DATA STRUCTURING IN TOPOGRAPHIC DATA BASES

M.M. Radwan*, J. Kure*, M. Al-Harthi**
International Institute for Aerospace Survey
and Earth Sciences (ITC)
350, Boulevard 1945
7500 AA Enschede
The Netherlands
Commission IV

1. Introduction

Digital technology was successfully introduced in the field of mapping in the late 1960's as a means of speeding up the production process and particularly the plotting and fairdrawing phases. It was not until the mid 1970's that the information perspective began to dawn and digital map data was seen to have a value by itself, besides being the data source for the traditional printed map. However, in order to be able to exploit this newly found value, the traditional topo map and the relationships seen between map elements with our human brain, had to be translated somehow so that they could be handled by a computer. This required that the digital topo map data be structured and that linkages be established between map elements for particular applications.

This paper outlines the general considerations to be taken into account in data structuring for multi-purpose applications, paying particular attention to the establishment of spatial relationships between data elements, both on a horizontal level within a single map layer and between separate map layers. An example is given of system implementation on an existing digital mapping system.

2. Need for Topographic Data Bases

National survey and mapping agencies are under pressure world-wide to reconsider their traditional roles, activities and organisational forms, as a result of general economic pressure, changing user requirements for more and more complex types of information and due to the technological developments over the past two decades that are either replacing or complementing established technologies in map and geo-information production.

In particular, the demand on these agencies to produce topographic data in digital form in order to satisfy the requirements of the large number of spatial information systems set up or being set up, is forcing agencies to a greater awareness of being in the information business, which includes mapping and not just the mapping business. This awareness in turn is leading to expansions of the missions of mapping agencies from conventional topo map production to the provision of the topo data bases on which a series of GIS-operations can be performed in different disciplines.

* ITC, Enschede, The Netherlands
** Geographic & Surveying Institute, MSD, Riyadh, Saudi Arabia

Trends which have been observed ([1], [3], [9], [16], [17], [19]) in this transitional process and in the assessment of the realistic requirements for topographic information are:

- a shift from the use of traditional topo maps towards the use of digital topo information
- as a minimum requirement, the digital topo data base shall include all features traditionally included on topographic maps
- a shift from the collection of digital topo data on an ad-hoc basis to satisfy specific project requirements towards the systematic creation and maintenance of a topographic data base that can be used continuously by all users
- a shift from the supply of map data in digital form towards the delivery of structural information, whereby the topological relationships amongst entities and the various attribute features are explicitly represented in the data.

In order to appreciate these differences, it should be borne in mind that digital techniques were introduced in mapping as long as 20 years ago. This was, however, to support the cartographic processes and directed towards speeding up production. As a by-product, the digital map data base, in which the map elements were grouped into map layers (map separates) and the data files were indexed for later use according to the map sheet numbering system.

Now, whilst this type of data base is effective for display purposes, it cannot be effectively used in GIS operations. First of all, it is based on a simple data structure (the line-map model) which only contains two geometric primitives, points and lines, and objects are defined by unrelated strings of coordinates. Secondly, it does not permit the creation of higher order structures or the recording of spatial relationships between map elements such as continuity, connectivity, adjacency, etc. These relationships, obvious to the human brain in the visual inspection of maps, are essential for digital techniques in spatial information analysis and modelling such as network analysis, highway routeing, stream flow modelling, polygon overlaying, cross-layer analysis, adjacency analysis, etc. The lack of these features in many of the existing commercial digital mapping systems is a serious shortcoming, which has lead to the emphasis on many occasions ([1], [8], [10], [12], [13], [17], [19]) of the need to establish a digital topo data base with a data structure and design parameters based on the multi-product concept i.e. be able to support the production and revision of a national map series in addition to GIS operations.

Measures currently being undertaken by mapping agencies to achieve this goal include:
- the establishment of standards for the storage and exchange of data
- the conversion of existing digital data into the required data structure. This includes the integration of data collected from various sources.
- the initiation of R & D programmes to tackle problems such as data organisation for GIS queries; quality models for the accuracy, resolution, completeness and temporal validity of data base information; economic models relating to the establishment of multi-purpose topo data bases; etc.

Whilst essential, these are only preparatory measures, since the implementation of the topo data base system in a particular environment requires that decisions be taken on the data base contents, the data model, the data organisation and the data format, and the decisions on these characteristics depend on the information needs of all potential users of the data base system.

The problems of determining the realistic information requirements of potential users and establishing a sufficiently flexible data model to support these requirements and even accommodate changes in user requirements without necessitating a major restructuring of the data base contents, are the subjects of this paper.

3. Considerations in the design of data models

Data models provide the abstraction, perception, fragmentation and description of the actual topography in a computer environment. This abstraction of reality will only incorporate those elements of reality considered to be relevant for the user applications under consideration.

Similar to the procedures adopted by others ([2], [5], [11], [14]), topographic data items that are of interest in a spatial data analysis are categorised here as follows:

3.1 Feature types

The two feature types that need to be distinguished are:
G (primitive objects): basic geometric primitives of map elements (points/lines/polygons)
O (compound objects): a higher order class of primitives, to which a user can assign information

3.2 Feature properties

Feature properties include both the attributes describing individual objects as well as the relationships or associations between them.

Attribute/feature descriptors (A):
A1 (spatial attributes): location (positional coordinates, orientation), geometry (object form)
A2 (phenomenological/functional attributes): non-graphical characteristics.

"Quality" parameters should also be included to allow users to evaluate the suitability of the information for their particular application. Examples of parameters used [7] are the information sources (type, resolution, date, reliability), the data derivation methods used, the decay rate, the completeness, the classification reliability, etc.

Relationship/association with other objects (R):
R1 (spatial associations):
RTH: topological relationships in order to connect physically contiguous features in the same layer/theme/class (connectivity of line networks, adjacency of polygons, etc.)
RTV: topological relationships between overlapping polygons in different layers/themes/classes, as well as references between objects of the same theme but crossing each other at different vertical levels (i.e. road crossing)
RP: proximity relationships to features that, although not contiguous, are declared neighbours on the basis of closeness. (RPH and RPV for relationships between features in the same or different layers respectively)
R2 (phenomenological/functional associates):

RF: relationships between features that are functionally related (RFH and RFV for relationships in the same or different layers respectively)
RD: dynamic phenomenological relationships between features, describing changes in the geometric or semantic properties of a feature as a result of changes in other features (RDH and RDV for relationships in the same or different layers respectively).

3.3. General considerations in data structuring:

Topographic modelling differs with the level of data organisation, starting from the actual real-world situation, and then passing through the application-oriented information structure (i.e. the mental construction of the real-world situation as viewed by system users) to the internal data structure of the system which manages the data (i.e. the data model for the data base management system) and the physical file structure in a specific computer environment ([13], [15]).

The initial modelling phase requires a conceptual understanding of the various user processes involved in utilising the data, since redundancy has to be avoided between different user perceptions of real-world phenomena in identifying and structuring the data elements. This structuring includes aspects such as arrangement of elements, analysis of linkages between elements, analysis of aggregation of elements into various classes, deciding on how attribute information can be assigned to the elements and finally deciding on how functional associations (non-graphical relationships) can be established between elements.

This application-oriented information structure is not necessarily that in which the data will be collected or finally supplied to the user. For ease of implementation, however, it should not be too abstract and it should have some resemblance to the structure of the data base management system. The latter is not a strict requirement, since some designers are able to establish their own DBM systems which allow them to map the conceptual model and its data structure directly into the physical file level. ([2], [20]).

There is a growing acceptance of the use of the topological model for the conceptual modelling of topography and a number of well-known data structures have been built up around this concept ([4], [6], [20]). A particular advantage is that in describing topographic features by sets of topological elements (nodes, arcs, polygons) and storing these without any redundancy, the spatial relationships of connectivity and adjacency between elements are explicitly expressed and are included as data items in the data structure.

For the sake of simplicity, most of the available topological structures deal only with one level of information. If layered information is available, the individual layers are structured independently of each other and with no explicit links between them. This approach allows referencing on the layer level (horizontal integrity) but no cross-referencing between layers (vertical integrity) and, as a result, manipulations such as polygon overlays for composite presentations and analyses have to either be derived visually or algorithmetically in a separate step.

The latter will undoubtedly be experienced as a serious shortcoming in all the user situations requiring more complex spatial analyses of data and will lead to the need to extend the simple, single-layer topological model to include the various relationships described earlier: topological, proximity, functional and phenomenological.

In principle, the topological relationships (RTH, RTV) are probably a minimum requirement for the expansion of the simple model. The inclusion of the other relationships will depend upon a full analysis of the user requirements in a particular situation, whereby it will have to be borne in mind that as more and more relationships are included in the data structure, the data base management system will become more complex and the computer overheads (storage and response time) will increase, particularly in the case of interactive user queries.

3.4 Multi-level data structuring

An efficient solution to the problem of reducing the complexity of the topological structure in a particular situation could be the application of a multi-level approach i.e. start with a structure (0), with minimum requirements such as the topological relationships RTH & RTV and then proceed to structures with increasing complexity. The structure (0) should at least include information in order to support the conventional topo map production and provide the geometrical reference to other structure-levels. This approach will be elaborated on in the case study.

4. Case Study

4.1 Introduction

The following report is on a study conducted to establish guidelines for the design of a data model for an existing mapping system, based on the above considerations ([17], [18], [19]).

The study was conducted from the two perspectives of:
- establishing a mechanism for the identification of the users, their requirements and their usage of topographic information.
- designing a data model which will be able to support these requirements, within the constraints of the available mapping facilities.

The available mapping system is based on the SysScan (GINIS) of Kongsburg. This system is a component in an integrated digital mapping system which can integrate both vector-based and raster-based data, in addition to having the capability for conventional line mapping.

4.2 User needs

Goals set up for the topo data base system were:
- it should be comprehensive and satisfy the requirements of as many users as possible, bearing in mind that the needs of all potential users can never be known completely and that the main challenge will be to ensure that the data base system has neither insufficient nor redundant information.
- it should be national in scope, making allowance for the fact that the information requirements for security and development purposes will vary from one region to another.
- it should be flexible and allow changes to be made to the contents, structure and topological links of the data base.

- it should be linked to the on-going national topo mapping programme in order to support map revision at the scales 1:50,000, 1:100,000 and 1:250,000.

The action plan drawn up to achieve these goals has as main elements:
- the identification of user requirements for topo information, including the detaildness of the classification required for each item, the degree of importance attached to each item and the urgency with which the various information items are needed. In collecting these requirements, an active approach is adopted of studying what users are doing and thereby evaluating what their real needs are, rather than concentrating on the map products users might request in order to satisfy their needs.
- reviewing the topo mapping situation, with particular regard to the requirements for a systematic revision programme.

- prepare proposals for the data base design (data base content, data format, data structure) that will allow users to evaluate the "quality" and thereby the suitability of the information for their particular application; allow users to view, to access and to aggregate information on freely defined criteria, be these spatial, thematic or functional in nature; allow users to navigate through the geometry of stored elements in single or multi-layers in order to establish relationships such as connectivity and adjacency; etc.

4.3 Design and implementation

It is beyond the scope of this paper to go into all the details of data base design. Instead, only a summary will be given of how the complex problem of data structuring was resolved within the constraints of the SysScan digital mapping system.

The GINIS-Graphic Data Structure (GDS) of the SysScan System is object related, whereby all objects and all information assigned to them (spatial or functional descriptors) are included in the hierarchical data structure. The system hierarchy and its logical levels can be extended or modified at any time and the system further allows the definition of functional relationships (pointers) between the objects themselves at any level in the hierarchy or between the objects and the associate descriptors.

Figures 1 and 2 shows the organisation of the "information structure", in which the 3 main classes of data items are:
- topographic entities in a logical hierarchy of first the theme/class data set (C), then the object data set (O) and finally the graphical primitives (nodes/arcs) (G) at the lowest level.
- attributes, in their two classes of spatial (A1) and non-spatial or functional (A2)
- relationships or associations, in their 4 classes: topological (RT), proximity (RP), functional (RF) and phenomenological (RD)

These data elements are stored in separate units or blocks: logical blocks for the data classes of the C type; objects blocks for the O type; geometric blocks for the G type primitives; associate-data blocks for the non-graphical descriptors A2 and relationblocks for the relationships of the RP, RF, and RD types. The logical blocks are linked to each other through up, down and side pointers, hereby providing a logical meaning to the other data blocks to which they are connected. This permits not only objects but also

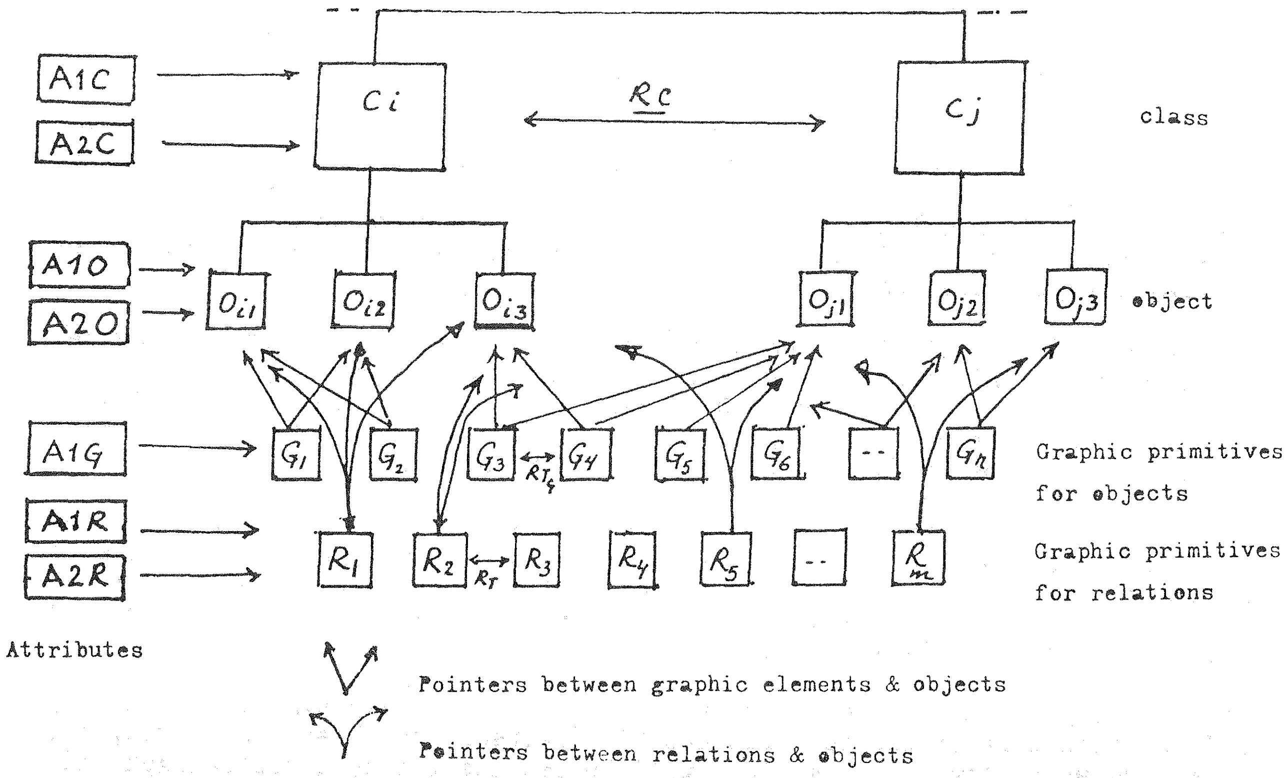

Figure 1

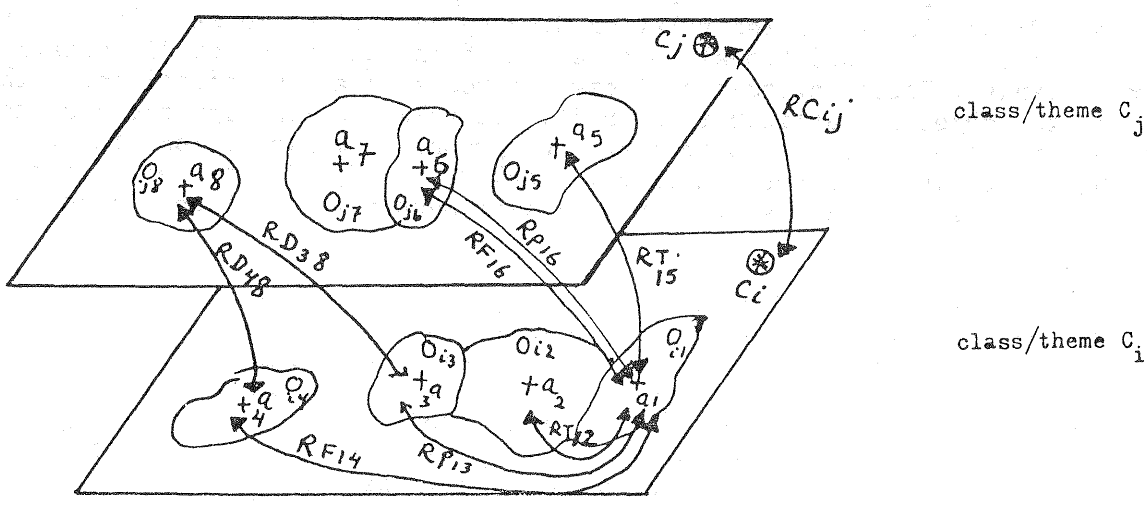

Figure 2

the spatial and functional descriptors and the relationships to be hierarchically structured. The "relation block" is not a SysScan term and it is introduced in this study as will be seen later.

If data items, and the blocks describing them, are linked to more than one data block on a higher hierarchical level, then they have various logical significances although they have only been stored once. This provides the possibility of handling common portions of layered data sets in an unredundant manner and the geometrical consistency supports cross-layer spatial analysis queries.

The above does not take care of queries related to non-topological relationships. This problem was solved by adopting an approach similar to that used in setting up the HBDS model [4], i.e. the geographic entities, their logical classes, their spatial and functional properties and their interrelationships are all represented in terms of objects.

This is illustrated in fig. 2, where the points Ci and Cj represent the class vertices for the class data sets Ci and Cj respectively; the reference points a1, a2,..... represent the object vertices for objects Oi1, Oi2,... of the types O or G and the arcs (links) represent the various types of relationships established between these elements.

When two objects are linked by more than one relationship e.g. the proximity and functional relationships RP16 and RF16 between the objects Oi1, Oi6, the links between the two vertices will be represented by a multi-graph [4]. Further, some of the links will refer to "virtual" objects that do not exist in reality. These are V-links, as opposed to the R-links that are associated with real objects.

These relation-graphic primitives can be structured in a similar way to the real graphic elements i.e. further grouped, classified and hierarchically organised in a "relation-hierarchy", comparable to the "object-hierarchy" described earlier. These structural elements will be stored in the GINIS Graphic Data Structure (GDS) in blocks of the geometric block type mentioned before. These blocks (called relation-blocks) have links to associate data blocks which carry additional information of the A2R type for "information carrier links". These information items include a relation identifier, the access to objects participating in it, its role and its significance etc. Furthermore, the various graphical and non-graphical manipulation functions of the GINIS system (access, extract, display, modify, delete, etc.) can be applied to these "relationships" in a similar manner to the "real" graphical elements.

However, in view of the complexity of the problem, it is inadvisable to include all of the relation sets in a single data structure, but rather to apply the multi-level data structuring approach outlined earlier. This involves starting from structure (0) that fulfils the minimum requirements of supporting the conventional topo map production and providing the geometric reference for the other structure levels that are then built-up independently to serve the needs of a particular user group.
To realise this, separate files are created for the various information categories:

File (1): Graphic file, including information about the themes/classes, the objects, the graphic primitives (arcs, nodes) and the topological relationships RTH and RTV

File (2): Relationship file, including information about the various types of relationships (RP, RF, RD), the graphic primitives representing these relations and the topological links between these graphic elements.

File (3): Attribute file, including information about the spatial and functional attributes. (A2C, A2O, A2R)

The various data structures (data models) are now built-up by transferring information of interest from these files into another file, using the GINIS.EXMAN package, and re-organising this depending upon the structure required for the particular application e.g. structure (0) for cartographic production will contain the information of file (1), re-classified into the conventional map overlays, whereas the application dependent structure (i) will contain the information of interest of files (1), (2) and (3) re-structured so that it can respond in the most effective manner to user queries.

5. Conclusion

The data structure and design parameters of a digital topographic data base should be based on a multi-product concept in order to effectively support the production of topo maps as well as GIS operations. The relationships between different data sets should be explicitly included in the data structure. This concerns topological, proximity, functional and phenomenological associations.

The graphical presentation of these relationships in a corresponding structure to the real graphical elements allows their manipulation within the digital mapping system in support of spatial analysis queries.

The multi-level data structuring approach permits different user groups to be served with minimum computer overheads in a multi-product environment.

REFERENCES

[1] Allam, M.: "The development of a national digital topographic data base", 40th Photogrammetric Week Stuttgart, 1985.

[2] Allam, M.: Lecture series at ITC, with reference to the strategic plan for the Topographic Survey of Canada, 1986

[3] Awda, G : " A study concerning the establishment of a digital data base system in Iraq", M.Sc. Thesis, ITC, 1983

[4] Bouille, F: "Structuring cartographic data and spatial processing with the HBDS", Harvard Papers on Geographic Information Systems, Cambridge, Mass., 1978

[5] Brassel, K.: "Topological data structure for multi-element map", Ibid, 1978

[6] Chen, P.: "The entity relationship model: toward a unified view of data", ACM Transactions on Database Systems, 1976.

[7] Chisman, N.: "An interim proposed standard for digital cartographic data quality" WG.II Report, National Committee for Digital Cartographic Data Standards for USA, 1985.

[8] Collins, S., G. Moon, T. Lehar: "Advances in geographic information systems", Proceedings of ISPRS Congress, Rio, 1984.

[9] Dangermond, J.: "Geographic Data base Systems", ACSM-ASP Fall Convention, 1984

[10] Deuker, K.: "Geographic Information Systems -Towards a geo-relational structure", Symposium on Spatial Data Handling, Zurich, 1984.

[11] Makkonen, K.: "Modelling a dynamic geo-data base - problems of data accuracy and structure conversions in data collection and processing". Proceedings of ISPRS Congress, Rio, 1984

[12] Moellering, H.: "The challenge of developing a set of national digital cartographic data standards for USA" National Committee Report, 1985.

[13] Nyerges, T.: "Representing spatial properties in geographic databases", 40th Annual Meeting of ACSM, 1980

[14] Nyerges, T.: "An interim proposed standard for digital cartographic data organisation", W.G.I Report, National Committee, 1985.

[15] Peuquet, D.: "A conceptual framework and comparison of spatial data models", Symposium on Spatial Data Handling, Zurich, 1984.

[16] Radwan, M.: "Establishment of a land data bank in Egypt", U.N. Regional Cartographic Conference, Cairo, 1983.

[17] Radwan, M.: "Digital photogrammetric systems in an integrated digital mapping environment", Integrated Digital Mapping and Information Services Project, MSD, Riyadh, 1986.

[18] Radwan, M.: "Establishment of guidelines and standards for digital map production", Ibid, 1987

[19] Radwan, M., J. Kure: "Design considerations in the establishment of topographic data bases" (In preparation, 1988)

[20] Shapiro, L., R. Haralick: "A spatial data structure", Geo-Processing Vol.1, No. 3, 1980