A Conceptual Framework For Automated Generalization and its Application to Geologic and Soil Maps

Stefan Steiniger and Robert Weibel

Department of Geography, University of Zurich
Winterthurerstrasse 190, CH-8057 Zürich
{sstein; weibel}@geo.unizh.ch

Abstract: Research in map generalization starts to focus on other types of maps than on topographic maps only. Therefore, new generalization concepts have to be developed, extending on the experiences and research efforts originally made in topographic generalization. This paper presents a new conceptual framework of map generalization with respect to thematic mapping based on an analysis of the literature. The framework consists of three phases: a structure analysis stage, the generalization stage and the visualization stage. The stages are first discussed by their objectives and subsequently with respect to more technical modeling aspects. Finally, as a case study, the framework is applied to conceptual modeling of the generalization of geologic and soil maps.

1. INTRODUCTION

In recent years automated map generalization has made significant progress in a number of ways. However, as these developments were mainly in response to demands put forward by national mapping agencies (NMAs), the core of these improvements focused on the generalization of topographic maps. An adaptation of these models and procedures to the requirements of thematic cartography is still largely outstanding. Thus, the objective of this paper is to establish a generalization framework for thematic maps with respect to vector data.

Research on automated generalization of thematic data has been reported by Fuchs (2002) and Galanda (2003). Whereas Fuchs (2002) concentrates on semantic aspects of the generalization of soil maps, Galanda (2003) developed a set of generalization constraints on categorical maps for use in an automated, agent-based generalization system (cf. Section 2.2. for explanation of constraints). Drawing from this work and the analysis of a number of thematic maps we propose a conceptual model that unifies the generalization of topographic maps with other, domain related models. The framework should be applicable to vector data, which are retrieved directly from GIS analysis or from a database, and should describe the necessary steps to a satisfactory visualization product. Here, no restrictions regarding the visualization output (e.g. paper map or screen) are made.

Starting with a literature survey on existing generalization frameworks and modeling techniques, we will afterwards introduce the new framework in general (Section 3). The following paragraph (Section 4) will explain more in detail some technical modeling aspects whereas in Section 5 we derivate the conceptual frameworks for generalization of geologic maps and soil maps.

2. EXISTING GENERALIZATION FRAMEWORKS AND MODELING TECHNIQUES

Several frameworks for automated generalization have been proposed over the past two decades. Some of the most influencing models will be explained shortly below. For the purposes of this paper, we distinguish these generalization frameworks into those that seek to conceptually structure the entire generalization process (process oriented models) and those that address the level of map objects (object level oriented models). Examples of the process oriented view are the frameworks of Brassel and Weibel (1988) and McMaster and Shea (1992). The feature-oriented view organizes the map objects into so called macro, meso and micro level objects and models the processes applied to objects of these levels. Macro level objects may represent the entire map, a section of a map or a specific object class (e.g. all roads). A group of objects, unified by some common cartographic criterion, would form a meso level object (e.g. an alignment of buildings). Finally micro level objects are single map objects like a single building or river. Examples of this type of framework include the one by Ruas and Plazanet (1996) and its enhanced version by the AGENT project (Ruas, 1999a; Barrault *et al.*, 2001).

2.1. Process oriented frameworks

The framework of map generalization by Brassel and Weibel (1988) distinguishes five stages of processing: *Structure recognition*, *process recognition*, *process modeling*, *process execution*, and *data display*. Generalization starts with a structure recognition of the source data which is constrained by the generalization controls (e.g. map purpose, scale, symbol specifications, etc.). The objective of that phase being a characterization of spatial and semantic relations of objects and their aggregates. Process recognition, as a second step, is also constrained by the generalization controls and aims to deliver information of process types and parameters to apply to the previously characterized data. In the third step, process modeling, the compilation of rules and procedures from a process library is made. The fourth step, process execution, is responsible for actually applying the data transformation (i.e. generalization) procedures, followed by the final step, data display, responsible for displaying the previously transformed data. While this conceptual framework advocated a sophisticated generalization process based on the characterization of the structure and meaning of map data, most generalization approaches, until the mid-1990s, used only simple built-in heuristics for data characterization (i.e. structure recognition). However, in recent years this has changed with generalization approaches that build on explicit data characterization, such as the one used in the AGENT project (Ruas, 1999a; Barrault *et al.*, 2001).

McMaster and Shea (1992) in their book give an review of generalization frameworks and extend the one by Brassel and Weibel (1988). They add missing parts and specify details for the process phases. Their framework is devoted to three main questions of generalization: *Why, when* and *how* to generalize? The first question (why) is answered by phi-losophical objectives, organized into *theoretical elements, application specific elements* (cf. Section 4) and *computational elements*. The question of 'when?' leads to geometric conditions and types of measures for data evaluation. Further, the transformation controls (operator selection, algorithm selection and parameter selection) are discussed. Thus, this part resembles the structure recognition phase of Brassel and Weibel (1988). The answer to the question 'How to generalize?' is given by a list of spatial and semantic transformations and their organization into so called *operators* (e.g. symbolization, aggregation, smoothing, etc.). Note that the term 'process' used by Brassel and Weibel (1988) is equivalent to the term 'operator' in McMaster and Shea (1992).

The generalization model used in the German ATKIS project (e.g. Grünreich, 1985; Morgenstern and Schürer, 1999) represents a different view of the overall generalization process. Here, a distinction into three types of generalization processes is made (cf. Figure 1). The first type is called *object generalization* which describes a mental generalization process (in the sense of abstraction and selection) from reality to data made by the data collecting person (surveyor, aerial photo analyst, or data analyst). A second type of generalization – *model generalization* – aims to reduce the resulting dataset of the *object generalization* under statistical control by reducing the number of object classes and the spatial accuracy. *Cartographic generalization* denotes the third process type. Here, the objective is a graphic representation of the digital data. Therefore symbolization specifications are applied and structures are modified locally and in an object-specific way (e.g. elimination, emphasis, shape distortion, displacement, etc.).
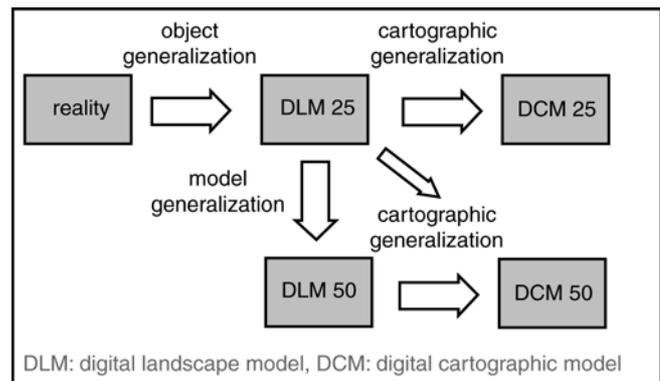

Figure 1: The generalization types of the German ATKIS project. The numbers denote the resolution scale in thousand (e.g. 25 equals 1:25'000).

2.2. Knowledge formalization and object level oriented modeling

Transferring the cartographer's knowledge to the digital domain is one of the difficult tasks of automating generalization. The cartographer integrates knowledge about cartographic conventions and human cognitive processes and its decomposition into an atomic set of facts and rules is difficult (Armstrong, 1992). A simplistic way of 'tranferring' knowledge is by providing a set of digital generalization tools which are interactively selected and applied by a cartographer. This way of using cartographer's knowledge is called *human interaction modeling* (Harrie and Weibel, 2005). For some of the tasks to be solved during the generalization process the formulation of conditions and actions is not as hard as for the artistic components of map making. Examples are legibility rules for ensuring minimal dimensions of object size and inter-object distances. Using rules such as *IF (area of building X < 200m^2) THEN (enlarge building)* the overall generalization process can be broken down into condition-action pairs, leading to an approach termed *condition-action modeling* (Harrie and Weibel, 2005), which was mainly pursued in the late 1980s and early 1990s.

Both types of modeling have their disadvantages. An evaluation of interactive generalization systems based on pure human interaction modeling showed a very low productivity gain and showed further that generalization results highly (and not surprisingly) depend on the skills of the user (Ruas, 1999b). Some weaknesses of rule based systems are on

one hand the difficulty of acquiring and formalizing cartographic rules and on the other hand the great amount of rules required to describe conditions and actions between map objects sufficiently well. Further problems arise from the sequencing of generalization operations, since the different operators may affect each other, potentially causing secondary conflicts. Thus, a need existed for a more flexible sequencing approach with respect to the dynamic problem context (Harrie and Weibel 2005, Holland 1986).

Induced by the drawbacks of rule based systems Beard (1991) proposed the use of *constraint based modeling* for automated generalization. Constraints formulate requirements of a generalized map, that is, conditions that a generalized map should adhere to. However, in contrast to rules, violation of such a condition is not bound to an action. Here, choosing an action to solve a problem is the result of a synthesis of conditions (Ruas and Plazanet, 1996). Beard (1991) initially suggested four types of constraints: graphic, structural, application, and procedural constraints. These types of constraints were revised by several authors, with a synthesis provided in Harrie and Weibel (2005), a source which also provides a more detailed description of process control and modeling in generalization than given above.

The object level oriented framework for automated generalization by Ruas and Plazanet (1996) realizes a constraint based modeling approach. Their model is based on one hand on the model of Brassel and Weibel (1988) using the parts of structure recognition and process recognition as well as the modeling and execution processes. On the other hand they assimilate the suggestion by Mackaness (1995) that only a constraint based and iterative refinement strategy with trial and error (i.e. backtracking facilities) results in a satisfactory map. The proposed framework, which was implemented in an experimental platform at IGN France, consists of three levels of processing. On the highest level a 'global master plan' determines a sequence of generalization tasks to apply on the level of the entire map (e.g. aggregate all connected objects of same class). On the second level a geographic situation is selected according to given task (e.g. the objects contained in an urban block). On the third, local level a generalization plan for every situation is developed and executed and the generalization result evaluated. These local plans determine the operations (e.g. simplification, smoothing) and their sequence and further determine appropriate solution algorithms and parameters (e.g. simplification with Douglas-Peucker algorithm and threshold of 5m) responding to constraint violations. If the evaluation phase reports an unacceptable solution three types of actions are possible: 1) other parameters are selected, 2) another algorithm may be used, or 3) a new local plan is chosen. This process is then continued until a satisfactory solution is obtained. This model has been further refined and re-implemented in the prototype generalization system of the AGENT project (Ruas, 1999; Barrault *et al.*, 2001) and its commercial successor software Clarity by Laser-Scan. Figure 2 shows the version of this iterative processing model as it was implemented in the AGENT prototype.
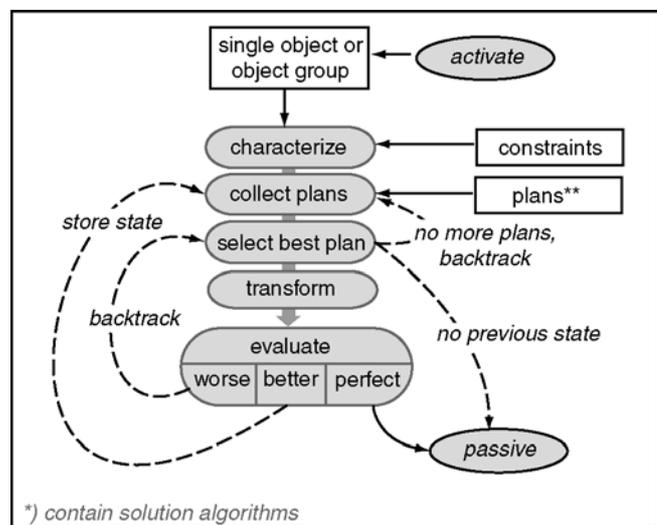

Figure 2: Generalization procedure for a single object or group of objects - after Barrault *et al.* (2001).

Peter and Weibel (1999) present a further constraint based framework, with particular focus on the generalization of categorical maps. It is strongly influenced by the frameworks of Brassel and Weibel (1988), McMaster and Shea (1991), and Ruas and Plazanet (1996). The paper discusses in detail generalization operations and constraints for categorical maps and proposes measures for their evaluation. A unique element is the data model specific view, distinguishing generalization operations and measures in the raster and vector domain, as well as hybrid approaches.

2.3. Models for solving generalizations conflicts with single and multi criteria

The discussion on modeling of solution processes is needed to understand the modeling approaches used in our generalization framework proposed below. Similarly as for the types of knowledge formalization (human interaction, rules, and constraints) a detailed discussion of the approaches for modeling single and multi criteria solutions in generalization can be found in Harrie and Weibel (2005).

To react on a single condition, or act on the violation of a single constraint, one can either interactively decide what to do or use rules. For instance, if the building area is a single condition, we can formulate:

IF (building area < x (= too small to be seen on the map)) THEN (enlarge building by scaling value x)

ELSE IF (building area < x) AND (building area < y) THEN (delete building)

ELSE (everything is fine, no action required).

Multi criteria solutions are necessary if more than one condition exist. For example, a building is too small to be seen on the map AND a large building is located 0.5 mm next to it. Here, a simple enlargement would lead to an overlap with the neighboring building. Several solution strategies for multi criteria problems are possible:

- solve *interactively*: the human decides what to do ⇒ this is not very effective (see Section 2.2);

- combine several rules in fixed order (*batch system*) ⇒ this is inflexible (see Section 2.2);

- combine several rules in dynamic order with respect to the situation, whereas the order is determined by a human, using *flowcharts* and *workflow modeling;*

- use of optimization procedures such as: (1) agent modeling, (2) continuous optimization, (3) combinatorial optimization, or (4) neural networks.

Workflow modeling and *flowchart modeling*, respectively is a technique used by many so-called data flow visualization systems and in the GIS domain by ESRI in the ModelBuilder®, providing an intuitive way of interactively chaining together several processing tasks (see below).

The solution of several interacting generalization conflicts, which can be seen as a multi criteria problem, by use of an *agent model* is accomplished in separate sub-processes. The overall solution is a pseudo optimal solution, since the partial solutions can be satisfactory or only acceptable, yet they balance each other in sum. An application of agent modeling is described in Barrault *et al*. (2001). *Combinatorial optimization* methods like simulated annealing (Ware *et al.* 2003) or genetic algorithms try to find iteratively an optimal solution by means of maximizing or minimizing a specific criterion. Such a criterion can be composed of others and the process of finding an extreme is influenced by a stochastic component. *Continuous optimization* methods are based on a descriptive function of the problem (e.g. line smoothing and line displacement in Steiniger and Meier, 2004). The optimization by least squares adjustment has been applied by Harrie and Sarjakoski (2002). Here, an optimal solution is found by minimizing the squared sum of differences among several optimal sub-solutions, one for each condition, and a balanced overall solution. *Neural Networks* are trained to deliver an optimal solution for a specific problem. Thus, up to now they have not been used for modeling entire generalization processes, but Sester (2005) used the technique for solving typification operations on building datasets.

Workflow modeling and agent modeling are generic techniques and have shown to be able to incorporate other kinds of multi criteria solution processes (e.g. rules-based approaches, or optimization techniques). In the next section our new generalization framework is introduced and in Section 4 both modeling types are proposed to be components of the framework.

3. A FRAMEWORK FOR AUTOMATED GENERALIZATION OF (THEMATIC) MAPS

The proposed framework draws from the previously discussed models, which were mainly inspired by topographic map generalization. The aim is to make our model more generic, in order to encompass the more general map generation process, directly from GIS data. Since it merges aspects of several models it is at the same time a process oriented and a object level oriented framework. Figure 3 gives an overview of the proposed conceptual framework. Three processing stages are distinguished: *(1) structure analysis, (2) generalization and (3) visualization.*

The first stage, **structure analysis**, is similar to the structure recognition phase in the model by Brassel and Weibel (1988). Here, however, this step refers only to the object class level (e.g. trees, roads, soil types, etc.) and the level of groups of map objects (e.g. alignment patterns, networks). Plus, structure analysis also results in a pre-selection of the map data in the form of a catalog that assigns retain/delete flags to map objects. Three objectives are assigned to this step:

- the selection of object classes and the definition of categories (e.g. number of soil map units; see below)

- the determination of an aggregation hierarchy for polygon elimination and aggregation, and

- detection of (visual) patterns to preserve them during generalization.

The resulting constraints from this step, here called *type A constraints*, are later introduced in the generalization phase (stage 2). Type A constraints stem from semantic and structural relations on the object class or group level. Visual patterns (e.g. a group of small lakes) or thematic dependencies (e.g. the association of a sandy soil type to a river bed) form structural constraints for the generalization process, since these structures should be maintained (Galanda, 2003). A systematic collection of possible structural and semantic relations in maps is given in Steiniger and Weibel (2005). Structure analysis prepares selection and classification operations (Mackaness, 1995) to be executed in the second stage (generalization). The resulting effects of generalization will take place at large, partially provoking radical changes to the map image.
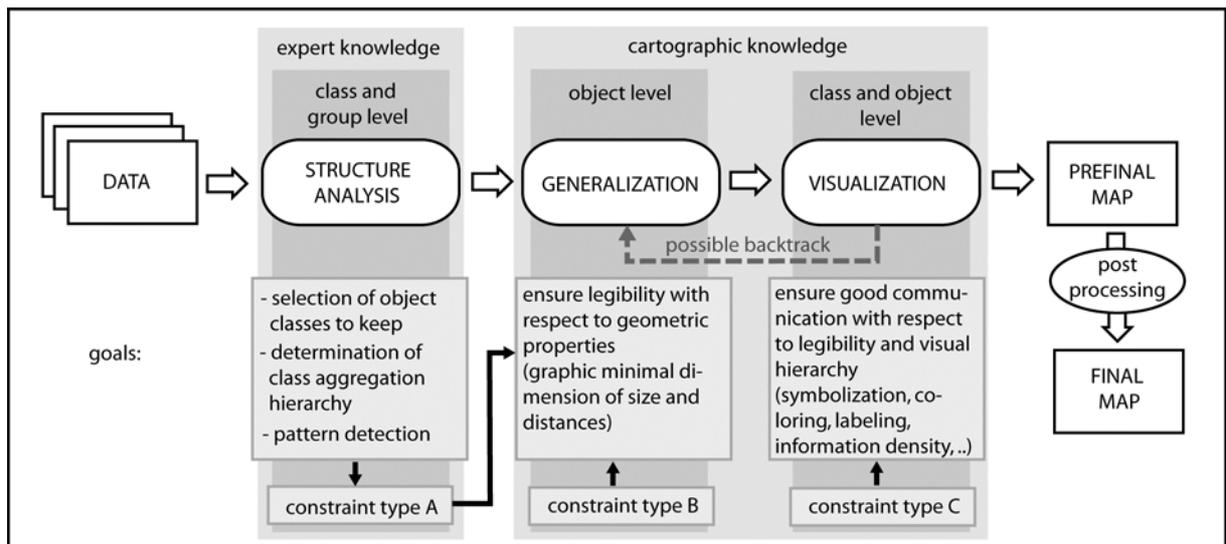
Figure 3: Framework for the generation of (thematic) maps from GIS Data.

The next phase, the **generalization stage**, is oriented to the map object level. Here, generalization will result only in local changes and therefore resemble a fine tuning process. Stage 2 executes the generalization process on the macro level (class/map), on the meso level (object groups) and on the micro level (single objects). The objective for this stage is to produce data whose resolution is adapted to the intended target map scale. The output of the complex generalization process should be a legible "raw map". Of course, the processing of the lower resolution data is influenced by a couple of different constraints:

- *Type B constraints*, which can be further distinguished into:
  - geometrical minimal dimension constraints, which depend on the target scale;
  - topological constraints;
  - procedural constraints (resulting from the knowledge of the cartographer – e.g. the order of processing: simplify first and then smooth – the selection of algorithms and parameters).
- *Type A constraints*: the pattern and dependency constraints of the first stage. This also includes a priority ordering of the various object classes compared to each other.

The resulting map, or better: the resolution adapted data, are further processed in the **visualization stage**. Tasks of this stage are (1) symbolization, (2) text labeling, (3) introduction of a visual hierarchy (cf. Dent, 1990) and (4) avoidance of information overload. Task (4) is accomplished by a density check, counting the number of objects per unit area. A cartographic rule is applied which says that recognition of map content is difficult if more than three layers of information exist (Spiess, 2002). The introduction of a visual hierarchy (3) includes on one hand a color and texture allocation for area objects, if no detailed specifications exist. On the other hand it includes the awareness of rules and effects of perception such as the figure-ground phenomenon. Therefore, the *constraints of type C* would cover rules for text labeling, the three-layer rule for information density, as well as rules for area coloring with reference to perception and to the map topic (e.g. consideration of international conventions for soil types or geologic formations).

Separation of generalization and visualization is not necessary for the generation of topographic maps since these maps have usually pre-determined colors and symbols that are equally applied across all maps of a series. In contrast, this does not hold for thematic maps. Every real-world situation represented by a thematic map has its own specificities which deserve to be highlighted by the cartographer. Furthermore, adaptation of the final map to user needs and output medium is a requirement of today's information society, particularly in thematic cartography. Examples for user adaptation include generation of different maps depending on age group (child, teenager, adult) or user experience (novice, intermediate user, or expert). Several output media for maps are imaginable, capable of displaying b/w or color, and ranging from different kinds of paper (requiring color adaptation) to screen, and web application. Recently, geologic services stated the objective to publish their maps on the web (among others: Saha *et al.*, 2003).

In Figure 3 the possibility of **backtracking** between the visualization and generalization stage, respectively, is indicated. This is due to the fact that information about the extent of symbols (particularly the width of line symbols) is often necessary for a satisfactory generalization result. Another reason are visual side-effects of color selection on geometry (cf. the Figure-Ground phenomenon), e.g. a new color for buildings in a city map results in a worse contrast and visual separation. Consequently the buildings may have to be displaced somewhat further from each other.

Application of the proposed framework: There is no requirement that every stage of the framework has to be fully executed. For example, the creation of a low resolution digital landscape model from a high resolution DLM (cf. Figure 1), called model generalization, would not need the visualization stage. Taking another example, the creation of a map of the number of births per county, we would start with the structure analysis stage, classing certain birth numbers. The classing can be done by optimization techniques as described by Armstrong (2003). Pattern detection and building of an aggregation hierarchy is not necessary, since aggregation of the county polygons is not needed to meet the map purpose. Hence, the following generalization stage can be restricted to operations of smoothing and simplifying the polygon outlines, if at all necessary. In contrast to the previous example, the visualization stage would be fully applied.

4. TECHNICAL ASPECTS OF THE FRAMEWORK

To illustrate some of the roots of the proposed framework an association of the steps of Brassel and Weibel (1988) and the philosophical objectives (*why* to generalize) by McMaster and Shea (1992) has been made. Table 1 shows the different objectives for the three stages of our model. All the steps of Brassel and Weibel (1988) and all theoretical elements and application elements of McMaster and Shea (1992) can be assigned. This shows that the new framework inherits elements of both global frameworks.

Model	Stage 1 - Structure Analysis	Stage 2 - Generalization	Stage 3 - Visualization
Objectives and steps of Brassel and Weibel (1988)	- structure recognition (on class and group level)	- structure recognition (on object level) - process recognition (on generalization algorithms) - process modeling (generalization) - process execution	- structure recognition (information density, free space for symbolization and labeling) - process modeling (symbolization, labeling, masking) - process execution
Philosophical objectives of McMaster and Shea (1992) TE: theoretical element AE: application element	- TE: reducing complexity - TE: maintain logical hierarchy - AE: map purpose and intended audience - AE: retention of clarity	- TE: reducing complexity - TE: maintain spatial accuracy - TE: maintain attribute accuracy - AE: retention of clarity (- TE: consistent application of generalization algorithms)	- TE: reducing complexity - TE: maintain aesthetic quality - TE: maintain logical hierarchy - AE: retention of clarity

Table 1: Assignment of the steps of Brassel and Weibel (1988) and the philosophical objectives of McMaster and Shea (1992)

In the following some more technical aspects of the framework are discussed. Figure 4 shows options for generalization process modeling. The different types of process modeling have been mentioned briefly in Section 2.2 and are described in more detail in Harrie and Weibel (2005). Starting again with the **structure analysis** stage the input data are usually
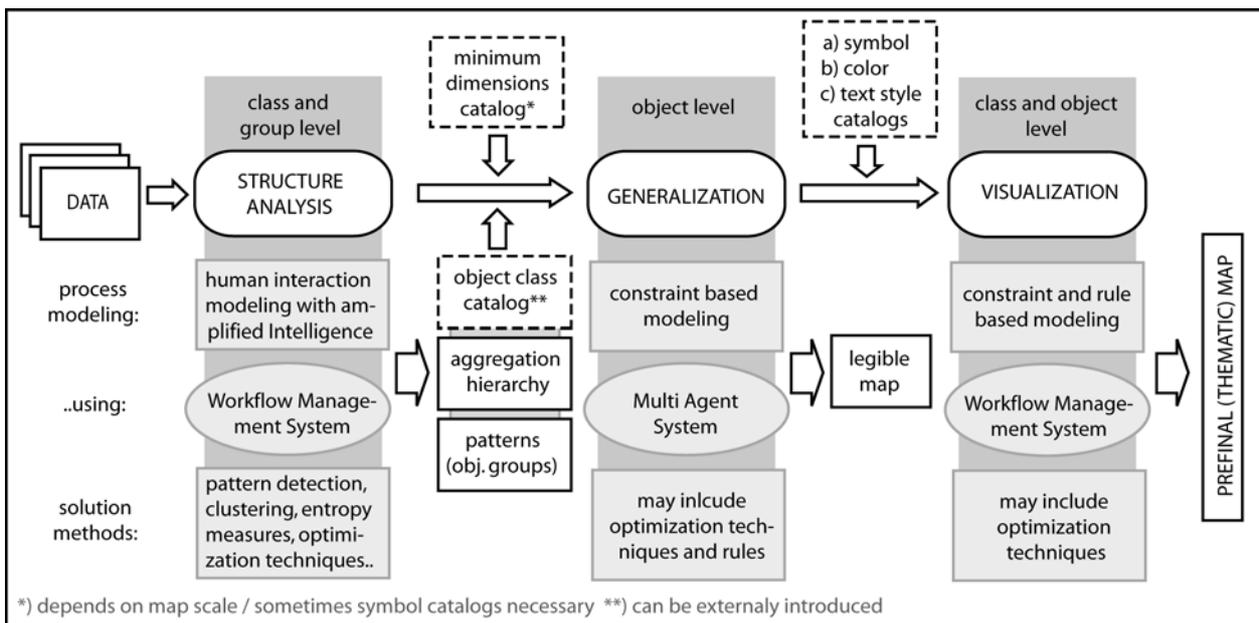

Figure 4: Proposed generalization process modelling options for the map generation stages.

of thematic type, complemented by topographic background information, e.g. administrative borders and rivers. Output data of this initial stage are an *object class catalog,* which contains information about the object class (topographic and thematic) to be retained as well as a reclassification scheme. Such information can be obtained from the analysis of the input data or by manual user input, but can also be dictated externally, e.g. by conventions. In the latter case, the catalog can be introduced after the first stage. Possible data analysis methods, used to obtain reclassification schemes, are optimization approaches (Armstrong, 2003) or clustering algorithms combined with entropy measures (Fuchs, 2002). If polygonal features are generalized, an *aggregation hierarchy* is necessary for the reclassification (e.g. merge 20 soil classes to 10) or to distribute the space of eliminated polygons among their neighbors. Both outputs (object class catalog and hierarchy) require expert interaction at least to choose among proposed aggregation schemes, resulting from data analysis, or validate the object class selection catalog. Hence, for the first stage the proposed type of generalization process modeling is *human interaction* with a focus on amplified intelligence (Weibel, 1991). *Flowchart modeling* techniques provide a flexible software approach to connect user interaction (decisions) and automation (e.g. to use the supporting data analysis techniques listed in Figure 4). The third output, information on detected *patterns* (cf. Fig. 4) should be confirmed by a domain expert if the patterns are of thematic nature.

The input and output data of the first stage are used and transformed in the **generalization stage**. Therefore an additional *minimum dimension catalog* is necessary, containing the legibility thresholds for minimum object sizes and separation distances with respect to the selected output scale. These values form the type B constraints and are used together with pattern and group constraints (type A) in a *Multi Agent System* (MAS) which performs the generalization process. How the iterative, agent-based generalization process works can be seen in Figure 2. After activation of an object or object group, now termed an agent, the agent characterizes itself against a set of constraints which act on the agent. The results of the characterization are hints on constraint violations, called generalization conflicts (e.g. an object is too small, is too close to another object, is too detailed, etc.). In the next step plans consisting of one or a set of algorithms to solve the conflicts are collected and ordered according to the weight of the affected constraint and severity of its violation. Now, the first of the ordered list of plans is executed and again an evaluation with respect to the constraints is made. If the evaluation result is positive since the conflicts are solved (= perfect state), then this state is stored and the agent returns to passive mode. If the conflict still exist either a new plan is chosen and executed or new plans are collected. This is repeated until a perfect state is reached or all plans have been tried. If the latter occurs the solution with the least severe constraint violations is chosen as the final state. The result obtained from the generalization stage is a legible map. To get a satisfactory solution it may be necessary for some map types to introduce already at this stage the final symbol dimensions (e.g. line widths) as constraints instead of simply using the minimum dimensions.

The legible map is passed on to the **visualization stage**, where the text labeling, the color and texture/pattern allocation and the legibility check is made. Therefore at least three types of catalogs are needed: a *text style catalog*, a *line style & point symbol catalog* and a *color & texture catalog* for the polygons. If no pre-defined catalogs exist definitions can be made on the fly by the use of a symbol brewer or color brewer (Schnabel, 2005, www.colorbrewer.org). If a color and texture catalog is created, several rules on coloring have to be considered. Some of them are listed in Brown and Feringa (2003) and could be realized as rules or constraints, such as the careful use of yellow for large areas. Thus, in the visualization stage, user interaction has to be combined with supporting algorithms (e.g. techniques for labeling; c.f. Petzold *et al.*, 2003) in a flexible way. This can best be achieved by the use of *flowchart / workflow modeling* techniques.

In conclusion it is obvious that in every stage **continuous decisions and evaluation** are made. In terms of process modeling we still need domain experts (e.g. soil scientists, geologists) in the first stage, where the specialist ensures that the purpose of the map is not missed. And we still need a (carto-)graphic expert, the cartographer, in the visualization stage to ensure a good communication of knowledge to the map user. Both stages have to combine interactivity with decision support algorithms, therefore the use of a workflow management system is proposed. The second stage is highly automated since we can formalize well knowledge on legibility constraints with respect to geometric minimum dimensions. Here, the automation may be achieved by the use of a multi agent system.

5. THEORETICAL USE CASES ON GEOLOGIC AND SOIL MAPS

Finally, we will discuss the application of the proposed framework to the generation of geologic maps and soil maps from GIS data. Whereas a range of literature exists on surveying and (field) mapping guidelines in geology and soil science (among others for geology: Schwarz *et al*, 2004; for soil: Rossiter, 2000), literature on automated map generalization in this fields is very sparse. The main reason is probably that in both fields national organizations currently focus on building up databases with digital (large scale) data. Thus, organizations are working on the data foundation for the proposed framework. An approach for semi-automatic derivation of soil maps of scale 1:100k and 1:200k from scale 1:50k has been described by Fuchs (2002). In geology, Artioli *et al.* (1995) made an attempt to automate generalization, which Downs and Mackaness (2002) comment with: *"Their encouraging results indicate that an automatic generalization procedure might be possible"*. Downs and Mackaness (2002) proposed a semi-automatic generalization approach, which is based on the Model of Brassel and Weibel (1988), developed together with the British Geological Survey. Drawing from this little literature and the visual analysis of maps the adapted frameworks for soil and geological map
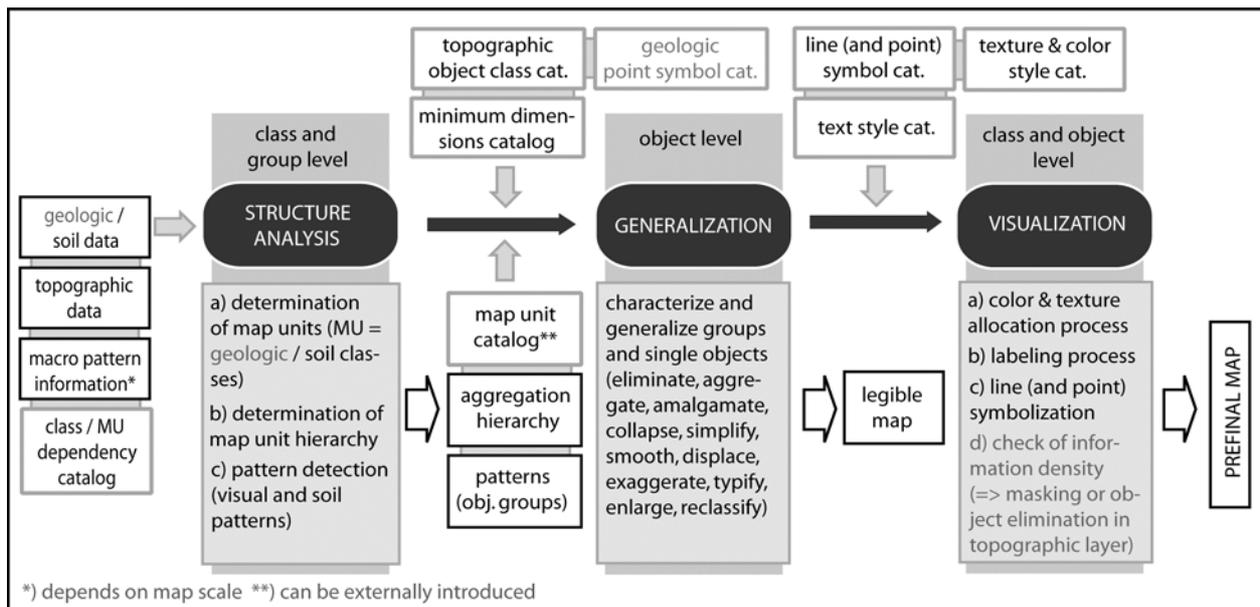
Figure 5: Proposed map generalization model for geological maps and soil maps. (Gray colored text applies to geology only).

generalization have been developed. Figure 5 shows both frameworks integrated into one model, due to the high degree of similarity. Note that text colored in gray applies to geological maps only.

The overall process starts with the definition of the **input data**, made up of thematic data and topographic base data. According to Steiniger and Weibel (2005) three other types of information have to be given. (1) Information on the existence of macro patterns is necessary if major scale changes should be accomplished. They describe the underlying subject specific model/structure, which the domain expert perceives in the raw map image, such as medium scale landscape structures in a large scale soil map. Furthermore, (2) class dependencies should be known (e.g. some soils might be intimately connected with others) as well as (3) inter-thematic dependencies (e.g. artificial 'soil' types induced by highway or railway tracks), which can be seen as thematic patterns. In the **structure analysis** module the number of map units and a map unit hierarchy must be defined. Also, visual and thematic patterns (here: groups of map objects) must be detected. The number of map units (categories) can be either determined by purpose (customer oriented soil survey) or conventions (taxonomies, soil series), or it is obtained from entropy measures through data analysis. Fuchs (2002) applied entropy, to measure the information (number of categories) per map area, and clustering, to obtain similarity information of the basic soil types. The definition of an aggregation hierarchy is closely connected to the previous task of defining the number of map units. Usually the results of data analysis are taken or conventions are used for this purpose. An example for the use of conventions in geology are categories by genesis, petrography and/or stratigraphy. In soil science the Soil Taxonomy, FAO System, or World Reference Base (cf. Rossiter, 2000) provide conventions for category formation.

In the **generalization** stage additional input is needed apart from the results of the first phase. A *topographic object class catalog* should be defined, since it is not necessary to use all information from topographic maps (e.g. camp sites, radio stations). Further, we need a *catalog of graphical minimal dimensions* (perceptual limits), which also can be adjusted to thematic map specifications like the minimum area size of soil patches (Fuchs, 2002). For geological maps the *point symbols* should be introduced at this stage. The reason is that symbols of certain thematic objects like boreholes, small moraines or springs should not overlap with the polygon boundaries after generalization. Also, note that polygons must be generalized with respect to point objects (particularly boreholes) to ensure that point symbols will still be inside the dedicated geological area. In contrast, for soil maps the *point symbol catalog* may only be introduced immediately before the third stage, if at all needed. Our visual analysis of existing maps showed that soil maps do not necessarily contain point information. With all the necessary information generalization can be performed as described in Section 4. Usually all the generalization operations (elimination, aggregation, amalgamation, collapse, etc.; cf. McMaster and Shea, 1992) will be applied.

The resulting data from the generalization stage are passed on to the **visualization** stage, accomplished by a number of catalogs. For instance in geology several local, national, and (informal) international standards have been established. Color standards for soil maps exist only for major soil groups, because dozens of different soil types are defined. Therefore, it would be useful to develop a '*color allocation system*', which creates map unit colors with respect to some constraints (traditions, soil hierarchy, overall area of map unit, output medium, perception). According to Rossiter (2002), in the USA no use of color and only labeling with symbols is made since the 1960s. Here, the visualization stage would consist of labeling process and line symbolization.

In geological maps a check for information density should be performed in addition to labeling, symbolization and color & texture allocation. An evaluation of geological maps of Switzerland of scale 1:25000 showed a very high complexity of the map image, more than that of comparable soil maps. These geological maps showed at least three information layers:

- the full topography (used as a base map, with some change in symbolization of forests),

- the lithology (polygon layer, visualized with colors *and* texture), and

- point symbols (moraines, springs, boreholes, etc.).

Additionally tectonic information (symbolized by lines) and text is possible. Thus, at certain places too many objects are visualized and information thinning is recommended. This can be done either by elimination of not really necessary (topographic) objects or by masking about the important information.

Finally, after the visualization stage, an interactive **post processing** step is necessary for unresolved conflicts of previous stages.

6. CONCLUSIONS

Research in map generalization has recently started to focus on other kinds of maps than on topographic maps only. Therefore, there is a need to develop new generalization concepts based on the experiences and research efforts previously made in topographic generalization. Following a review of the literature, as well as additional evaluation of sample thematic maps and rules of map production for soil and geological maps such a general framework for (thematic) map generalization has been proposed in this paper. This conceptual framework is organized into the three main stages of structure analysis, generalization, and visualization. The human is kept in the loop, and will be in future, in the stage of structure analysis, where the control of the domain expert ensures that the purpose of the map can be met. The human expert is also critically needed in the visualization stage, where the cartographer ensures a good communication of information. Separating the generalization and visualization stages is reasonable with a view to different output media (screen, web, paper) and user adaptation (user age, experience, etc.).

We have shown how the proposed conceptual framework can be applied to the generation of soil and geological maps. Other use cases have been made – but are not reported – for a political election map, a hazard map, model generalization of topographic data, and the production of a topographic map. They all could be matched to and expressed by the elements of the proposed framework. In some of the examples a proper separation of the generalization stage and visualization stage is hard to achieve, since symbols, with their extents, are sometimes needed to reach a satisfactory result in the generalization process (e.g. if displacement is induced). The need to introduce symbols (or their extents) in the generalization stage increases proportionally with the intended scale change.

An implementation of the framework in a comprehensive generalization system is still outstanding. However, on one hand parts like the major parts of the generalization stage, modeled with a multi agent system, are already available in commercial software and show encouraging results. And on the other hand, workflow modeling has shown its applicability in GIS and promises to be a flexible approach to chain together necessary interactivity with algorithmic processing tools also for generalization purposes. Thus, the implementation of the proposed framework seems to be feasible and future generalization research may draw on it if maps of different thematic domains should be generalized.

REFERENCES

ARTIOLI, G. P., G. CIARDI, M. L. GARBERI and L. MARTELLI; 1995: The digitised geological cartography of the Emilia Romagna Region to scales 1:10 000 and 1:50 000. The cartographic generalization process. *Proceedings of ICA Conference,* Barcelona, pp. 860-871.

ARMSTRONG, M. P.; 1991: Knowledge classification and organization. In: B. Buttenfield and R. McMaster (eds.), *Map generalization: making rules for knowledge representation.* Longman, London, pp. 86–102.

ARMSTRONG M. P., N. XIAO and D. A. BENNETT; 2003: Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Annals of the Ass. of the American Geographers*, Vol. 93, No. 3, pp. 595-623.

BARRAULT, M., N. REGNAULD, C. DUCHÊNE, K. HAIRE, C. BAEIJS, Y. DEMAZEAU, P. HARDY, W. MACKANESS, A. RUAS and R. WEIBEL; 2001: Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalization. *Proceedings of 20th ICA Conference*, Beijing, pp. 2110–2116.

BEARD, M.; 1991: Constraints on rule formation. In: B. Buttenfield and R. McMaster (eds.), *Map generalization: making rules for knowledge representation.* Longman, London, pp. 121–135.

BRASSEL, K. E. and R. WEIBEL; 1988: A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems*, Vol. 2, No. 3, pp. 229-244.

BROWN A. and W. FERINGA; 2003: *Colour basics for GIS users.* Pearson Education Ltd., Harlow, England.

DENT, B. D.; 1990: *Principles of thematic map design.* 2nd edition, Addison-Wesley, Wm. C. Brown Publishers.

DOWNS, T. C. and W. A. MACKANESS; 2002: An integrated approach to the generalization of geological maps. *The Cartographic Journal*, Vol. 39, No. 2, pp. 137-152.

FUCHS, M.; 2002: *Methoden zur objektiven Ableitung von Bodenkarten im Folgemaßstab.* Ph.D. thesis, Free University of Berlin.

GALANDA, M.; 2003: *Automated Polygon Generalization in a Multi Agent System.* Ph.D. thesis, Department of Geography, University of Zurich.

GRÜNREICH, D.; 1985: Ein Vorschlag zum Aufbau einer grossmassstäbigen topographisch-kartographischen Datenbank unter besonderer Berücksichtigung der Grundrissdatei des ALK-Systems. *Nachrichten aus dem Karten- und Vermessungswesen*, Vol. 1, No. 95, p. 55.

HARRIE, L. and T. SARJAKOSKI; 2002: Simultaneous graphic generalization of vector data sets. *GeoInformatica*, Vol. 6, No. 3, pp. 233 – 261.

HARRIE, L. and R. WEIBEL; 2005: Modelling the overall process of generalisation. In: W. Mackaness, A. Ruas and T. Sarjakoski (eds.), *Challenges in the Portrayal of Geographic Information: Issues of Generalisation and Multi Scale Representation.*

HOLLAND, J. H.; 1986: Escaping brittleness: the possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In: R. S. Michalski, J. G. Carbonell and T. M. Mitchell (eds.): *Machine Learning: an artificial intelligence approach.* Morgan Kaufman, Los Altos, pp. 593-623.

MACKANESS, W. A., 1995: A constraint based approach to human computer interaction in automated cartography. *Proceedings of ICA Conference*, Barcelona, Spain, pp. 1423-1432.

MCMASTER, R. B. and K. S. SHEA; 1992: *Generalization in digital cartography.* Association of American Geographers: Resource Publications in Geography. Washington, D.C..

MORGENSTERN, D. and D. SCHÜRER; 1999: A Concept for model generalization of digital landscape models from finer to coarser resolution. *Proceedings of ICA Conference*, Ottawa, Canada.

PETER, B. and R. WEIBEL; 1999: Using vector and raster-based techniques in categorical map generalization. *Third ICA Workshop on Progress in Automated Map Generalization*, Ottawa.

PETZOLD, I., G. GROEGER and L. PLUEMER; 2003: Fast screen map labelling – data structures and algorithms. *Proceedings of ICA Conference*, Durban, South Africa.

ROSSITER, D. G.; 2000: *Lecture Notes & Reference: Methodology for Soil Resource Inventories.* ITC Lecture Notes SOL.SRI. Enschede, the Netherlands: ITC. 2nd revised version. http://www.itc.nl/~rossiter/

RUAS, A., 1999a: *Modèle de généralisation de données géographiques à base de contraintes et d'autonomie.* Doctoral Thesis, Université de Marne-la-Vallée.

RUAS, A., 1999b: *Synthesis of a test on generalisation.* Report from OEEPE Working Group on Generalization, Draft version, OEEPE Publication in 2001.

RUAS, A. and C. PLAZANET; 1996: Strategies for automated generalization. In: M. J. Kraak and M. Mollenaar (eds.): *Advances in GIS research II* (Proceedings 7th International Symposium on Spatial Data Handling). Taylor & Francis, London, pp. 6.1-6.18.

SAHA, A., D. GANGULY, S. R. SENGUPTA, B. MUKHOPADHYHAY and N. HAZRA; 2003: Digital cartographic standards for 1:50,000 geologic map: some important considerations. *Map India Conference 2003*.

SCHNABEL, O.; 2005: Map Symbol Brewer – a new approach for a cartographic map symbol generator. *Proceedings of ICA Conference*, La Coruña, Spain.

SCHWARZ, C., L. KATZSCHMANN and K-H. RADZINSKI; 2004: *Guidelines for the compilation of geological maps – general principles.* E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

SESTER, M.; 2005: Optimizing approaches for generalization and data abstraction. Accepted for publication in: *International Journal of Geographic Information Science*.

SPIESS, E.; 2002: Map Compilation. In: R. W. Anson and F. J. Ormeling (eds.), *Basic cartography for students and technicians.* Elsevier Science Publishers, London, pp. 25-84.

STEINIGER, S. and S. MEIER; 2004: Snakes: a technique for line smoothing and displacement in map generalisation. *The 7th ICA/ACI Workshop on Generalisation and Multiple Representation*, Leicester (GB).

STEINIGER, S. and R. WEIBEL; 2005: Relations and structures in categorical maps. *The 8th ICA Workshop on Generalisation and Multiple Representation*, La Coruña.

WARE, M., N. THOMAS and C. JONES 2003: Resolving graphic conflict in scale reduced maps: Refining the Simulated Annealing technique. *Fifth Workshop on Progress in Automated Map Generalization*, Paris.

WEIBEL, R. 1991: Amplified Intelligence and Rule-Based Systems. In Buttenfield, B. P., and R. B. McMaster (eds), *Map Generalization: Making Rules for Knowledge Representation*, Longman Group, pp. 172-186.

CURRICULUM VITAE

Stefan Steiniger

Stefan Steiniger received a diploma degree in geodesy from the Technical University Dresden (Germany). As a minor he also studied cartography and obtained an undergraduate diploma. Since April 2004 he is research assistant and PhD student at the "Geographic Information Systems" division at the University of Zurich. The working title of his PhD thesis is "Data Enrichment for Control of Generalization Process". His current research interests are map generalization, data enrichment, pattern detection and process modeling.