CONCEPTUAL MODEL FOR GENERALISATION OF LITHUANIAN SPATIAL REFERENCE DATA

Lina Papsiene¹, Ausra Kalantaite², Kestutis Papsys³

¹, ²Vilnius Gediminas Technical University, Saulėtekio ave. 11, LT-10223 Vilnius, Lithuania.
E-mails: ¹lina@rytai.lt; ²ausrak@zum.lt
³Vilnius university, M. K. Čiurlionio str. 21, LT-03101 Vilnius, Lithuania. E-mail: kestutis.papsys@rytai.lt

Abstract. Small scale spatial data are widely used at regional and national level – not only for mapping, but also for assessment of environmental conditions for purposes of planning, forecasting, etc. Therefore it is necessary to prepare such data professionally. Generalisation of large or medium scale spatial data is the most efficient process to produce smaller scale data. Of course, simple transfer of information is almost never suitable to satisfy the requirements for small scale maps. Additional transformations (generalisation) are necessary. During the process of generalisation complexity of spatial information may be significantly be reduced in terms of number of objects, geometry, etc. But the main spatial, non-spatial and topological characteristics of the objects have to be preserved. Process of reduction is irreversible, therefore it is necessary at first to clearly define the requirements for small-scale spatial data (for example, density of spatial objects, minimal allowed area, width and length of object, minimum length of the edge of object, spatial links between the objects). Given those requirements it is possible to develop a conceptual model and procedures of generalisation between particular data sets. Such model will also describe the appropriate selection, aggregation, simplification or alignment algorithms.

Keywords: generalisation, spatial data, reference spatial data, conceptual model, scale, GIS.

1. Introduction

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

Generalisation of spatial objects means selection of the main, essential and typical for specifically location objects and generalisation of their qualitative and quantitative characteristics. During this process, complexity of provided information is decreased, but important characteristics of the object are retained and unimportant features are skipped (Urbanas, 2001).

Generalisation is an irreversible process, so it must be thoroughly planned in advance so that the result obtained after its performance satisfies set requirements. At the primary generalisation process designing stage, it is necessary to prepare the conceptual model which would allow defining the exact stages of generalisation, principles, algorithms and verification rules for spatial data in relation with the quality requirements raised to spatial data. However, in order to obtain a realistic model, it is necessary to perform analysis of:

− Requirements raised to spatial data,
− Methods of generalisation and results reached by using them.

This article analyzes the principles of generalization based on the research of Müller (1991), Peng (2000), Başaraner (2002), Cecconi (2003), etc.

2. Review of development of Lithuania reference spatial data

In Lithuania, use of methods of generalisation during developing spatial data in small scales is not popular due to a few reasons:

− Limited understanding or even ignorance of algorithms of generalisation suggested by software;
− Striving that the obtained result complies with the requirements raised to traditional maps, so, the result obtained by using algorithms of generalisation does not always satisfy expectations;
− No time is invested in a analysis or development of other algorithms.
On the other hand, there are no strict requirements what kind of the result of spatial data must be obtained in the different map scales. There are specified only the general requirements to objects, such as their type or attributes. However, accuracy of representation of spatial data or thickness of objects is not set. Due to these reasons, development of spatial data is based on examples of spatial data which are mostly based on the principles of classical cartography.

The Lithuania reference spatial data are collected at tree main scale levels – 1:10 000 (base scale), 1:50 000 and 1:250 000. The high quality reference spatial data of the basic scale are developed and updated by using the latest digital orthographic maps. Accordingly, the base scale spatial data are perfect for automatic or semi-automatic preparation of spatial data at smaller scales.

3. The analysis of the spatial data

While analyzing spatial data, the requirements raised to it at different scales which impact its qualitative and quantitative parameters related with accuracy and representation should be analyzed.

The scale mostly defines resolution of spatial data. Resolution is defined as "smallest object or feature which is included or discernible in the data" (Goodchild, 1991): Geometric resolution indicates the geometric abstraction of the spatial data. Peng (2000) distinguished four aspects: a) geometry type, b) minimum object size (for example, minimum area or minimum length of object), c) minimum distance between two neighboring objects, and d) minimum object granularity (for example, minimum length of edge of object). The indicated permissible values are essential for planning of processes of generalisation in order to be able to choose suitable methods of generalisation, algorithms and parameters. In case the requirements for geometric resolution of spatial data are not defined, there are necessary personally to set optimal requirements for spatial data considering the cartographic expression of representation of spatial data. For that reason the person’s ability of distinguishing 0.5 mm objects or their changes on a digital map is usually evaluated (Table 1).

Table 1. The smallest spatial objects seen in different scales

<table>
<thead>
<tr>
<th>Object Type</th>
<th>1:10 000</th>
<th>1:50 000</th>
<th>1:250 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>5 m</td>
<td>25 m</td>
<td>125 m</td>
</tr>
<tr>
<td>Polygon</td>
<td>0,0025 ha</td>
<td>0,063 ha</td>
<td>1,56 ha</td>
</tr>
</tbody>
</table>

The requirements for accuracy should be not lower than those shown in Table 1. Spatial data is often applied requirements of higher accuracy. For example, the minimal distance between the vertexes of edge is 50 m and the minimum accepted area is 0.6 sq.km for reference spatial data in a scale 1:250000.

When keeping on analyzing reference spatial data, it is obligatory to evaluate its representation in different scales described by:

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

Reference spatial objects in different scales usually reflect the same phenomena of the real world and are represented in all scales. The exception in Lithuanian reference spatial data could be build-up territories represented in scales 1:10 000 and 1:50 000 which become objects of populated localities (cities, towns and villages) in a scale 1:250 000. Therefore, it is very important to evaluate the qualitative parameters applied for spatial objects in different scales. Thus, on the basis of attribute information of spatial objects, we will be able to select the data which is only essential for that scale according to its specific features, importance or size and avoid excess information.

After analyzing the types of geometric objects used for representation of reference objects in different scales (Table 2), three cases can be distinguished:

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

Table 2. Geometric representation of reference spatial data depending on their scale

<table>
<thead>
<tr>
<th>Geographic Objects</th>
<th>1:10 000</th>
<th>1:50 000</th>
<th>1:250 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roads</td>
<td>Line</td>
<td>Line</td>
<td>Line</td>
</tr>
<tr>
<td>Streets</td>
<td>Line</td>
<td>Line</td>
<td>Line</td>
</tr>
<tr>
<td>Railways</td>
<td>Line</td>
<td>Line</td>
<td>Line</td>
</tr>
<tr>
<td>Watercourses</td>
<td>Line</td>
<td>Polygon</td>
<td>Line</td>
</tr>
<tr>
<td>Lakes and pounches</td>
<td>Polygon</td>
<td>Polygon</td>
<td>Polygon</td>
</tr>
<tr>
<td>Buildings</td>
<td>Polygon</td>
<td>Points</td>
<td>Points</td>
</tr>
<tr>
<td>Built-up areas</td>
<td>Polygon</td>
<td>Polygon</td>
<td>Points</td>
</tr>
<tr>
<td>Agricultural areas</td>
<td>Polygon</td>
<td>Polygon</td>
<td>Polygon</td>
</tr>
<tr>
<td>Forests</td>
<td>Polygon</td>
<td>Polygon</td>
<td>Polygon</td>
</tr>
<tr>
<td>Geodetic points</td>
<td>Points</td>
<td>Points</td>
<td>Not Applicable</td>
</tr>
</tbody>
</table>
4. Generic principles of generalisation

The principles of spatial data generalization were thoroughly researched.

Before choosing methods and principles of generalisation, it is necessary to choose the degree of generalisation depending on the scale which defines the amount of provided information compared with the amount of the primary spatial data. The easiest way of describing the degree of generalisation is to apply the parameters which can be divided into two logical subtypes:

− Quantitative parameters,
− Qualitative parameters.

Quantitative parameters determine the quantitative characteristics of generalisation, such as the amount of points, lines or polygons elements. For example, the amount of rivers per unit of territory (area) is thickness of rivers. This parameter can be easily expressed in a mathematic way (Чурилова Е. et al. 2004):

\[
SD_T = \frac{n_T}{A} \quad (1)
\]

\[
SD_T = \frac{\sum_{i=1}^{n} l_i}{A} \quad (2)
\]

\[
SD_T = \frac{\sum_{i=1}^{n} A_p}{A} \quad (3)
\]

where \(SD_T\) – thickness of spatial data, \(n_T\) – amount of point objects per defined territory, \(l_i\) – sum of lengths of linear objects per defined territory, \(A_p\) – sum of areas of area objects per defined territory, \(n\) – number of objects per defined territory. Thus, it is possible to determine \(SD_T\) of 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). \(SD_T\) is measured by units (amount, sum of lengths or areas) in territories of the set size (for example, sq.km). In order to have a proper amount of generalized objects, it is necessary to set thickness of spatial data for every scale, every type of object and specific territories in advance (Table 3) (Чурилова Е. and Колособа M. et al. 2004).

<table>
<thead>
<tr>
<th>Objects</th>
<th>Type</th>
<th>Scale</th>
<th>Territory types</th>
<th>(SD_T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roads</td>
<td>Lines</td>
<td>1:50 000</td>
<td>Cities</td>
<td>196</td>
</tr>
<tr>
<td>Roads</td>
<td>Lines</td>
<td>1:50 000</td>
<td>Towns</td>
<td>89</td>
</tr>
<tr>
<td>Roads</td>
<td>Lines</td>
<td>1:50 000</td>
<td>Rural</td>
<td>23</td>
</tr>
<tr>
<td>Roads</td>
<td>Lines</td>
<td>1:250 000</td>
<td>Cities</td>
<td>10</td>
</tr>
<tr>
<td>Roads</td>
<td>Lines</td>
<td>1:250 000</td>
<td>Towns</td>
<td>7</td>
</tr>
<tr>
<td>Roads</td>
<td>Lines</td>
<td>1:250 000</td>
<td>Rural</td>
<td>4</td>
</tr>
</tbody>
</table>

\(SD_T\) must be determined considering the optimal amount of information which will be representing. By using this information the common possible \(SD_T\) has to be calculated for all specific territories (for example, build-up or rural areas) which is distributing to all spatial objects represented in specific territory.

After determining limitary thicknesses of spatial data and considering generalisation of bigger scale spatial data is always used for preparing a smaller scale spatial data, i.e., by decreasing \(SD_T\) (except dominant area objects which must become dominant in the generalized territory when diminishing the scale), the iterative process of selection of spatial data must be performed as long as the determined \(SD_T\) is reached (Fig 1).

Fig 1. Example of selection process using parameter of \(SD_T\)

The selection is one of the methods of generalisation of spatial data use to reach the limitary meanings of \(SD_T\). A search for the limitary meaning, calculation of the sum of lengths of objects and other are methods requiring especially high powers of calculation. Therefore, the information technologies and specialized GIS software must be involved for generalisation process for using the functions of automated clipping, calculation of length and area, SQL sentences of database queries and modelling environment.

Qualitative parameters determine the qualitative characteristics of generalisation of spatial data, such as the length of a river, the population of a town, the class of...
geodesic points and etc. In this situation, the recommended qualitative parameters for special determination of the type of spatial data, its scale or represented territory must be clearly defined. For example, river \( l > X \). During data selection according to its qualitative parameters, external generalisation 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 generalisation of objects must be applied during every type of generalisation. This is objects simplification. After performing selection during generalisation, i.e., after only choosing necessary data and excluding that not complying with the criteria, the internal generalisation must be performed. During this process the shapes of objects must be simplified. Simplification of the shape is performed as the cartographic expression of representation of spatial data decreases along with a decrease in the scale (Shea, K.S. et al. 1989), i.e., as the scale decreases, fewer and fewer elements can be represented in a unit of 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 representation of spatial data. During simplification, both the amount of vertexes per unit of length and other parameters, such as indication of fixed points (points of start and finish) and selection of the method of rarefaction of vertexes, must be considered. The methods of rarefaction of vertexes can be the following:

- Methods of elementary selections which just eliminate excess points (Fig 2). They are easy to implement and fast to perform.

- Logical methods which try to deviate from the simplified curve as little as possible (Fig 3). These methods are more complicated and slower, but they result in the image of a spatial object which is less different from the original.

It is also necessary to indicate the minimal area of the object obtained after simplification in algorithms of simplification of area objects. Objects which do not comply with this condition are removed and the extra selection is performed in this way.

The highest degree of simplification is performed in case of changing the type of geometry of objects. In case:

- Polygon objects (for example, buildings) are converted into points objects, 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, the central line of the object must be calculated.

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

When performing generalisation of spatial data, combination is applied as a separate case of the method of simplification. So, the simplification is in general performed not only inside objects but between the adjacent objects also. The adjacent objects must be combined in case the distance between uniform objects becomes less than the set cartographic expression of representation of 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. Objects of different geometric types are applied different types of combination:

- Combination of point objects means the aggregate of uniform points are defined and represented as one area object.

- Combination of linear objects is mostly performed by combining two lines represented in parallel. For example, two traffic lanes are combined into one street or street-limiting pavement lines are reformed into the central line of street.
Combination of area objects is performed by combining area objects located in the set distance into one object. The type of combination should be different for natural (Fig 4) and anthropogenic objects (Fig 5). When combining anthropogenic objects, angularity must be kept and the algorithm must intend such kind of combination of objects which includes as many straight angles as possible.

Fig 4. Example of nature objects combination

Fig 5. Example of anthropogenic objects combination

In order to obtain the result of the process of generalisation which satisfies the defined requirements as much as possible, certain preparative works must be performed which are related with generalized spatial data.

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

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

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

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

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

Fig 6. Example of primary, generalised and smooth river objects

5. Concept of generalization of reference spatial data

The proposed concept for generalization of spatial data was prepared regarding to result of analysis reference spatial data in Lithuania. This conception will allow designing of consistent processes for automated generalization in subsequent stages.

Generalisation of reference spatial data can be divided into three main stages:

- preparation of parameters of generalisation,
- preparation of generalized data,
- generalisation.

Preparation of parameters of generalisation must first be started from a detailed analysis of technical specifications of reference spatial data (scales 1:10 000, 1:50 000 and 1:250 000) by indentifying the structure of spatial data, its representation and requirements raised to accuracy. In case the requirements for accuracy of spatial data are not defined, there are necessary personally to set optimal requirements for accuracy of spatial data, otherwise there will not be possibility to prepare exactly model processes of generalisation. According to these parameters:

- \( SD_T \) have to be set for each type of object and specific territories which are different in their specificity.
- Certain algorithms and parameters of generalisation have to be chosen which will help to obtain a suitable result.

In order to obtain suitable result after generalisation, often need preparation of the primary data. Thus, the further process of generalisation of reference spatial data must be performed in the following order (Fig 7):

- Combination of the primary spatial data with the same qualitative unique characteristics.
Combination of spatial data according to the requirements of accuracy (minimal distance between objects) in a certain scale during which the adjacent spatial objects with the same qualitative characteristics are combined.

Division of spatial data into specific territories.

Decrease in thickness of divided spatial data by applying the cyclic process of selection and qualitative characteristics and comparing the obtained result with determined $SD_T$.

Determination of solidity of selected spatial data in case of spatial data breaks during a decrease in thickness of spatial data.

Combination of selected spatial data with the same qualitative characteristics.

Simplification of spatial objects according to the requirements of accuracy (for example, minimal distance between the vertexes of the line or boundary of polygon, minimal area or length) set for spatial data of every type in a certain scale.

Smooth of simplified spatial objects in order to improve visual characteristics of an object. This action is optional.

Restoration of topologic relations among spatial objects in case of having damaged them during generalisation.

In order to properly decrease thickness of spatial objects, $SD_T$ must be both calculated for each type of object and for the specific territory. However, it is not possible to state specific territories distinguished for one type of objects (for example, a few distinguished urban and rural territories) will also be specific for other ones (for exam-

Correctness of generalized spatial data is determined by the qualitative and quantitative requirements which are raised to the expected result of generalisation in advance and make presumptions to choose suitable methods of generalisation for obtaining a suitable result and applied parameters.

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

References


