GEODESY AND CARTOGRAPHY
ISSN 2029-6991 print / ISSN 2029-7009 online
2011 Volume 37(4): 143–148
http://dx.doi.org/10.3846/13921541.2011.645310

UDK 528.91

POSSIBILITIES OF UPDATING SMALL-SCALE BASIC SPATIAL DATA IN LITHUANIA USING GENERALIZATION METHODS

Lina Papšienė[1], Kęstutis Papšys[2]

[1]Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania
[2]Vilnius University, M. K. Čiurlionio g. 21, LT-03101 Vilnius, Lithuania
E-mails: [1]gkk@vgtu.lt (corresponding author); [2]kestutis.papsys@gf.stud.vu.lt

Received 12 September 2011; accepted 21 November 2011

Abstract. Small-scale spatial data are widely used at regional and national levels not only for mapping but also for the purposes of planning, forecasting, etc. Therefore, a professional preparation of such data is necessary. The generalization of large or medium scale spatial data is the most efficient process to produce and update smaller-scale data. Certainly, a simple transfer of information is almost never suitable to satisfy requirements for small-scale maps. Additional transformations (generalization) are necessary. Spatial information complexity may be significantly reduced in terms of the number of objects, geometry, etc. However, the main spatial, non-spatial and topological characteristics of the objects have to be preserved. The process of reduction is irreversible, and therefore it is necessary at first to clearly define requirements for spatial data (for example, the density of spatial objects, the minimal allowed area, the width and length of an object, a minimum length of the edge of an object, spatial links between the objects). The above imposed requirements provide a possibility of defining procedures for generalization and a conceptual model between particular data sets.

Keywords: generalization, basic (reference) spatial data, density of spatial data, combination of spatial data, simplification of spatial data, small-scale, GIS.

1. Introduction

Presently, spatial data has not only used for preparing traditional or digital maps; along with the development of geographic information systems (GIS), it is widely applied for various spatial tasks related to planning or prognostication (Başaraner 2002). Therefore, there is a stronger need for the most essential varying constantly updated spatial data at different scales. In order to decrease the costs of works, the creators of spatial data have to progress to the automated development of such information. Müller (1991) considers that generalization is promoted by economic requirements. Thus, it would be enough to invest in the creation of –large or middle-scale qualitative data and later use the efficient automated principles of generalization for the purpose of creating smaller-scale spatial data.

The generalization of spatial objects means the selection of the main, essential and typical for a specific location objects, including their qualitative and quantitative characteristics. The complexity of provided information is decreased during this process, however, the most important characteristics of the object are retained and unimportant features are skipped (Urbanas 2001).

Generalization is an irreversible process, and therefore it has to be thoroughly planned so that the result obtained after performing it should satisfy the set requirements. At the initial stage of designing the generalization process, it is necessary to describe a model that would allow defining the exact stages of generalization as well as principles, algorithms and verification rules for spatial data in relation with the quality requirements raised for spatial data. However, in order to obtain a real model, the analysis of requirements for spatial data, the methods of generalization and results reached by using them are necessary.

This article analyzes the principles of generalization based on research conducted by Müller (1991), Peng (2000), Yaolin, Molenaar, Tinghua, Yanfang (2001), Başaraner (2002), Cecconi (2003), Foerster, Stoter, Kobben (2007) etc.

2. Updating Basic (Reference) Spatial Data in Lithuania

Basic spatial data in Lithuania are collected with reference to tree main scale levels – 1:10000 (basic scale), 1:50 000 and 1:250 000. High quality basic-scale spatial data are developed and updated using the latest digital orthographic maps, field measurements, etc. Accordingly, basic-scale spatial data are perfect for automatic or semi-automatic preparation of smaller-scale spatial data. However, when updating basic spatial data at the scales of 1:50 000 and 1:250 000, initial automatic generalization, manual generalization and vectorisation are used.

Full automatic generalization of spatial data in Lithuania is not popular because the obtained result does not always satisfy expectations. However, it is related to:

- a limited number of generalization tools in used software;
- a limited understanding of algorithms for generalization;
- no time is invested into the analysis or development of other algorithms.

On the other hand, there are no strict requirements and united opinion of what results of spatial data must be obtained at different map scales. Only general requirements for objects, such as their types or attributes are specified. However, the accuracy of representing spatial data or the thickness of objects is not set. Due to these reasons, the development of spatial data is based on the examples of spatial data mostly based on the principles of classical cartography.

3. The Analysis of Spatial Data

While analyzing spatial data, the requirements set for different scales and affecting its qualitative and quantitative parameters related to accuracy and representation should be examined.

The scale mostly defines spatial data resolution defined as the "smallest object or feature included or discernible in data" (Goodchild 1991): Geometric resolution indicates the geometric abstraction of spatial data. Peng (2000) distinguished four aspects: a) geometry type, b) minimum object size (for example, a minimum area or minimum length of an object), c) a minimum distance between two neighbouring objects and d) minimum object granularity (for example, a minimum length of the edge of an object). Thus, an important point is to describe the main requirements for geometry and the representation of spatial data. The indicated permissible values are essential for planning the processes of generalization in order to choose suitable methods of generalization, algorithms and parameters.

Usually requirements for a minimum size of an object, a minimum distance between two neighbouring objects or minimum object granularity are not strictly defined in basic spatial data in Lithuania. They are usually related to the personal ability of distinguishing approximately 0.05–0.01 mm objects or their changes in a digital map. However, spatial data are sometimes applied to the requirements of higher accuracy. For example, a minimal distance between the vertexes of the edge is 50 m and the minimum accepted area is 0.6 sq. km for basic spatial data at a scale of 1:250000.

Further analysis of basic spatial data shows it is obligatory to evaluate its representation at different scales described by:

- the represented phenomenon of the real world (geographic object) (for example, lake, road);
- qualitative parameters for the represented object (for example, the lakes bigger than 4 ha, highways and state roads);
- the types of geometric objects used for representation (for example, area, line).

Basic spatial objects at different scales usually reflect the same phenomena of the real world and are represented at all scales. The exception of the basic spatial data in Lithuania could be the build-up territories represented at the scales of 1:10 000 and 1:50 000 that become the objects of populated localities (cities, towns and villages) at a scale of 1:250000. Therefore, it is very important to evaluate the qualitative parameters applied for different-scale spatial objects. Thus, on the basis of attributive information about spatial objects, we will be able to select data that is only essential for that particular scale according to its specific features, importance or size.

Having analyzed the types of the geometric objects used for representing different-scale basic objects (Table 1), three cases can be distinguished:

- the same type of a geometric object is always applied for spatial data at all scales (for example, roads are always represented by lines);
- the same types of geometric objects are always applied for spatial data at all scales chosen according to the qualitative characteristics of the represented object (for example, though all waterways are represented by lines, they are also represented by areas depending on their width);
- different types of geometric objects are applied for spatial data at different scales (for example, buildings are represented by polygons at a large scale and by points at a small scale).

Table 1. Geometric representation of basic spatial data in Lithuania depending on the scale

Geographic Objects	1:10 000	1:50 000	1:250 000
Roads	Line	Line	Line
Railways	Line	Line	Line
Watercourses	Line Polygon	Line Polygon	Line Polygon
Lakes, pounds	Polygon	Polygon	Polygon
Buildings	Polygon	Points	Points
Built-up areas	Polygon	Polygon	Polygon Points
Vegetation areas	Polygon	Polygon	Polygon
Geodetic points	Points	Points	Not Applicable

4. Generic Principles of Generalization

Follow the various classifications of generalization operators (Yaolin *et al.* 2001; Cecconi 2003; Foerster *et al.* 2007) four basic types of generalization can be identified: a) decrease in density, b) simplification, c) combination and d) smooth.

Before choosing the methods and principles of generalization, it is necessary to choose the degree of generalization depending on the scale and the type or territory of a spatial object. This parameter defines the amount of provided information compared with the amount of primary spatial data. The easiest way of describing the degree of generalization is to apply the parameters that can be divided into two logical subtypes: quantitative and qualitative.

Quantitative parameters determine the quantitative characteristics of generalization, such as spatial distribution density. Traditionally , it is measured as the number of objects per unit. For example, the amount of rivers per unit of a territory (area) is the density of rivers. This parameter can be easily expressed mathematically. The calculation of density depends on the geometry of the object type (Zhang *et al.* 2008):

$$SD_D = \frac{n_P}{A_{(point)}}, \quad (1)$$

$$SD_D = \frac{l_L}{A_{(line)}}, \quad (2)$$

$$SD_D = \frac{A_P}{A_{(polygon)}}, \quad (3)$$

where SD_D – the density of spatial data; n_P – the amount of point objects per defined territory; l_L – the sum of the lengths of linear objects per defined territory; A_P – the sum of the areas of area objects per defined territory; A – area of territory. Thus, it is possible to determine SD_T of the objects of each type for every scale, every type of spatial data and every type of the represented territory (in case the amount of geographic objects is significantly different in different territories). SD_D is measured by units (the amount and sum of lengths or areas) in the territories of the set size (for example, sq.km). In order to have a proper amount of generalized objects, it is necessary to set the thickness of spatial data for every scale, every type of an object and specific territories in advance (Table 2).

Table 2. An example of an average density of roads (polylines) in 25 sq. km

Scale	Territory types	SDD
1:50 000	Cities	196
1:50 000	Towns	89
1:50 000	Rural	23
1:250 000	Cities	10
1:250 000	Towns	7
1:250 000	Rural	4

SD_D must be determined considering an optimal amount of disclosed information. When using this information, common possible SD_D has to be calculated for all specific territories (for example, build-up or rural areas) and distributed to all spatial objects represented in a specific territory.

After determining a limitary density of spatial data and considering the generalization of a bigger scale, spatial data are always used for preparing smaller-scale spatial data, i.e. when decreasing SD_D (except dominant area objects that must become dominant in the generalized territory when diminishing the scale), the iterative process of selecting spatial data must be performed as long as determined SD_D is reached (Fig. 1).

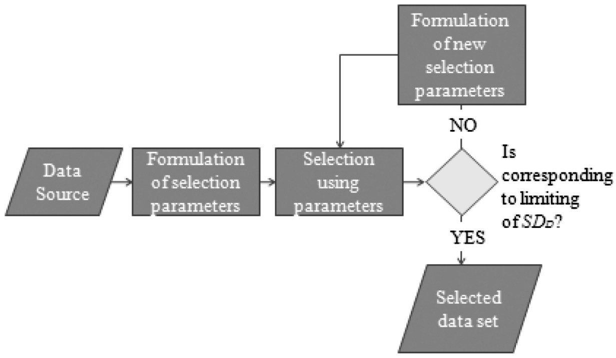

Fig. 1. A model of the selection process using the parameter of SDD

Various methods are used for reaching the limitary meanings of SD_T. Search for limitary meaning, the calculation of the sum of the lengths of objects, etc. are the methods requiring especially high powers of calculation. Therefore, information technologies and specialized GIS software must be involved in the generalization process for using the functions of automated clipping, calculating the length, area, SQL sentences of database queries and modelling the environment.

Qualitative parameters determine the qualitative characteristics of the generalization of spatial data such as the length of a river, the population of a town, the class of geodesic points, etc. In this situation, the recommended qualitative parameters for a special determination of the type of spatial data, its scale or represented territory must be clearly defined, for example, river $l > X$. When selecting data according to its qualitative parameters, the external generalization of spatial data, which also uses the method of selection, must be performed. However, in order to apply this method, it is necessary to compile qualitative parameters as attributive information in databases.

Except the method of selection, the internal method of object generalization must be applied to every type of generalization, which is object simplification. After performing selection during generalization, i.e. choosing only necessary data and excluding that not complying with the established criteria, internal generalization must be made. During this process, the shapes of objects must be simplified. The simplification of the shape is performed as the cartographic expression of representing spatial data decreases along with a decrease in the scale (Shea, McMaster 1989), i.e. as the scale decreases, fewer and fewer elements can be represented in a unit of an area of the real world. When performing simplification, the main parameter is the amount of vertexes (bends) per unit of length. It must be set for every scale, every type of spatial objects and every type of the represented territory in advance. However, it cannot be lower than the cartographic expression of representing spatial data. During simplification, both the amount of vertexes per unit of length and other parameters, such as the indication of the fixed points (points of start and finish) and the selection of the method for eliminating vertexes, must be considered. The methods of eliminating vertexes can be as follows (Fig. 2):

methods of elementary selection eliminating excess points are easy to implement and fast to perform;

logical methods deviate from the simplified curve as little as possible, are more complicated and slower, but result in the image of a spatial object which is less different from the original one.

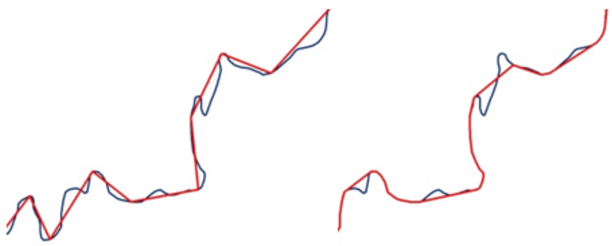

Fig. 2. An example of line simplification using elementary and logical eliminations

It is also necessary to indicate the minimal area of the object obtained after simplifying the algorithms for the simplification of area objects. The objects that do not comply with this condition are removed and extra selection is performed.

The highest degree of simplification is performed in cases of changing the type of the geometry of objects:

- polygon objects (for example, buildings) are converted into point objects and the inner centroid of the object must be calculated;
- polygon objects drawn along one axis (for example, roads or rivers) are converted into linear objects and the central line of the object must be calculated.

The conversion of objects from one type of geometry into another must be performed along with the method of selection, i.e. the objects satisfying the conditions are first selected and only then the type of their geometry is changed. For example, the selected rivers narrower than 6 m and represented as polygon objects at a large scale are converted into linear objects by which they are represented at a small scale.

When performing the generalization of spatial data, combination is applied as a separate case of the method for simplification. Thus, simplification is generally performed not only inside the objects but also between the adjacent objects that must be combined in case the distance between uniform objects becomes less than the set cartographic expression of representing spatial data in relation with a decrease in the scale. For example, the group of small marshes is combined into one big group, small sierras are combined into one big formation and separate build-up territories are combined into a solid build-up territory, etc. The objects of different geometric types are applied to different types of combination:

The combination of point objects means that the aggregate of uniform points is defined and represented as an object of one area (Fig. 3).

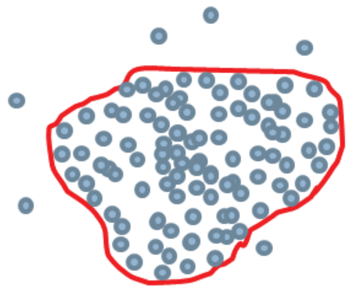

Fig. 3. An example of the point aggregation of the area object

The combination of linear objects is mostly performed by collapsing two lines represented in parallel. For example, two traffic lanes are collapsed into one street or street-limiting pavement lines are reformed into the central line of the street.

The combination of the area objects is performed by combining the area objects located in the set distance into one object. The type of the combination should be different for natural and anthropogenic objects (Fig. 4). When combining anthropogenic objects, angularity must be kept and the algorithm must intend the kind of combining the objects that include as many straight angles as possible.

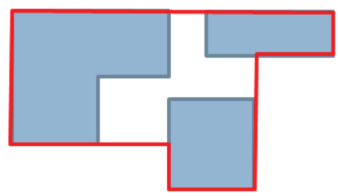

Fig. 4. An example of an anthropogenic combination of the objects

In order to obtain the result of the generalization process satisfying the defined requirements as much as possible, certain preparative works related to the generalized spatial data must be performed.

Before performing works of decreasing the thickness of spatial data, the following procedures in relation with spatial data must be adopted:

- dissolving according to unique attributes (for example, the name and area of a river) in order to properly evaluate the thickness of spatial data;
- clip according to territories where a decrease in the thickness of spatial data is performed in all determined territories.

After decreasing the thickness of spatial data and prior to performing other works of generalization, it is necessary to perform the dissolution of divided spatial data according to the main unique characteristics of the object and restoration of solidity (in case the object breaks caused by a decrease in the thickness of spatial objects in order to perform even generalization in the whole object).

After finishing the generalization of spatial data, the following additional works can be done:

- smoothing spatial objects mostly becomes angular and unlovely after using the algorithms for generalization. The main aim of smoothing is to improve visual characteristics of the objects. However, it is to note, the parameter representing a minimal distance between the vertexes of the line or boundary of the polygon can be damaged while smoothing;
- the restoration of topologic relations among spatial objects in case of having damaged them during generalization.

5. Concept of Basic Spatial Data Generalization in Lithuania

The proposed concept of spatial data generalization was prepared regarding the results of analyzing basic spatial data in Lithuania. This concept will allow designing the consistent processes of automated generalization at subsequent stages.

The generalization of basic spatial data can be divided into three main stages (Papšienė *et al.* 2011):

- modelling the processes of generalizations;
- the reorganization of basic spatial data;
- the generalization of basic spatial data.

Modelling the processes of generalization include:

- the determination of requirements for spatial data (density, geometry resolution);
- the selection of generalization methods;
- the determination of generalization parameters;
- the determination of priority in using methods.

The determination of generalization parameters must first be started from a detailed analysis of technical specifications for basic spatial data (scales 1:10 000, 1:50 000 and 1:250 000) by identifying the structure of spatial data, its representation and requirements raised to accuracy. In case the requirements for the accuracy of spatial data are not defined, the process of setting them is necessary; otherwise, no possibility of preparing exactly the model processes of generalization exists. According to these parameters:

- SD_T have to be set for each type of the object and specific territories different in their specificity;
- certain algorithms and parameters of generalization have to be chosen to help with obtaining suitable results.

In order to obtain proper results after generalization, the preparation of primary data is frequently needed. Thus, the further process of the generalization of basic spatial data must be performed in the following order (Fig. 5):

- a combination of primary spatial data having the same qualitative unique characteristics (*merge*);
- a combination of spatial data according to the requirements of accuracy (minimal distance between the objects) at a certain scale where adjacent spatial objects with the same qualitative characteristics are combined (*aggregate or collapse*);
- a subdivision of spatial data into specific territories;
- a decrease in the thickness of divided spatial data applying the cyclic process of selection and qualitative characteristics and comparing the obtained results with determined SD_D (*selection*);
- determining the solidity of selected spatial data in case spatial data breaks during a decrease in the thickness of spatial data;
- a combination of selected spatial data with the same qualitative characteristics (*merge*);
- the simplification of spatial objects according to the requirements of accuracy (for example, a minimal distance between the vertexes of the line or boundary of a polygon, minimal area or length) set for spatial data of every type at a certain scale (*elimination of vertexes*);
- smoothing simplified spatial objects in order to improve visual characteristics of an object (optional action);
- the restoration of topologic relations among spatial objects in case of having damaged them during generalization.

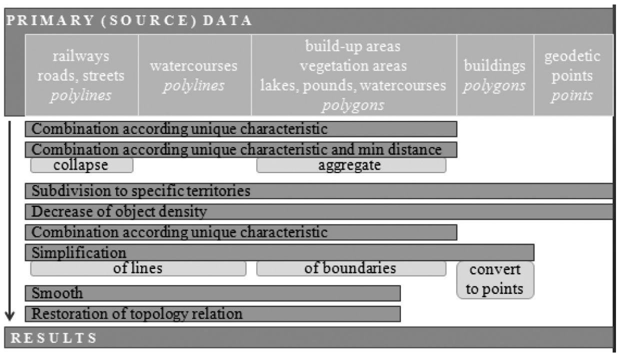

Fig. 5. A concept of basic data generalization

6. Conclusions

In order to properly decrease the thickness of spatial objects, SD_D must be both calculated for each type of an object and for a specific territory. However, it is not possible to state specific territories distinguished to one type of objects (for example, a few distinguished urban and rural territories) will also be specific to the others (for example, the specificity of an urban and rural territory does not impact the thickness of waterways). Thus, specific territories for which SD_D is determined must be separately distinguished for each type of an object.

The correctness of generalized spatial data is determined by qualitative and quantitative requirements raised to the expected results of generalization in advance and makes presumptions of choosing proper methods of generalization for obtaining suitable results and applied parameters.

Striving for efficient results of the generalized spatial data is ensured by a logical sequence of the selected methods of generalization, which consistently intends the succession and periodicity of the performed works of generalization. Therefore, by describing common sequences of generalization processes, the conceptual model of generalization gives possibilities for a further elaboration of a logical sequence of generalization processes and for the selection of certain algorithms and applied parameters.

References

Başaraner, M. 2002. Model Generalization in GIS, in *Proc. of the International Symposium on GIS*. September 23–26, 2002. Istanbul, Turkey.

Cecconi, A. 2003. *Integration of Cartographic Generalization and Multi-scale Database for Enhanced Web Mapping*: Ph.D. Dissertation. Zurich University, Switzerland. 155 p.

Foerster, T.; Stoter, J. E.; Kobben, B. 2007. Towards a formal classification of generalization operators, in *23rd International Cartographic Conference (ICC 2007).* Aug 4–10. Moscow, Russia.

Goodchild, M. 1991. Issues of Qualiy and Uncertainty, in *Advances in Cartography*. Edited by Müller, J. C. International Cartographic Association (ICA), London: Taylor & Francis, 113–139.

Yaolin, L.; Molenaar, M.; Tinghua, A.; Yanfang, L. 2001. Frameworks for generalization constraints and operations based on object-oriented data structure in database generalization, *Journal of Geo-Spatial Information Science* 4(3): 42–49. doi:10.1007/BF02826923

Müller, J. C. 1991. Generalization of Spatial Databases, in *Geographical Information Systems: Principles and Applications*. Edited by Maguire, D.; Googchild, M.; Rhind, D. Longman. London, 457–475.

Papšienė, L.; Kalantaitė, A.; Papšys, K. 2011. Conceptual model for generalization of Lithuanian spatial reference data, in *The 8th International Conference "Environmental Engineering": Selected papers*, vol. 3. Ed. by Čygas, D.; Froehner, K. D. May 19–20, 2011, Vilnius, Lithuania. Vilnius: Technika, 1402–1407.

Peng, W. 2000. Database generalization, concepts, problems, and operations, in *The International Archives of the Photogrammetry and Remote Sensing*, vol. 43, Part B4. Amsterdam, 826–833.

Shea, K. S.; McMaster, R. B. 1989. Cartographic generalization in a digital environment: when and how to generalize, in *Proc. of 9th International Symposium on Computer-Assisted Cartography*. Baltimore, USA, 56–67.

Urbanas, S. 2001. Objektiškai orientuotos duomenų bazės kartografinės generalizacijos procese, *Geodezija ir kartografija* [Geodesy and Cartography] 27(2): 68–73.

Zhang, X.; Ai, T.; Stoter, J. 2008. The evaluation of spatial distribution density in map generalization, in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, Part B2. Beijing.

Lina PAPŠIENĖ. PhD student at the Department of Geodesy and Cadastre, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania. Ph +370 5 274 4703, Fax +370 5 274 4705, e-mail: *gkk@vgtu.lt.*
A graduate from Vilnius Gediminas Technical University (MSc in measurement engineering / geodesy and cartography, 2000), actively participates in the processes of developing infrastructure and producing basic (reference) spatial data at the scale of 1:250 000 in Lithuania. Research interests: generalization and harmonization of spatial data, SDI.

Kęstutis PAPŠYS. PhD student at the Centre of Cartography, Vilnius University, M. K. Čiurlionio g. 21, LT-03101 Vilnius, Lithuania. Ph +370 5 272 4741, Fax +370 5 373 7723, e-mail: *kestutis.papsys@gf.stud.vu.lt.* MSc from Vilnius University, 1999. Research interests: GIS methods for the management of natural, social and ecological hazards, spatial data generalization.