphail | Mango | 4 |

Again two pitfalls must be avoided. The first is to skip citing the host tree or trap type again and again, but at the end of the season, when a yearly summary of capture for each type of host tree is required, there will be no way to extract this unless every capture has that detail listed. The second pitfall is to organize the capture data in advance per trap, or per trap type. Again, this will lead to much unnecessary work when information is needed that spans all traps or all hosts.

So a flat-file database is conceptually easy to understand, and easy to design, but requires a lot of work maintaining all the details up to date. Large scale data collection, in an area-wide pest management programme for example, will require a more efficient database scheme. A relational database design is the preferred choice.

**2.2. Relational database**

When choosing a suitable software package for implementing a database, the administrator must consider two important, long term concerns: a robust foundation, and scalability. An area-wide pest control programme with, for instance, 500 traps which are monitored every other week, will accumulate about 13 000 rows of capture data per year. While this amount of raw data can be handled by a simple spreadsheet table, within three years, and anticipating additional traps, the quantity of data could be well over 50 000 rows, an amount which is out of bounds for a 'flat' spreadsheet table. An administrator concerned whether the chosen data storage method will be able to expand to maintain several years' worth of data, and to cover additional traps, will choose a relational database.

This choice offers the added benefit of indexed tables, which are linked together by the database management system (DBMS), thus affording rapid search and retrieval. Figure 3 is a schematic representation of tables related within a DBMS.

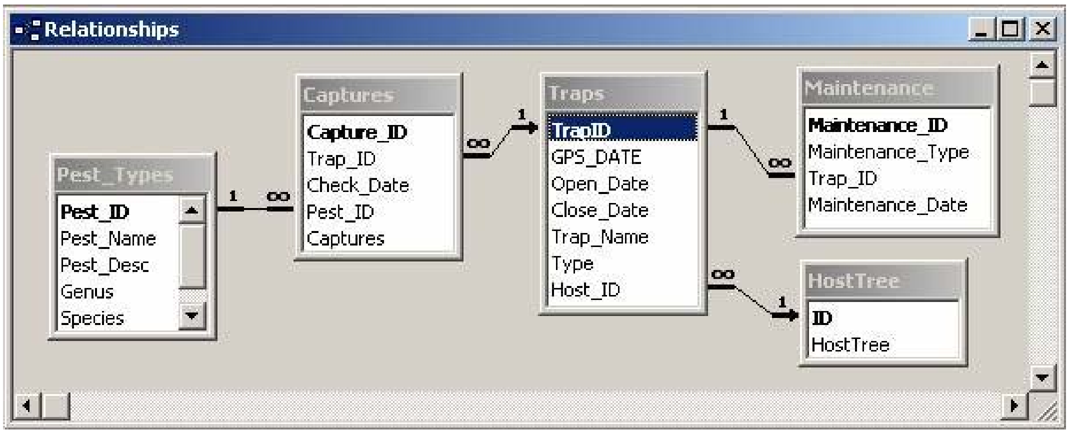



*Figure 3. Diagram of table relationships.*

Again, we focus on the example of trap monitoring in a pest control programme. Looking at the central table "Traps", note that several tables are linked ('related') to it through ID numbers. Each trap can be placed on a host tree. But the host tree name does not appear in the traps table; instead the Host ID (a number) is used to link the two. We might need to keep a record of the maintenance of traps. The maintenance chores are listed in the "Maintenance" table, and each entry will have the Trap_ID of the particular trap that was maintained. On the left is the table holding all capture data, "Captures". Here we see the Trap_ID as before, which hooks into the table of traps, and another index, Pest_ID, which gives the identity of the pest captured.

Now some of the tables in the above scheme will have only a few entries. The host tree table might contain 10 or 20 varieties of host tree. The table of traps will contain as many traps as we have in the project. Clearly the captures table will be the largest. Each surveying cycle it will grow by hundreds of rows — tens of thousands of rows at the end of each season. Yet, if we set up proper indexing as outlined in Fig. 3, the DBMS software will be able to query, summarize or filter very large datasets almost immediately. Moreover, there is no possibility of misspellings or typing mistakes: when we enter a new trap, we chose its trap type from a predetermined list, thus almost eliminating error. Likewise, each capture event will include the index number of the pest captured, which automatically links to the table of pests, giving us the pest's name, description, scientific name, and any other details that we keep in that table. Later on in the project, we might add a photograph of each pest to the database, or give each pest a severity level. Once these new attributes are added, they immediately become available from the traps table also, due to the linked ID numbers. There is no simple way to implement these goals in simple 'flat' data tables.

### 2.3. Characterizing the database

Every effort should be made to complete a detailed design for data collection before an IPM programme gets under way. This includes sketching a flow chart with channels of communication, transfer of files, preparation of weekly reports, etc. But first, programme organizers need to draw up a characterization of the information they expect to gain from their data. It bears repeating that the time for this analysis is before data start accumulating. When characterizing an information system, the following issues should be addressed:

— A database 'scheme' must be laid out which establishes data tables, indexes, links between tables, and exactly what attributes should appear in each table.

— Queries and reports will be prepared individually for different management levels. What averages should be offered to each management level? Field managers will need weekly reports, while the central office might prefer monthly summaries.

— In a GIS, data will be aggregated, or summarized both temporally and spatially. Thus capture data, for example, can be summarized and displayed on a regional basis for a short period of time, or alternatively, a specific location can be isolated and data for that spot averaged over an extended period of time.

— How often are the data updated? How 'fresh' must the data be for each management level? Field managers, for example, must receive their weekly reports immediately upon finishing a survey cycle of traps. Even a short delay will render the information useless to respond to problems in time.

— Data accuracy must also be defined as part of the characterization process.

Answers to all these questions lead directly to the design goals of the database. Now the structuring of data tables, communication channels, reports and final maps all become straightforward.

### 2.4. Ensuring error free data entry

Software intended for database application development comes with a tool set to both structure the data tables (see Section 2.3) and to design data entry forms. Anyone with a basic familiarity with computers is familiar with typing details into a 'text box' or choosing an option from 'list boxes'. These data entry elements offer two important advantages over simple tabular lists: efficiency and avoidance of errors. Choosing from a list is always easier than typing in values for each row of data and makes it impossible to enter an erroneous value. If we must type data into a text box, that text box should be formatted so that no false values can be entered. If a text box should contain numerical values (i.e. for pest captures), then that text box must be designed to accept only numbers. Tricks such as these are contained in all application development frameworks, and they should be taken advantage of to ensure the quality of the entered data.

Further thought should also be given to the finer points of data entry forms. First of all, most data elements should have a default value — the value that will most often be entered. If the majority of traps in a programme are of the Mcphail type, set that as the default in a form for entering new traps. When filling in a date, we will most often enter today's date. So today's date should appear automatically as the default. Next, if the computer operator is selecting a choice from a list box containing hundreds of values, time will be wasted scrolling through the whole list for every data entry. It would be wiser to break the values into 'sublists', thus increasing efficiency. For example, if we have hundreds of traps throughout an IPM programme, the data entry form might first allow the user to choose a region, then display the list of traps for that region only, making the scrolling quicker. Every computer form has a 'tabbing order': which element of the form does the cursor move to when the user hits the TAB key. This should work in an intuitive manner, so that the user can hit the TAB to quickly move to the next field for each data item. Finally, consider which items are required and which are not. Saving of the form should be enabled only when all required items are entered. These mechanisms should be employed to ensure that the daily routine of data collection and entry results in information that is reliable and error free.

# Chapter 3

## GIS APPLIED TO PEST CONTROL PROGRAMMES

The following sections review some 'tips and tricks', i.e. GIS techniques that have direct relevance to pest control programmes.

### 3.1. Planning trap locations

Monitoring is an integral part of nearly every IPM programme. Traps are placed over an extensive geographical area, often baited with food attractants or pheromones. The project administrator must be sure that the data are collected from a reasonable distribution throughout the programme region.

The first method used to enforce an even distribution of monitoring traps involves creating a rigid matrix of potential trap locations in advance, then using these proposed locations to actually place the traps. Simple tools exist within the GIS to produce a spread of points, spaced at even intervals. If traps are placed in a fixed grid array and an interval of 700 m is chosen, then each trap 'covers' about 50 hectares ($700 \times 700$ m = ~50 000 $m^2$). Next, these points can be converted into a file of GPS waypoints and uploaded into a data logger GPS. Now the GPS is used to navigate to each of the proposed locations. Obviously some of the locations will be unusable (on top of a building, in the middle of a lake, etc.), so the final trap locations will not stick rigidly to the original waypoints. But once a final location is chosen, it is immediately entered into the GPS, including all relevant attributes (see Section 1.4). In the end, the final permanent trap locations, captured in the GPS, are transferred to the GIS computer and compared with the original proposed waypoints. This will give a clear picture of the distribution of traps across the region. In Fig. 4, the regular grid of proposed trap locations is used as a guideline to permanently positioned traps. Note that several of the final trap locations are chosen exactly on the proposed grid, some are shifted by less than 50 m, some are shifted as much as a few hundred metres (the red arrows), and some of the proposed locations, which are out of the agricultural area altogether, are ignored.

Often, locations for monitoring traps are chosen based on other considerations — easy access for surveyors, susceptible crops, etc. A programme administrator might need to analyse the distribution of traps after the fact, to ensure that they still give reasonably good coverage. Again the GIS comes to his aid with a procedure known as 'Theissen polygons' (Fig. 5).

This process overlays on top of a given point layer a new polygon layer so that each point is enclosed by exactly one new polygon. Furthermore, the new polygon areas are created in a way such that everywhere within each polygon the closest point is always the point that it contains. Each polygon thus becomes a kind of 'area of influence' for each point. For example, if a Theissen polygon layer is built surrounding trap locations, it is reasonable to expect that pests anywhere in a certain polygon will be captured in the trap within that polygon, since all other traps are farther away. Now, once the Theissen polygons are built, using the GIS we calculate the area of each of the polygons, and it becomes immediately obvious which traps 'cover' an area larger than the recommended trap density.

Remember that the Theissen procedure is only a mathematical calculation. Administrators and surveyors will always be expected to use their knowledge of the region and personal judgment in trap placement. This tool gives administrators one additional, objective view of their trap distribution to supplement other subjective concerns.

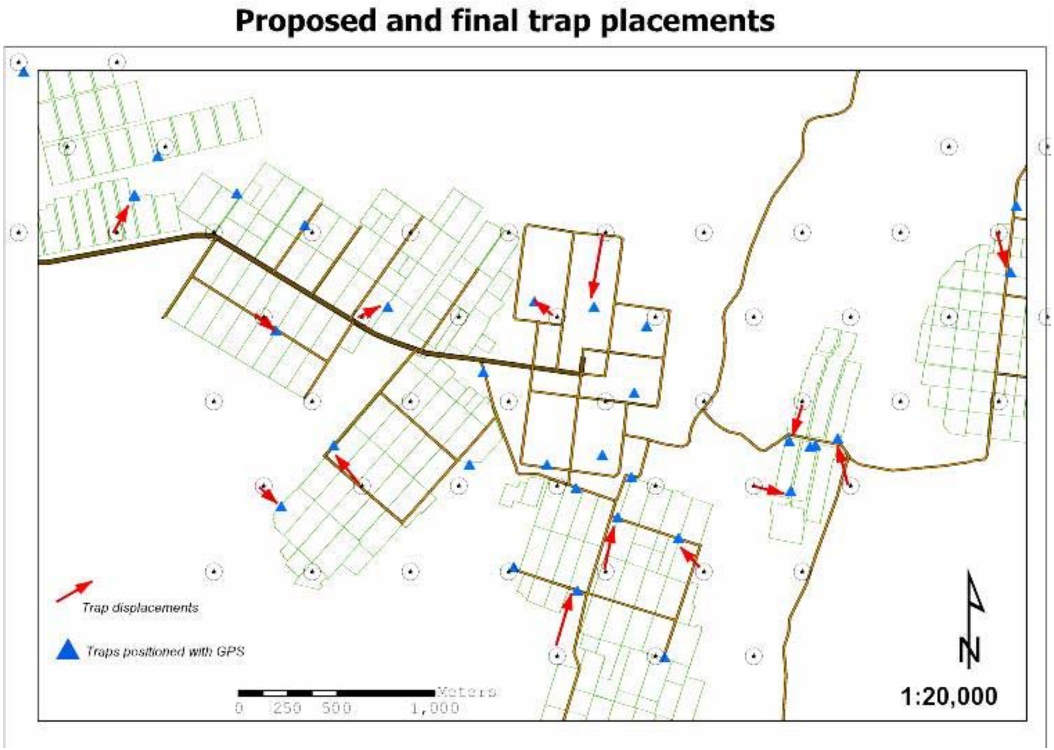



*Figure 4. Proposed and final trap locations.*

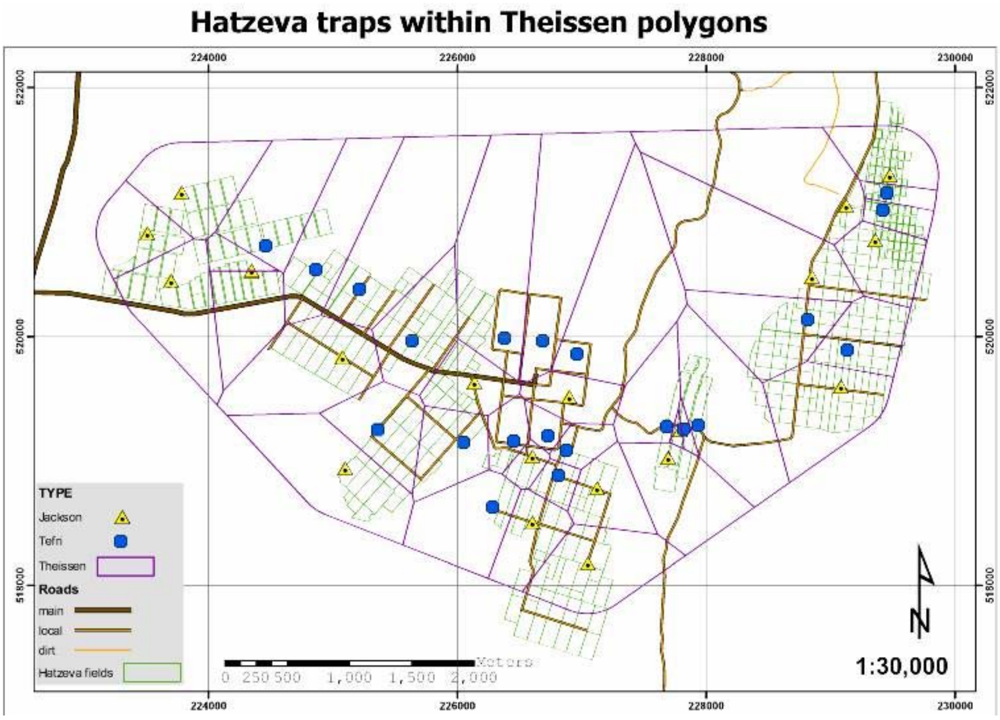



*Figure 5. A trap layer enclosed with Theissen polygons.*

### 3.2. Linking capture data to trap locations

As previously explained, a GIS binds data to their geographical location (see Section 1.2). The information can be linked to a geographical layer either permanently or dynamically. Permanent data attributes are stored together with their location as part of the tabular data within the GIS file formats. For example, the standardized shapefile format includes a *.dbf file (see Section 1.2) which holds the attributes for that layer. As previously explained, static information will typically be stored in this way. For example, a geographical layer of agricultural fields might contain information regarding the farmer who owns the field. This kind of information would not normally change over time. By the same token, traps of the same type might be permanently hung on a certain host tree, so we would include this information together with the trap layer.

On the other hand, capture data will change week by week as the monitoring continues through the season. Crops grown in each field might change each season. This kind of dynamic information is best handled in a database, and then linked to the geographical layer (Table 4). With this arrangement we can take advantage of the ample data entry and query tools that a DBMS offers (see Section 2.4). In the context of a pest control programme, consider the necessity to extract FTD as a measure of relative abundance of insects. Traps are visited approximately every X days. But for each individual trap, X might vary from one monitoring cycle to the next. We must have a query designed within the DBMS to keep track of the visiting date for each trap, and to calculate the FTD for each visit based on the interval since the previous visit. Furthermore, traps might be skipped during the off season for certain crops, then revisited in the coming season. The database application must take into account an option to render certain traps 'disabled' when they are not being checked, then to re-enable those traps again when the crop enters a new growing season. A disabled trap will not accept capture data, and will not be counted in the FTD calculations. A DBMS accomplishes these kinds of queries most efficiently.

As previously explained (see Section 1.3), a common synchronized indexing system must be maintained between the geographical layers and the dynamic tables of data. Each new trap placed in the field must have a trap number or ID, which will be duplicated in the tabular database. Indexes can be copied from the database to the geographical layer, or from the geographical layer to the database. However, one of these two procedures must be chosen. If traps are first numbered in the field, then we need a form in the DBMS to enter the new trap numbers into the tabular database. If, on the other hand, we first enter new traps into the database, then a method must be devised within the GIS software to label each new trap location with its correct ID from the database. Either way works as long as it is used consistently.

One additional note for fruit fly pests: traps are sometimes systematically rotated within a specific site following the phenology of fruit hosts. The programme administrator must decide if he considers this rotation as a single trap whose location changes slightly, or whether each new location constitutes a new trap. If the rotation shifts the trap a small distance, the administrator may choose to retain the original trap ID and thus maintain consistent trap data for a location, while the host tree changes through the season.

In any case, a reliable indexing scheme is a prerequisite to every successful GIS. Once the geographical layer can be joined to an external table of data, then the GIS computer operator can use the whole gamut of cartography tools, the most important being symbology.

### 3.3. Symbolizing pest infestation data

GIS software offers a rich collection of symbology for displaying numerical data. First is the 'colour ramp' or, in mapping terminology, a choropleth: a scale of colours which represents different values.

For example, a ramp from light pink to dark red might indicate low to high captures of a certain insect. This method is most applicable to area features. In other words, if we have a measure of pest infestation in agricultural plots throughout a region, we give each plot a different hue based on the level of infestation.

When the infestation data relate to discrete points (i.e. traps), we choose ‘graduated symbols’ to represent capture data (Fig. 6). Here the size of the symbol for each trap grows larger and larger with increasing capture rates. We might see five sizes of symbols, each indicating a higher range of FTD. This method of symbology gives a very dramatic graphical image of capture rates and immediately isolates hot spots. Note that special consideration should be given to data values of zero. In any plant protection effort, a level of zero is important information and should appear separately on the map. Do not leave the zero value to be lumped together with the low capture rates. Zero captures might be given a colour of grey, and then FTD rates from 0.01 to 0.1 get a small red dot, 0.1–0.5 a slightly larger red dot, 0.5–1.0 an even larger dot, etc.

A special note for SIT programmes: Release of huge quantities of sterilized males, and their subsequent capture in monitoring traps, must be displayed in a way that correctly portrays their role in the project. High rates of capture of sterile males indicate a successful release; low or zero captures mean that the region was missed. So the programme administrator might request an ‘inverse’ graduated symbol — high capture rates shown as small dots, and low capture rates as large dots! Figure 7 demonstrates this approach by representing traps with zero sterile male captures as a yellow square, while traps where sterile males were captured are shown as smaller and smaller orange circles. The larger circle (low or zero males captured) indicates a spot not well covered by the release. On the other hand, the smaller circles show a high capture rate, suggesting a successful release of sterile flies near those traps.

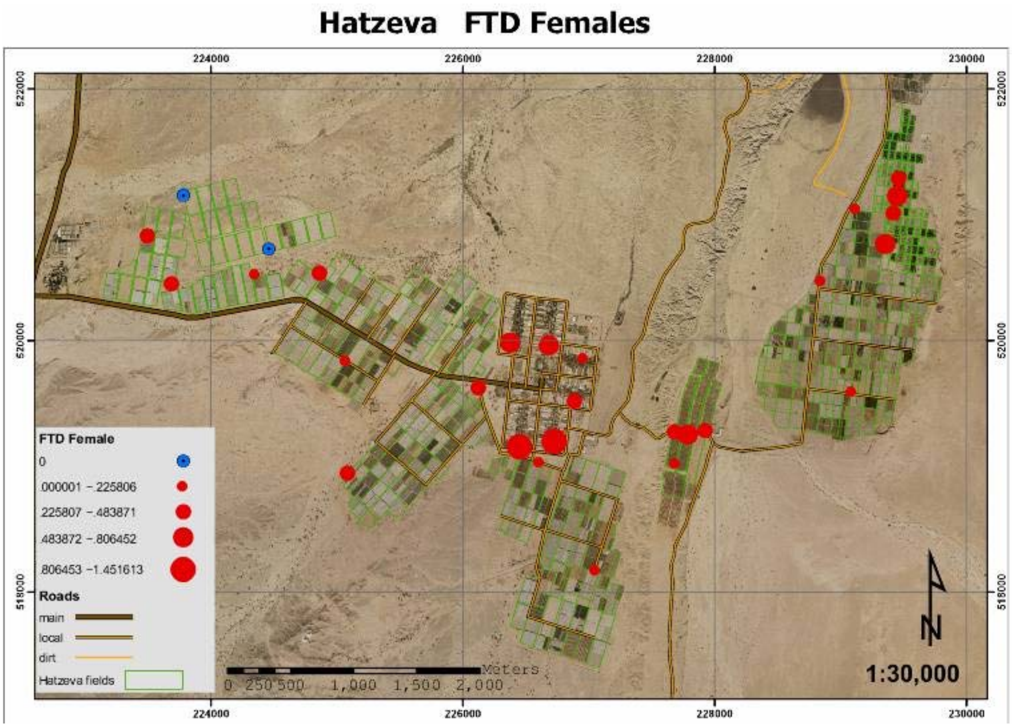



*Figure 6. Capture data displayed as graduate symbols.*

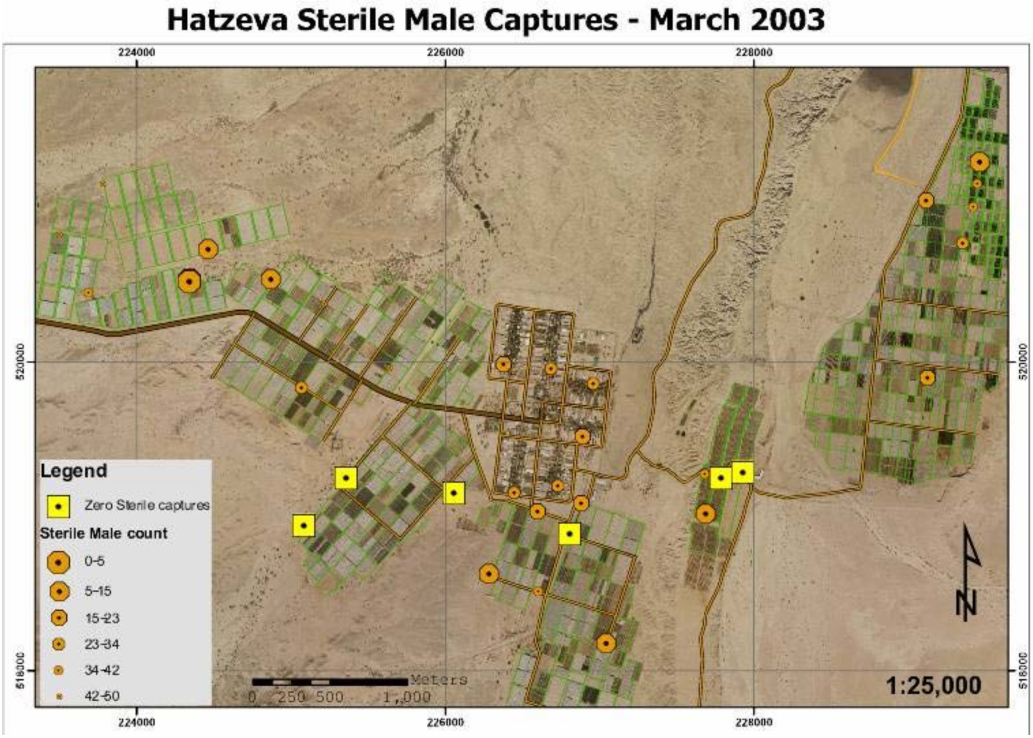



*Figure 7. Displaying sterile male capture data from a SIT project, using inverse graduated symbols.*

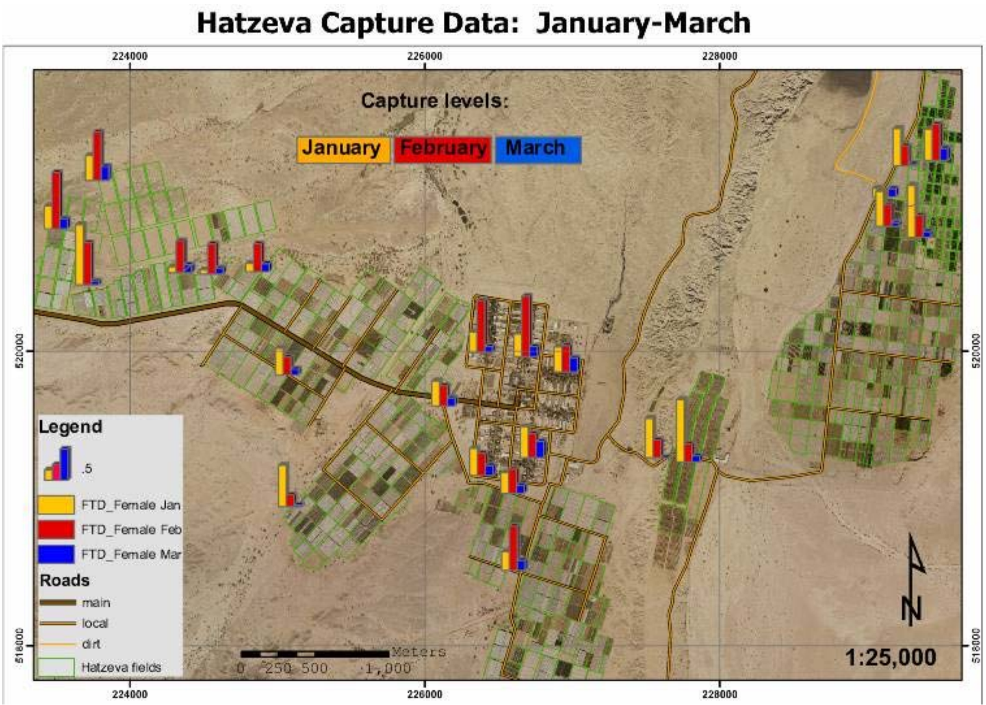



*Figure 8. Display of capture data as bar charts.*

If the need arises to show infestation by several pests on the same map, we can apply the 'chart' symbology option. Here each insect will appear as a bar or pie slice in a small chart at each trap location. In the same manner, chart symbology can be applied to area features — fields — to represent several different pest or disease levels in each field. A bar or pie chart will then appear at the centre of each field, with each bar or slice of the pie showing the infestation level of a different pest.

The chart symbology can also show temporal changes in capture data by displaying each time slot as a separate column. The map in Fig. 8 illustrates average FTD over a three month period, where each column represents one month.

### 3.4. Interpolating point data over a region

Interpolation is the mathematical technique for guessing values in between known measurements. Some classic applications of interpolation in GIS include creating a layer of precipitation for a whole region based on several scattered rainfall measurement stations. Another use is building topographic contour lines from many individual elevation survey points.

A word of caution is in order before applying interpolation techniques to plant protection. Interpolation works only when you can expect the values to be more or less evenly distributed throughout the interpolated region. This is usually the case with rainfall data, or with elevation measurements. It is reasonable to expect the topography to change gradually from one spot to the next (sophisticated interpolation software can receive as input a layer of cliff lines and take these into account to increase the accuracy of the resulting topography layer). Agricultural pests and diseases in general do not fit this model. Local changes in environment — residential areas, fruit orchards — cause drastic changes in the levels of diseases and pests. So interpolation will be applicable only over large homogeneous regions (same crops) with a high density of traps.

After verifying that a project is suitable for interpolation (based on the above reservations), an administrator can employ the technique to obtain a 'surface' of pest infestation. This surface is identical in concept to an elevation layer, only instead of height above sea level, the values represent pest data. The surface can be viewed using a continuous colour ramp (an 'isopleth') or as contour lines representing lines of equal levels of pest infestation. Figure 9 demonstrates this technique by interpolating precipitation data (over two drought years) from many rainfall collection stations, creating a set of contours of average precipitation per year.

The three mathematical methods for interpolating a surface from discrete point values are IDW (inverse distance weighted), spline and kriging. Both the second and third methods create a 'smooth' surface, and can be applied to rainfall data, where there is not necessarily a connection between the value at one point and the next. Splining forces the final surface to go through each of the original data points (best suited for elevation points), while kriging does not, thus smoothing the surface even more. In a plant protection situation, IDW is the best choice, as it gives greater weight to closer points when creating the surface. We would assume that if a certain trap gives a high capture rate, then any location near that point should also show a high capture rate in the resulting surface. This is the assumption IDW is based on. The resulting surface is more 'jagged' than the two other methods, but that should give us an estimate that is closer to the real infestation levels.

To round out this topic, we will review the option of displaying pest capture data for each individual field or plot in a region. A GIS can perform spatial queries, so we can average trap capture data for all the traps within each plot in an agricultural region. This average value can be used to display infestation (see Section 3.3) for each plot using a colour ramp. While this is not strictly an interpolation of data, we should not ignore this simple, straightforward procedure as a means of obtaining another area-wide representation of average infestation.

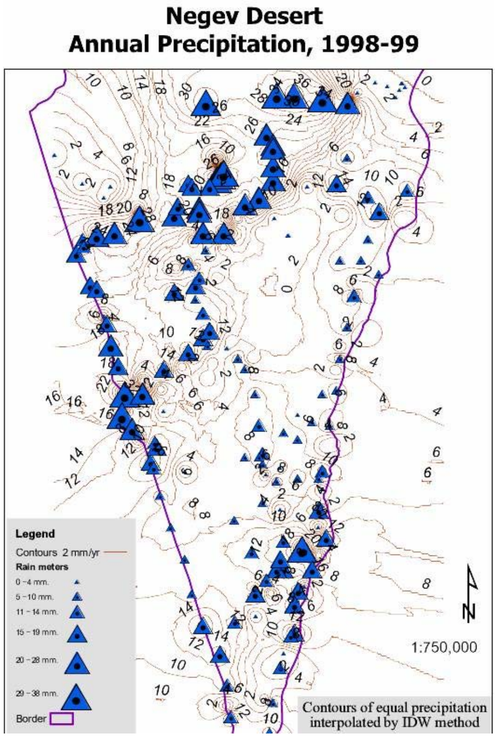



*Figure 9. Annual precipitation based on rainfall data, averaged over the Negev Region using the IDW method.*

### 3.5. Background layers

When searching for and acquiring geographical background layers, the IPM administrator must balance several considerations. If the area covered is small (less than 50 km long), a map of scale 1:50 000 will print out a plot of 1 m in length. In this case, all data layers should be suited to scales of 1:50 000 or better. A country-wide IPM programme covering hundreds of kilometres will require printing at a scale of 1:500 000, and the detail in high resolution geographical layers will be lost. In that case, there is no reason to bear the expense of high resolution imagery and data layers.

The administrator needs to be aware that all layers should be projected into the same coordinate system. This is often the case, since all sources of geographical data within one country will usually standardize on one national coordinate system. However, there might be exceptions to this rule: for instance, geographical data obtained from international sources will usually be non-projected, i.e. degrees latitude and longitude, also known as a “geographic coordinate system” based on the WGS84 Datum. And large countries will keep data in local coordinate systems for each region (in the USA, each State has its own coordinate system based on different projections). All modern GIS software supports reprojection of data layers, provided the original projection is known. So it is important to know both the resolution and the projection of any data source.

Searching for and collecting background layers begins with the inception of the GIS. Depending on budget restraints, aerial and satellite imagery may be purchased. Among the various background GIS layers, aerial photographs are by far the most impressive. First, they give an air of reality to the finished map which cannot be matched by any vector layers. Second, aerial photos can be used to create vector layers by tracing outlines of features on the photo. This process is known as ‘heads-up digitizing’. While not as accurate as true digitizing done by a surveying shop (photogrammetry), for some uses it can be sufficient. Aerial photographs must be ortho-rectified before they can be applied to

GIS use. A non-rectified aerial photo is useless for mapping! Ortho-rectification includes two separate processes: correcting distortion due to the angle of the photograph and anchoring the corrected image to a local coordinate system. The end result is that other layers, obtained from other sources, should overlay the aerial photo in the proper location.

An important consideration when acquiring aerial photography is the resolution required. The higher resolution images are much more expensive, so take into account the scale of maps that will usually be required, and request aerial photos to match. By the same token, satellite imagery (see Section 1.5), which is usually of lower resolution than aerial photographs, can be acceptable for very small scale maps. For example, a colour aerial photo with a pixel size (the small spots that make up the digital image) of less than 1 m is perfect for 1:10 000 scale maps. Landsat 7 images have a pixel size of about 30 m, which will give a sharp view at scales of 1:500 000 or smaller. The French SPOT satellite images offer better resolution (pixel sizes down to 5 m).

Some understanding of computer file compression techniques will help when dealing with aerial photography or satellite images. These images give very high detail and thus create quite large computer files. The standard format for high quality graphical images is TIFF. A colour aerial photo covering 10 km × 10 km can become a 250 MB TIFF file or larger. A mapping project that uses several such images will take a long time to load and will be slow at each redrawing of the screen. Two popular compression systems are in wide use for graphical files brought into GIS software. The company LizardTech, Inc., offers the MrSID format, and ER Mapper (Earth Resources Mapping) has the *.ecw format. Compressing the original TIFF files with either of these formats will greatly ease the management and display of large image files. It is important to note that the popular *.jpeg and *.gif formats used in digital photography, web pages, etc., are NOT applicable to GIS. Both of these are 'lossy' compression techniques which achieve high compression by changing the original image, either by reducing its original colour scale or by changing the pixel resolution. In either case the image will be altered, rendering it of poor quality for mapping.

Often, topographic maps are scanned and brought into the GIS as background imagery in a similar fashion to aerial photography. The same considerations regarding scale, rectification and file compression apply to this source of background layer also.

Every pest control programme will require a set of vector layers including district boundaries, roads, rivers, towns, etc. Often these geographical data can be obtained from the government survey office. Furthermore, the pest control programme staff will want to supplement the above standard vector layers with data directly relevant to the programme: trap locations, fields, crops, monitoring routes, flight paths of the release planes (in the case of SIT). These data will be obtained from a variety of sources. Next, a GPS can be used (see Section 1.4) to map out features that do not exist elsewhere. This is a quick way to add vector layers, which would otherwise be unavailable. Drawings done in CAD format such as cadastral layers, land use maps, surveys for building permits, etc., can be converted into GIS layers. Often the Department of Agriculture will have maps of crops or soil types. The meteorology department might have rainfall or climatic layers. Thus programme administrators will have to seek out and collect the needed geographical data, while weighing the importance of each layer against the cost of purchasing, or time required to map out the features with a GPS. All these steps take time, and need to be put in motion as part of the initiation of the GIS department.

## Chapter 4

## STRATEGIES FOR IMPLEMENTATION

The initial steps taken when first setting up your GIS will determine to a large extent the success of the data collection and management efforts as a whole. Getting started 'on the right foot', weighing each design issue and choosing the proper tools and methods from the outset, leads to a fruitful implementation of the GIS. This final chapter presents in a more 'cookbook' form those early steps to implementing a successful GIS within an area-wide IPM programme.

### 4.1. Planning the format for maps

Always in characterizing an information system, the expected final results, the end product of the efforts, will dictate much of the process which will lead to those desired results. Thus a program administrator must first define several parameters regarding the output of the GIS. These includethe following points:

— Which background layers will be used (see Section 3.5);

— How large an area each map will cover;

— What scale, and thus how large a printout, is needed;

— How often maps will be printed;

— What data will appear, and how they should be presented (see Section 3.3);

— What tables of numerical data should be included with maps;

— Who should receive copies, how often.

### 4.2. Choosing equipment and software

GIS hardware requirements begin with a stronger computer than is normally used for regular office work, for two reasons. First, many GIS functions are quite computationally intensive, involving many mathematical operations when displaying and querying geographical layers. Second, GIS layers (especially aerial or satellite imagery) are often quite large computer files. So the demands placed on the computer system include a fast CPU, large RAM memory, a good video card and a large, fast access hard disk. Most computer manufacturers offer a 'workstation' class model which is built especially for graphics design or CAD uses and addresses all of the above requirements. At the time of writing this manual, the current technology includes a Pentium IV processor running at 2.4 Ghz, with 1 GB of DDR RAM. The latest hard disk technology, ATA-200, offers very fast access, and disks of 80 GB or more are common. And of course, a large, high resolution monitor (with a dot pitch of 0.21 mm or better) should be chosen. A CDRW for burning compact disks is pretty well standard today.

A means of making reliable backups (see Section 4.3) should be purchased together with the computer. For many years the accepted method for backup storage was on magnetic tapes or 'zip' drives. The tapes were relatively cheap compared to other media. Today, the cost of hard disk drives has dropped to the point that purchasing a secondary hard disk just for backups makes the most sense. An additional disk is more reliable than magnetic tapes, and it can be installed in an external case which is disconnected from the computer for 'off-site' storage after each backup run.

Every GIS operator wants to print out presentable colour maps. A good inkjet printer, capable of printing A3 sized paper (30 × 42 cm), is a requirement. A GIS office serving many areas, with a heavy demand for printouts, may choose a large format plotter for producing full-size maps. Plotters are classified by the width of the paper roll that they use (60, 90 or 120 cm) and the speed of throughput. Project administrators should be aware that a plotter is a heavy initial investment and the maintenance (ink cartridges, etc.) is also expensive.

As explained (see Section 1.4), GPS instruments can be roughly divided into two categories: navigation and data collectors. The distinction between the two is not clear-cut, but generally the cheaper units are designed for recreational use, i.e. navigation only; these allow saving point locations without the ability to save attribute data. A high-end data collection GPS will interface directly with the GIS program and enable the operator to save a wide variety of information together with the location of geographical features. Both types of unit have a place in IPM programmes. The simple navigation units can be employed for spot checks: surveyors can keep a GPS on hand all the time, and quickly record unusual problems. A data collection unit will be chosen when the goal is to map out new features with their attributes or add new data to existing features.

Administrators of IPM projects need to choose three software packages when implementing a GIS data collection system. First, a database program is chosen which matches the extent of the data to be collected. The popular MS Office suite comes with the Access database which can easily handle all but the largest IPM programmes. Since many users purchase this software suite for regular office use anyway, the Access database is often a reasonable default for maintaining trap collection data. Larger, country-wide programmes will need to consider a more efficient, multi-user database management system (DBMS). This choice of database software should be taken in collaboration with a database administrator and a hardware/software solution provider. The decision will include a server capable of handling the database, as well as a long term contract with the solution provider for maintenance and support.

Next, the GIS software itself should be chosen based on availability of support and software upgrades. In this highly dynamic field, problems come up all the time, and software vendors often supply patches and upgrades. Administrators should assure themselves that they will have good service and support for whatever GIS software they choose. The most popular GIS package available worldwide is Arcview, marketed by ESRI. Due to its popularity, a very broad base of Arcview users has developed over the years, and they freely exchange methods and solutions to many common problems over the Internet. However, Arcview is not cheap, and by no means the only GIS software available. An administrator with a compelling reason can successfully employ other GIS programs, provided there is access to support when the need arises.

Finally, a software program must be chosen to allow the GPS units to interact with, and transfer data to, the computer. High end data logger GPS units will come packaged with sophisticated software for connecting and transferring data to and from the GPS, as well as programming various functions in the GPS itself. Many of the simple navigation units (notably the Garmin units) can also connect with a PC using free software available for download from the Internet.

### 4.3. Dictating a backup strategy

The importance of proper backups cannot be overstated. Procedures for making backups need to be determined and tested before data collection begins. A well designed backup strategy should cover:

— Choice of a suitable backup medium. First the total amount of data, geographical and tabular, to be collected over an extended period (two seasons' worth, for example) should be estimated. Then the backup medium can be chosen with a capacity for all the data. Some smaller projects can dump all their files on rewritable CDs which can store 640 MB. Currently, with the dropping cost of conventional hard disk drives, the best choice for a small GIS office would be an additional hard drive in an external case, dedicated to storing backups. This medium is the most reliable of all means of data storage, and, with a detachable external case, the drive can be safely stored off-site. For larger GIS installations, data will be kept on a network server which will often have a tape drive holding tens of gigabytes of data. Even larger networks will install a special backup server with an array of tape drives, and software to carry over backups from one drive to the next.

— Balancing daily 'incremental' backups and periodic 'full' backups. In order to guard against faulty backup media (or lost media), several 'generations' of backups are kept. At least one copy is often kept off-site to ensure availability of the data in the unfortunate case of theft, fire, etc. So if a full backup of the data is run weekly, a careful plan will require maintaining additional separate copies, covering past weeks. A thorough plan will also include daily backups of just the new and altered files, and these daily runs will also be kept for several generations. So in the worst case scenario of total loss of data, first the most recent full backup must be restored, and then the daily incremental backups since that full run must also be restored. If, for whatever reason, the most recent tapes/disks are not available, then restoring from the two week old full backup and the previous incremental runs will result, at the worst, in the loss of one week's worth of information.

— Appointing a backup administrator. While it is convenient and desirable to automate the backup process with scheduling software, someone must be in charge of switching tapes and checking that each run completes successfully. The administrator must understand the schedule of full and incremental backups, and must have available off-site a storage location for one copy of each run. Backups are useless unless they can be restored! Insist that the backup administrator rehearse the restore process by staging an unexpected loss of data and asking that the 'lost' files be restored.

### 4.4. Planning the data flow

Based on strategic decisions (from Section 4.1) regarding programme goals, a flow chart should be drawn up to clarify to the whole staff who transfers what to whom, and when. First a form (whether paper or computer based) will be designed for collecting raw pest data from the field. Surveyors need to understand exactly how to fill in the information, and how often the form is transferred to data collectors and by what means (paper forms will be faxed, computer files sent by email, etc.). Of course, a computer based collection scheme requires choosing a standardized file format (see Section 1.2).

Next, data collectors must enter the new data, merging them into the growing database. They must understand the database tools for querying, averaging, etc., to extract the required reports. They should be held to a schedule of periodic reports based on the criteria in the program characterization. These periodic queries will be transferred to the GIS operators for map production. Again, file formats will be agreed in advance, so that importing fresh data into the GIS will be seamless.

GIS operators will have explicit instructions as to what maps to produce, how often, and to what scale. They will know when to expect new data files, and how to symbolize the results. Thus they will be capable of providing printed maps immediately upon receiving new data tables. Advanced GIS software packages offer a free viewer for GIS layers. The GIS staff should be well versed in installing and using these viewer programs, in order to make distribution of finished maps to the program staff more efficient.

Finally, everyone who needs to see the results should be aware of the time schedule, so that each distribution of new maps can be anticipated. Field surveyors might receive partial, detailed maps of their region every week, whereas programme administrators will expect bi-weekly maps of the overall IPM programme. It is the programme administrator's responsibility to plan a complete data flow cycle in the very early stages of implementation.

### 4.5. Refining the process

We close with a recommendation to review and refine the data collection and display process every so often. Bottlenecks or inefficient procedures will always creep up; they should be identified and corrected quickly. New techniques will become available, such as new features in the software to improve map design or distribution. As an IPM programme grows in extent, methods that were sufficient when the treated area was small may have to be replaced. Perhaps the detail or frequency of map printing needs to be changed as the programme matures. A mechanism should be established to examine the data flow and GIS operation once or twice a year, to pinpoint weak spots and draw the necessary conclusions.

### 4.6. Recommended equipment

To be well equipped, a GIS office would require at minimum the items listed in Table 6.

TABLE 6. MINIMUM EQUIPMENT FOR A GIS OFFICE

| Category | Item | Model/version* |
|---|---|---|
| GPS | Trimble hand-held GPS | Geo XT |
| | Trimble Terrasync® | 2.5 |
| | Trimble Pathfinder Office® | 3.0 |
| GIS software | ESRI Arcview® | Arcview 9.1 |
| | ESRI extension | Publisher |
| | ArcReader | (Installed on non-GIS computers) |
| Database | Microsoft Office® | MS Access 2003 |
| Computer and peripherals | Workstation | Pentium IV 3.0 Mhz, Windows XPPro<br>1 GB DDR2 memory<br>120 GB SATA disk<br>Video card with 128 MB memory<br>Large, high quality monitor<br>Internet access |
| | Backup drive and media | CD/DVD writer, or<br>external hard drive, or<br>zip or tape backup drive |
| | Printer | Colour inkjet for up to A3 sized paper |

* Model or version details as of October 2005.

Table 6 represents the basic requirements for a standard GIS office. These products are widely used in mapping and professional GIS/GPS applications. However, there are many alternative software packages and GPS instruments available, offering specific features and covering a wide price range. Additional information and product reviews can be viewed on the Internet at the following web sites:

GIS Software:

http://software.geocomm.com/reviews/
http://gislounge.com/library/blsoft.shtml
http://gislounge.com/ll/gissoftware.shtml
http://dmoz.org/Science/Social_Sciences/Geography/Geographic_Information_Systems/

GPS instruments:

http://www.digitalgrove.net/GPS.htm
http://dmoz.org/Science/Earth_Sciences/Geomatics/Global_Positioning_System/

## BIBLIOGRAPHY


GLEEN, D., Designing Information Systems, Open University of Israel (1989) (in Hebrew).

MINAMI, M., Using ArcMap, ESRI Press (2000).

MITCHELL, A., The ESRI Guide to GIS Analysis, ESRI Press (1999).

SHORT, N.M., Remote Sensing Tutorial (1999), http://rst.gsfc.nasa.gov/Front/overview.html

THURSTON, J., POIKER, T.K., MOORE, J.P., Integrated Geospatial Technologies, Wiley and Sons (2003).

TRIMBLE NAVIGATION LTD (Sunnyvale, CA)

PathFinder Office Getting Started Guide (1999).

GeoExplorer Series Getting Started Guide (2003),
http://www.trimble.com/geoxm_ts.asp?Nav=Collection-37915

Terrasync Operation Guide (2003),
http://trl.trimble.com/dscgi/ds.py/Get/File-140657/TerraSync.pdf