The I.G.N. small scale geographical data base (1:100000 to 1:500000)

François Salgé, Daniel Piquet-pellorce

Institut Géographique National France

2 avenue Pasteur 94160 Saint-Mandé France

ABSTRACT

Small scale geographical informations have always been expressed through graphics. The new arising digitization technics allow to turn the geographical information concept into a more complete one, rid of some constraints such as the scale, the projection, the precise symbolism. Consequently, the french **Institut Géographique National** decided to define and to constitute a Small scale geographical data base.
Three main goals are assigned to this data base:

-answering the needs of digital geographical information and becoming the skeleton of a general distributed geographical data base,
-answering a raising demand of specific maps,
- modernising the I.G.N. own small scale series production.

It is essential to know the small scale information user needs, which data he can supply us with and how he can take part in the data base construction. These are the aims we have assigned to an investigation to some thirty people which are representative of the user community. Then the data base starting content will be determined.

The origin and the basic material from which the data capture is performed, depend on the theme to be merge in the data base. Roughly, the main way to digitize the geographical information geometric aspect is the use of a scanner. The selected-feature geometry is digitized from fundamental layers of our present medium scale maps (1:25000 to 1:50000). The SPOT imagery will constitute one of the main geometric updating source and one of the main land use and land cover source.

The data capture is forecast to be performed from the first of january 1986 to the end of 1990. The map production from the data base will begin during 1988.

The setting up of this data base, data capture and exploitation, is such a technical and fundamental change that it means that the cartographer, the draught man and the user have to deeply modify their concepts. The proposed challenge is to turn their static cartography into a new dynamic one, to turn their monolythic graphical expression into all possible expressions of the geographical

information itself, for itself, information which becomes the prime information again.

INTRODUCTION

The needs of small scale maps (from 1:100.000 to 1:1.000.000) are extremely various, for a country like France. The French Institut Géographique National (I.G.N.) partially replies, at the moment, with several regular general topographical series at the scales of 1:100.000 (green series), 1:250.000 (red series and JOG), 1:500.000 (OACI) and 1:1.000.000 (CIM and serie "900"). These series, which are essentially derived from each other, go through frequent updating (1 to 3 years).

Two main characteristics have to be outlined:
-these cartographical series lowered themselves after several reprintings (accuracy and consistency damage due to successive and relative transfer, graphical quality damage due to the different copies of the original)
-The updated data collection is single and should be coherently used at the same time for the different series.

A complete remaking of this cartographical set have to be seriously considered and for that purpose a deep modernization and the resort to digital cartography are necessary.

For several years and more precisely now because of a governmental policy of decentralization, a strong demand of varied maps for local ("departement" or "region") needs have been expressed: various maps in different systems of reference, with a specific symbolism, drawn at the most suitable scale (from more than 1:100.000 to less than 1:500.000), according to various cutting (geographical or polygonal or administrative cutting). All those parameters will be chosen by the user and the map semiology will be specially and precisely adapted to the pecular problem which is to be mapped.

At the same moment, thematic cartographies spread out and ask for map background production (specially adapted), and imply every time a specific symbolism, color and selection.

At last, for the last few years, more and more establishments are managing themselves located informations in a digital way. This implies a new demand for general files concerning the topography. Every body is then capable of handling his own data according to general information (roads, hydrography, urban centers, geonames, orography...). It is possible for him to produce some document for study ("blue papers"). Some of those structures have already begun the digitization of roads, rivers, administrative limits... This is done from heterogeneous and incoherent documents,

without any aims of accuracy, projection, updating... This will jeopardize the exchange of located data and increase the communication problems.

Within this context, trying to offer a coherent answer to those problems, and directly taking its inspiration from the "commission nationale de l'information géographique" conclusions (CNIG 83), the french Institut Géographique National has been studying since 1983 the construction of a small scale geographical data base (previously called BDPE, its french acronym, presently called B.D.Carto, which stands for cartographical data base and which should be called decametric geographical information data base). Its main aims are to be able

-to solve IGN needs of derived mapping at 1:100.000 to 1:500.000

-to give to potential users a topographical skeleton which enable them to locate their own data.

Before the data capture description and the information structure evocation, the data base content has to be defined. For that purpose an inquiry beside potential users has been set up and a rapid description of its modality will be done. But firstly the main principles which led us and the main date will be explained.

SMALL SCALE GEOGRAPHICAL DATA BASE PRESENTATION

This paragraph gathers together some informations about the principles to be used and the planned schedule.

The B.D.Carto principles

The topometric accuracy to be offered should be between 10 and 20 meters, in order to ensure a strict inter-themes superposition and to be compatible with the french Spot-satellite panchromatic-band, which will be one of the geometrical source of information. Those SPOT images, either in the multispectral mode, will be one of the main sources for the updating of the data base and for the land resource part of the data base.

The known needs of B.D.Carto-like data and the data capture constraints let us planify the digitization of the whole France within the next six years. The data updating will be performed each year for the main themes (to be defined) and every two or three years for the secondary ones. Consequently, the themes to be digitized inside this data base must be perenious and must be coherent with the data-collection means, according to that aimed temporal accuracy.

An other fundamental principle is to archive and manage data which must be less distorted as possible by the output

technics limitations (generalization, conventional symbolism and cartographical displacement of objects) This means that topographical information, instead of "cartographical" one (1), must be digitized.

One of the proposed challenge is to modelize the topographical information in term of data base structure: get the as "natural" as possible description of the geographic data. The choice of the topographical interpretation of the data versus the cartographical one means to be able to get rid of the traditional conventions, inherited by the survey history, which reflect the insufficiency of the traditional technics in expressing the plurality of the countryside.

If we are interested in managing only one data base able to create multiple cartographies, it came to our mind that the best data representation is the one which is independant of every output representation. In our case, the topographical data, which describe the countryside reality, free of cartographical constraints, is the only possible choice. This principle preserves also the compatibility of information created by various sources like maps, remote sensing imagery, telemeasurement ... Per contra, this choice implies some important software developments concerning the automatic cartography, which have to solve the constraints of the cartographical redaction.

About the planning

The main dates to be announced are those of the next table.

THE POTENTIAL USERS INQUIRY

The objective of this paragraph is to offer a brief overview of the planned inquiry which must deliver its conclusions by the end of this summer.

Inquiry objectives and modality

Due to the decreasing cost of the computer-science products (software and hardware), and to the increasing memory capacity, the use of digital geographic data will be generalized within the next decade. Because the data base system, B.D.Carto, studied by I.G.N. for its small scale products, aims at being useful for most of the users, it seemed to be essential to ask those peoples about their needs of such data and the way they intend to use those geographical informations.

(1) "cartographical information" stands for information which is interpreted by the cartographers

year	to be done
1986	-data capure software inprovement -remote sensing imagery use study (for land use) -begining of the communication and hydrography network digitization -use of PISTIL (Bernard 86) -cartographical station definition -inquiry and data base content definition -data structure definition
1987	-SPOT images use for land use -updating procedures study (including SPOT) -end of data structure definition -choice of the data base machine -output production study
1988	-begining of the updating -begining of the cartographical output production
1990	-end of the communication and hydrography network digitization -full operational use of the data base
1991	-end of the complete digitization
...	-full operational use of the data base -toward a further generation!

For that purpose an inquiry had been set up lying upon two kinds of documents:

-the first one is a general explanation of the B.D.Carto, mostly what had been written in the last paragraph,

-the second one is an illustration set of what could be the data base content. Those documents, drawn in a "cartographical way" (maps), are not some proposals for a new and different mapping, but only a way to describe and to suggest, without any inventory, what could be the B.D.Carto data and to allow the potential users to express what could be their needs and to incite them to react in front of this suggested content.

The proposed questions belong to three different domains:

-the B.D.Carto content: is this content suited to their needs? And if not, what are the other data to be collected? If I.G.N. is not capable of collecting them, who can afford to get those complementary informations and who can update them? Are those users able to assume the management of those kind of data, linked to the remaining of the data base, or do they prefer letting the B.D.Carto System supply this management?

-how to inquire and use it: what kind of demands have the users to employ for obtaining the useful information from the B.D.Carto? What are the ways they intend

to use it for their own applications?
-general questions: for example the problems linked with confidentiality, confidence and responsability of the information.

The inquiry modality lies upon three phases:
-phase 1: a first letter has been sent to every people liable, in our opinion, to be interested in answering the inquiry. The letter gave the B.D.Carto "general explanation" and asked for a meeting between them and I.G.N.
-phase 2: for every meeting an inquiry purpose explanation is developped and an illustration set presentation is performed. An explanation is joined to each illustration. It depicts what are the data represented on the "map" and what its partial content can be useful for.
-phase 3: when those users get enough time (1 or 2 months) to study those illustrations, and to think to the utility and the practical use of the data base, a second I.G.N. engineer's visit is done for a constructive discussion about the B.D.Carto content, use ...

The consulted users

The B.D.Carto potential users, who are to be visited, belong to four main classes (central administration, local administration, public services, private company). This set of users ought to offer a sample as representative as possible of the different domain to be investigated (communication, hydrography, tourism, "map background", orography, remote sensing users...)

The B.D.Carto inquiry illustrations

As shown in the next table, the 15 illustrations made for the consultation may be matricially divided into six categories according to the themes and two categories depending on the scale. The themes, which have been held back, are those which seemed to us the most eye-catching and propulsive. Those 15 illustrations will be roughly described during the oral presentation. Some slides will show them figuratively.

theme/scale	1:250.000	1:100.000
A:communication	A1:simplified road map A2:road and route map	A3:"départemental" road map A4:road map for car drivers
B:tourism	B1:cultural wealth B2:ecological wealth	B3:touristic map
C:background	C1:road-hydrography	C2:background for electric purposes C3:orography and hydrography
D:administration	D1:administrative boundaries	D2:administrative informations
E:hydrography	E1:simplified hydrography	E2:hydrographic map
F:remote sensing		F1:"spatio-carte"

Hot reactions

In general, the consulted users gave us some informations during the first visit. But, because up to now we have not met every one, the synthesis of these reactions will be presented orally in september.

Synthesis of the answer

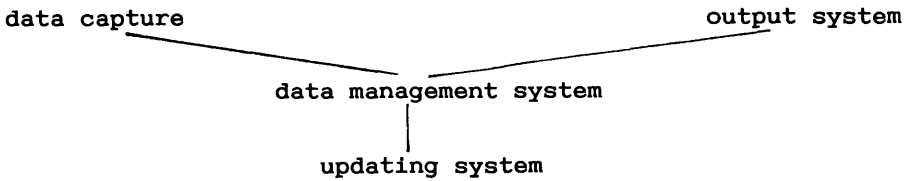
The result of this inquiry will allow us to define precisely the content of the B.D.Carto. By the end of this summer, this will be done and will form one of the subjects of the oral presentation.

STRUCTURING THE DATA BASE

This part aims to show what the basic options which had been chosen for the system are and attempts to propose a first solution for the data structure.

Basic options

The B.D.Carto system is described by four main subsystems (Salgé & Piquet-Pellorce 85) as shown in the next scheme.



-The first one, let us call it "data capture", is involved in the organisation of all the data capture. It has three main functions:

*-paper-information compilation and graphical document preparation: this function, which is mainly manual, squares with gathering all the documents (films, maps, ...) useful for the selection of the data which is to be digitized. It includes also some work on the graphical films which are afterwards digitized by a scanner.

*-selected informations digitization: This function set two computer systems in action. The fisrt one, known as "automatic digitizing system", is in charge with the scanner digitization, the raster-to-vector conversion and the automatic correction processes (Pernot 85) and supply a file to be edited and tagged. The latest, the interactive workstation PISTIL (Bernard 86), let the operator edit and tagg the geometric informations in order to produce the ready to insert data. A control is assumed by this function in order to get rid of "all" the faults.

*-insertion into the data base: this function is nothing but an interface program between the data-capture system and the data-management system of the data base. This will be done via a network between the workstation and the data base machine.

-The second one, let us call it "data management", has to manage the data. This is the heart of the system. A data-management system has to be found, and has to be able to store, update and retrieve the data. It lies upon a data structure modelizing the reality. It will be finally defined after the consultation and must be able to be extended if necessary.

-The third one, the "updating system", is in charge with the data extraction from the data base for updating them. This will be performed on the interactive graphic workstation (PISTIL). This subsystem has to be defined.

-The fourth one, the output system (which stands for master-building system), is involved in the generation of all the possible cartographical outputs from the

data base, including generation of the master-films necessary for the map printing. It is also divided into two main computer system.

*-The first one is able to build a draft file where informations to be mapped are automatically symbolized. This system implies the study of methods for information symbolization and generalization and have to propose for editing an as-perfect-as-possible drawing with "cartographical accuracy". This automatic system will be modular in order to include little by little the most performant algorithms (comprising automatic geonames placing)

*-The latest edits this automatically-proposed drawing files. The operator interactivally modifies the file, computer-mistake corrections, cartographical displacements, geonames positioning ... At the end of this editing process it turned this drawing file into the final master file to be processed on a laser camera (for example) creating the master films for printing.

The geographical extent of a working set during a special working session, is generally very limited. The data capture, for example, is performed from existing maps (1:50.000 sheet) or from remote-sensing images (SPOT). In these cases the geographical extent is only 600 square-kilometers wide. By comparison with the size of the data base (presently reckoned as 1 Gbyte for 550.000 square-kilometers) we think that the working-set average-size is less than 40 Mbyte. The presently estimated average-time for editing and tagging this working-set (40 to 80 hours) suggests us to define the interface between the four subsystems as a geographically and thematically limited working-set on which the editing (or the cartographical work or the updating work) is performed before the back-insertion into the data base (Salgé & Piquet Pellorce 85). This means that between the management-system and the three other ones the data base transactions do not imply any time constraints, the rapid query transactions are locally processed on a limited working-space.

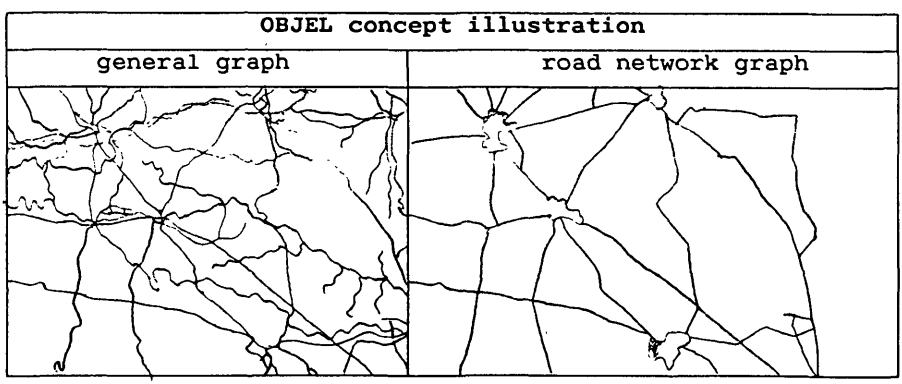
Structuring the B.D.Carto

Referring to (Salgé & Piquet-Pellorce 85) and (Motet 86), the B.D.Carto structure, which once more will be definitely fixed after the consultation results, is split into two very tightly linked different parts:

-Thematic and topologic part: The objects or entities found here represent the geographical objects of the universe we have to describe. We think that the data base relational model is well suited to depict this geographical information and a relational data base management system is able to manage that part.

-Geometric part: in this part we found the arcs (or with the GKS terminology Polylines) nodes and polygons. Here are the coordinates. This geometric part is seen as the general-graph (planar or not planar that is the question (Shakespeare 1600)). The links between those two levels are made by the concept of OBJEL which stands for elementary objects (in french OBJet ELementaire) and which is a set of arcs, nodes or polygons depicting an elementary phenomena of the geographical world.

Let us take the road-network as an example. The first entity to be pointed out is the OBJEL. The road-network is considered here as a non-planar graph (according to the graph theory), where the nodes are the crossings between roads, highways... (the interchange of two highways is seen as a node), and the arcs are made by the roads parts from a crossing to another crossing (we assume that the point where there is a characteristic change (width for example) is seen as a two-branches node). This road-network graph is built on the general graph and, by the way, the road-network nodes and the road-network arcs, which are some of the OBJELs, appear as a set of general-graph arcs, nodes or polygons. The OBJEL concept is the first abstraction of an element set of the geometrical-part . The OBJEL bears the basic geometrical information of an elementary geographic entity in the same way that the PIXEL bears the basic geometrical information of an elementary image entity. Lying upon this OBJEL concept the entity-relationship scheme of the geographical information is then possible.

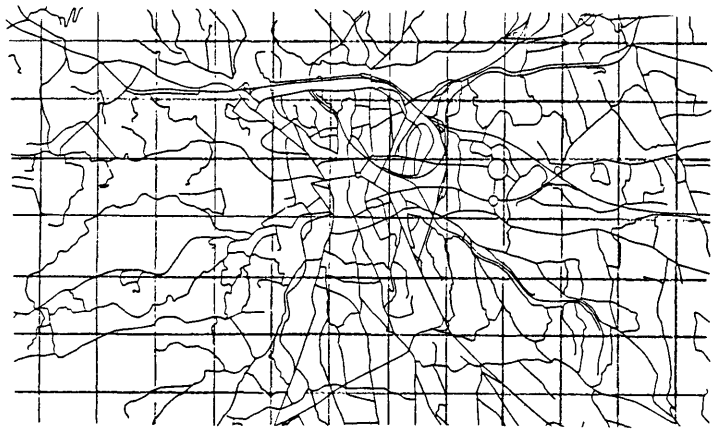


For the data retrieval purpose it is essential to adopt a geometric-level structure which keeps the inherent proximity relationship of the geographical space. The "small scale" information (1:100.000) is a two-dimensional one (X and Y for example) and the problem is to find a method which manage together points which are neighbours. Among the systems proposed in the bibliography (peano keys , Quad tree ,

or mixed method (Laurini 86)) we have chosen the tiles concept (in French: Dalles). The geographical space (or the plane) is divided into regular rectangles (about 1 or 2 km wide). This defines a grid which is superimposed upon the geometrical-level. The arcs, nodes and polygones which cross a same tile are then managed together (in fact only the part or arcs and polygons inside the considered tile). Due to the small size of the tiles, the number of the inner elements is limited (less than 100). Then the mechanisms for the data retrieval and sort are simplified. The demanded-data determination (for a working-set) is nothing but a sort among a small amount of elements.

an illustration of the tile concept

all the arcs, nodes (and polygons) inside a tile are managed together.



DATA CAPTURE

The communication and hydrographic network digitization has begun in January 1986. It is focused on the geometry and the basic attributes. The process is described in the next table. The main source for the geometric level is the conventional 1:50.000 scale map.

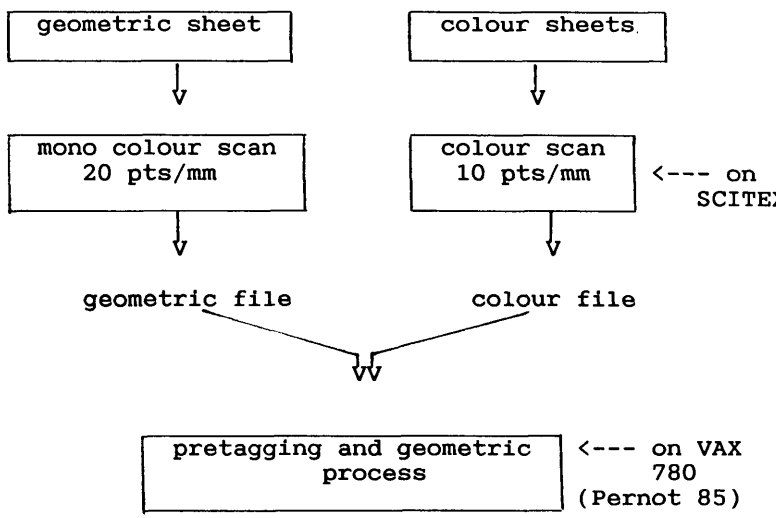
information collecting

COMMUNICATION		
geometric sheet	road network	aperture mask or selected roads redraw (1)
	railways and lanes	selected item redraw
color sheets	tagging colour preparation (administration and width)	

HYDROGRAPHY	
geometric sheet	selected hydrography redraw (2)
color sheets	pretagging sheet (width, channels and areas)

(1) the interchange and the towns are simplified
(2) hydrographic areas such as lakes, marsh, "textural hydrography" are drawn as boundaries.

Scanning and automatic process



Work on the interactive graphic workstation where is performed the complete tagging and the geometric

correction. Presently the system used is a prototype (CYTHERE on Vax 780) of the future interactive graphic workstation PISTIL (Bernard 86)

SPOT use The two informations caught from SPOT imagery are land-use data and geometric update. These uses are now on study and the main results will be presented during the oral presentation.

CONCLUSION

Information quantity is on the increase, so do its need. Communications are going easier and easier. Informations are exchanged, compared, superimposed. Computer science is turning graphical, allowing everyone to display and relevently handle located data. Within such a context, the I.G.N. mission is not only to answer the needs, we call forth in the introduction, but also to supply everyone with the unique reference, unique security for coherence. Because the map is fundamentally a communication-tool and the background-maps are the reference, the datum and by the way the common structure for all thematic cartographies, the B.D.Carto must be The Reference for all, real skeleton for a national geographical data base, which let the tremendous amount of geographical data, belonging to one or another, be coherent and really open to all. Terms for the valorization of the caught-by-everyone informations, the B.D.Carto and the SPOT imagery, both in coherence, have to federalize the whole medium-accuracy (decametric) geographical information, concerning the French territory, and answer, by the way, one of the wishes of the "Commission Nationale de l'Information Geographique (CNIG 83).

This data base is also a real challenge for the cartographers. Its setting up, data capture and exploitation, is such a technical and fundamental change that it means that the cartographer, the draught man and the user have to deeply modify their concepts. The proposed challenge is to turn their static cartography into a new dynamic one, to turn their monolythic graphical expression into all possible expressions of the geographical information itself, for itself, information which becomes the prime information again.

REFERENCES

(Salgé & Piquet-Pellorce 1985)	Les projets de bases de données topographiques et géographiques à l'Institut Géographique National problèmes et solutions techniques	convention informatique 16-20 Sept 1985 Paris
(CNIG 1983)	rapport de la commission Nationale : l'information géographique	
(Bernard 1986)	A workstation for handling located data : PISTIL	Auto-carto London 1986
(Motet 1986)	(To be edited)	Euro-carto V Paris 1986
(Pernot 1985)	Correction géométrique de la geomètrie après scannage	Euro-Carto IV 14-17 Oct 1985 Frankfort
(Laurini 1986)	Les clefs de Pèano: un nouveau modèle pour les bases de données multidimensionnelles et les bases d'images	2ème journées Bases de Données Avancées Giens (Var) France 22-25 Apr 1986