# LOW COST SYSTEM FOR MAP INCORPORATION INTO A RASTER OR VECTOR GIS

M. Campos[1], F. Diego[2] & X. Pons[3] [2]

[1] Departament de Ciències Ambientals. Facultat de Ciències Experimentals i de la Salut. Universitat de Girona. Pl. Hospital núm. 6 17071 Girona. Catalonia.
[2] Departament de Biologia Vegetal, Biologia Animal i Ecologia. Unitat de Botànica. Universitat Autònoma de Barcelona. 08193 Bellaterra. Catalonia.
[3] Centre de Recerca Ecològica i Aplicacions Foretals, CREAF. Facultat de Ciències. Universitat Autònoma de Barcelona. 08193 Bellaterra. Catalonia.
Spain. E-mail: ibbt0@cc.uab.es

## Abstract

This paper explains the design and setup of a method for digitizing and georeferencing a large number of thematic, A0, maps. The basic hardware is a PC and a scanner, used in the first step of the process. The software used is composed by a general raster editor (paint facility) and by different programs that accomplish specific tasks (many of those written for this work).

The methodology is based on several steps:
STEP 1: Map scanning (copying the map in a transparent film sheet -acetate- and/or adding georeferenced points if necessary)
STEP 2: If the map contains polygons, creation of a file containing thematic attributes for each polygon.
STEP 3: Manipulation of files coming from the scanning process. Editing and skeletonizing.
STEP 4: Creation of a file containing the control points for geometric transformations.
STEP 5: Raster-to-vector conversion or raster completion.
STEP 6: Georeferencing and estimating planimetric accuracy.
STEP 7: GIS integration.

The main advantages of this methodology are: a) The low cost of the required equipment (a PC is enough because the scanning process can be made in some specialised establishments). b) The high geometric fidelity of the features of the obtained map (even if they are copied over acetate, because this process is more accurate than the polyline conversion produced when the edges are manually digitized). c) The savings in terms of time and money, specially when the map complexity increases.

## 1 Introduction

The input of cartographic data into a raster or vector GIS is an aspect that can take a very important part of the time of persons and computers involved in cartography. Digitizing thematic maps of polygonal nature is usually done with a table digitizer (tablet) although semiautomatic systems based in scanners already exist (it may take 20-40 person hours to digitize the boundaris on a 60x40 cm size 1:50 000 soil map [2]). These systems are usually oriented to vector line generation, which later will be incorporated to a vector GIS [1, 2]. The use of scanners for entering maps into raster GIS is not very common and, when exists, it is based on image processing techniques, usually linked to equipments with special graphic memories, not easily available for large capabilities [7, 10]. The enormous cheapening of microcomputers and the increase of their RAM speed allows making these processes possible on the central memory of these machines, bypassing the need of A0 digitizing tables and expensive equipments. Moreover, if the result is not intended for a vector GIS it is also possible to avoid the process of building topological relationships.

The main goals of this work are: a) Develop a methodology for entering printed maps into a GIS by scanning the maps. b) Obtain a low cost methodology and, usually, a saving of operator time. c) Obtain a geometric quality as least as good as the obtained with the traditional digitizers.
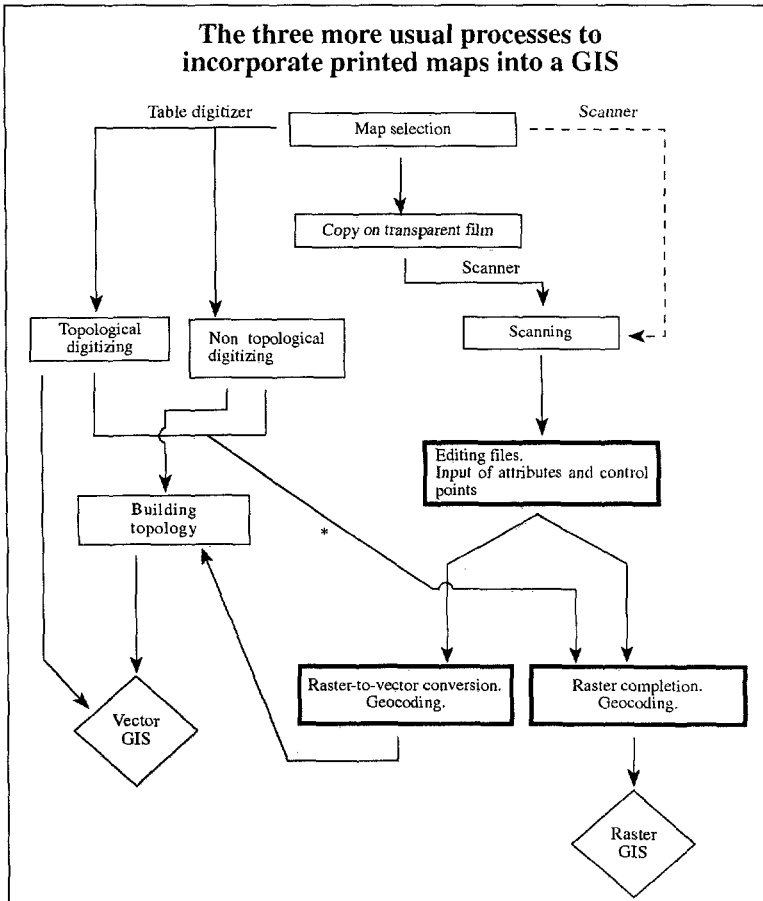


**The three more usual processes to incorporate printed maps into a GIS**

**Figure 1:** The three more usual processes for incorporating printed maps into a GIS. Dashed lines indicate unusual processes. Bold-face boxes contain processes specially developed in this work.
* Indicates that a vector-to-raster conversion of polygon limits should be done.

## 2 Methodology

The methodology we expose allows digitizing and georeferencing printed maps with the aid of a scanner and usual techniques from image processing and cartography. These data are then entered into a raster o vector GIS [3, 4].

The essential hardware is just a 32-bit personal computer (386 or better) with VGA and high capacity hard disk and a scanner (that can be hired). On the other hand, the essential software is a raster drawing program for handling the binary images, a program for vector editing able to display a background raster file, and a set of specific programs (skeletonizing, vector-to-raster and raster-to-vector conversion, map generalization, georeferencing,...).

The steps of this methodology are explained below (Figure 1 and 2).

### 2.1 Map scanning (copying the map in a transparent film sheet -acetate- and/or adding georeferenced points if necessary).

Obviously, this process begins with the selection of maps that contain the geographical information we want to enter into the GIS; once selected, maps will be scanned. In the scanning process it is important to register the lines that define the geographic objects as well as the attributes of each one of them. For that reason, and for the maps matching these conditions, before scanning we usually draw these lines and attributes over a transparent film sheet (acetate, vegetal, etc...), which will be scanned. We will also draw in the transparent film sheet (with a special symbol) several control points having known map coordinates; so that, thesey marks will be represented in the digital file produced in the scanning process, allowing the georeferencing process.

### 2.2 Creation of the file with the thematic attributes from the inner parts of each polygon.

At this step we use a software that allows opening the scanned (raster) file as background while digitizing and editing a vector files. Then we digitize, inside each polygon, the corresponding attributes (as text entities), so creating a vector-point file. Having more than one attribute for each polygon (if they are equivalent) will not be a problem, because the rasterizing software we use takes it into account; on the other hand, this can help the labeling process.

This is a manual process, but it can be quite fast if we scan the polygon boundaries and the polygon attributes together. In the case of maps with linear entities, it is not essential to enter the attributes in this step.

### 2.3 Handling the files produced in the scanning. Editing and skeletonizing.

The raster file produced in the scanning contains many elements that must be removed or modified. The involved processes are: a) Removing symbols identifying polygon attributes (they are already digitized). b) Closing the polygons that remain opened because they have not been well drawn in the scanning process. c) Opening the parts of the polygons strangulated by the excessive nearness of their boundaries. d) Removing the points or lines that do not correspond to any line or polygon boundary (noise). We need a raster drawing editor with paint and erase facilities to perform this manual process.

The final result will be a binary raster file that just contains lines of the geographical elements (usually with more than one pixel width, although it depends on the resolution chose in the scanning process). These characteristics are essential to perform the skeletonizing process. So now we can already perform the skeletonizing, an automatic process that reduces all the linear elements to have one pixel width (always remaining the most centered pixels of the initial line) [9].

### 2.4 Creating the control points file for georeferencing.

We have to make a control points file if we want to correct the planimetric faults that could be produced by the scanner in the image and to convert the image coordinates to map coordinates. At this step, we digitize (on screen) a point vector over the central part of each control that we had drawn in the map. The file created in this process must have a particular structure (coordinate x image, y image, xMAP and yMAP separated with blanks and joined in two blocks, the equations adjustment block, and

the test block) so that it will be used, at a later step, for georeferencing and correcting.
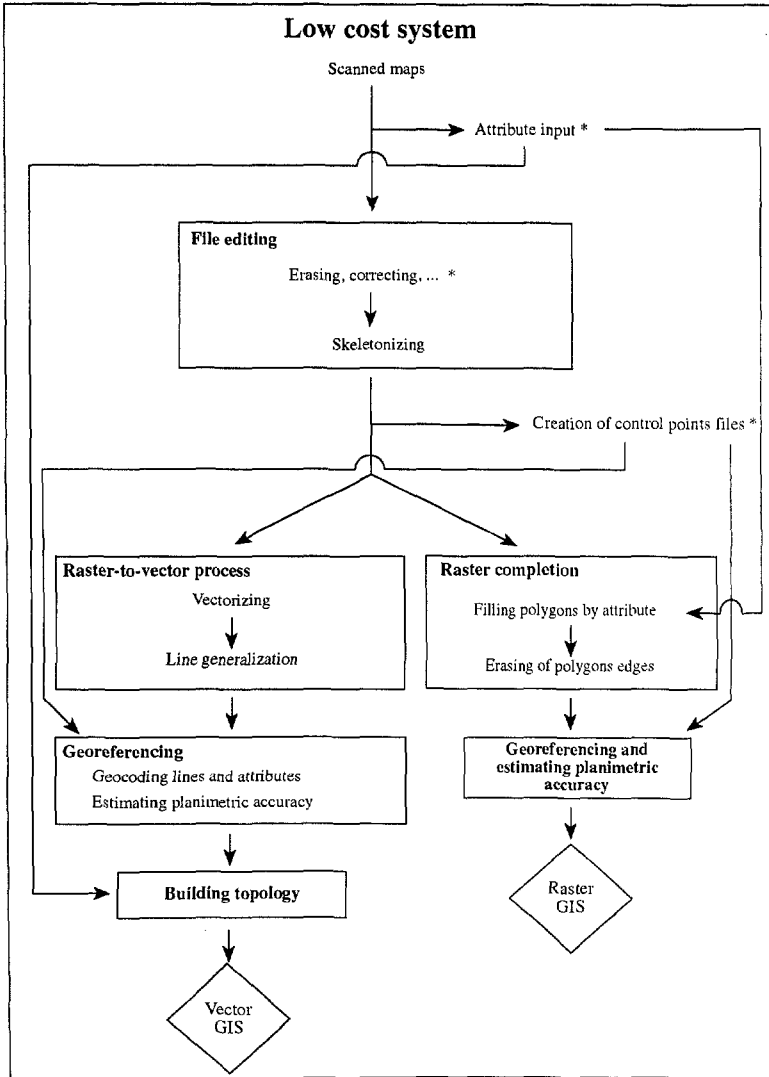
# Low cost system

Scanned maps

Attribute input *

**File editing**

Erasing, correcting, ... *

Skeletonizing

Creation of control points files *

**Raster-to-vector process**

Vectorizing

Line generalization

**Raster completion**

Filling polygons by attribute

Erasing of polygons edges

**Georeferencing**

Geocoding lines and attributes

Estimating planimetric accuracy

**Georeferencing and estimating planimetric accuracy**

**Building topology**

Raster GIS

Vector GIS

**Figure 2:** The basic diagram of the proposed low cost system to incorporate printed maps into GIS.
* Manual processes.

2516

## 2.5 Raster completion (for raster GIS).

The skeletonized file with the edges of the polygons already has a raster structure, so it can be directly imported (or maybe with some change in the format) into a raster GIS. But when we use a digitizing tablet we start rasterizing the digitized lines.

With an automatic system based in the polygon filling [9], we can fill the inner part of the polygons. The process starts from the file containing the points holding attributes that give their values to the pixel located at the point coordinates. This pixel acts as a seed, spreading this value towards the neighbouring pixels, and this process is repeated until it finds the pixels of the edge of the polygon. Note that, previously, edge pixels should be assigned to a value different from any attribute and from the background (e.g., 255 to the edge pixels and 254 to the background ones).

The same process is aware of the faults in the creation of the attribute file; these faults can be of three kinds: a) polygons without attributes in their inner part; b) two or more different attributes inside the same polygon; c) attributes located over edge pixels. For the first kind of fault, the program creates a vector point file. This file contains the central coordinate of all the pixels with background values (e.g., 254), which are pixels without an assigned value (so that their polygon did not have any attribute). This file can be displayed over the raster file, correcting then these omissions by digitizing the missing attributes. For the second and third kind of faults, another point vector is created, containing information about the erroneous attributes: value, value of the pixel they placed over, and coordinates. With these files we can correct the faults occurred in previous steps.

At a second step, and also with an automatic process, we assign to the edge pixels some of the surrounding thematic values. This assignation is done regarding the more frequent value of the neighbouring pixels.

## 2.6 Raster-to-vector conversion (for vector GIS).

The incorporation of the skeletonized file into a vector GIS is done with an automatic vectorizing process. Note that this process is not needed when we digitize using a tablet. The process is done by creating a vertex in the center of each pixel and drawing the limits of the geographic objects by joining the vertices. Afterwards, a line generalization is done (also an automatic process); this allows smoothing the thousands of right angles of the lines related with the initial shape of the pixels as well as with the consecutive and aligned vertices. So many vertices that do not report relevant information are removed, reducing the size of the file and getting shapes more similar to the original map. Note that this vector file will be constituted by non-topologically structured lines. These lines will represent polygon edges when the map is formed by polygons. Also note that the vector point file with the attributes can be directly imported into a vector GIS because it is already in vector form.

## 2.7 Georeferencing and estimating the planimetric accuracy.

Georeferencing and estimating the error will be different depending on which kind of GIS (raster or vector) we want to use (Figure 2). In the case of raster file georeferencing, we use a typical algorithm like the explained in [8], in its nearest neighbour variant. When we want to integrate the files produced in the raster-to-vector process into a vector GIS we have to georeference the lines file and estimate the error in addition to georeference the attributes file. In this case we use a minimum square test and each vertex is relocated with the resulting function. In both cases we use a set of control points to calculate the first or second degree polynomials, and another set to do a test of the outcome RMS.

## 3 Discussion and Conclusion

In Table 1 we discuss the main advantages and disadvantages of the different processes shown in Figure 1.

| Topological digitizing | Non-topological digitizing | Low cost system |
|---|---|---|
| **Advantages** | | |
| Topology is built "on the fly", giving a very consistent way for map digitizing. Errors are less common than in the non-topological digitizing. | There is only one manual process. | Copying maps over transparent films does not require special hardware and software and can be done wherever. |
| There is only one manual process. | There is no need of waiting for building topology during digitizing because this process is done automatically in a later step. | Once the map is scanned, operators always work in a single environment: the screen, avoiding switching between the tablet and the screen. |
| | | Performance increases when map complexity does. |
| | | The process is simple if working in a raster GIS. If working in a vector GIS, all remainder processes are automatic. |
| | | Software and hardware have a lower cost. |
| **Disadvantages** | | |
| For large maps, it requires very expensive hardware and software. | Topology is built in a step separated from digitizing. Errors are, then, more common, both affecting geometry (gaps, slivers, undershoots, overshoots, weird polygons and dead ends [2, 4, pp. 399 and 400]) and contents (labels). | Copying on transparent films is necessary except in very particular cases. |
| Digitizing is an enervating work. Fatigue affects more easily than when copying maps on film. | | Dependence on large scanners, not always available for rent. |
| A digitizing tablet is always needed. | Digitizing is an enervating work. Fatigue affects more easily than when copying maps on film. | |
| Non-topological digitizing | Exporting data to raster GIS is slower than in the low cost system [4, pp. 416] | |
| | A digitizing tablet is always needed. | |
| | It requires expensive hardware and software. | |

**Table 1:** Main advantages and disadvantages of the three more usual processes for incorporating printed maps into a GIS.

(Almost all the software used in this project has been developed by X. Pons and runs under DOS and Windows. If you feel interested, please contact the authors of this paper.)

## References

[1] Aronoff, S., 1989. Geographic Information Systems: A management perspective. WDL Publications. Otawa.

[2] Burrough, P. A., 1987. Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press. Oxford.

[3] Campos, M. & Pons, X. 1994. Informatització cartogràfica del Parc Natural de la Zona Volcànica de la Garrotxa. Informe tècnic. Departament de Medi ambient de la Generalitat de Catalunya. Barcelona.

[4] Diego, F., Martín, J. & Ribas, J. 1994. Connexions biològiques entre els espais d'interès natural del Vallès. Criteris de conservació. Informe tècnic. Departament de Medi ambient de la Generalitat de Catalunya. Barcelona.

[5] Douglas, D. H., Peker, T. K. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, Vol. 10 (2), pp. 112-112.

[6] Foley, J. D., van Dam, A., Feiner, S. K. & Hughes, J. F., 1990. Computer grafics. Principles and Practice. Addison-Wesley. pp. 979-982. Reading.

[7] Jackson, M.J. & Woodsford, P.A., 1991. GIS data capture hardware and software. In Maguire, D.J., Goodchild, M.F. & Rhind, D.W. (eds.). "Geographical Informations Systems. Principles and aplications." Vol. I, pp 239-249. Longman Scientifical & Technical. Essex.

[8] Lillesand, T. M. & Kiefer, R. W., 1987. Remote sensing and Image Interpretation. John Wiley & Sons. New York.

[9] Pavlidis, T., 1982. Algorithms for graphic and image processing. Springer-Verlag. pp. 180-189, Berlin.

[10] Pons, X., 1992. Aportacions a la teledetecció i als Sistemes d'Informació Geogràfica en estudis de vegetació. Tesi doctoral. Universitat Autònoma de Barcelona. pp. 391-422. Bellaterra.