GEODEZJA I KARTOGRAFIA
t. XLIX, z. 1, s. 7–20 (2000)

Tadeusz Chrobak

Institute of Geodesy and Cartography
Academy of Mining and Metallurgy
(30-059 Kraków, Al. Mickiewicza 30)

Modelling of spatial data in a database for the needs of cartographic generalization

The article presents a spatial data model describing a geographical space which makes it possible to automatically generalize data once the scale of a map has changed. To that end the Molenaar FDS data structure, extended by the geometrical rule, has been applied. The so formed spatial data model has been checked upon the example of the road network of „Downtown Cracow", confirmation of the accepted rules for the FDS structure having been acquired.

1. *Introduction*

Spatial data modelling is the process of transition from the researched fragment of reality to its informative representation, most often in the form of the map as a virtual image or the one that has been recorded upon a permanent base.

The process refers to the following stages:
— specification of position of beings (objects, phenomena, events, etc.) within real space,
— digital record of beings as a model of the researched fragment of reality,
— structure of data model describing the researched fragment of reality,
— conversion of the model to graphic image (to the form of signatures, lines, curves, patterns and colour patches),
— stimulation of images on the presented fragment of the researched reality, meant for knowledge on geographical space to be enriched.

The stages of the modelling process comply with the approach suggested by Molenaar, i.e. apart from reality — made up by geographical objects and their features — they also specify a data model, data structure of data collection, the possibility of generalization having been taken into account as well. For each of the foregoing categories selection will be carried out, in accord with the conditions specified in the data model. The model is based upon defined sets of geographical objects and their mutual relationships, this is why it characterizes reality with no deformations resulting from the standard modelling process.

This is a *digital model of the researched fragment of reality*. In order to pass on the information it contains, the model should be processed to a *digital cartographic model*. The latter is linked with a series of instructions for the plotter or printer, specifying the manner of the formation of point, linear and surface symbols.

Data specified as "geometrical" comprise information on the position of objects. These might be: geographical or geodesic coordinates, reference record, code numers referring to statistical units, topological reference (e.g. B is located between A and C) or descriptive terms, such as addresses, streets, etc. Spatial nature of beings is most often expressed by means of conventional symbols, whose character dependes on the scale or resolution.

Both from the viewpoint of data collection and presenting procedures upon the map, it is exteremely significant whether the beings (objects, phenomena, events) of the real world are approched as scattered or continuous ones. Continuous representations refer to the phenomena whose variability is of a continuous character (e.g. precipitation, temperature, pressure). Scattered objects are limited, the coordinates of the limits thereof to be determined immediately upon the map.

From the viewpoint of visualization, specification of the character of attributive information is of primary importance. Attributes may refer to visible objects (e.g. buildings), or invisible qualities (e.g. atmospheric pressure). While defining attributes of objects, we classify them according to their qualitative or quantitative character, upon an appropriate level of the assumed measuring scales (nominal-qualitative scale, ordinary-quantitative scale, interval-quantitative scale, quotient-quantitative scale). Classification of objects alongside the topology thereof, these are the necessary conditions for the generalization process to be carried out; the spatial data model should be open to them.

All spatial data change in time; this pertains not merely to information on attributes of objects (e.g. the pollution degree of an area), but position of objects as well (e.g. continental drift). Time may by considered, alongside geometry and attributes, as the third important component while investigating into variability of processes in time.

On the map for example the following symbols are meant for presentation of objects and their interrelationships: dots, lines, patches of different shapes, sizes and brighntess (valour), colours, granularity (patterns) and orientation. Visual variables applied on the map refer to real objects and relationships, the user may specify relationships between point, linear surface and volumetric objects in many (one, two, three and four) dimensions [1].

Specification of the data structure is the next step of the modelling process. This is connected with deciding whether data are represented in screen, vector or hybrid (i.e. screen/vector) structure. Within a vector structure data are organized according to the character of objects, i.e. the geometrical features of the data are determined by a set of coordinates which on the map are linked by lines (vectors) and are connected by means of attribute links. Within a screen structure data are organized on the basis of spatial addresses, geometrical features are determined by position of the cells of the network, cells' addresses are combined with attributes. Further on the vector approach will be discussed, for such is Molenaar's *formal data structure* (FDS).

Geographical object is the elementary unit of vector data structure. There are a few kinds of vector data structure. Two of them have been shown in Fig. 1.

	code	coordinates	attribute
point	116	(x, y)	city
line	34	$(x, y), (x, y), \dots (x, y)$	limit
line	37	$(x, y), (x, y)$	limit
line	36	$(x, y), (x, y)$	limit

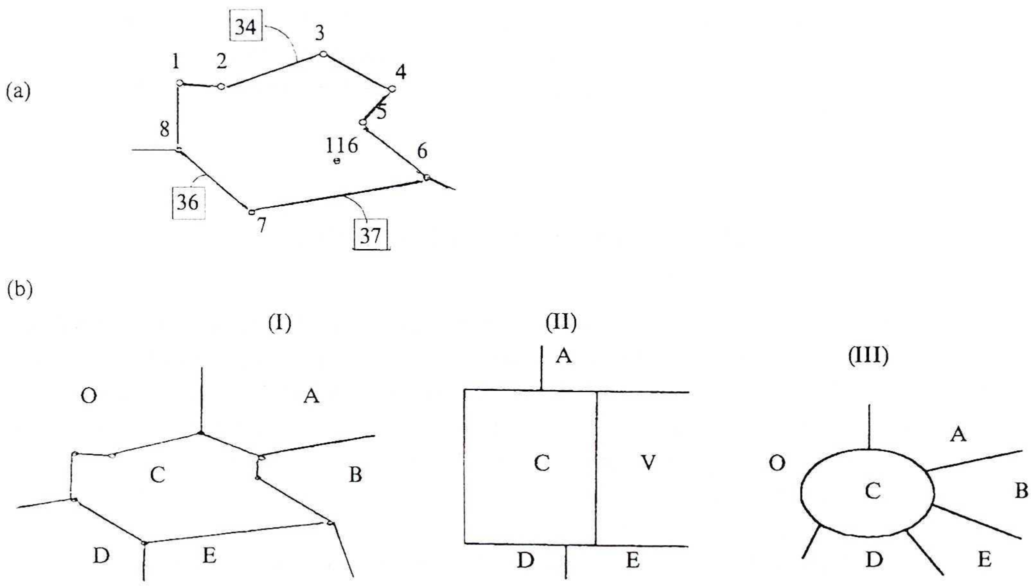


Fig. 1. Vector data structure of (a) spaghetti type, (b) topological type

The simple type of vector structure is called spaghetti (Fig. 1a). All objects in it are defined as singular items, due to which the topology thereof is not determined. Whereas Fig. 2 presents a topology which makes it possible to specify relationships between spatial objects. Graphically it may be presented by means of different images of the same area — Fig. 1b I. 1b II, 1b III.

2. *Data modelling*

Within a spatial databse the generalization process should be divided into notional and graphic ones. In the former (notional) one, it is manipulation (transformation) of the data model's recording structure that is being done on it together with data being prepared for visualization. In the latter-graphic generalization is being accomplished.

In oreder to present geographical data in the context of notional generalization, hierarchy of classes and classification of objects on the basis of the attributes thereof to be created for data, will be applied. Hierarchy, with the use of FDS for single-valued vector maps developed by Molenaar [2], combines different aspects of data models objects-oriented and spatially oriented. In this model (FDS) point, linear and surface objects are presented together with their geometrical and thematic aspects. On the basis of geometry and topology of objects and the thematic description theoreof (making up an arranged set of objects), it is possible to accomplish automatic cartographic generalization of geographical data according to different map scales. Hance structures of the data model are of primary importance while defining principles (functions, operations) for generalization of spatial information for different levels of generalization.

2.1. The formal data structure — FDS

The formal data structure (FDS) for single-valued vector maps, developed by Molenaar [2] is a topological data model which meets the requirements of computer generalization. It is capable of managing the geometrical and thematic aspects of geoinformation within an object — or sematically-oriented data model. FDS will be applied at objects of elementary types of geometrical data such as points, lines, areas, and mutually related thematic (attributive) sets.

FDS satisfies the following conditions:

a) classes of objects must be mutually exclusive, i.e. each object has just one class (of attributes),
b) a class of objects comprises geometrical data of a single type,
c) approaching a map as a graph, all points meant for description of geometry will be treated as knots of univocal localization within the accepted system of reference,
d) edges of the graph will be represented geometrically as sections of straight lines,
e) for each pair of knots there is at the most one edge that connects them; besides, knots may be combined into one chain or more,
f) edge of the graph has an arc on its right and left sied,
g) for each arc $l_{p,q}=\varphi[a_p, a_q]$, with $a_p \neq a_q$,
h) for each geometrical data type there is just one event of connection between objects; e.g. an edge may be at the most one linear objects and have one area on its left side and one area on its right side.

2.1.1. Topological relationships

By means FDS it is possible to identify a few topological relationships owing to which the model is a useful analyzing instrument. For exapmle, topological relationships occurring in FDS control such properties as: interior of the polygon, crossings, links, beginning, end of a edge, etc. These might be applied for different operations which are fundamentally prerequisite to generalizations. Object shift algorithms might be simplified by such relationships as an area having been limited by a line, particulary on detailed topgraphic maps of urban development areas. Object linking algorithms might be also thus simplified, e.g. an area touching another one. The relationships in question have been presented in Fig. 2. Last but not least — just like most conceptual models, FDS may by presented upon object-, net- or logically-oriented models.

Fig. 2 presents a formal data structure thanks to which it is possible to generate logical, topological relationships between objects and classes.

1. **Classes** (of points, lines, surfaces), hierarchy resulting from rooted planar graph having been preserved*.
2. Objects of point, linear and surface type.
3. **Knots, arcs, edges** applied in terminology of the planar graph.
4. Relationships between knots, edges and arcs:
   a) an edge has an arc on its **right** side,
   b) an edge has an arc on its **left** side,
   c) an arc has its **beginning**,
   d) an arc has its **end**,
   e) a chain of arcs makes up a **polygon**,
   f) lines **cross** in a knot,
   g) lines intersect: **upper** line; **lower** line.
5. Logical data structures make it possible to generate:
   a) links between classes of the **IS** type (downwards the tree of the graph),
   b) linking hierarchy of components of the **PART** type (upwards the tree of the graph).

2.1.2. Semantic dependences within FDS structure

Within computer environment digital structurization of hierarchy in semantic networks by means of the model concept is connected by relationships of the **IS** and **PART** type. E.g. classes are connected by relationships **subclass — superclass,** in which a class has one direct superclass at the most. The concept holds true in the simplest taxonomic hierarchies, in which each class that has one direct superclass at the most, is an acyclic rooted graph (no cycles, hierarchy of objects preserved).

* Graph G, which has its geometrical graph G′, capable of being presented upon a plane without intersections of edge lines, is called a planar graph, whereas geometrical graph *G′* corresponding to it upon the plane is called flat graph [4].

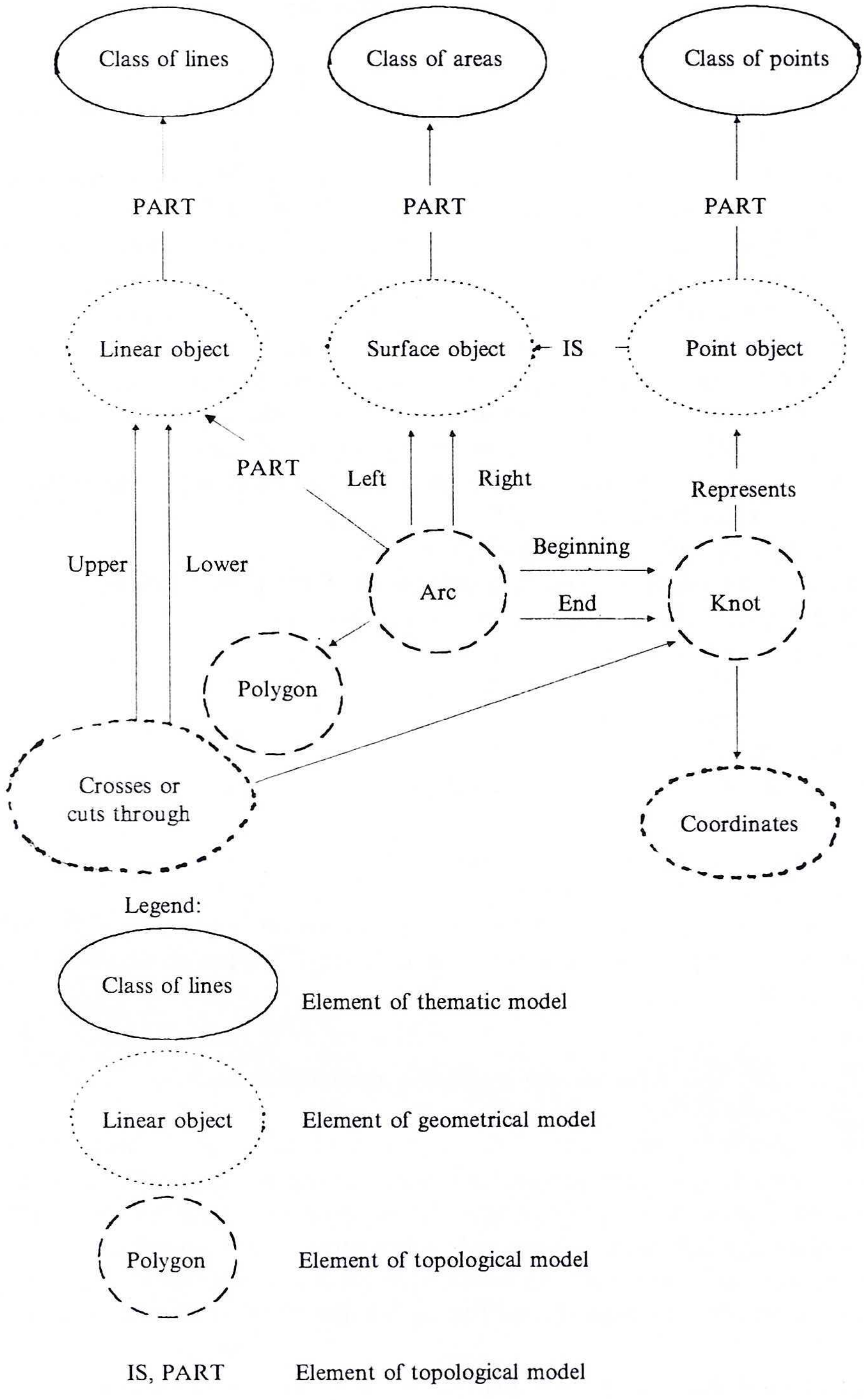


Fig. 2. FDS for single-valued vector maps within a twodimensional space

Linking hierarchy differs from classifying hierarchy by the mere fact that it refers to the omitting in which relationship between objects is an object of a higher level. Linking hierarchy enables omitting connected with complex objects, built from elementary objects of a higher level.

In classifying hierarchies inheriting the structures of object attributes comes about upwards the tree, which enables a more detailed and specialized thematic description of objects within a single course down to lower hierarchic branches. Linking hierarchy of structures of object attributes has a down-the-tree character; starting from the elementary object level, objects of a higher complexity are built. Complex objects inherit values of attributes after the parts that they are composed of.

Connections between classes are usually called links of the **IS** type and may stand for the fact that a certain type of objects is a generalization of a different kind, e.g. motorway **IS** a road network, Carcow **IS** a city, it **IS** a populated area. Links of the **IS** type are components of classifying hierarchy.

Connections of the **PART** type are components of linking hierarchy, e.g. the Rudawa River is a **PART** of the Vistula's system, it is a **PART** of hydrographic network. Connections of the **PART** type combine a certain set of objects into a complex object which in turn generates other complex objects, etc.

There is no standard computer terminology for the components of classifying and linking hierarchies. Herein teminology applied by Richardson [4] has been accepted.

Fig. 3 presents links between object identifier, geometrical data and thematic (attributive) data on a considerably general level.

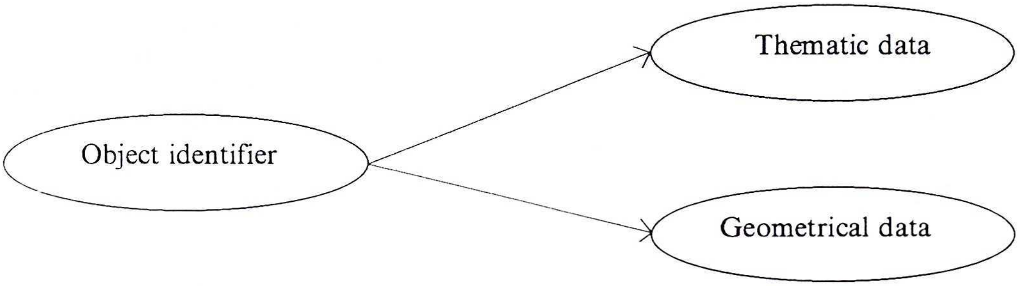


Fig. 3. Elementary FDS

In order to make FDS (Fig. 3) meet the requirements of cartographical generalization, data must satisfy the following rules:

— Molenaar's thematic rule [2]:

$$\begin{gathered}\forall_{C \in SC} I(C_i) > I(C_i + 1), \\ \forall_{A \in C} A_p(O_1) > A_p(O_1 + 1)\end{gathered} \tag{1}$$

where:

I — is a set of attributes characteristic of class C,

C— is a class, due to which for optional class C_i belonging to superclass SC, class C_i in class hierarchy is superior to class (C_i+1) etc.,

A— is a set of attributes belonging to class C, subset of attributes A_p defining object O_1 which hierarchically is superior to object O_1+1.

— Chrobak's geometrical rule [5]: if attributes of linear objects take into account class hierarchy and object classification, objects having been defined as a planar graph, the choice and elimination of linear object L_j for a 1:M scale map are determined by the following conditions:

1) drawing presentation of L_j upon a map,

2) threshold measure of recognizability of the enviroment of L_j.

On a map, scale 1 : M there is drawing L_j, the length of its chord satisfying the following inequality:

$$L_j \in \mathrm{L}, \text{ if: } L_j\ (a_{j1},\ a_{j2}) \geqslant 0.6\,\mathrm{M},\ [\mathrm{mm}] \tag{2}$$

where:

L — set of linear objects of a superclass within cartographical space for scale 1 : M,

L_j — length of edge (chord — broken line) j, $j=1, 2, 3, \ldots n$,

a_{j1}, a_{j2} — opening and closing knots of the edge of line L_j,

M — denominator of the scale of the map in question.

Thershold measure of recognizability of the environment of L_j (thickness of line — 0.1 mm) will be determined by two triangles whose dimensions are closest to the elementary triangle. Vertices — knots represent the objects of those two triangles which belong to set — L of the cartographical space. Object L_j connects triangles (Fig. 4) in which the two remaining vertices are knots of the already defined linear objects — on the basis of object classification and class hierarchy, and their minimum dimensions — in the space of the 1 : M scale map.

In the first tringle (Fig. 4), its vertices having been marked as: a_{gj1}, a_{gk2}, a_{gi1}, vertex a_{gj1} is the researched opening knot of object L_j. In the triangle the other two knots represent objects L_j and L_k connected with L_g. The connected objects (making up a side of the triangle) already belong to the cartographical space of the map in question, for in object classification and class hierarchy they are superior to object L_j.

In the second triangle (Fig. 4), its vertices having been marked as: a_{mj2}, a_{m1}, a_{m2}, vertex a_{mj1} is the researched closing knot of object L_j. In the triangle the remaining vertices meet the requirements of object hierarchy just like in the first triangle. The sequence of the choice of the triangles and knots does not affect the legibility of L_j's environment.

The markings of the edges of the second triangle differ from those in the first one. This is a purposeful presentation of an event (second triangle, Fig. 4) in which two control knots are the beginning and end of a object that ranks higher in classification. This is pointed out by notation of edge index L_{m0j}, Fig. 4 in which

m — object ranking highest in the researched triangle,
0 — (second item of index) — opening knot of an object that already exists on a map (third position of index 0 stands for closing knot, e.g. L_{mj0}, Fig. 4),
j — researched object.

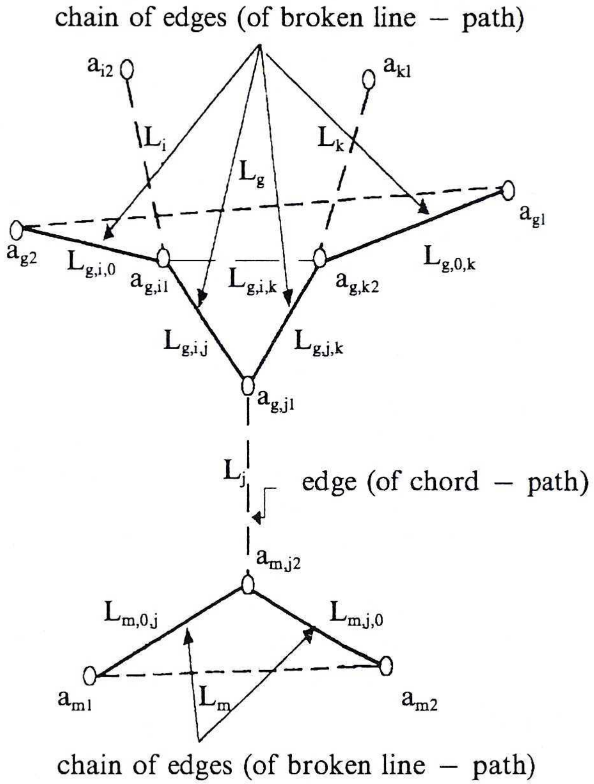


Fig. 4. Analysis of the threshold recognizability measure of environment of linear object L_j

– – – edges (of paths): L_g, L_i, L_j, L_k, L_m; ——— edges of path chain L_g: $L_{g,0,k}$, $L_{g,i,0}$ — opening and closing, $L_{g,j,k}$, $L_{g,i,j}$ — intermediate ones; ——— $L_{g,i,k}$ — edge in researched triangle ("closing"); 0 — knots, e.g. a_{g1} — knot of object, L_g — opening knot, $a_{g,i1}$ — opening knot of object L_i connected with object L_g

The threshold measure of regonizability of the environment of object L_j (drawing line 0.1 mm thick) — within cartographic space, scale 1 : M — has been presserved in the lengths of the side (edges) of the two triangles (their minimum dimensions possibly closest to the elementary one) connected with researched object L_j (Fig. 4) satisfy the following inequality:

$$\Delta(L_{g,i,j}, L_{g,i,k}, L_{g,j,k}) \geqslant 0.6 \text{ M [mm]}, \; i \neq j \neq k, \tag{3}$$

where:

$L_{g,j,k}$ — length of edge in triangle for three different objects (L_i, L_j, L_k) whose knots, beginning and end of an edge, belong to object L_g. Each of the triangles is determined by: two vertices — knots (representing two objects already occurring on the map) and the third one — researched knot of the object — L_j,

or

$$\Delta(L_m, L_{m,0,j}, L_{m,j,0}) \geqslant 0.6\,\mathrm{M}\ [\mathrm{mm}],\ m \neq j \tag{3a}$$

where:

L_m — length of edge (chord of broken line),

$L_{m,0,j}$ — length of edge between opening knot (second item of index — 0) of object L_m and knot of object L_j,

$L_{m,j,0}$ — length of edge between closing knot (third item of index — 0) of object L_m and the researched knot of object L_j.

The foregoing rules — geometrical and thematic (attributive) — specify requirements of a data in order model in order to make it possible to automatically present linear objects upon a map at an optional scale.

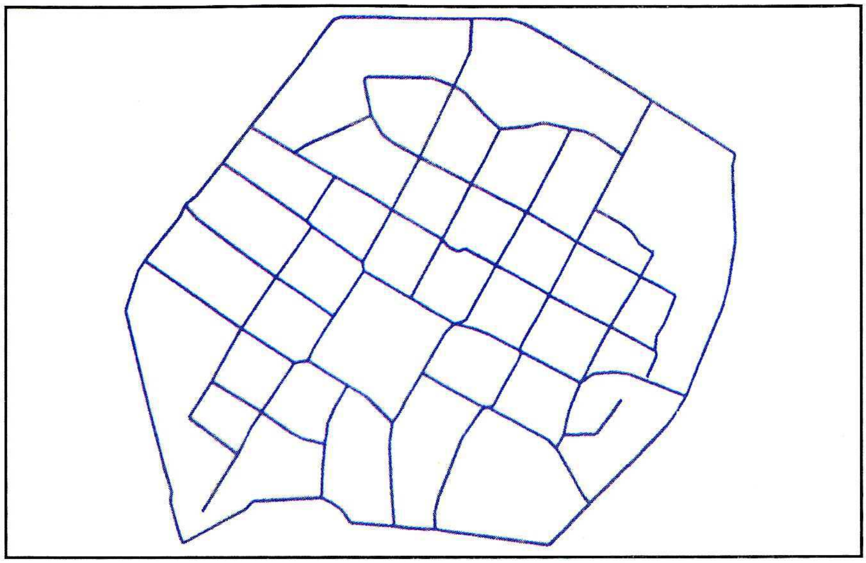
The foregoing minimum dimensions of drawing pertain to line thickness 0.1 mm. If linear dimension is changed to surface one (thickness of line having been changed) up to a length of 0.6 mm., a fixed thickness will be added, for 0.6 mm is minimum measure of the length of an edge of a triangle which guarantees for its "recognizability".

The rule does not apply at objects belonging to a superclass (closest to the root of the tree — graph), they are chosen by the map's editor. Molenaar defined that choce-and-elimination process on the basis of defined objects of a superclass, as structural generalization [6].

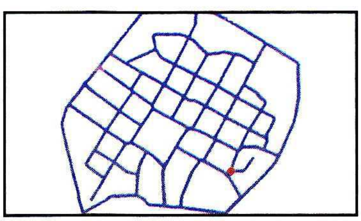
3. *Practical verification of the spatial data model*

The presented rules — attributive and geometrical — of the FDS model have been applied at the process of object elimination, at an optionally changing scale (from larger down to a smaller one) for the esteblished Cracow Downtown region (Fig. 5). Voivodeship roads which are part of the limits of the region have been specified on the basis of Polish standards. They have been classified in the folowing way:

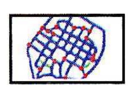
1. **Basztowa St.** opening knot *1* closing knot *3*,
2. **Straszewskiego St.** opening knot *5a* closing knot *5*,
3. **Westerplatte St.** opening knot *3* closing knot *4a*,
4. **Dunajewskiego St.** opening knot *6* closing knot *1*,
5. **Podwale St.** opening knot *5* closing knot *6*,
6. **Area limit** opening pseudo-knot *4a* closing — *5a*.



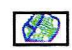
Scale 1:10 000 thickness 1.0 mm



Scale 1:25 000 thickness 1.0 mm

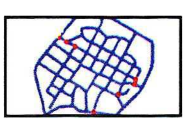


Scale 1:75 000 thickness 1.0 mm

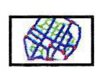


Scale 1:150 000 thickness 1.0 mm

Scale 1:200 000 thickness 1.0 mm



Scale 1:50 000 thickness 1.0 mm



Scale 1:100 000 thickness 1.0 mm

Scale 1:300 000 thickness 1.0 mm

Fig. 6. "Cracow Downtown" after generalization, if scale and width of roads change
Colour: red – vertex to be shifted
blue – road after generalization
green – road to be eliminated

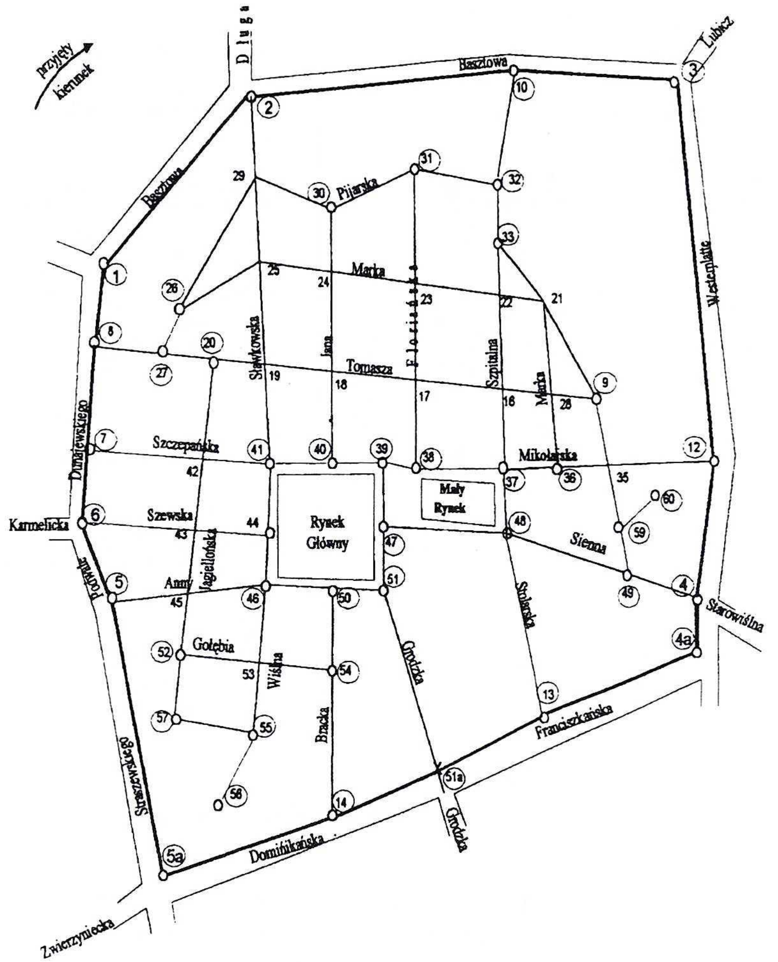


Fig. 5. Street networks of "Cracow Downtown"

③ – knot number of road of a class superior to local one, ⑰ – number of opening knot or closing knot of local road, 18 – number of intersection of local roads, 4a, 5a – numbers of pseudo-knots, ○ – opening or closing knot of a road, × – pseudo-knot of an object belonging to two regions

Construction of regions*, generation — by means of topology — of complex objects from elementary ones, as well as class and object classifications have been carried out in accord with [5]. The road network in Cracow Downtown after generalization, if the scale of the map in question or widths of roads change, has been

* Region is a part of a plane limited by point knots and edge lines of a geometrical graph upon a plane which make up a cycle together with those points and knots, yet without all other knot points or edge lines [4].

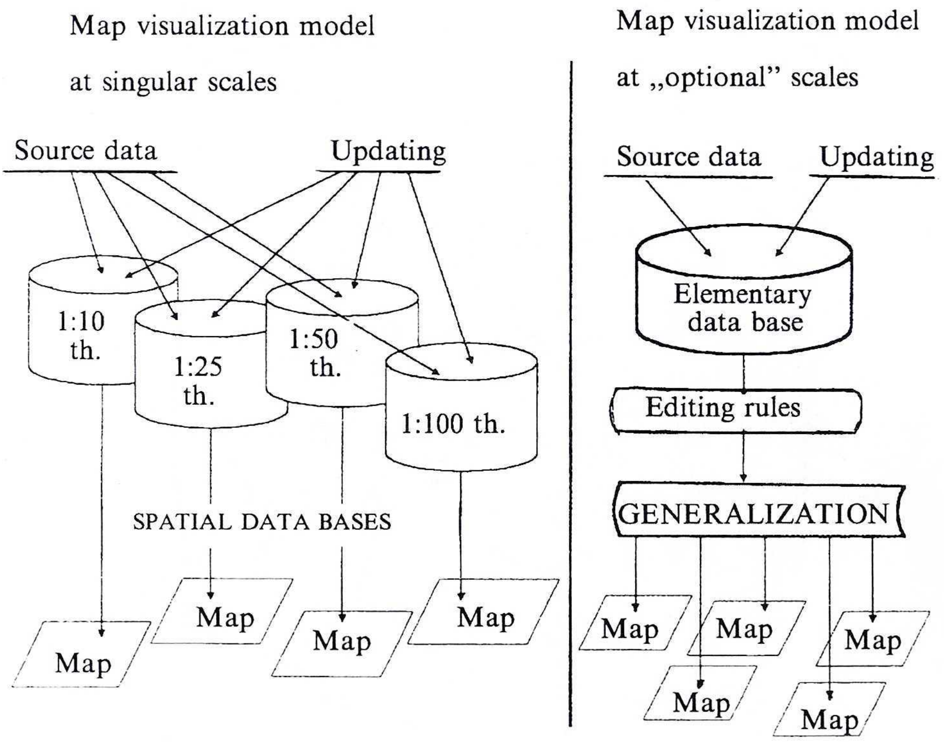


Fig. 7. Database with regard to generalization process dependent on scale [7]

presented in Fig. 6. The obtainded results of automatic elimination of linear objects with the use of FDS make is possible to formulate the following conclusion: *generalized data bases may be automatically generated from a detailed database (Fig. 7) which preserves topology, class hierarchy and object hierarchy.*

REFERENCES

[1] Kraak M.-J., Ormeling F. *Kartografia, wizualizacja danych w przestrzeni*, PWN, Warszawa 1998

[2] Molenaar M., *Single valued vector maps — a concept in GIS.* Geo-Informations Systems, 2 (1), s. 18–26, (1989).

[3] Richardson D. E. *Automatic spatial and thematic generalization using a context transformation model.* Doctoral dissertation, Wageningen Agricultura University. R&B Publications, Ottawa, Canada, 1993.

[4] Kulikowski J. L., *Zarys teorii grafów*, Biblioteka Naukowa Inżyniera, Warszawa 1986.

[5] Chrobak T., *Badania przydatności trójkąta elementarnego w komputerowej generalizacji kartograficznej*, AGH, Uczelniane Wydawnictwo Naukowo-Dydaktyczne, Kraków 1999.

[6] Molenaar M., *The role of topologic and hierarchical spatial object models in database generalization*, Netherlands Geodethic Commission, Nr 43, Delft, s. 13–35, (1996).

[7] Lee D., *Generalization Cartography Intergraph Corporation*, Huntsville 1992.
[8] Molenaar M., *Object hierarchies and uncertainty in GIS or why is standardization so difficult.* Geo-Informations Systems, 6(3), s. 34–48, (1993).
[9] Pachelski W., *Aspekty metodyczne projektowania baz danych przestrzennych. Modelowanie danych przestrzennych.* Materiały z seminarium naukowego, IGiK, Warszawa, (1997).

Tadeusz Chrobak

Modelowanie danych przestrzennych w bazie danych dla potrzeb generalizacji kartograficznej

S t r e s z c z e n i e

W pracy przedstawiono model danych przestrzennych do opisu przestrzeni geograficznej pozwalający na uogólnienie danych w sposób automatyczny, gdy zmienia się skala mapy. W tym celu wykorzystano strukturę danych FDS Molenaara rozszerzoną o regułę geometryczną. Tak stworzony model danych przestrzennych sprawdzono na przykładzie sieci dróg „Centrum Krakowa" uzyskując potwierdzenie przyjętych reguł do struktury FDS.

Тадеуш Хробак

Моделирование пространственных данных в базе данных для потребностей картографической генерализации

Р е з ю м е

В работе представлена модель пространственных данных для описания географического пространства, дающая возможность обобщения данных автоматическим способом, когда изменяется масштаб карты. Для того была использована структура данных FDS Моленара, расширена о геометрическое правило. Образованная таким способом модель пространственных данных была проверена на примере сети дорог „Центр Кракова", получая подтверждение принятых правил для структуры FDS.