IMPLEMENTATION OF A KNOWLEDGE DATABASE FOR THE GENERALIZATION OF TOPOGRAPHIC MAPS IN GIS SYSTEMS

Iwaniak, A.[1] and Paluszynski, W.[2]

[1]Institute of Geodesy and Photogrammetry, Wroclaw Academy of Agriculture, Poland. E-mail: iwaniak@ar.wroc.pl
[2]Institute of Engineering Cybernetics, Wroclaw University of Technology, Poland.
E-mail: Witold.Paluszynski@ict.pwr.wroc.pl

ABSTRACT

The goal of this work was to determine whether the existing commercial software for map generalization permits automating this process and applying it to produce Polish civilian topographic maps of 1:50,000. The scope of the project was mainly limited to buildings and communication, two most important and most difficult to generalize map elements.

The first step was to work out the knowledge base from the existing technical standards for topographic maps and consulting it with the authors of the concept of the Polish topographic map (1,2). Then a data model had to be created for the topographic maps of 1:10,000 and 1:50,000 scales, according to the technical standard for the digital cartographic models (DLM) for 1:10,000 and 1:50,000 scale maps.

Then the knowledge base was implemented within the DynaGEN environment. All the generalization steps had to be defined and configured in the GIS environment. We determine what and where to generalize by making spatial analyses, and determine how to generalize by selecting the appropriate algorithms and their parameters.

The whole generalization process is then evaluated, including the results obtained and the time and effort it took. This leads to many observations and conclusions which can be used to design an optimal digital data generalization process. Drawing conclusions from these experiments a comparison is made between relative efficiency of performing map generalization according to this procedure, and its alternative: the standard CAD tools.

1. INTRODUCTION

The generalization of topographic maps has some specific features. The appearance of both the source and target scale map is precisely defined by an official technical manual, which sets forth the editing standards for the topographic maps. This, however, does not make the process totally objective (3).

To describe the issues arising in generalizing topographic maps, the concepts of Digital Landscape Model (DLM) and Digital Cartographic Model (DCM) introduced by Gruenreich (4) are useful. Figure 1 presents the scheme of cartographic product processing currently in force in Poland.

It is worth noting that the process of generalization from DLM at 1:10,000 to DCM at 1:50,000, can alternatively be divided into two processes:

- DLM/1:10,000 to DLM/1:50,000, including mainly elimination, aggregation, collapse, classification, and simplification,
- DLM/1:50,000 to DCM/1:50,000, mainly connected with displacement, extension, symbolization, smoothing, exaggeration, and squaring.

This approach is interesting since the first process requires the preservation of proper topological relations, while in the second process this is not necessary. Additionally the development of modern GIS technologies, such as dynamic pipes, permits on-line generalization of temporary DCM's. Although this is by far imperfect, from the point of view of the technical manual for generalization, but much improving the readability for the end user. Some discrepancies from the manual are not critical since the user has access to the same database containing the source data.

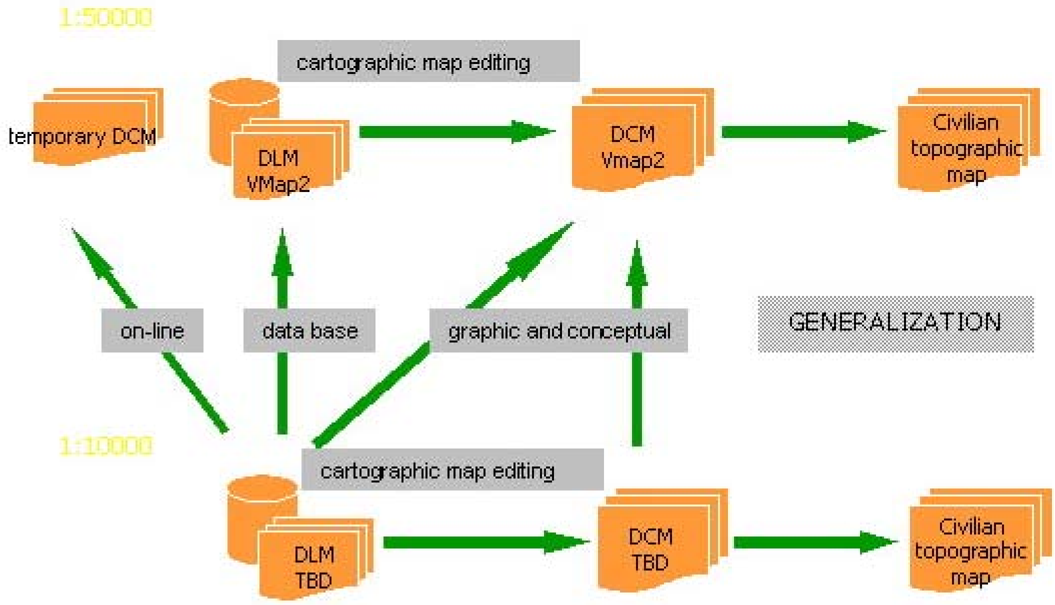

Figure 1. Cartographic product processing currently in force in Poland.

2. COMMERCIAL SOFTWARE FOR MAP GENERALIZATION

In 1992 Intergraph launched the first commercial program for map generalization, the Map Generalizer. It operated within the MGE environment and had a number of useful generalization algorithms. Its main shortcoming was the lack of the control of topology. This was fixed in the next map generalization program from Intergraph, the DynaGEN, which worked within the Dynamo environment. Dynamo is a highly specialized, object-oriented GIS environment. It incorporates many features important for the generalization process, such as: on-line topology, objectivity, and hierarchical structure of the data bases. Additionally, Dynamo has the map editing capability in the CAD mode as well as in the GIS mode.

DynaGEN, introduced in 1997, is another program from Intergraph which assists in the process of cartographic generalization. It was built according to the idea of amplified intelligence. DynaGEN is an application functioning as a subsystem of Dynamo, which means allows using the graphical environment, topological functions, and data models of Dynamo. The program offers two modes of operation: the batch mode (automatic), the assisted mode (under a human cartographer's supervision), and the manual mode. In the experiment the automatic and assisted modes were used. The program offers a "dynamic" operation, which means the operator can change any parameter values using sliders, and visually inspect the dynamically changing results. Only after intentionally accepting the results, any changes are made in the data base.

3. THE DESCRIPTION OF THE EXPERIMENT

The experiment proceeded by building the knowledge base concerning the generalization of Polish civilian topographic map from 1:10,000 to 1:50,000 scale in the DynaGEN environment. The 1:10,000 and 1:50,000 maps constitute the basic edition of Polish civilian topographic maps.

The work was divided into the following phases:

- working out the data model for the 1:10,000 and 1:50,000 maps in the Dynamo environment,
- acquiring vector data by scanning and vectorizing selected urban areas and removing geometric errors in order to establish proper topologic relations,
- accumulating the knowledge base about generalizing topographic maps from the existing official instruction for editing such maps and interviews with experts in the domain,
- defining generalization operators as well as their algorithms and parameter ranges, associating specific objects as inputs and outputs for the operators, including the ways of determining the attributes of newly created object, and defining the disallowable topological changes (for automatic and interactive modes),
- performing actual generalization of the selected areas of topographic maps.

4. WORKING OUT THE DATA MODEL FOR THE 1:10,000 AND 1:50,000 MAPS

Based on the existing official specification for topographic maps, a hierarchical data model was worked out for the 1:10,000 and 1:50,000 scales. In the Dynamo environment this leads to the creation of a dictionary (.dd extension files), which is used in the whole process of creating and generalizing maps. Figure 2 presents the model diagram for the settlement layer at 1:50,000 scale.

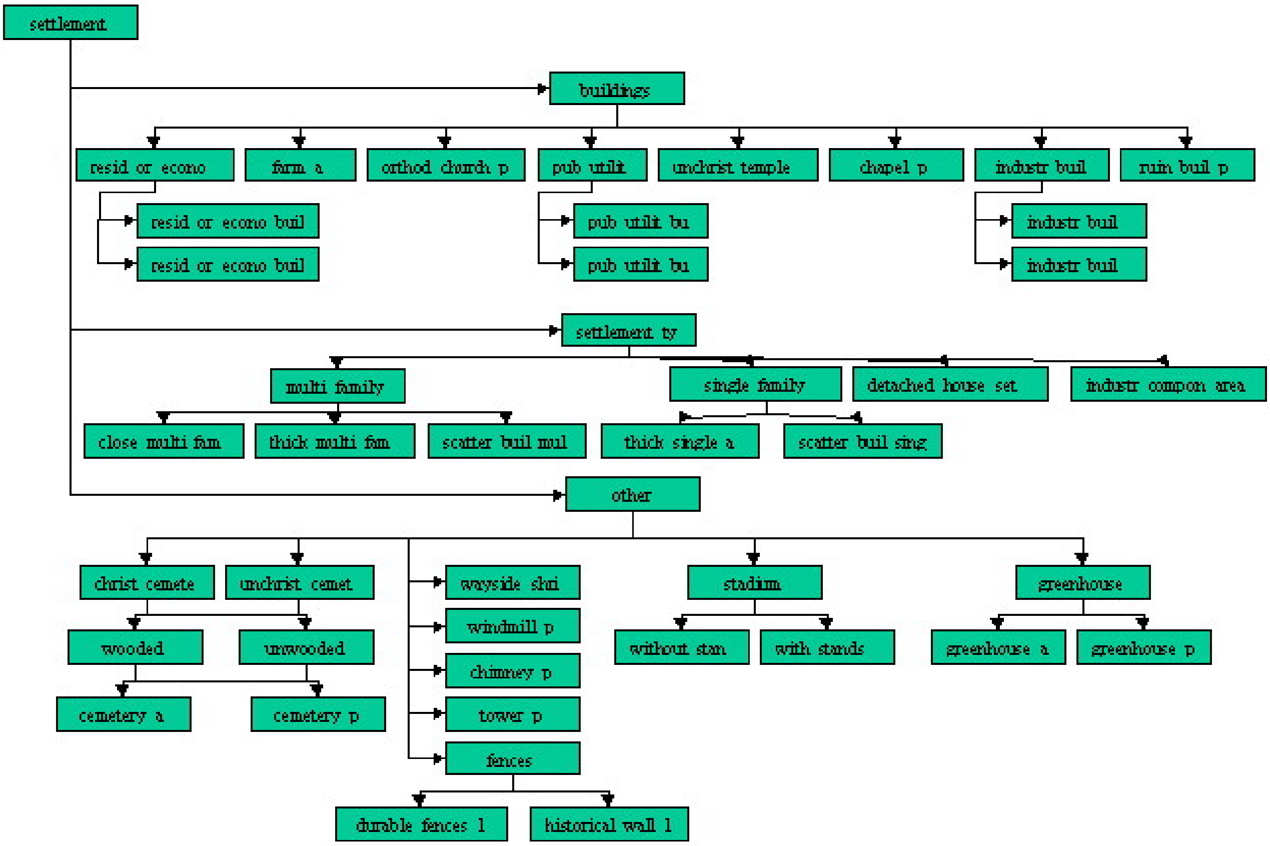

Figure 2. Object structure diagram for topographical map scale 1:50000, settlement layer.

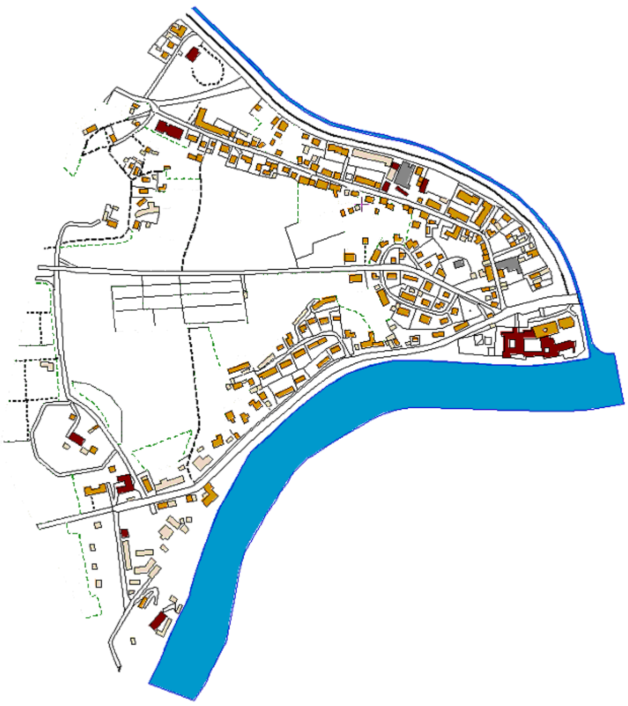

Figure 3. A section of a map of the city of Cracow at 1:10,000 scale, selected for generalization.

5. VECTOR DATA ACQUISITION

For the purpose of the experiment several fragments of a 1:10,000 topographic map of selected urban areas have been scanned and converted to vector format, according to the data model worked out. The data have in turn been edited to remove geometric errors to set up proper topological relations. Some of the data have been worked out within the scope of the Master's Thesis work of the students: Patryk Stelmaszak, Katarzyna Rant, and Iza Chebicka. Figure 3 presents an example of a section of such a map at 1:10,000 scale.

6. KNOWLEDGE BASE ACQUISITION

The knowledge base for generalization was worked out based on the existing official specifications for editing the topographic maps at the 1:10,000 and 1:50,000 scales, and after having gone through consultations with Dr Wieslaw Ostrowski, a domain expert and the chief author of these specifications.

Presented below is a summary of the rules worked out for the generalization of buildings into single-family settlement areas.

- Areas with dense cottage or urban residential settlement are marked with a sign for dense single-family settlement.
- Developed parcels are aggregated into single-family built-up areas if the distance between the buildings is less than 30 meters and the parcels adjoin each other.
- If the distance exceeds 30 meters but is less than 50 meters then the area can also be aggregated provided that farming buildings are located on the parcels to be aggregated.
- A minimum of three cottages or residential buildings can be aggregated into a single-family built-up area or a minimum of two cottages or buildings if they are part of a larger built-up area and are separated from the rest by a road running across that area, Figure 4.
- Typically, the whole parcel is included in the built-up area, unless there is no building on it, in which case it is altogether omitted.
- The minimal width of a strip of settlement is 3 millimeters. Narrower settlements are widened to 3 millimeters, Figure 5.
- If the undeveloped part of a parcel has more than 3 millimeters width, then it is excluded from the built-up area, .
- If two single-family areas are closer than 2 millimeters, then they are merged, Figure 6.
- Breaks in the line of settlement are shown down to 2 millimeters. If a parcel sticks out less than 2 millimeters outside of the line of settlement, then it is truncated to the line of settlement, Figure 7.
- The minimum distance between the settlement outline vertices is 30 meters.

Figure 4. Three cottages or residential buildings can be aggregated into a single-family built-up area.

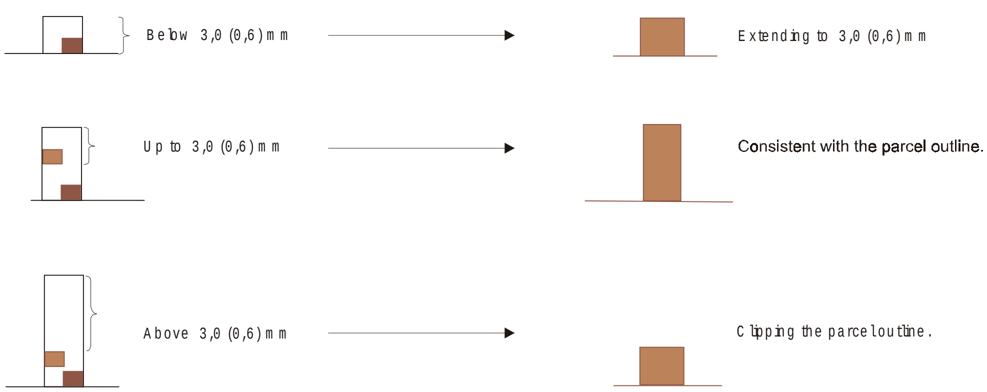

Figure 5. Consistency with the parcel outline.

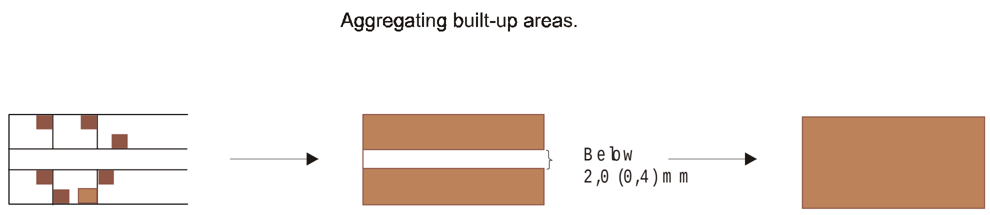

Figure 6. Two single-family areas closer than 2 millimeters are merged.

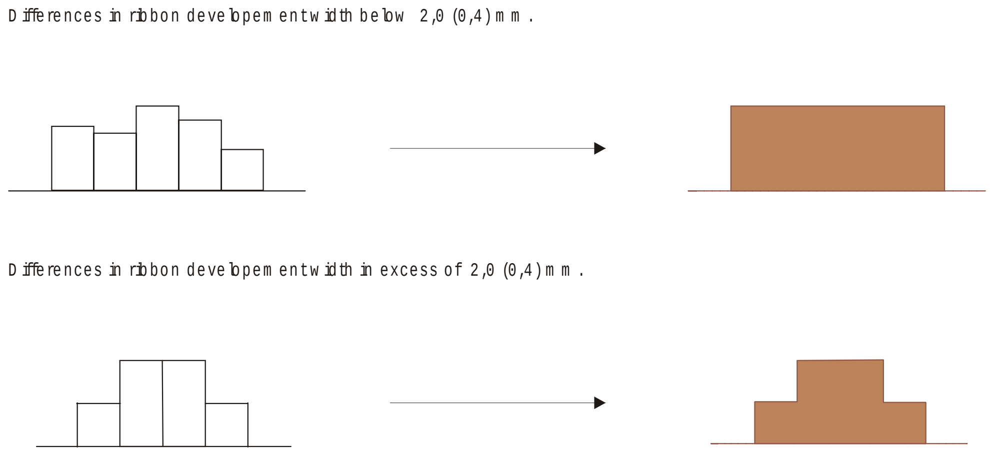

Figure 7. Breaks in the line of settlement are shown down to 2 millimeters.

7. WORKING OUT THE KNOWLEDGE BASE

In the previously mentioned AAI system for urban area generalization the implementation consisted of writing textual rules. In the present system the knowledge base is implemented by: (i) defining generalization operators to be used, (ii) describing the conditions to be met by the generalized objects, (iii) describing the attributes of the objects resulting from generalization, and (iv) defining topological constraints which should be observed across generalization. These elements are technical and specific to the DynaGEN system.

Additionally, the sequence of particular generalization steps needs to be defined, and this is presented now.

7.1 Adjacent parcel aggregation

- removal of undeveloped parcels and parcels with public or industrial buildings
- selection of adjoining parcels to be aggregated satisfying items 2 or 3 above by creating around them buffers of a suitable diameter
- adjacent parcel aggregation using the DynaGEN's adjoining area
- aggregation operator (newly created built-up areas have associated with them the number of aggregated parcels)
- removal of areas resulting from aggregating two parcels and are located a distance greater than the width of a road (0.6 millimeters) from other aggregated areas
- individual inspection of other areas made up of two aggregated parcels to see whether they are separated from a larger aggregated area by a road
- individual inspection of all single parcels to see whether items 7 or 9 above aren't violated; if so then adjustments are made

7.2 Disjoining areas aggregation

- drawing 30-meter buffers around the free-standing buildings, which are not located on parcels
- analyzing the intersections of the buffers created in the previous step with the areas aggregated in step 1, and putting the intersecting buffers onto a queue
- processing the queue and manually aggregating corresponding buildings with the areas aggregated in step 1, using the "Nonorthogonal Aggregate Area" algorithm and manual selection of parameter values to ensure correct shape and appearance
- finding the aggregated areas located within 20 meters from one another, such pairs are then individually aggregated using the "Nonorthogonal Aggregate Area" algorithm with manual selection of parameter values

7.3 Exaggeration

- inner buffers are created in the aggregated areas in order to find the areas of less than 3 millimeters width (on the 10:10,000 map)
- areas narrower than 30 meters are manually corrected to ensure the minimal settlement width (item 5)

7.4 Parcel elimination

- all remaining parcels containing single-family buildings are removed

7.5 Area smoothing

- smoothing of the boundaries of aggregated areas using the "Smoothing" operator with the "simple-average" algorithm and the value of 3 for the "look-ahead" parameter

7.6 Boundary line simplification

- the boundaries of the aggregated areas are processed using the "Simplify areas" operator, "Area Preservation" algorithm, and the value for the "Area Change Allowed" parameter individually selected in the range of 5 to 25

7.7 Boundary extend

- the boundaries of the dense single-family settlement areas are extended to the street lines using the "Boundary Extend" operator, "Areas to Lines, Areas" algorithm, and the parameter values individually adjusted in the following ranges: "Threshold Tolerance": 0.10-0.20, "Zone Tolerance": 0.10-0.20, "Hole Retention": value of 20
- after this operation the area boundaries require minor corrections which are done using the "Modify Element" tool

As the result of the above steps, a map is obtained like the one presented in Figure 8.

Figure 8. A map fragment obtained by generalization in the DynaGEN environment, a: the outcome, b: reduced 5 times.

8. RESULTS EVALUATION

The important questions, and the motivation for this project, are whether the process of generalizing maps can be automated, how much of this can be achieved under the DynaGEN environment, and whether this is a good environment for building the systems for map generalization.

The authors are accepting the common belief that this process is currently not possible to fully automate. Therefore the approach adopted was to use a commercial package for map generalization and use its specialized operators together with applying some changes manually as in a CAD-like environment.

The following observations and conclusions, concerning the issues and problems arising in the process of generalizing Polish civilian topographic maps, according to quite stringent requirements set forth in the technical manual, can be formulated:

- In order to achieve the correct result it is necessary to individually adjust the parameters of the specific algorithms of the generalization operators. In most cases this is contingent on the distances between objects and their shape. Therefore in many cases and individual inspection is required.
- The algorithms and their implementations of DynaGEN, in some cases are not perfect, and, due to high complexity, it is not possible to adjust their parameters to achieve the required result. Consequently, manual corrections are necessary. Even if the appearance of a map is satisfactory at a given scale, in some cases a geometry correction is necessary for preserving topological constraints (which is necessary for the correct operation of DynaGEN).
- The collection of generalization operators available in DynaGEN is not sufficient for executing all the steps for all possible cases, for example, see item 3 – exaggeration.
- The specification of the map editing rules given in the official technical manual is often not possible to translate into the terms and parameters of a GIS or related system, such as the DynaGEN, and often is generally not possible to formulate in any formal representation language, e.g. to make some changes "to preserve clarity".
- There is not enough information in GIS systems about the generalized objects. E.g., when generalizing the street network it is important to consider the historical background of the development of a city. In principle, such additional information can be stored in GIS systems for the purpose of more proper generalization, but this is questionable whether this is a viable solution.

A good example of the problem of individual selection of parameter values is the disjoint aggregation of buildings 2(c), which can be directly converted into a sequence which allows a higher degree of automation and a significant shortening of the process. However, this would require some relaxation of the requirements presently in force.

The proposed implementation of the knowledge base likely is not the only possible one. Some changes in the order of steps and the degree of automation are possible.

9. EFFICIENCY

The time required for processing a map according to the procedure described in this paper, biased toward the maximum use of the assisted mode of DynaGEN, is comparable to the processing time for the same map in a CAD-like environment by a trained operator thinking in a holistic way. This comes from the fact that some basically simple tasks must be divided into several steps, with individual parameter adjustments at each. For example, such is the process (omitted in this description) of the simplification of complex building outlines, e.g. churches, apartment buildings, etc. On the other hand, significant savings result from the capability of performing one step automatically for a larger area.

Processing maps in GIS systems in general simplifies and speeds up making decisions of object or generalization step selection by allowing the operator to make spatial queries. Therefore it is reasonable to use the approach described here for only some generalization steps, which takes the best advantage of its features.

10. CONCLUSIONS

The generalization of the Polish civilian topographic map from 1:10,000 to 1:50,000 scale is a very complex process. At 1:50,000 scale there are selected buildings marked on the map as well as built-up areas. The built-up areas are divided into: compact multi-family areas, dense multi-family areas, sparse multi-family areas, dense single-family areas, sparse single-family areas, sparse farm house areas, summer house areas, and industrial/storage areas.

Generalizing each of these areas has its own idiosyncrasies. The description of the experiment given in this paper presents the process of generalization of only one type of settlement which, as can be seen, has many steps. The knowledge base for the whole process is significantly longer, and its implementation requires many steps and substeps of generalization. The operator is required to perform these steps individually, adjusting parameter values for specific objects. For that reasons, executing all the steps of generalization strictly adhering to the interactive mode will likely convert the difficult and time-consuming process of manual generalization into a likewise difficult and time-consuming process of digital generalization.

Since the fully automatic generalization still seems impossible at the current level of technology, the optimal compromise seems to be to use the DynaGEN system in all: automatic, interactive, and manual modes.

The manual mode is necessary to allow the operator to have the freedom of skipping several painful steps by making one arbitrary and comprehensive change in the shape of some object(s).

It is worth noting, that the example generalization rules presented in this paper are defined in the technical manual in clear and unambiguous terms. Unlike that, there are other fragments there, which are quite simple intuitively, but are hard to put in objective mathematical categories. For example: "*do something if space allows*", or "*if possibilities exist*", or "*without generalization, if possible*".

In spite of what has been said here, the experiments conducted have lead to the following conclusions:

- the DynaGEN system is a universal and advanced tool assisting in the process of generalization, having a rich collection of generalization algorithms,
- delivers a coherent methodology for the whole process,
- has excellent tools which allow one to define illegal topological changes, which transparently, but in a clean and coherent way define the knowledge base for generalization; eg. in the presented example, an "Area surrounds gained" disallowable topological change was introduced, which prevented the generalized areas from overlapping one another.

One of the conclusions from this work concern the way that generalization should be defined. In order to simplify the generalization within GIS systems, and especially automating much of the generalization process, some compromise should be made in writing the technical specification in the map editing manual, between achieving a desired appearance, and making the rules easy to implement using typical spatial queries and applying common operators.

In 2003 two new products have been introduced in Poland: the topographic data base and the VMAP2 (military edition of the topographic maps). This way it will be separated to separate the generalization process into two parts: the DLM generalization and the DCM editing, according to the scheme in Figure 1.

The topographic data base can be identified with the DLM, and VMAP2 with the DCM, which has proper topological relations. The edition process of a military topographic map is much simpler than that of the civilian topographic map. Therefore it should be expected that in the nearest future there will be a move toward generalization the TDB into the VMAP2.

11. ACKNOWLEDGEMENT

This work was supported as a research grant from the funds of the State Committee for Scientific Research.

12. REFERENCES

[1] The principles of the topographic map editing at 1:10,000 scale, Technical Manual (in Polish), Chief Office of Geodesy and Cartography, Warsaw (1999)

[2] The principles of the topographic map editing at 1:50,000 scale, Technical Manual (in Polish), Chief Office of Geodesy and Cartography, Warsaw (1998)

[3] M.J.Kraak and F.Ormeling, Cartography: Visualization of Spatial Data, Addison-Wesley Longman Limited (1996)

[4] Gruenreich, Konzeption und erste Erfahrungen aus der Ausbauphase des digitalen Landschaftsmodell 1:25000 (DLM25), In: Proceedings XIX FIG Congress, Helsinki (1990)

[5] A.Iwaniak and W.Paluszynski, Knowledge-Based Generalization of Topographic Maps, In: GIS2001 Conference, Vancouver (2001)

IMPLEMENTATION OF A KNOWLEDGE DATABASE FOR THE GENERALIZATION OF TOPOGRAPHIC MAPS IN GIS SYSTEMS

Iwaniak, A.[1] and Paluszynski, W.[2]

[1]Institute of Geodesy and Photogrammetry, Wroclaw Academy of Agriculture, Poland. E-mail: iwaniak@ar.wroc.pl
[2]Institute of Engineering Cybernetics, Wroclaw University of Technology, Poland.
E-mail: Witold.Paluszynski@ict.pwr.wroc.pl

Biography

Adam Iwaniak is Assistant Professor at the Department of Geodesy and Cartography of the Agricultural University of Wrocław. His 1997 Ph.D. thesis was devoted to the application of expert systems in GIS. His interests were originally in developing software for geodesy, and since 1990 he has been involved in GIS. Since 1994 he also directs the LIS division at one of the largest mapping and surveying companies in Poland. His present interests concentrate on: application of expert systems to map generalization, data standards in GIS, and implementing industrial GIS systems. Since 2002 he directs the GISLab at the Agricultural University of Wrocław. Since 2002 he is also a GIS advisor with the Surveyor General of Poland.