TOPOLOGY FOR CADASTRAL SYSTEMS: DO WE NEED IT ?

By G.Carrai(*), R.Favilli(**), L.Morandini(***)

1 SUMMARY

An emerging issue in GIS/LIS technologies applied to cadastral systems is the question whether a full topologic structure is really needed or if modern technologies and algorithms are able to handle geometrical primitives at a lower level.
The answer to this question leads to different base and application software solutions and different final uses of cadastral data.
Authors explain what topology is and which advantages it brings in dealing with geographic data, using examples of topologic data model actually used by commercial GIS software; non-topologic data models are also reviewed as these become more and more commonplace.
Data modelling issues, specific to cadastral applications, are also analysed and advantages/disadvantages of using topology are reviewed.

2 WHAT IS TOPOLOGY

According to the dictionary topology is: *"a branch of mathematics concerned with those properties of geometric configurations (as point sets) which are unaltered by elastic deformations (as a stretching or a twisting) that are homeomorphisms"* (1994 Merriam-Webster's Collegiate Dictionary).
Which, in more mundane terms can be explained as: topology is the set of relationships between geographic features which are not affected by projections and transformations. An example of a relationship which belongs to a topology is the convergence of four roads to a cross-roads; this convergence remains true whatever transformation we apply, be it a projection (from geographic to metric coordinates) or a change of ellipsoid (from, say, International 1924 to GRS80). On the contrary, many relationships between geographic features change when a transformation is applied: the distance between two points or the shape of an area change when we project geographic features.
The storing of a subset of these relationship in a set of files is what, more properly, GIS people intend when they talk about "topology". Some GIS, for instance, store the fact that four roads converge to a cross-roads by storing the coordinate of the point which represent the cross-roads together with the identifications numbers of the four lines representing the roads; hence it is fast to retrieve which roads are connected to which, and this is of utmost importance for building efficient best-path algorithms.
Computing all relationships of which topology is composed requires great computing power, so, traditionally, this step is done once for all and all relationships are permanently stored as links between records in such database. Just an example to clarify: to check which roads are connected every line's end should be matched with each other line's end; and this is a task which increase exponentially as the number of line increases.

2.1 Some aspects of topological world

<u>Plane topology model concepts</u>. Conceptually, plane topology can be defined as a planar graph, where geographic reality is decomposed into a finite set of 0 cells (nodes), 1 cells (edges), and 2 cells (faces). This terminology is defined by an algebraic topology that establishes rules for decomposing continuous three-dimensional objects into representations of finite models. Once this topologic mapping has been performed, a system can be modelled in a way that permits more complex relationships between objects to be established.
The purpose of topology is to capture and retain knowledge concerning a cell's spatial and thematic relationships with its neighbouring cells. For a topologic model to be valid, these relationships must remain constant regardless of changes in scale, shape, or size. With topology embedded in a data model, very useful relations can be established, such as adjacency and connectivity. Topologic and geometric relationships (such as size, angle, and shape) provide powerful resources that allow geographic reality to be fully modelled.

(*) Managing Director of SVALTEC – Florence Italy - E-mail: svaltec@ats.it
(**) Managing Director of Mappamondo Informatica – Florence Italy - E-mail: R.Favilli@fol.it
(***) GIS Consultant – Terni Italy – E-mail: l.moran@ita.flashnet.it

Topologic objects. Plane topology extends graphic models of nodes and edges to the development of a more powerful and expressive model that contains spatial relationships. In addition to metric capabilities (distance, shape, or size), topology determines spatial neighbour relations. By defining more rigorous and complex relationships in the data model, the true properties of a system can be more effectively represented. Various mathematical models are available. In addition, simpler models can be used to create more complex ones. The most simple is the graphic model. More complex and useful surface-based models are built upon the graphic, using topology to capture additional analytical information.

A graphic model represents geometric information based on node and edge primitives. This model provides the base for other models that define more complex relationships. The node primitive is composed of a location in an established coordinate space. Most representations are two dimensional (x,y) or three dimensional (x,y,z). An edge is composed of a minimum of two nodes, with more details concerning linear or spline interpolation between the end points.

Because of the complexity of geographic information and the limitations of a graphic model, plane topology provides a better model for defining relationships. The use of a planar topology model (based on the two-dimensional manifold), for instance, maps more concisely in the two-dimensional space that current computing systems can manage. VPF, for instance, provides four distinct levels of topology: full planar topology (level 3); a linear planar graph (level 2); a nonplanar linear graph (level 1); and no topology defined, indicating a geometric model (level 0). Many standards use the notion of topology as a constraint to enforce integrity rules upon the feature definitions. As the entities require fewer topological relationships, the rules can be relaxed. For instance, if linear features in a transportation network are being modeled, then the requirement of full planar topology may be relaxed because it is not necessary.

The concept of topology held by the geometric primitives is carried upward to features and their associated thematic information. An area feature is usually labeled, or contains information pertaining to the enclosed region. For instance, an area can have a category (soil class, surface material type) or a numeric value (population size, number of airports). Topology is used to provide operations and information to distinguish between these thematic objects. Thematic relationships can exist between features without requiring the primitive geometry. For instance, a set of the islands in the Pacific Ocean (Oahu, Maui, Hawaii, Molokai, etc.) can be defined as different features with differing geometric primitives, but they can be related to one another as the Hawaiian Islands.

Topological operations. Topological operations are based on the single notion of adjacency—that is, if two objects are next to each other, it is necessary to maintain the adjacent relationship between them. To distinguish the topological aspects from the geometric aspects of geography, we are only concerned with whether two objects, A and B, are adjacent to each other and not with whether A is bigger than B, or one is to the north of the other, or the length of their common boundary.

Many complex topological operations can be derived from adjacency alone. In the georelational data model, two topological operations are paramount: boundary and coboundary. For example, an edge has a start node and an end node; the nodes are the boundary for the edge. The edge, in turn, is the coboundary of the node. Of course, the coboundary of a node can have more than one edge if many edges meet at a node. Similarly, faces have edges as boundaries. The coboundaries of edges are maintained in the left and right faces.

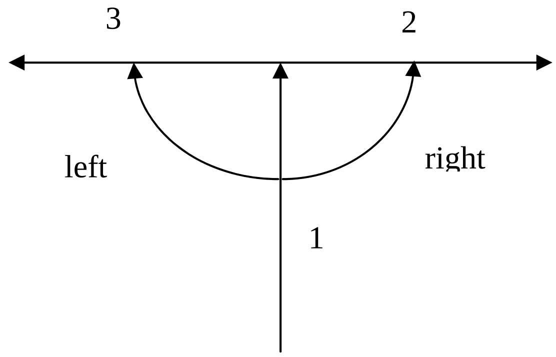
Topological rules. The integrity rules of the topological model are contained in the definition of the objects themselves. A plane model restricts itself to planar geometry, where all entities must lie in the same plane. In addition, all faces must be mutually exclusive and not overlapping. These constraints allow the objects to be defined in context and allow operations to be performed in a consistent manner. While these rules may seem to restrict the system model, they define the data model's domain, taking advantage of the underlying structures. By restricting the faces to be constructed of nonoverlapping regions, powerful set operations can be applied to the objects (such as union, intersection, or join).

The georelational data model. The use of a combination of the relational and the planar topologic data models provides a hybrid model for geographic data management, analysis, modeling, and display. The georelational model provides the data structure foundations for a spatial database, and software provides the rules and operators that manipulate topology, geometry, and relational objects in the form of tables. Whenever an operation requires thematic information, the use of relational and topologic table operations are used to supply the result. If the operation is spatially related, geometry and topology together will be used. This triad of categories (geometry, topology, and relational tables) provides a robust database architecture.

The georelational operations and algebra are not part of the standard, but rather are implemented in software.

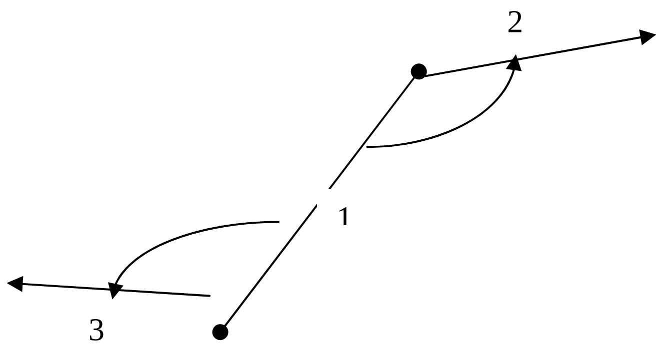
2.2 Navigation in Topology

Topology is a “self-sustaining” structure in the sense that an ASCII file topological structured might allow to draw the features without any specific software application but only a “tracing” system.
Full topology gives in fact a manner to follow lines to reconstruct the geometric feature: each arc bears at its terminating nodes the indication of the “route” to follow to find the next arc “coboundary” at the node to depict the feature.
However different standards have different systems to navigate; from this, the need of a “header” containing the criteria of navigation.
Having a close eye to two different approaches: in CEN standard, the spatial relations between edges is stick to each node, so that a left edge of edge A is the edge on the left of terminating node of edge A, and the right edge is the edge at the right of terminating node of edge A. Optionally it could be given also the previous right and left edge.



Edge 2 is the next right Edge with respect to Edge 1.
Edge 3 is the next left Edge with respect to Edge 1.
Edge 1 is the previous right Edge with respect to Edge 2.
Edge 1 is the previous left Edge with respect to Edge 3.

In VPF standard the right and left edges are those, respectively: the edge which is encountered counterclockwise with respect to the terminal node and the left is the one encountered counterclockwise with respect to the beginning node.



Edge 2 is the right with respect to Edge 1.
Edge 3 is the left with respect to Edge 1.

In both cases the edge bear also the information of left and right face sharing the edge.
More information is required when an edge match the boundary of the sheet or as technically is called, the “tile”. In such a case, besides the right/left faces and edges, more fields have to be add to indicate the Tile, the external face (the one of the adjoining tile) and the External edge (the one continuing the feature in the adjoining tile).

3 TRENDS IN COMPUTING

3.1 Higher capacity mass-storage, faster processor

We were used to processor which doubled in power every one year and half, but now we are seeing the same happening for mass storage and main memory. A low-end PC doesn't come with less than 1 Gigabytes of hard disk space and 8 Megabytes of RAM, this means that the overall performance of a computer has much improved because the relative high-cost of disk or RAM was a bottle-neck.

Therefore, now we have, at very low price, powerful computers, with a main memory which can store huge geographic data sets and plenty of mass storage, on which we can store a serious raster data base: what can we do with this equipment ?

One thing could be to use a data structure which consumes more disk space but deliver faster performance, another one could be to have some topological calculations done on-the-fly and only when needed. As we have seen, topological calculations are complex, and they should be done whenever a geographic feature is changed, this leads to a data structure which is hard to manage, and useful only when checking the correctness of geometric data or when making some complex data analysis (topology overlay, buffering, shortest path search, ...), but for all the displaying and query, topology is useless, and this kind of simple tasks are the vast majority of operations done with a GIS.

Hence, why not to skip all the topology building making simpler data structures in order to gain quicker displaying and query ? This is precisely what some GIS makers have done .

3.2 Convergence of information media

Computers used to crunch numbers, and they still do, what is new is that they can make users see, hear, fell (virtual reality) those numbers. Databases used to be filled only with alphanumeric data, for which the relational data model was built; but now images, drawings, video and audio clips, not to mention hypertext links, are an increasing part of an organization's data, so it makes sense to store all this different stuff with the same DBMS which uses to manage the "ordinary" data.

Therefore, the current trend in DBMS is to make the "universal database" which can store and manages data in many forms (multimedia data) and make them available to specialized programs for editing. An universal database is built upon an extension of the relational data model and exposes an extension to SQL to let application software retrieve multimedia data; apart from storing, multimedia data need specialized software for editing, an image needs a photographic software to be manipulated, while a drawing needs a CAD program and an audio clip needs a sampler; hence, a universal database is completed by many specialized add-ons, usually made by third party vendors.

4 TOPOLOGY AND GIS SOFTWARE

4.1 A system built on topology: Arc/Info

Arc/Info is one of the leading GIS software, it was originally developed during the '70 and much improved and rewritten ever since, it is owned by ESRI Inc, (an US firm) and, being multiplatform, run on almost every operating system: from Windows NT to UNIX or VMS workstations made by IBM, Sun, HP, DEC, Silicon Graphics. Arc/Info is meant to be an all-in-one product, it has plentiful of features ranging from vectorization of raster images to 3D analysis to map algebra.

Arc/Info uses a so-called "geo-relational" data model, in which geometric data are stored as graphical files, while alphanumeric data are stored as tables in a database. Topologic relationships are accessible by users because they are stored in tables, but after editing of a geometric feature these relationships have to be rebuilt.

The kind of relationships which make topology depends on the kind of feature, here's a list of topologic relationship divided by kind of features:

* Point: No relationships, every point is by itself
* Line: Which line is connected to which
* Polygon: Which lines make a polygon, which polygons a line divide (every line has a polygon on its left and one on its right)
* Route: Which lines (or portions of lines) make the route

* Region: Which polygons make the region

This set of relationships makes it easy find the best-path through a roads network, or to find little polygons which are produced by data capturing mistakes or polygons' contours which don't close (the first point is not the same as the last), line which intersect at points other than their ends, and so on. Despite its usefulness, it has some drawbacks, to display a polygon means to retrieve every line which belong to it (relationships are build on top of each other, so there are no redundant data) and to display them: this means to access many records scattered in different tables just to display a feature. Moreover, if a user wants to edit the polygon he has to lock (to make not editable by other users) it and every neighboring polygons, make changes, rebuild topology.

4.2 Non topologic approach: ArcView

ArcView is another product from ESRI Inc. (the same maker of Arc/Info) and was made in 1994, it's available on the same platform as Arc/Info and was originally meant to be a low-end viewing product, but its capabilities, in some areas, are quite powerful and rivals those of full-fledged GIS packages.
ArcView can read topologic data produced by Arc/Info, but can't create such data, its native data format is not topologic. In essence, this means that every geometric feature is by itself: every feature is stored on a record with no links to other features. Data are redundant, polygons which share a common line have this line replicated for each polygons, this lends to consume more disk space. This lack of topology, also, means that to modify a feature is easy, no topologic computations are involved, and that retrieve and display a polygon means just to read one record.
Lack of topology means that if the use wants to find the best path through a road network, an additional, topology building, step is to be carried out, moreover, control over the correctness of geometric data is difficult, so multiple users can modify features easily, but at the cost of unproved data.

4.3 Intergraph face to topology

Intergraph Modular GIS Environment, better known as MGE, let to its users to choice if use topology (that is topologically structured files) or not.

Basic MGE modules - MGE Basic Nucleus, MGE Basic Administrator, MGE Base Mapper - still using Microstation as graphic engine - simply employ Microstation 32 format to implement geographic data management. This doesn't absolutely mean a lack of efficiency in areas handling; even the lowest level member of MGE family, the personal oriented MGE-PC, is provided with advanced tools for the construction and the management of polygons. Main functions to built polygons from a certain boundary lines set are:

- automatic creation of nodes relative to all intersections
- automatic flag or correction of free endpoints within an established tolerance
- automatic deletion of spurs, that is the "small" overshoots and dangle arcs
- automatic merging of duplicated lines
- automatic placing of centroids
- automatic creation of polygons generated by a line set

Also sophisticated spatial analysis is allowed combining MGE spatial operators (inside, intersect, overlap) and spatial criteria (selected elements, external or internal bands and rings etc.). Spatial constraints can also be combined with attributes-based query.
Intergraph developed a specific MGE-based application, named Parcel-Manager, specialized in cadastral data handling. The user is guided by a simple icon panels and forms interface trough all basic cadastral operations, like creation of the parcel, change of its ownership, splitting, merging, calculation of area, perimeter, frontage etc.. All functions are implemented fully integrating geometric information with data-base registration data. The combination of Parcel-Manager with Parcel-Vec- the product for automatic vectorisation from raster images of both property boundaries and parcel numbers- represents Intergraph ready-made, non topological proposition for massive, quick cadastral data input and maintenance.
On the other side, if you anyway decide to employ topology, MGA- MGE Spatial Analyst- provides you the capabilities to create, query, analyze, display your topologically structured geographic files (topofiles in MGA jargon) , generated from Microstation 32 and relational database files.
In this case you have the advantages of a completely integrated spatial/attribute query language which allows the combination of boolean operators - such as AND, OR, NOT - set operators - such as INTERSECT, UNION, MINUS, DIFFERENCE - with 20 spatial operators - such as Overlaps, Entirely Contains, Entirely Contained By

etc. The system allows you also to define customized spatial operators, logically combining the primitive ones. The result is that GIS operator manage his geographic and descriptive information in a way similar to traditional database users: query products, instead of simple reports, are thematic maps, illustrating high sophisticated territorial phenomena relationships.
The following query (retrieving all buildings in a certain district and all fire stations within 5 km of that district) is only a banal example of such a powerful instrument:

FROM Buildings; CONTAINED_BY FROM District WHERE DistrictNumber=<nnn>;
UNION
(FROM Firestations; WITHIN 5 Km; FROM District WHERE DistrictNumber=<nnn>)

4.4 MapInfo SpatialWare: Topology Manager

MapInfo, one of the leading GIS vendor has recently shipped the new product SpatialWare which inherit the methodologies of a high level specialised GIS: System 9 by Unisys. The SpatialWare solution is based on a gateway to a powerful Relation DataBase Management System (Oracle): it seamlessly integrates spatial data with the core business data, adding powerful spatial analysis capabilities to IT environment. The main strong points are a maximisation of the value of data because make use of a central single location of database so that existing applications can be quickly enhanced with complete spatial analysis capabilities. It leverages the Oracle capabilities and existing application investments as well as the server capabilities for a more efficient environment.
Topology Manager (TM) forms and creates features of any topologic type from a set of primitives and a rule to apply against the primitives. Topology Manager assembles the primitives into features and write them to the database as features classes, to an external Oracle table, or to the clipboard as a data flow. One of the most important uses for Topologic Manager is for processing data from and exporting data to Data Interchange Modules (DIMs).
The functions of Topology Manager are:

- ensuring geographical consistency in the database;
- building topology fabric from diverse input sources;
- uploading/downloading any type of primitives to/from multiple databases or clipboards;
- assembling geometries with different topologies to form a single topology type based on a given aggregation type;
- creating data transfer tables which may be used by applications that transfer geometries via the clipboard as data flows or via SpatialWare as class tables or external spatial data tables;
- preserving data and referential integrity of existing output features in a multi-user access environment.

The storing and retrieving of data is according to ISO SQL/MM standard, designed to handle multimedia data types, including text, image, video, audio, and spatial information.
The peculiarity of the solution is that to allow spatial information to be managed and queried inside a single database, Oracle; SpatialWare will also be available for the use in Informix and Sybase environments. The client is basically MapInfo Professional, but other data browser will be certificated in the near future (ArcView is one of them).

5 GEOMETRIC DATA ISSUES FOR CADASTRAL APPLICATIONS

Narrowing the field of investigation to those cadasters which are logically related to immovable properties, the geometry can range from a simple fabric of parcels to a spatially complex layering of thematic information. Since the meaning of Cadastre is “ an official register showing details of ownership, boundaries” it comes without saying that a Cadastre should be maintained by the an official agency, must have a sufficient detail to depict the spatial extension of a property in general by defining the relative or absolute position of property’s boundaries. Although there exist cadasters of punctual entities (i.e. cadastre of springs/wells) the common factor of cadastral geometric structures is the polygonal organisation of entities; sorted by themes it is therefore quite common that said polygons are adjacent one to another enforcing the criteria that a same piece of land cannot be covered by two different ownership thus two parcels cannot overlay. In general a polygon should have a code of belonging to one theme or coverage, and a set of attributes which broadly define and describe the entity; in any theme it cannot be possible that the same entity is dually identified at the same time and in the same position and this condition made it logically controlled.
It then can be said that the reality can be schematised following the flow: i) conceptual formalism supplies “rules” for modelling data; and, ii) formal description techniques supply “grammar” for representation. The reality is therefore embedded into a schema where geographic data are stored according to defined criteria; the application

that make use of these data return a set of services (enquiry, drawings, reports) and a set of information about data, that in IT terminology is called “metadata”, for the reusability of data.

The main aspects that are basic to a geoinformation cadastral model are:

- Spatial aspect: this class can be specialised into two sub-classes which in turn reflect the “grammar” for representation and “rules” for its modelling. The first is Geometric Primitive (position and interpolation method) and the second is Structured Primitive (based on Adjacency, Shareness, Connectivity).
- Quality aspect: based on Accuracy, Completeness, Up-dateness
- Semantic aspect: Identifier, Attributes, Relations
- Temporal aspect: entering date, up-dating date, deleting date

To the Semantic aspect it can be referred the Land Registration System which mutually links throughout the Parcel Identifier Key.

The Quality together with the Temporal aspect give information about the reusability of data; one accessory but sometimes very helpful aspect is also the so called lineage where the source and capturing criteria of data is seldom stored.

Enquiring the system follows in general two key-points: the relation with the identifier of the property or/and a spatial location.

As many themes are overlaid more and more spatial relations are growing in importance; in a “seamless” geodata with themes at different scale, a research based on an adjacency matrix is faster than any other algorithm to “select” features which position is, for instance, along a river or a road.

In treating geometric data, topology has different stages of applications:

- data are analysed making use of no topological methods but with spatial operators;
- data are structured into topology and in this way are stored into the system;
- data are controlled through topology but are stored in a relational database together with their attributes.

Some database producers have developed their own Spatial Database Operator which by the fact have no relation with topology and geometric manipulation but try to identify each point not making use of a couple of columns but its position in a “cellar” system.

6 CONCLUSIONS

Trying to achieve some conclusions about the opportunity to utilize topological organization for cadastral applications is not an easy task; in absolute there is no concrete boundary between a non topology and a full topology environment; topology seems to be a synonymous of geometric integrity although those GIS that don’t make use of topology are able to ensure a geometric correctness which in turn might be synonymous of integrity.

The phase of cadastral data collection (including parcel codes labels, parcels and buildings boundaries and other cadastral partitions) often consists of massive map digitalization or raster images vectorization. During these activities it is generally more effective to enter cadastral boundaries lines only once, to create nodes and “clean” lines automatically, to consider parcel codes labels as centroids and to use topological organization to build polygon coverage.

If cadastral data are only one of several geographic features and layers and the system is focused on spatial relationship between different objects, topological organization can increase efficiency in data maintenance (simplifying update and integrity check operations), archive organization (avoiding data repetition) and analysis (allowing the implementation of sophisticated spatial query languages)

On the contrary, if the aim of a cadastral GIS is the maintenance and the management of cadastral geometric data and their link with the registration system to produce maps, sketches and certificates, topology is an useless burden. Modern software packages allow efficient tools for easy and fast geometric editing (split and merge parcels) and geographic indexes grant a sufficient implementation of all spatial queries generally requested in such applications.

Bibliography

(1) *Oracle7 MultiDimensional* – Oracle Corporation 1995

(2) *Towards a 3-D GIS Database* - M. Kofler, GIM, 1997

(3) *Expression of 3-D Space* -T. Ohtsuka – GIM, 1997

(4) *GRASS 4.0 Reference Manual* - USACERL (U.S. Army Corp of Engineers Research Lab., 1991

(5) *Feature level topology for Arc/Info using the Arc/Info-Oracle interface* - J. van Smaalen, M. Molenaar – ISPRS, 1994

(6) *Towards 3D-GIS: experimenting with a vector data structure* - V. Bric, M. Pilouk, K.Templfi - ISPRS, 1994

(7) *Modelling update propagation in spatial databases* - H. Kemppainen – ISPRS, 1994
(8) *Spatial Information takes up residence in the RDBMS* - D. Thomson - Database Programming & Design,1997
(9) *Mapping the future of Geomatics* - O. Ostensen - ISOTC211, 1995
(10) *Intergraph technology for National Cadasters* - Intergraph, 1996
(11) *Digital Representations of Topographic Surfaces* - J.Carter- PE&RS, 1994
(12) *Views on Iteroperability and Spatial Data Infrastructures in Europe -GSDI and ESDI* -P.Burrough – ASITA, 1997
(13) *Towards a European Policy Framework for Geographic Information* - DGXIIIe Luxembourg, 1997