Towards a formal classification of generalization operators

Theodor Foerster, Jantien Stoter & Barend Köbben
Geo-Information Processing Department (GIP)
International Institute for Geo-Information Science and Earth Observation (ITC)
Enschede, The Netherlands
{foerster,stoter,kobben}@itc.nl

Abstract

Classification of generalization operators is one of the challenges in current generalization research in order to provide generalization functionality meaningfully on the web. This paper proposes a classification framework for these operators by integrating the commonly accepted generalization model of Gruenreich with generic geo-data models (ISO 19109 & OGC GO-1 Application Objects). The operators that are classified in our framework are based on literature research and will be classified according to Gruenreich's model. The impact of the operators will be described by applying these geo-data models. Future research will show if the list of operators needs to be extended.

1 Introduction

Within the last decades, generalization research carried out different operator classifications based on their different characteristics as for instance within the Agent project (Lamy et al. 1999) or others (e.g. McMaster & Shea 1992; Yaolin et al. 2001). These classifications are project-driven, i.e. they are based on one specific application such as map production. Therefore these classifications do not aim to be general to serve any generalization application, nor do they aim to be consistent. The classifications are not transparent, as they cannot be reconstructed and are not based upon a formal model. Additionally they are incompatible to each other, as some classifications point out different operators than others. But they are also inconsistent internally, as they do not apply the same criteria for each of the operators.

However a comprehensive, unambiguous classification of operators is essential in the context of web-based generalization process. In this paper we describe a distinct set of operators, based on an inventory of existing operators extracted from published research on generalization, where we will try to be as complete as possible.

So the proposed set of operators is a result of harmonizing the different descriptions available in literature as well as the used operator names in literature. The operators will be classified based on the well-known and broadly-accepted model for generalization of Gruenreich (1992). We choose Gruenreich's model as it describes the generalization process completely. Gruenreich identifies two types of generalization, namely model generalization used to obtain a data model at a lower level of detail and cartographic generalization used to obtain a readable map at a certain scale taking cartographic constraints into account. Using these models allows us to classify the operators and to define the impact of the operators.

The impact of model generalization operators will be described using the formal *ISO 19109 General Feature Model* (ISO 2003). This commonly used model for GI-related data modeling (INSPIRE 2003) provides a means to describe the impact of model generalization in a generic way as will be shown in this paper. The impact of cartographic generalization operators will be defined using the *OGC GO-1 Application Objects* model, which describes a generic cartographic object model in an object-oriented way. The combination of all three models builds the *classification framework* for generalization operators (Figure 1) as proposed in our study.

Goals	**Impact**	**Operators**
Model Generalization	ISO 19109 General Feature Model	
Cartographic Generalization	OGC GO-1 Application Objects	

Figure 1: Classification framework for generalization operators.

Describing the impact of the operators in this classification framework on an object-oriented basis allows us to check the operators for consistency and to ensure their atomic nature. This is important as the determination of generalization operators is known to be highly subjective (Rieger & Coulson 1993).

Overall our approach towards the classification of the operators is top-down as it bases the definition of the operators upon generic data models. This top-down approach differs thereby from the common bottom-up approach for such operator classifications, based on map generalization studies. However the bottom-up approach did not result in any generic operator classification by now.

Section 2 will give a short overview of the terminology and will review the related literature about operator classification in generalization research. The paper will introduce the different components of the proposed classification framework (Section 3) and will then come up with the set of identified operators and link them into the classification framework (Section 4 & 5). Also a clear definition of the operators will be given and a distinct view on these operators is carried out. In Section 6 the paper will discuss the relation between model generalization and schema translation (Lehto 2007), as the output of both approaches are familiar and their operators may overlap or related to each other. Finally we will give an outlook to further research.

This paper contributes to generalization research, as it links the operators to the popular model of Gruenreich and as it introduces a way to present these operators described by two formal models in a consistent and transparent way. The classification provides a common understanding of the operators for both interactive and automated processes. Finally after the operators are defined consistently and linked to a formal model, the formalization of the operators is the next step.

2 Terminology & related research

Before introducing the classification of operators we want to define the terms *generalization operator*, *generalization algorithm* and *constraint*.

The idea of generalization operators evolved within the early generalization research by extracting abstract descriptions of single actions of the cartographer during manual generalization. Thus a generalization operator is an abstract description of atomic generalization functionality. It is atomic in the sense, that it only affects well-defined and isolated aspects of a feature in an undividable way. Nevertheless, being atomic does not imply that such functionality is not without any side effects.

Generalization operators have been identified as a key abstraction in order to compare and classify different generalization algorithms. An operator is thereby implemented by different algorithms. These algorithms are outside the scope of this paper.

Constraints have been introduced by Beard (1991) to replace complex rules for cartographic generalization. Constraints define the state, which is assigned to single and groups of cartographic objects and should be maintained or reached in order to produce a readable map. Weibel & Dutton (1999) state that model generalization is a formal process, which does not address any graphic aspects and includes thereby no constraints.

Regarding topology, model generalization can influence it and lead to topological errors in the produced data set. Thus topologic consistency is an important property (sometimes called a topologic constraint) in model generalization, which should be maintained as far as possible. However we see topologic consistency as a characteristic of the implementing algorithm or the applied data model: there are some algorithms and data models, which support topology checking[1] and some that not. So we do not consider topologic consistency in our framework.

2.1 Literature review on classification of operators

McMaster & Shea (1992) introduced a first classification of generalization operators, which consists of twelve operators and two categories. Their introduced categorization into spatial transformations and attribute transformations is trivial in the sense as it classifies classification and symbolization as the only attribute transformations and the others as spatial transformations. Additionally the classification does not seem to be sufficient to reflect the current aims of data production, as symbolization is mentioned as a generalization operator. However in current GI research the visualization and the data are separated to reduce complexity and avoid redundancy.

The Agent project (Lamy et al., 1999) focused on enhancing automated cartographic generalization for map production. The aim of the project was to develop a hierarchy of communicating objects (so called agents), which try to solve cartographic conflicts on the level of single features and groups of features. Thus they subdivided the operators in these two groups. However the classification does not consider model generalization, because it focuses on cartographic generalization.

The classification of Yaolin et al. (2001) aims at an object-oriented framework for model generalization operators but it mixes up the concepts of constraints (for cartographic generalization) and model generalization. Additionally some important operators for geometry type transformation are not covered such as combine and collapse (but which are covered by the other classifications). An overview of all these mentioned classifications is provided in Figure 2.

[1] For instance Saalfeld (1999) proposes a topology aware simplification algorithm.

McMaster & Shea	
spatial transformations	Simplification Amalgamation Refinement Displacement Smoothing Merging Exaggeration Aggregation Collapse Enhancement
attribute transformations	Symbolization Classification

Cecconi et al. (Agent)	
<unspecified>	Thematic selection Thematic aggregation
individual objects	Weeding Unrestricted simplification Enlargement Exaggeration Fractalization Smoothing Rectification
individual or groups of objects	Selection Elimination Displacement
groups of objects	Amalgamation Combine Typification

Yaolin et al.
Simplification Merge Amalgamation Aggregation Classification Selection

Figure 2: Overview of operator classifications merged and simplified after McMaster & Shea (1992), Cecconi (2003), Yaolin et al. (2001).

To be complete, we want to point to the recently published book by Li (2006), in which he includes a review of operators based on a geometry-oriented view.

3 The classification framework

As mentioned in the introduction the classification framework that we propose consists of three established models (Figure 1). The overall setting is given by the model of Gruenreich. It provides a comprehensive view on automated generalization as it separates the data from the maps and proposes a multi-stage generalization approach from reality to a dataset or to a map (Figure 3).

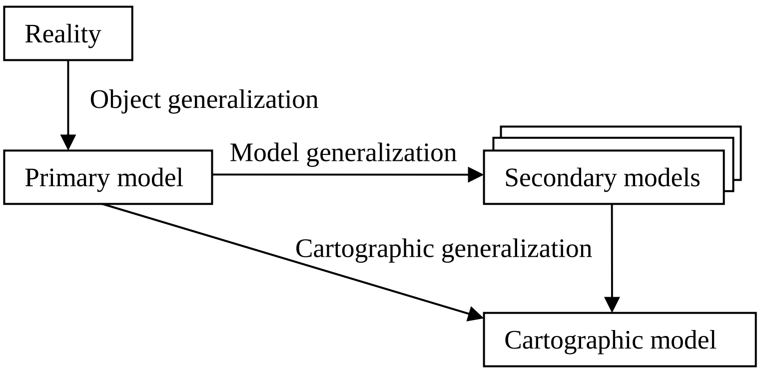

Figure 3: Generalization model of Gruenreich (1992).

This separation is most suitable for the current web-based and on-demand dissemination approach for data and maps of National Mapping Agencies. Other models do not provide such abstract view on generalization processing but describe more a fine-grain analysis of the logical and sequential (McMaster & Shea 1992) or philosophic (Brassel and Weibel 1988) aspects of generalization processing. It is important to note, that model generalization might be a pre-process for cartographic generalization, in which the user model for the visualization will be derived. The cartographic generalization is then applied upon the already symbolized map data to satisfy the constraints. So we consider to have symbolization as a pre-process of cartographic generalization as well (see Section 5). According to the Gruenreich model we classify the operators into *model generalization operators* and *cartographic generalization operators*.

The General Feature Model (ISO 19109) specifies the generic object-oriented structure of feature types, their properties and their interrelations (Figure 4). It provides comprehensive guidelines how to model geographic phenomena by linking especially the spatial schema, which specifies the geometric and topologic model (ISO 19107).

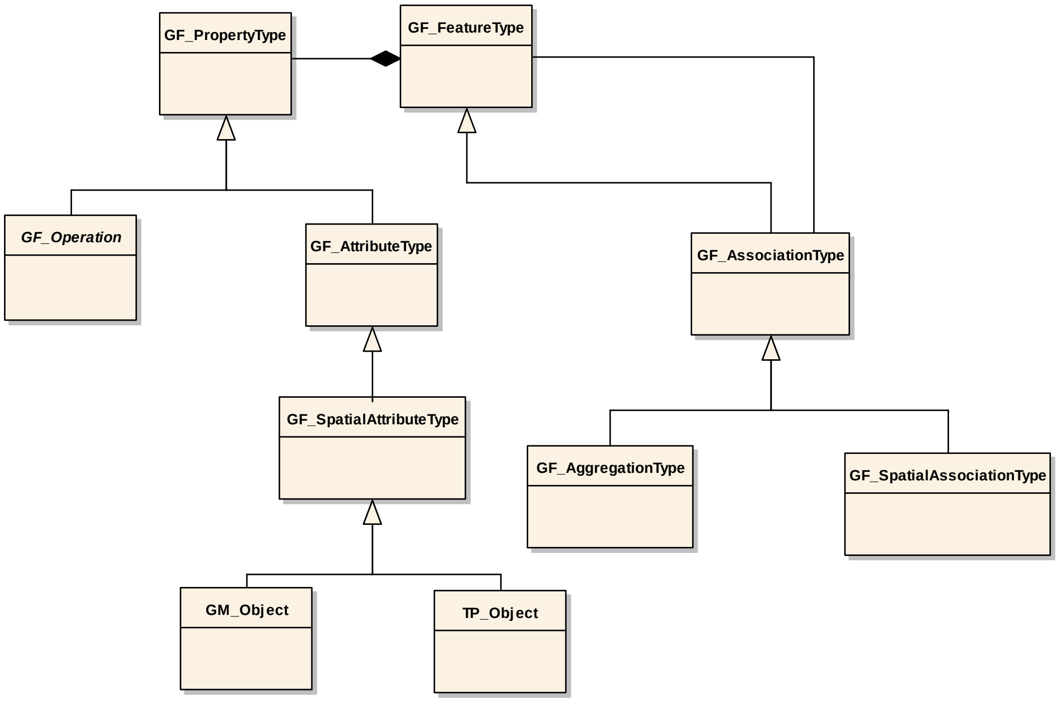

Figure 4: Overview of the ISO 19109 model.

The OGC GO-1 Application Objects model specifies an object-oriented view on graphic objects such as cartographic objects and is the basis for the definition of the cartographic generalization operators. It consists of three types of graphic representations for a 2-D graphic environment `GraphicLineString`, `GraphicPolygon` and `GraphicPoint` (Figure 5). Additionally the types have attached a certain `GraphicStyle` and sometimes a `Symbology`. For cartographic generalization purposes these properties are immutable (Section 5).

Each of the graphic representation types are indirectly linked via the ISO 19107 (geometry model) to the General Feature Model, upon which the model generalization operators are being defined by our approach.

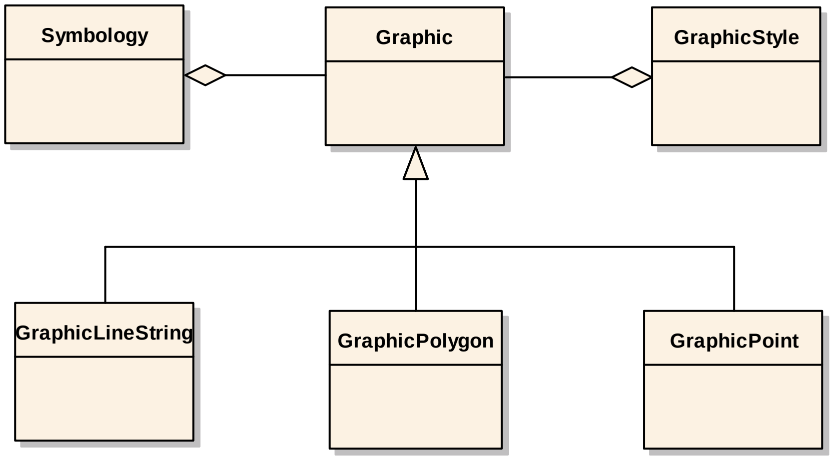

Figure 5: Overview of the OGC GO-1 Application Objects model.

3.1 Proposed classification

To build the classification each of the operators has to be positioned in the model of Gruenreich. The operators we propose are based on literature review (Section 2.1). The basic criteria to decide about its primary affiliation in the Gruenreich model is, if an operator can be defined and applied homogeneously on a feature type or feature instance level. An operator is only applied individually on the feature instance level, if a conflict has to be solved. This is only the case in cartographic generalization. Model generalization operators are applied globally upon a dataset. Cartographic generalization operators are guided by globally defined constraints, but are applied individually upon a group or single instances of features. However it is important to note, that in case of both operator types their impact is always local.

Our proposed classification is summarized in Table 1. The following two sections will explain the different operators using the General Feature Model and the OGC GO-1 Application Objects model.

Model generalization	**Cartographic generalization**
Class Selection	Enhancement
Reclassification	Displacement
Collapse	Elimination
Combine	Typification
Simplification	
Amalgamation	

Table 1: Operator affiliation to Gruenreich model.

4 Model generalization operators

It is important to note, that some of the proposed operators (reclassification & class selection) do not inherit any spatial aspect, but are assignable to rules, which are defined upon a (derived) spatial attribute. These operators are incorporated in all previous proposed operator classifications. Additionally we want to mention, that the term “spatial” explicitly includes topological objects (such as edge and node) as also described in the General Feature Model (Figure 4). Thus the operators could also be defined for such topological types (`TP_Object`). Covering such topological types in generalization is in line with the current approach of National Mapping Agencies to provide a comprehensive product line of datasets for various purposes (Lawrence 2004), and goes beyond the traditional goal of generalization for map production.

In the following paragraphs we want to lay down our definitions of operators for model generalization. For each operator, we define its impact by referring to the General Feature Model classes (Figure 4) indicated by `typewriter` font.

Class Selection

This operator selects the specific instances of a specific feature type, which should appear in the target data model. It also includes some filtering of the feature type properties according to the target data model (such as a database query). However it does not influence the feature type hierarchy such as reclassification. A popular example of class selection could be: select all the features with a geometry part of a specific geometric pattern. This operator is closely related to schema translation, as if there would no spatial rules involved in the class selection, it would be just equal to the filtering operator of schema translation specified by Lehto (2007). It is important to note, that this class selection has no impact on the spatial attribute of the feature itself.

- Selecting features according to the specified `FeatureType` and `AttributeTypes` as specified in the target model. To this class selection are certain rules attached specifying (derived) `SpatialAttributeType` or/and (derived) `SpatialAssociationType` which have to be fulfilled according to the target data model.

Reclassification[2]

[2] Sometimes reclassification is also called classification. However we prefer the word reclassification as this name reflects that this operator is always based upon an existing data model.

This is an elementary generalization operator, but it does not address spatial aspects by definition (i.e. has no impact on the geometric attribute). However it is an important operator, as it can cast certain instances of features to become member of other feature types according to the target data model, based on derived spatial characteristics. Additionally it can change the attributes of features according to the target model. Reclassification drives or is followed by operators such as amalgamation, combine and collapse, because they can reflect the reclassification also for the geometric attributes (amalgamation) and change the geometric attribute according to the target data model (combine and collapse). Also this operator has an equivalent in schema translation, if the reclassification does not base or require any transformation of the geometric attributes.

- Casting a group of features (`AggregationType`) according to a specific rule (involving `SpatialAttributeType` or `SpatialAssociationType`)to another `FeatureType`.

Collapse

This is a highly complex operator which involves spatial aspects. It is triggered by reclassification, if the target data model specifies a feature type, which has a spatial attribute with a decreased dimensionality (i.e. requires collapsing the geometry from polygon to line or to point).

- Changing the `GM_Object` of a `FeatureType` from polygon to line or to point.

Combine

Combining a group of features with lower dimensionality to one feature with higher dimensionality has a heavily invasive impact, which not only changes the attribute type, but also goes along with a change of the feature type as well. Combine is thereby the result of a reclassification, in which the geometric attribute type of the object is changed. For example reclassify sites of type *Leisure* (modeled as point) to feature type *tourist attraction (modeled as area)*. As the type of the geometric attribute of the feature has been changed, combine is involved. This operator is related to amalgamation, but it is more invasive, as it has to create a new geometric type based upon the geometric attributes of the original features. We separate that operator from amalgamation also according to the literature.

- Based on the `SpatialAssociationType` a set of features of the same feature type and same type of `GM_Object` (mostly of type Point) will be combined to a new feature with a new GM_Object of a higher dimensionality (mostly of type Polygon).

Amalgamation

This is a special operator, as it can be applied globally upon feature type level (model generalization) and locally upon a group of features (cartographic generalization). It is about amalgamating a group of spatially adjacent geometries (of the same geometric type and member of the same feature type) into a single geometry. This operator constructs a new outline boundary for the new geometry. In the context of model generalization it mostly goes in line with reclassification, as the geometric attribute should also reflect the applied classification. So for instance several adjacent forests of different type (e.g. coniferous & deciduous) are reclassified to forest area, it is necessary to amalgamate the geometries of the original features to a new geometry and assign this geometry to the reclassified feature. We do not make a distinction between amalgamating connected (Fusion) or non-connected features (Merge), as this is highly dependent upon the data situation and the data model.

- Based on GM_Object of the original features of the same FeatureType sharing a certain SpatialAssociationType a new GM_Object will be generated.

Simplification

This is an operator, which is also used to reduce the amount of data. However as modeling may aim at reducing the data volume, we suggest to keep it as a model generalization operator. Simplification is a not that invasive upon the feature, because it only deletes aspects of a geometry based on a certain criteria.

- Based on a certain geometric threshold function several Points of a GM_Object will be deleted, however the remaining feature will not be modified (as defined in Saalfeld (1999)).

Looking at the structure of the General Feature Model, a spatial collapse might be interesting to be modeled as an additional model generalization operator, which would be different from the normal collapse. Therefore we want to extend the classification by adding a spatial collapse which transforms any GM_Object to TP_Object (e.g. transforming roads consisting of lines to a road network consisting of edges and nodes). Also the special case of network simplification by weeding out unimportant topological aspects is not classified in the literature. Such operators did not appear, as model generalization has been used mostly as a pre-processing step for cartographic generalization.

The proposed classification shows, that the model generalization operators are dealing with spatial aspects (either as `SpatialAttributeType` or `SpatialAssociationType`) but to fully transform data from a target to a source model, *schema translation* mechanisms have to be applied. Merging of several non-spatial features is necessary when a group of features to one feature of another feature type has to be classified. So besides amalgamation you would apply a *schema translation merge* as described in Lehto (2007). Model generalization and schema transformation are complementary in order to perform geo-data transformation. The line between schema transformation and model generalization is quite vague. A generalization operator is always about to analyze derived attributes related to geometry (e.g. shape classification, topological relations) and to transform it. Schema translation operators are related to the analysis of already existing (mostly non-spatial) attributes and their transformation. So it seems always a question of deriving (model generalization) and matching (schema translation) attributes.

5 Cartographic generalization operators

In order to motivate the separation of the operators, we want to lay down our view on cartographic generalization and its embedding in the map production process. Cartographic generalization aims at reducing cartographic conflicts. These conflicts are caused by the map position of the cartographic objects and their applied symbology and styling. Such conflicts can be addressed by constraints that describe the conflicts, that should be avoided. Note, that cartographic generalization is applied after symbolization (Figure 6).

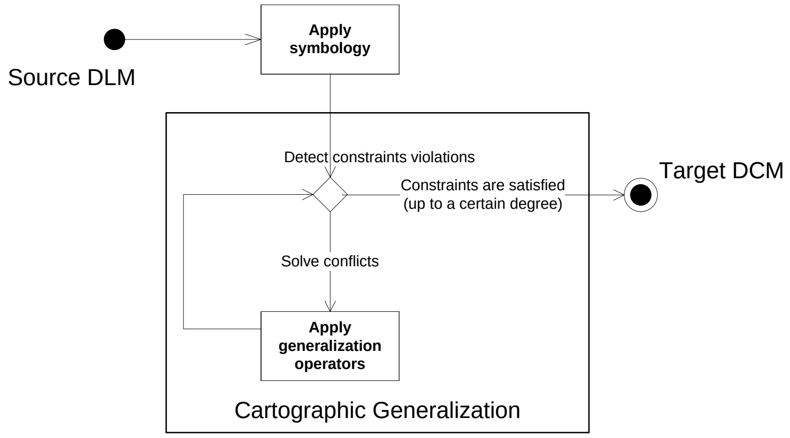

Figure 6: Cartographic generalization process.

This is an important aspect, because it prevents a lot of operators of model generalization applied being in this context, such as classification. Cartographic generalization changes the position, alters the shape or amalgamates cartographic objects.

We are aware of the fact, that cartographic conflicts can be solved by applying various sequences of operators. So it is impossible to assign types of constraints to specific operators in general, because it depends on the data situation (and also on the implementing algorithm), which operator is applicable. Anyhow it is possible to define a prioritized list of operators attached to each constraint for a specific data situation (Ruas, 1998). Looking at the description of operators, we can define their impact on the different cartographic objects. As an object representation model of cartographic objects we utilize the GO-1 Application Objects specification (Figure 5).

Enhancement

This operator modifies specific geometric parts of a graphic object to produce a pleasing representation or to emphasize an object. This includes smoothing of lines, squaring of building or enlarging/exaggerating of features. So on an object-level such enhancements modify specific parts of the geometry (i.e. set of coordinates).

Displacement

This operator moves the complete graphic object by applying the same vector to each part of the graphic. The final result is an object with a changed location but still preserving the original shape, also in absolute terms.

Elimination

The operator removes the graphic object from the map display. This operator is somehow the equivalent of the class selection operator, as both operators result in a set with a reduced number of objects. However their level of definition and application is different. Elimination is performed upon a feature instance level and not on a global level such as class selection.

Typification

This is a combined operator, which is highly complex. It replaces a set of graphic objects with a smaller set of graphic objects. The operator has to determine the applicable set of new graphic objects and then arrange them in a pleasing way. We are aware of the fact, that this is not atomic in the original sense, but it is impossible to separate the operator in an appropriate way, as the actions applied are highly depending on each other and have to be performed as a whole. Also generalization literature has identified this operator as a separate one.

Amalgamation

This operator does the same as for model generalization, but just on the cartographic object level. It merges different graphic objects (representing the same feature type) to one, by preserving the original shape of the outer geometries.

Our classification of operators highly depends on the granularity of the applied model. In case of the General Feature Model it seemed to be appropriate (as it even include it topological aspects), but in case of the OGC GO-1 Application Objects model it turned out to be too coarse grain. For instance it could be interesting to separate the enhancement operator. However this is impossible as such fine-grain aspects were not be reflected. So in order to have a consistent view on the operators we decided to put them together, as they are not distinguishable based on the current available cartographic models. This is a problem, which has been stated Muller et al. (1995).

6 Conclusion

The proposed classification framework provides a first attempt for a comprehensive mechanism to classify the operators and describe their impact appropriately. The overall setting by Gruenreich's model covers the complete generalization process and allows thereby classifying all possible generalization operators. The applied models connected to model and cartographic generalization provide consistency and transparency for the definition of the impact of the operators and allow demonstrating the operator's atomic nature. However the cartographic model does not provide a sufficient mean to reflect all characteristics of cartographic generalization as shown by the example of the enhancement operator.

As the classification is based upon such commonly used models we assume that it is easy to adopt it and integrate the operator descriptions into existing generalization applications of NMAs. This allows validating the classification with the current practice of NMAs.

In order to formalize the generalization operators, the descriptions based on the applied object models provide a first step. However the mathematical and formal description has still to be done.

Also of high interest is to investigate the link to schema translation. Linking schema translation and model generalization is an interesting approach as certain operators seem to depend on each other or are overlapping. To fully transform a dataset from a source to a target data model both types of schema translation and model generalization operators are necessary. However establishing the link and identifying the dependencies between both types of operators is also part of future research.

Acknowledgements

We would like to thank the partners of the DURP consortium and the people of COSIT laboratory of IGN France, who gave fruitful feedback on our ideas and thereby improved the value of this paper.

References

Beard, K. M. (1991), Map Generalization: Making Rules for Knowledge Representation Constraints on rule formation, In: *Map Generalization: Making Decisions for Knowledge Representation*, Buttenfield, B. & Mcmaster, R. *(ed.), Longmans*, 121-135.

Brassel, K. E. & Weibel, R. (1988), A review and conceptual framework of automated map generalization *International Journal Of Geographical Information Systems*, 1988*, 2,* 229-244.

Cecconi, A. (2003), Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping*, University of Zurich.*

Gruenreich, D. (1992), ATKIS - a topographic information system as a basis for GIS and digital cartography in Germany *From digital map series to geo-information systems*, 1992, *Geologisches Jahrbuch Reihe A*, 207 - 16.

INSPIRE (2002), INSPIRE Architecture and Standards Position Paper, *JRC - Institute for Environment and Sustainability, Italy.*

ISO (2003), Geographic information - Rules for application schema *ISO.*

Lamy, S.; Ruas, A.; Demazeu, Y.; Jackson, M.; Mackaness, W. & Weibel, R. (1999), The Application of Agents in Automated Map Generalization, *19th International Cartographic Conference*.

Lawrence, V. (2004), The Role of National Mapping Organizations *The Cartographic Journal*, *41*, 117-122.

Lehto, L. (2007), Schema Translations in a Web Service Based SDI, Wachowicz, M. & Bodum, L. *(ed.), 10th Agile International Conference on Geographic Information Science 2007.*

Li, Z. (2006), Algorithmic foundation of multi - scale spatial representation, *CRC.*

McMaster, R.B. & Shea, S.K. (1992), Generalization in Cartography.

Mueller, J. C.; Weibel, R.; Lagrange, J. P. & Salge (1995), Generalization: state of the art and issues, In: *GIS and generalization*, F. Mueller, J. C.; Lagrange, J. P. & Weibel, R. *(ed.), Taylor & Francis.*

Rieger, M.K. & Coulson, M.R.C. (1993), Consensus or confusion: cartographers' Knowledge of Generalization, *Cartographica*, *30*, 69-80.

Ruas, A. (1998), OO-constraint modelling to automate urban generalization process, *SDH 98*.

Saalfeld, A. (1999), Topologically Consistent Line Simplification with the Douglas-Peucker Algorithm, *Cartography and Geographic Information Science*, *26*, 7-18.

Yaolin, L.; Molenaar, M. & Tinghua, A., D. (2001), Frameworks for Generalization Constraints and Operations Based on Object-Oriented Data Structure in Database Generalization, *20th ICC*, Du, H.L. *(ed.)*, *3*, 2000-12.