A geographic data model based on HBDS concepts: The IGN Cartographic Data Base Model

François Salgé - Marie Noëlle Sclafer

Institut Géographique National - France
B.P. 68 - 2, Av. Pasteur
94160 Saint-Mandé
FRANCE

ABSTRACT

The cartographic data base (BDCarto) is a major IGN-France priority. Its main aim is to deliver structured geographic information with a 1:100,000 level of abstraction. It will be completed by the end of 1992 and different output products will be available for users: files, customized maps and map series (1:100,000 - 1:500,000).

The content of the data base has already been defined and the BD Carto glossary is composed of some 700 terms.

In order to define the relational form of the data base the content definition went through a structuration process. The first step was to define a geographic data model using HBDS concepts.

Roughly this model lies on two levels : the geometric level gives the geometric topology and the coordinates whilst the semantic level gives all the information about the geographic objects and their semantic topology (when necessary).

The paper describes the geographic data model and its application to the BD Carto content. It explains how the structuration process allows us to model the ground truth and get rid of the constraints of a graphical model.

THEORETICAL INTRODUCTION

GRAPH THEORY

A graph G=(X,U) is the pair constituted by a set $X=\{x_1,x_2,...x_n\}$ of vertices and a family $U=(u_1,u_2,...,u_n)$ of elements of the cartesian product X x X called arcs.

A subgraph of G is a graph having all of its vertices and arcs in G. A subgraph of G is said to be a connected component of G if for each pair of vertices of the subgraph there exists a path joining them together using only arcs of the subgraph.

A graph is said to be planar if it is possible to represent it on a plane such that the vertices are distinct points, the arcs are simple curves and two arcs do not meet beyond their extremities.

A face of a planar graph is by definition a plane region bounded by arcs such that two arbitrary points in that region can always be linked by a continuous line in the plane which does not encounter vertices or arcs.

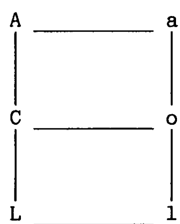
A layer is a planar subgraph of G, generated by $A \subset X$ such that $\forall a \in A$, $\forall b \in X-A$, there is no path in U linking a and b. In other word, a layer is a connected component of G or is the union of connected components of G.

HBDS CONCEPTS

HBDS stands for Hypergraph Based Data Structure. There are six basic concepts. The object concept 'o' makes it possible to define the basic entities of the considered domain. Those objects may be grouped by their characteristics; this leads to the concept of class 'C'. Those characteristics are known as attributes 'a', this defines the class-attribute 'A'. Links 'L' between classes are possible whose occurrences are known as object-links 'l'.

One can also define an hyperclasse which is a set of classes. Figure 1 is a graphical summary of HBDS concepts.

Fig 1 Graphical representation of HBDS



MODEL OF THE BDCarto GEOGRAPHIC DATA

The terrain elements modelled in the database are described by two information levels: a geometric level which specifies their positions and a descriptive level which specifies their characteristics and their non-metric description. The entire set of data, making it possible to describe each elememt of the terrain modelled in the BDCarto, will be called

The following is a list of synonyms which have been used in the literature:
vertex = point = node = junction = 0-simplex
arc = line = edge = branch = 1-simplex

geographic data.

GEOMETRIC LEVEL

The geometric level is constructed by definition on n independent layers. Each layer lies on a planar graph constituted by arcs, vertices and faces.

For a given layer there are three classes: arcs-class, vertices-class and faces-class.Those three classes will also be called geometric classes.

Two topologic relations exist between the arcs-class and that of the vertices:

- "have for initial vertex"
- "have for final vertex"

Likewise two topologic relations exist between the arcs-class and that of the faces:

- "have for the left face"
- "have for the right face"

It is the geometric level which supports the coordinates specifying the positions of the elements of the database.

Each arc is represented by a broken line.

DESCRIPTIVE LEVEL

The object corresponds to the descriptive part of the terrain elements to be represented. They are grouped in "object-classes" (sets of objects) which form a partition of all the objects contained in the database.

An object-class is distinguished by a class name which indicates the nature of the objects contained in the class and by a list of the class-attributes, such that any object contained in the class takes for each class-attribute one and only one value amongst the possible values of the attribute (definition domain). The complete set of values taken by an object for each class attribute constitutes a description of the object.

Objects can be linked one to another. Objects can be linked to geometric elements. That dependence is expressed by relations between classes (relation in its mathematical sense, of which the arrival and departure domains are classes). There are two types of relations:

- Construction relations which make it possible to reconstruct the geometry of all the elements of the terrain modelled in the database. They link object classes amongst themselves or object classes and geometric classes.
- Semantic relations which make it possible to express a link between two objects not conditioning their geometric reconstruction. There are no semantic relations between object classes (of the descriptive level) and geometric classes.

One can distinguish two types of objects. 1) Elementary objects are directly in a construction relation with the geometric level. An element of the geometric level can be related with several elementary objects and reciprocally an elementary object can be related with several elements of the geometric level. 2) Complex objects are constructed with elementary objects or less complex objects.

A theme is constructed by classes of objects covered by the description of part of the terrain reality. There will be semantic relations between classes of objects belonging to different themes.

A theme can correspond to several layers and a layer can correspond to several themes.

Certain classes can be regrouped for convenience into sets of classes, e.g. in order to use simplifying generic names. The HBDS diagram in Fig 2 describes the types of entities which make up the model of the BDCarto geographic data.

INFORMATION ON THE GEOGRAPHIC DATA

CONSTRAINTS ON THE GEOGRAPHIC DATA

This is the logical assertion concerning the data base entities whose value must be TRUE.In other words, the constraints are postulates which, as a rule, verify all the data in the base. There are constraints called integrity constraints in the database, which can only be true and must therefore be verified by the acquisition sub system (without exceptions).

The other constraints, semantic constraints, which can have exceptions, must preferably be evaluated before insertion.

Integrity constraints

uniqueness: any object, any element of a geometric class, any construction relation occurrence, only exists once (for a given date) in the database.

definition domain: any attribute of any object of any class exists, is unique and belongs to its attribute-class definition domain.

positioning: any object of the database is connected to the geometric level through a set of construction relations in order to produce an unique geometric representation of the terrain element which it is modelling.

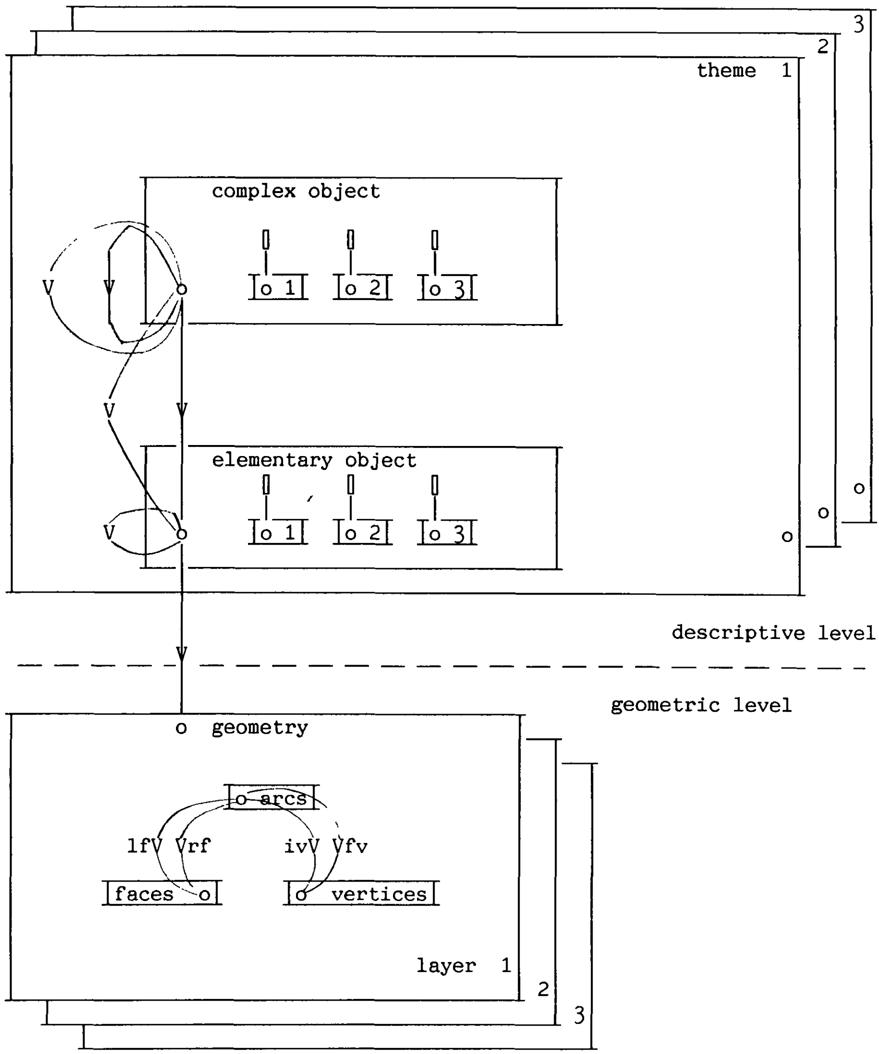
layer: the set of elements of a geometric class in the same layer constitutes a planar graph.

Semantic constraints

These constraints vary for each entity and must be compulsorily specified for each one.

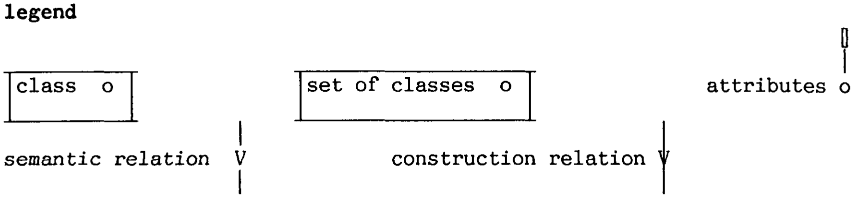
object-class: constraints concerning possible combinations of attribute values for an object of the class (subset of the cartesian product of the definition domains of all the attributes of the class of objects)

Fig 2



3
2
theme 1
complex object
o 1
o 2
o 3
elementary object
o 1
o 2
o 3
o
o
o
descriptive level
geometric level
o geometry
o arcs
lf
rf
iv
fv
faces o
o vertices
layer 1
2
3

legend



class o
set of classes o
attributes o
semantic relation
construction relation

relation between classes: constraints concerning restrictions on the departure domains and arrival domains and its properties

geometric class: constraints concerning the minimum size of arcs, faces ...

geometric construction of elementary objects: constraints on the type of geometric description of elementary objects

other constraints: constraints concerning a combination of classes and/or relations.

GENEALOGY (ORIGIN) OF GEOGRAPHIC DATA

The genealogy specifies for each geographic datum:
-the source used for the data capture, i.e.
  - -type of source
  - -scale of source
  - -identification of the source (number or title, date of publication)
  - -date of collection of the information in the field

-the process which leads to the integration of the digital data into the database, including the coordinate transformations. This process is described by reference to the specifications.

The genealogy is an information which concerns the classes of objects (from where and how does one know that the objects which constitute it really exist), the relation between classes of objects (from where and how does one know that such an object is related with such other object), the geometric construction relations (from where and how does one know that such an object can have such a geometric description), the attributes (from where and how does one know that such an object takes such a value for that attribute).

It is attached to each geographic datum.

QUALITY

Position accuracy

It is defined by two parameters: the mean error and the standard deviation between the coordinates read in the database and those of the real field point that they are supposed to represent.

That information concerns each geometric construction relation or each relation occurrence.

Semantic information accuracy

One is interested here in the difference between the values taken by the semantic information in the data base and their reality in the field. That measurement is given in the form of a probability [P(condition)]. That information concerns:

-the object-classes: accuracy of the identification of the object of the class

$P(\exists\, o \mid o \in Co \text{ and } o \notin \overline{Co})$

probability that an object exists which belongs to class Co in the database and whose real class is not Co.

-the relations between classes: identification accuracy of the relation occurrence expressed by:

$P(\exists\ o,o' \mid \neg(\overline{oRo'}) \text{ and } oRo')$

probabiility that two objects exist which are related in the database and are not related in the reality.

-the attributes: accuracy of the value taken by each object for that attribute expressed by

$P(\exists\ o \mid o.A \neq \overline{o.A})$

probability that an object exist whose database attribute is o.A and its real attribute is not o.A.

Logical consistency

It verifies the respect of the semantic constraints. It is attached to each constraint C expressed in the form of a probability:

P(objects, relations, attributes exist | C=FALSE *)*

It can be forced to 0.

Completeness

This concerns the exhaustivity of the information effectively acquired with respect to the database content specifications and the field reality. That measurement is given in the form of a probability. That information concerns:

-the object-classes: existence of all the objects of the class expressed by

$P(\exists\ o \mid o \notin Co \text{ and } o \in \overline{Co})$

probability that a real object exists which is forgotten in the database.

-the relations between classes: existence of all the occurrences of a relation expressed by:

$P(\exists\ o,o' \mid \neg(oRo') \text{ and } \overline{oRo'})$

probability that an occurrence of a relation has been forgotten.

-the attributes: number of the unknown value taken by that attribute expressed by

$P(\exists\ o \mid o.A = \text{unknown})$

probality that an object exists for which the value of the attribute A is not known.

Quality assessment

A deductive estimation based on a knowledge of the errors at each stage of the process, on the calibration of the geometric sources and on the hypotheses concerning error propagation can be deduced from the genealogy information and can be attached to the construction relations and to the occurrences of those relations.

The evaluation of the quality, due to its statistical nature, only has a sense with respect to a given set of information. Consequently, a quality estimate is undertaken by means of a quality report specifying:

the zone concerned by the evaluation

the content elements having been the subject of the evaluation

the date of the measurement

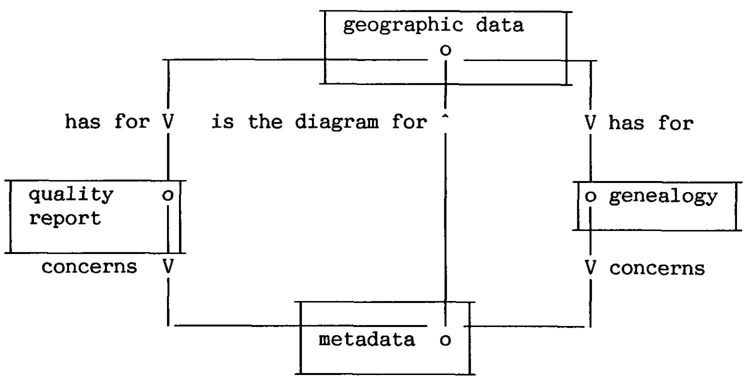
the description of the method of measurement
the description of the external reference
the reference validity date
the value of the result
the objects to which one can allocate the result.

The geographic data are structured in conformity with the model previously discussed. The description of that data structure, with which one associates the content and quality specifications, constitutes the metadata.

SUMMARY HBDS DIAGRAM

Two sets of data are distinguished: the geographic data themselves and the information on the data. The later breaks down into three groups: the metadata and constraints, the genealogical information, the quality reports. The HBDS diagram which rules those data sets is as follows (Fig 3).

Fig 3 HBDS summary



Ref. :1 The IGN cartographic data base : From the users'needs to the relational structure - F. Salgé - Marie Nöelle Sclafer.
in Euro Carto 7 Environmental Applications of DIGITAL Mapping - 1988
2 A survey on the HBDS methodology applied to cartography and land planning - F. Bouillé
in Euro Carto 6 - 1987
3 The proposed standard for digital cartographic data
in the American Cartographer Volume 15.No.1 - January 1988