François Bouillé
Institut de Programmation, Tour 55-65, Université Pierre et Marie Curie
4 Place Jussieu, 75230 Paris cédex 05
FRANCE
336-25-25 or 329-12-21, poste 53-90

STRUCTURING CARTOGRAPHIC DATA AND SPACIAL PROCESSES WITH THE
HYPERGRAPH-BASED DATA STRUCTURE

INTRODUCTION

Cartographic and spacial phenomenons present a complex structure which
contains many intrinsic informations and is the key of almost all problems
requiring data processing.

Data structures models like trees, nets and boxes are not convenient because
they do not respect this natural and essential phenomenon structure. More
especially, it must not be confused with a storage structure; file and
record concepts must not appear in the data structure. The data structure
is the part of the phenomenon structure which has been discovered, studied
and understood. Such a definition does not take into account the nature
of the problems to be solved, the storage and retrieval of data, the
characteristics of the computer... A more general model is thus necessary,
which takes into consideration the topological aspects before those concerning
the metrics, as being the main features of the information.

Such a model is shortly presented in this paper, which is based upon the set
theory and the hypergraph concept. It is presently used for various
applications, mainly in cartography and simulation, which often are imbricated,
for instance in environmental studies and space planning.

THE HYPERGRAPH-BASED DATA STRUCTURE (HBDS):

According to the set theory, a set is composed of elements which have
properties and may present relations or not (BOUR 39). Though a property is
nothing but a particular relation, we keep this old distinction. Using the
abstract data type concept (LISK 74), we consider four abstract data types
respectively named: class, object, attribute and relation. They must be
associated with distinctive graphical concepts. Graph theory is generally
used, but cannot correctly represent the difference between a set and its
elements. We include here the hypergraph concept (BERG 70) as the main
component of the structure. The HBDS is composed of partial subgraphs, some
of them being hierarchized substructures, others being not.

The Hierarchized Part of the HBDS:  

The basis is a directed graph which is an arborescence $A_c$, the vertices of
which are classes $C_i$, so defined:

$$ C = \{ C_i \}, \quad S = \{ S_{i,j} \} \in C^2, \quad A_c = \langle C, S \rangle $
With each vertex $C_i$, we associate a group $A_i$ of valuations $A_{ip}$ corresponding to the attributes of the class; no previous hypothesis is done about the nature and number of these attributes. $C_i$ also carries virtual valuations which correspond to the attributes of all its possible predecessors in $A_c$.

We now consider a set of vertices; a partition is done and each part is associated with a vertex $C_i$; this part is represented by an edge of an hypergraph $H$ around the vertices of the part. The edge is also named "class"; the vertices are the "objects" of the class $C_i$ and indicated here by $O_i^n$. They carry valuations named "attributes of objects" corresponding to the properties of the subset they belong to, namely the attributes of their classes, but they are particular values of these properties. For instance, if COLOUR is an attribute of the class, GREEN may be an attribute of an object of this class. Attributes of objects are indicated by $A_{ip}$ (correspond to $A_i$). Considering the attributes $A_{ip}$ of all objects of all $i_p$ classes, objects may be hierarchized by $i_p$ a forest $A_o$, the connected components of which are only arborescences, possibly degenerated. Thus, each object may have a predecessor in the upper class and successors in the subclass(es).

Attributes may be empty values, corresponding to uncomplete arborescences

$$O = \{O_i^m\}, S' = \{ O_i^m O_j^n \} \in O^2, A_o = <O, S'>$$

The Non-Hierarchized Part of the HBDS: (Figure 2)

A data structure must be able to represent any relations. We now define two multigraphs $G_{cc}$ and $G_{co}$. A relation may concern some components but not obligatory all of them; it is limited to objects belonging to particular classes. Though a relation is verified by two objects, respectively $O_i^m$ and $O_j^n$, $O_i^m R O_j^n$, it is specific of the both classes the objects are respectively belonging to. Consequently, the classes $C_i$ and $C_j$ are joined by an arc, named "link", which carries the relation $R$ that the objects of these classes may verify or not. Two classes may be joined by more than one link, their graph thus being a multigraph. A relation carried by a loop on a class concerns objects of the same class. With each link going from $C_i$ to $C_j$ we associate another one going from $C_j$ to $C_i$, which is the symmetrical link and carries the reciprocal relation. The $k$-th relation is indicated by $R_{ij}^k$ and is carried by a link $L_{ij}^k$. The multigraph $G_{cc}$, directed and symmetrical, is thus:

$$C = \{C_i^j\}, C_{ij}^{L^k} = (C_i, C_j), L_c = \{ C_{ij}^{L^k} \} \in O^2, G_{cc} = <C, L_c>$$

The second multigraph $G_{co}$ concerns the objects which verify the relations carried by the first one. When $O_i^m R_{ij}^k O_j^n$, there is a link from $O_i^m$ to $O_j^n$ materializing the verification of a relation, and a symmetrical one corresponding to the reciprocal relation, even if this last one is symmetrical. The second multigraph is thus defined:
Figure 1 -- The Hierarchized Part of the HBDS

Figure 2 -- The Non-Hierarchized Part of the HBDS
\[ O = \{ o_i^m \}, \quad O_i^{m_k} = \{ c_i c_j^o \}, \quad L_o = \{ c_i c_j^{o_2} \}, \quad s_o = \{ o, L_o \} \]

Data Concept in the HBDS:

The 4 fundamental components of the set theory are taken into account, represented with distinctive manners, and we may propose a definition of "data" concept:

\[
\text{<data> ::= <class> | <object> | <attribute> | <link>}
\]

\[
\text{<attribute> ::= <attribute of class> | <attribute of object>}
\]

\[
\text{<link> ::= <link between classes> | <link between objects>}
\]

The Skeletal of Structure:

Now, the skeleton of structure is a partial subgraph of the data structure, composed of the arborescence \( A_c \), the edges of the hypergraph \( H \), the multigraph \( G_{cc} \) (links between classes carrying the relations), and the attributes of classes, generally represented by square shapes attached to their class. The skeleton-structure sums up the main information of the data structure, and all other data are inserted in or around this skeleton, but in a second time. It is the part of the data structure which is deduced of the analysis of the phenomenons, before the concrete processing of data, and even before their capture. The skeleton may also be updated, but it mainly represents the most important characteristics which determine a data structure. Adding or suppressing one-hundred objects in a class is a minor event compared with the suppression of a class, even if this class is empty.

Some HBDS Extensions:  
(Figure 3)

When the number of classes of a data structure is thus progressively increasing, it sometimes happens that classes belonging to a first group have links with all the classes belonging to a second group, and reciprocally. If these links carry the same relation, a simplification is possible. Using a new time the hypergraph concept, a new partition is done on the classes, and two edges of an hypergraph are considered, the vertices of which are the classes of each group. This edge is named "hyperclass" and the links carrying the relation are replaced by a special link joining the hyperclasses and named "hyperlink". We define the concision-degree of an hyperlink as the number of links thus replaced. In some concrete cases this degree goes beyond one-hundred.

A simplification is the multilink: when two classes are joined by several links, those may be replaced by a multivalued link carrying all these relations and named a "multilink". Hyperlink and multilink concepts may be combined to give an hypermultilink which considerably simplifies some complex cases.
Figure 3 -- HBDS Extensions: Hyperclass, Hyperlink

Data Definition and Manipulation in the HBDS:

A language is built allowing to define and manipulate the data, which is an extension of the SIMULA 67 language. This very high-level programming language is the most convenient and well-adapted to this type of data structure and is based upon the same tools. It contains many concepts which are used in the HBDS model, like the class, object, attribute concepts. The extension, nevertheless, does not depend upon a particular language and may be considered as a package requiring a pre-processor.

The SIMULA 67 language is also very convenient for space processing and naturally for simulation, because it was at the early beginning a simulation language. Now it is a universal language which also can be used for simulation (DAHL 71).

When studying the algorithms which are necessary to solve a problem, it often is easier to use mathematical notations, without depending upon the syntactical details of a particular programming language. Thus, all cartographic algorithms are firstly written in the EXEL language (ARSA 74); likewise, the whole system allowing data manipulation and definition (HBDS language) has been firstly written in EXEL and translated afterwards in a high-level language.

EXEL, SIMULA 67 and the HBDS language allow well and structured programming, quick and efficient processing.

SPACIAL STRUCTURES USING THE HBDS MODEL:

The first structures early studied concerned the earth sciences cartography, and more especially geological and topographical maps (BOUI 75). But now more than twenty themes are considered and we intend to study new types of cartographic documents. Among the themes already studied, there are:

- orography
- topography
- geology
- geomorphology
- geochemistry
- hydrogeology
- gravimetry
- magnetism
- resistivity
- seismic
- pedology
- meteorology
PIEZOMETRY
BATHYMETRY

and another category of information, composed of:

ADMINISTRATIVE BOUNDARIES
TRANSPORTATION NETWORKS

etc.

The resulting data structure is presently composed of

150 classes

approximately and many links. But at the end of year 1978, it will probably reach 500 classes. New extensions are foreseen which are found but not completely formalized.

In the following pages, we present some simple examples, and firstly, a small application limited to the saddles in the topography. Figure 4 is proposed to sum up the concepts of the HBDS.

The HBDS is convenient to process data which are defined by topological properties, because it is a topological data structure. A matrix may be superimposed, but the basis is the topology. To illustrate this main aspect, we have chosen a simplified case: an unknown island in the Pacific Ocean which is discovered by a navigator; he immediately sees that it has two tops and a col, thus there is an object of the class 'COL' and two objects of the class 'TOP'. In a second step, the navigator measures the heights of these three elements: 67.5, 43.2, 34.8. We know the values of the attributes of the three objects, now. Consequently, we also know the data structure, with a skeleton composed of classes like 'MAP', 'LEVEL', 'CURVE'; though we do not know the coordinates of the isolines, we may generate the objects corresponding to these isolines and the links corresponding to three possible types of relations between a curve and another one, which may be:

- of same height (thus neighbor),
- of upper height,
- of lower height. (third step)

The graph representing the relations between the objects is nothing but the dual of the graph associated with the isolines (BOUI 75). In a fourth step, when the coordinates of auxiliary points of the isolines are digitized, we may generate the attributes of the objects of the class 'CURVE', in fact a numerical bidimensional array.

Now, we introduce the reference points on a map. Such a point may be at a top, or only between two isolines of different heights. We consider the class 'REFERENCE POINT' and its links with the classes 'TOP' and 'CURVE'. On Figure 6, two points are drawn, and the corresponding objects in the data structure are represented, with their links.

On this picture we have not differentiated two types of relations between a reference point and a curve, more exactly with an upper or a lower one. More, there is not only one type of point, but several. We define a hyperclass 'POINT' which is linked with the class 'CURVE' by an hypermultilink carrying two relations. Likewise the hyperclass 'POINT' is linked with the
Figure 4 -- This very simple example allows one to sum up the HBDS components. There is a class 'COL' with an attribute 'ALTITUDE'. An object of this class, the 'COL DES GOULES' is situated between tops 'A' and 'B'; its attribute has the value '997' (its height). There is a link between the class 'COL' and the class 'TOP' expressing the fact that a col has a relation with its neighboring tops. Because this particular col is between 'A' and 'B', there are links going to these tops.
Figure 5 -- Building a Data Structure in Four Steps
Figure 6 -- Including the Reference Points in the Data Structure
class 'TOP' but it is only by a hyperlink carrying one relation, because if there was an upper curve neighboring it, it would not be a top.

Another important phenomenon is the fact that a map has generally neighboring maps and that elements intersecting the border of the map are theoretically the next map. The problem is not the research of a method to connect the corresponding elements, which would be an error of methodology, but only to recognize the structure of such an artificial phenomenon. Finally, two classes are defined hierarchized and respectively named: 'PERIPHERY' and 'BORDER'. Some other links are recognized which are loops on classes. The structure is represented on Figure 7. Such a connecting structure concerns all the types of thematical maps.

Another kind of map, rather different, is composed of administrative areas and administrative boundaries. Figure 8 shows the three types of French subdivisions, respectively named 'DEPARTEMENT' (department), 'CANTON' (district) and 'COMMUNE'. Relations between neighboring areas are evidently carried by loops on the corresponding classes. Relations between areas and boundaries are represented by links, for instance, between 'DEPARTMENT' and 'DEPARTMENTAL BOUNDARY', then at the same level. But, as the upper drawing lets appear a canton may be on the border of a department, or even a commune. Finally, all classes must be linked, and we replace nine links by a hyperlink between two hyperclasses indicated on the picture by dotted-lines. The concision-degree of such a hyperlink is then equal to nine.

Geology is more complex. A geological map is composed of areas limited by boundaries. They may be represented by a directed multivalued multigraph (BOU 73, 74). The method is detailed in these papers, and we only recall here the structure composed of seven classes, three classes for the areas, and four classes for the boundaries, with several types of links. Figure 9 shows this structure, but in fact it is linked with the topography, with the fails, etc. When building geological models for simulation, there are generally 16 classes and three hyperclasses necessary.

**PROCESSES STRUCTURE WITH THE HBDS:**

The HBDS is well-suited for simulation, the processes being considered as objects of processes-classes belonging to a temporal structure. Because of adoption of SIMULA 67 to the HBDS and of the capabilities of this programming language, it becomes very easy to associate spacial components with processes, in order to study the evolution of a cartography after a crowd of events.

In the preceding paragraph, we deal at last with geology and this is a science which has been firstly used to test the HBDS and the connection between space and time. Figure 10 shows in the left-hand side upper corner a very simplified map where geology is superimposed on topography. The reader must immediately deduce that the corresponding structure is composed of the three classes of the topography and the seven classes of the geology (and some others...). Intersections between geologic boundaries and topographical isolines are materialized by links and allow to define some curvilinear areas indicated as "ZONE" on the picture in the left-hand side lower corner. The zone thus represented is composed of two isolines and boundaries, but there are many real cases rather complex. Now, we consider the process-
Figure 7 -- A substructure composed of two classes, 'PERIFERY' and 'BORDER', is connected to the topography. Though it has been only conceived to represent this artificial phenomenon, it allows one to consider the topography as if the elements were not limited by the border of the map. A same substructure concerns other types of thematical maps.
Figure 8 -- The three types of French administrative subdivisions, their data structure, and the simplification of this one by using two hyperclasses (Administrative Boundary and Administrative Area) and a hyperlink.
Notations are those introduced for map digitisation by using a directed multivalued multigraph (BOUI 73, 74).

Figure 9 -- The Substructure Corresponding to the Geological Maps
A very simplified map with the topography and the geology which are superimposed.

Geology and topography intersections allow to define a zone in which the factors may be considered as constant; such a zone may be associated with a specific process.

Figure 10 -- Spacial structure and process structure linked together, allowing the simulation and modelling of the real world with an accuracy better than the models based upon matrix for the geometry and on sequential processes. Here, the objects of the process classes are considered as working in quasi-parallelism and interaction. Such features are easily obtained with a language like SIMULA 67, and especially the SIMULATION class. (Pictures from BOUI 76).
structure, drawn in the right-hand side upper corner, comprising three classes named 'EROS', 'TRANS' and 'SED' for erosion, transportation and sedimentation. Two subclasses are distinguished, according to the nature of the material which is concerned (limestones or sandstones). Three objects are represented. SIMULA 67 allows the quasi-parallelism and the interaction of processes, and for instance, an erosion-process may activate a sedimentation process, such an interaction corresponding to a link in the HBDS representation. But, a process may be considered as a constant in each zone, because the conditions are well-defined. So, we associate a process with a zone, and this association requires the linking of the spacial structure with the temporal one. In the right-hand side lower corner, such a correspondence is done, a process (erosion of sandstone) being linked with the cartographic elements defining the "zone", namely two isolines and two boundaries. Then a process may modify some cartographic elements, even make them disappear; reciprocally, some features of cartographic components may interrupt a process which concerns them. For instance, an erosion process may modify the coordinates of the borders of a zone; but if the erosion makes the zone disappear off the map, this specific process is stopped. Likewise, processes may generate new cartographic components; for instance, a differential erosion on a top with a limestone between two sandstones is able to generate a saddle in the landscape (BOUI 76).

Other examples are presently studied, and the system seems to be convenient for environmental studies and other domains for which cartography and processes have an equal part to play.

VERY LARGE CARTOGRAPHIC DATA BANK:

The HBDS is also the kernel of a new model of data bank (BOUI 77) which is protected, simultaneously shareable, portable and distributed. The HBDS is used as the external concentrical abstraction of the whole system, the only one which is seen by the user. This user describes, manipulates, and displays the data through the HBDS concepts. But because of the universal capabilities of the HBDS it is also used to structure the data bank system, the components of which are hidden classes, objects, attributes and links, with a boot-strap mechanism. The implementation will be operational at the end of year 1978.

The first applications which will be processed are the Paleontological Collections Data Bank of the Université Pierre et Marie CURIE (800,000 specimens, 40,000,000 data) and naturally the cartographic classes (presently 150 and probably 500 at this time). In fact, the HBDS may be used without the data bank, if the number of data does not require such a sophisticated system.

CONCLUSION:

The purposes of this paper were only to present the HBDS and some simple applications in cartography and simulation. The set theory and the hypergraph concept allow to define new tools which are more efficient for complex scientific data processing.

The HBDS is not isolated, but has the help of two powerful tools, EXEL and SIMULA 67, the first one for the algorithmic expression, the second one for
the programming (which is essentially different and is often confused).

In the whole paper we have dealt with the natural structure of the phenomenon we try to recognize, and never with a particular problem. I want to emphasize such a method: if the structure is well-recognized, then all problems have a solution in a partial subgraph of the data structure, else there are two cases: either the problem is a nonsense (user error) one, or the structure has been badly recognized or has been incompletely studied (manager error).

The most important characteristic of the HBDS is perhaps the fact that it has been defined on a topological basis which is the main topic of this symposium and will be the final conclusion of this paper.

REFERENCES:


(BOUI 76) BOUILLÉ, F., 1976: Cartographic reconstructions by displaying and simulation, 8th Int. Cartographic Conference, Moscow, preprint 13p.


(BOUR 39) BOURBAKI, N., 1939: Théorie des ensembles, Livre I.
