

ARCHAIC DATA MODELS
OR HARDWARE AS A CONCEPT KILLER

Jean-Philippe GRELOT
Institut Géographique National
136 bis, rue de Grenelle
F - 75700 PARIS - FRANCE

ABSTRACT

Scientific thinking is supposed to go from observation depictions to laws, experimentation, validation and observation again. In this respect, automated cartography looks more like a hardware science than like a geographical one. The archaic data models we currently use do not really depict located information but only fit the working method requirements of graphic devices : digitizers, display consoles, plotters. Available computers have not been designed for two-dimensional processing and there is a mathematical obstacle for simulating a two-dimensional world in a one-dimensional system. And any system which cannot solve in a certain way this problem will fail when trying difficult processes such as name placement and generalization; the failure is usually called interactive editing. Automated cartography has to be thought with a mathematical approach and not only as automated drawing. Several tools have to be explored. The paper illustrates these topics and proposes several concepts for designing geographic information systems. It does not pretend to give solutions but wishes to widen the data models range.

INTRODUCTION

Automated cartography is now widely spread around the world as anyone can see in international conferences. But if we look closer, we ascertain that it essentially fulfil two functions : graphic display on the one hand, statistical analysis and modelling on the other hand. On the contrary it fails with other fundamental topics such as feature recognition, name placement and generalization. Now these topics share common characteristics : they require managing a bi-dimensional graphic or cartographic space. Scientific thinking is supposed to proceed along several stages :

- (a) observation phase;
- (b) observation depiction;
- (c) research for laws deduced from observations;
- (d) experimentation;
- (e) validation;
- (f) observation again.

If we accept this statement, we have to admit that automated cartography is not so much linked to geographical sciences : it is rather a daughter of computer science and it derives its data models from computer science. I agree that it improved them through continual approach.

ARCHAIC DATA MODELS

Which are these models which are now so much familiar to us and as a consequence we use them as cartographic data models ? We usually distinguish between vector models and raster models, and I will describe them here on the way they are processed on computers.

Vector models

Basically vector models consist in sampling points which define or delineate features when they are linked with an adequate interpolation method. They define it when it is a point or linear feature and they delineate it when it is an areal feature. The basic model known as "spaghetti model" has been quite improved and the data structure now currently used defines a hierarchy with three levels :

- (a) features which address
- (b) arcs or lines which address
- (c) points and end-points called nodes.

Sometimes links have been created between all the arcs ending in the same node and between all the objects sharing the same arc : it is a somewhat topological part in the structure.

Discussed like this, vector models give us cause for satisfaction. But we must keep in mind the way data are written and can be accessed. They are not written in a two-dimensional space but they follow one each other in a linear mode : computer processing links them one after the other with some jumps according to pointers. We do not really describe features but the way we draw them with a drawing device or we digitize them with a digitizing table and a cursor.

Raster models

Raster models superimpose to space a generally regular grid whose basic elements are pixels. Each pixel is associated one or several attributes referring the object or objects it belongs. Once again the grid model has a deceptive aspect because computing transcribing of pixels eliminates contiguity among pixels from a single feature. More often than not feature connexity is lost not only because the feature has been cut down row by row, but mainly because no link between these various pieces does exist. No better than with vector mode, we do not describe features but the way they are coloured on TV screens and raster plotters. Let us have an example. Since 1978 IGN has been using for areal data raster files run-length encoding methods, entirely describing ^{each} row of the image after the other. In 1985 a new system has been undertaken whose name is "standard raster file" : the basic record is generally a square tile, which is itself run-length encoded. It is not a conceptual change, at least as far as feature models are concerned; at the utmost it may be a change in working methods. As a matter of fact the first row system fits raster mode input/output automated devices such as scanners and plotters which process whole row by whole row, while the second tile system fits interactive consoles for which display can efficiently be managed tile by tile. The made change outlines a preference for interactive processing to the detriment of automated processing.

Two points have to be reminded from this discussion :

- (a) classical data models derive from graphic devices and not from cartographic considerations even if those induced a few improvements;
- (b) computing is linear as far as both data storage and processing are concerned.

The issue is projecting a two-dimensional space onto a one-dimensional space, which forces in breaking down spatial contiguity and connexity relationships in contempt of the most elementary topology. Automated cartography has been designed as a drawing being made and this results in a mathematical dead-end : let us make just the opposite and have a look to a few mathematical concepts to be used in cartography.

A MATHEMATICAL APPROACH

A map is a compilation of features from four sets :

- (a) areal features;
- (b) linear features;
- (c) point features;
- (d) place names.

Several compilations may be superimposed : for instance, thematic maps have topographic backgrounds. But we may process each compilation individually, which means that collections are independent one from each other or at least that they can be separated.

Then areal features constitute a partition : each point in the plane belongs to one and only one feature. In this respect, area boundaries are artefact used by cartographers in order to hide approximate merges between colours.

On this partition are superimposed linear networks which are magnified according to graphic legibility requirements. For our discussion we obviously consider them as linear networks.

Then point features are added and cartographers draw them with means of magnified conventional symbols. or to other features

Place-names are caught to these features/which have a geographic nature but no precise geometric delineation. A set of geographic named features do not constitute a partition : there are holes and overlaps. Place-names requires specific processes : we do not study them here.

This approach is made with respect to features : any system which is not able to process a features as a whole and which forces in separately considering its components, will forbid many manipulations. In a vector-made land-use map, how can we shift a polygon in order to establish a row ?

Vector models have been improved when taking into account the graph concept. A graph is not a set of lines which one digitizes and which has a metric component of its own. A graph is an image through a function with its own metric component which reflects topological properties in the initial space. Graph models have been adopted in many organizations (Guptill 1986, etc) but the explicit function associating features to the plane and its properties have not been studied up to now.

The partition concept had fruitful applications such as cutting down files into small units easy to handle because having only few features (Morehouse 1985 , etc). These small units or tiles generally are addressed by their coordinates and are stored whole row of tiles after whole row of tiles : a storage and address using Peano curves or keys may be an alternative.

Describing linear features by polygon lines restraints extraordinarily geometry to straight lines. Parametric curves and shapes can be used as well (Saalfeld 1986). More generally analytical geometry has to be considered and even algorithmical geometry. Both consist in describing features according to several of their properties and in finding an alternate solution to the map any language shows for describing irregular shapes (Hagège 1985). The types of depictions have to be related to natures of features : spline functions for railroads, straight lines for real-estate data, fractals for river patterns, etc.

Replacing in this way a decomposition into elementary simplistic segments by an analytical expression leads to consider other mathematical concepts which take into account characteristics of geographical features instead of being restraint by their graphic appearance. One builds then a general frame which can accept another essential notion : fuzzy sets. Cartography made us familiar with pure taxonomies, and on the contrary satellite images only have mixed pixels. These views are two translations of a probabilistic depiction of the studied phenomenon : cartography keeps a manichean thresholding from it, remote sensing starts with average values and tries with many difficulties and builds up a cluster. The probabilistic aspects has to be considered; it is interested either in delineating features with relatively continuous transitions or in defining classes with appropriate names : a wooden area at small scale integrates many glades, while glades are separate from wooden areas at large scales. It is an evidence that probabilistic depiction is closely related to analytical geometry.

Cartographic data models define entities and links between entities. In the graph model links carry on a few topological stammers. And the choice of any model gives of course preference for particular applications. This restraint comes from the linearity of computer processing which is not compatible with the two-or three-dimensional nature of studied phenomena. The multi-dimensional building can be made by multiplying links but the multiplication of pointers make computers inoperative. It is then possible to build up embryonic links such as procedures only calculating pointers in a dynamic way when required. Another method for considering them lies in the algorithmic aspect of links in trees, like quad-trees and B-trees.

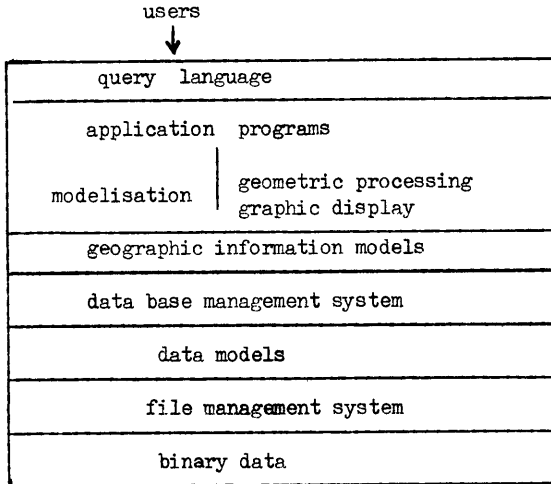
Such embryonic links may contain gateways between data models - General principles for designing data bases forbid duplications which locks up-dating. But on-line cartographic data bases are intended for applications, and usual up-dating delays make it possible to up-date data outside the data base with data structures accorded to workstations and software packages : one copies a box from the data base, edits it and puts it back again. With that respect several data models can be used for the same phenomenon, and they can be partially derived one from the other; for

instance, relief can be described

- (a) by contour lines in vector mode plus spot heights;
 - (b) by a digital elevation model in regular raster mode;
 - (c) by a Delaunay triangulation in irregular tessellation mode;
- etc...

A PROPOSAL FOR GEOGRAPHIC INFORMATION SYSTEMS

Geographical reality has several faces. A complementarity between data structures results in a proposal for designing geographic information systems along seven successive layers (Grelot 1985). At the lower layer stand data, single collections of binary characters; they are physically stored in files accessed through a file management system. The logical composition of data, whose knowledge is necessary to read them, constitute the data models. A data base management system covers this group. The next layer is a main conceptual layer which describes geographic information models. This design formally separates them from data models : a geographic model is an abstract modelling of an object (e. g. this mountain - geomorphological element of relief-), which can have at the very same time several representations built on different and complementary data models (contour lines, digital terrain models, triangular tessellation, etc), each of which requires a specific mode of spatial data distribution (vector mode, raster mode, quad-tree, etc). Application programs process objects and may use for particular reasons one or several data structures for a single object. Lastly a query language gives users the access to the information system through normal language.



A 7-layers geographic information system

CONCLUSION

My aim here is not to give actual solutions to automated cartography shortcomings. I only wanted to focus on some fundamental aspects. Automated cartography has to process with features and objects, and not only with lines and pixels. We have to move from a hardware science to a cartographic science applying mathematical notions : otherwise we will not have enough workstations and technicians to digitize and process our expected data.

REFERENCES

- Grelot J.-Ph. 1985, Principles and theoretical aspects of computer aided cartography and standards, CERCO Intensive Course in Computer Aided Cartography
- Guptill S.C. 1986, The National Digital Cartographic Data Base - Next Generation, 1986 ACSM - ASPRS Annual Convention vol. 1 pp 218-221
- Hagège C. 1985, L'homme de paroles, Fayard - Paris
- Morehouse S. 1985, Arc/Info - A Geo-Relational Model for Spatial Information, Auto-Carto 7 Proceedings pp 388-397
- Saalfeld A. 1986, Shape Representation for Linear Features in Automated Cartography, 1986 ACSM-ASPRS Annual Convention vol 1 pp 143-152.
