The Cartographic Journal Vol. 47 No. 1 pp. 41–54 February 2010

REFEREED PAPER

Challenges for Automated Generalisation at European Mapping Agencies: A Qualitative and Quantitative Analysis

Theodor Foerster[1], *Jantien Stoter*[2] *and Menno-Jan Kraak*[1]

[1]International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands. [2]Section GISt, OTB, TU Delft & Kadaster, Apeldoorn, The Netherlands
Email: foerster@itc.nl

The automation of generalisation is an important issue at National Mapping Agencies (NMAs) to reduce data production costs and to improve data maintenance. This paper presents the challenges for automated generalisation at European NMAs integrating a qualitative and quantitative analysis. The qualitative analysis focuses on the current strategies for automated generalisation at NMAs. The quantitative analysis extends these findings and measures the status of automated generalisation functionality at NMAs using the required and missing generalisation operators as indicators. The results are interesting for the research community, the software vendors and NMAs to streamline their efforts to accomplish full automation of generalisation processes.

Keywords: automated generalisation, National Mapping Agencies, requirements of NMAs, generalisation functionality, generalisation operators

INTRODUCTION

Producing and maintaining topographic data is one of the main responsibilities of National Mapping Agencies (NMAs). Within this context, automated generalisation is an important goal for NMAs to increase efficiency of data production at multiple scales and to enable customized data products. Ideally automated generalisation is the way to implement the single scale database and to improve thereby the maintenance of all data products.

However, automated generalisation is still mostly subject to research and only specific research results have found their way into practice as for instance shown by Lecordix *et al.* (2007) and Regnauld and Revell (2007). This is due to the data complexity of complete datasets (in contrast to selecting specific layers from datasets as usually studied in research), incompatibility of data models, even within the same NMA, lack of generic view on NMA requirements and limited processing facilities.

The objective of the research presented in this paper is to get an insight into the limitations that NMAs encounter with respect to automated generalisation. This will identify areas for future research and for software developments regarding automated generalisation. To meet this objective, an integrated qualitative and quantitative analysis has been carried out, which identifies the urgent and fundamental problems of NMAs' daily work related to automated generalisation. The analysis updates and integrates the work of Stoter (2005a, 2005b) and Foerster and Stoter (2008).

Stoter (2005a, 2005b) reports on the results of a workshop that took place in 2005 and was attended by twelve European NMAs. The workshop focused on two questions: what are the trends and policies for automated generalisation within NMAs, and which topics need further study to improve generalisation processes within NMAs? Although the workshop took place a few years ago, the participants of the workshop updated the findings at the end of 2008.

The second research source for this paper is Foerster and Stoter (2008). This study presents preliminary results of a survey that was conducted at the end of 2007 and completed by 11 NMAs from eight European countries. The survey aimed at capturing a common view of NMAs on missing generalisation functionality. While the workshop provided initial findings about the NMA perspective on automated generalisation, the survey tested and quantified these findings.

The NMAs that participated in both the workshop and the survey are considered to be representative of NMAs that have good quality data available and that already have seriously considered the automation of generalisation within their processes. Some of them have made major steps towards full automation as will be demonstrated in this paper. Consequently by concentrating on this focus group, the research aimed at exposing outstanding generalisation problems. The result is an integrated analysis on the challenges of automated generalisation at European NMAs.

The paper will first provide an overview of previous research on challenges of automated generalisation for NMAs (the section on 'LITERATURE REVIEW'). The section on 'QUALITATIVE ANALYSIS OF CHALLENGES

DOI: 10.1179/000870409X12525737905123

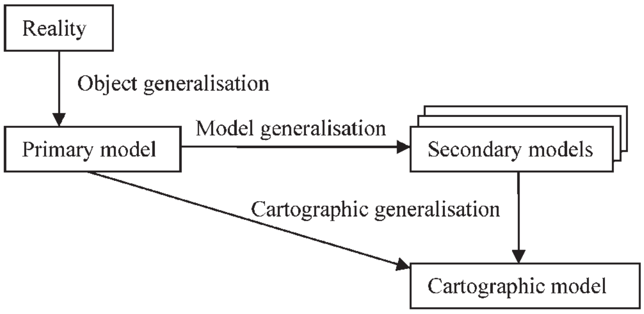

Figure 1. Generalisation model of Gruenreich (1992)

FOR AUTOMATED GENERALISATION' presents the challenges for automated generalisation within NMAs based on the workshop findings. An important outcome is the need for generic applicable generalisation functionality. This requires a common view on what this generalisation functionality should encompass. The quantitative analysis of the survey presented in the section on 'QUANTITATIVE ANALYSIS OF RELEVANT GENERALISATION OPERATORS FOR MAP PRODUCTION' provides such a common view by concentrating on generalisation operators. It studies how specific operators are implemented in current map production with respect to importance and problematic (i.e. lacking) characteristics. In a second step, operators are analysed with respect to the importance of feature types on which they are applied. The paper ends with a conclusion that discusses the findings in relation to challenges for automated generalisation as recently identified by Mackaness *et al.* (2007b).

It should be noted that this paper specifically focuses on datasets that are maintained by NMAs in order to represent topography. Consequently, datasets such as cadastral data (in some countries also maintained by NMAs), orthophotos and digital terrain models are not considered.

LITERATURE REVIEW

Research in the area of automated generalisation has yielded a lot of concepts and applications in the last 20 years. A good overview can be found in McMaster and Shea (1992), in Weibel and Dutton (1999) or in the most recently published book of Mackaness *et al.* (2007a). Different views on generalisation have been developed, such as the generalisation model by Gruenreich (1992), which separates generalisation into model generalisation and cartographic generalisation (Figure 1). Model generalisation is concerned with the transformation of data according to a target model and cartographic generalisation aims at producing usable maps out of data while avoiding cartographic conflicts.

Besides the concepts for automated generalisation, different initiatives studied the challenges of automated generalisation for NMAs empirically. For instance, Muller and Mouwes (1990) studied existing map series to get an insight into challenges for automated generalisation. They identified two types of generalisation knowledge to be automated: superficial knowledge and deep knowledge. Superficial knowledge is written down in map specifications meant for interactive generalisation. Deep knowledge is more important and much more complex to automate and is used by cartographers when superficial knowledge does not suffice. Rieger and Coulson (1993) carried out a survey among a group of cartographers performing interactive generalisation and found that the classification of generalisation operators differs depending on the specific cartographer. Additionally, they discovered that a consensus on such a classification does not exist. Besides, those studies several interviews were carried out with experts to learn more about requirements for automated generalisation. Examples are McGraw and Harbison-Briggs (1989), Nickerson (1991) and Kilpelaïnen (2000). Also examples of reverse engineering are available aiming to collect generalisation knowledge from comparing map objects across scales (Buttenfield, 1991; Leitner and Buttenfield, 1995; Weibel, 1995). Other studies generated rules from interactive generalisation carried out by a cartographic expert (Weibel, 1991; McMaster, 1995; Reichenbacher, 1995; Weibel *et al.*, 1995). Several studies applied machine learning techniques to convert expert knowledge into specifications. Examples are Weibel *et al.* (1995), Plazanet *et al.* (1998), Mustiere (2001, 2005) and Hubert and Ruas (2003). Ruas (2001) investigated within the OEEPE project the state-of-the-art of generalisation by evaluating different interactive generalisation software packages. The tests performed within this project were specific to cartographic conflicts, generalisation operations and algorithms, and some test datasets. Brewer and Buttenfield (2007) ran map exercises with students on different datasets at various scales. The results of the exercises were compiled to the so called ScaleMaster, which provides guidelines for generalisation processes. Within the currently ongoing EuroSDR research project on the state-of-the-art of generalisation, the capabilities of several generalisation systems are being tested on a selection of test cases (Stoter *et al.*, 2009).

QUALITATIVE ANALYSIS OF CHALLENGES FOR AUTOMATED GENERALISATION

The challenges of NMAs towards automated generalisation were analysed in a 2 day generalisation workshop organized in April 2005 at the International Institute for Geo-information Science and Earth Observation, Enschede, The Netherlands. The workshop studied the following two questions:

1. What are the trends and policies on automated generalisation within NMAs?
2. Which topics require further research?

This section describes the most important results of the workshop. The section on 'Trends and policies on automated generalisation' addresses the first question and summarizes the generalisation process within the participating NMAs. A more extensive description of the workshop findings can be found in Stoter (2005a, 2005b). The section on 'Necessary research to improve current generalisation practice' answers the second workshop question and identifies topics that need further research to better serve practice. As mentioned before, the results of the workshop were updated by the participants at the end of 2008.

Trends and policies on automated generalisation

All participating NMAs maintain vector datasets at different scales to support their production processes. Either one, seamless database is maintained per scale or several databases are maintained for one scale based on (old) map sheets (one database per map sheet). All participating NMAs recognize the importance to introduce automated generalisation (or at least as automated as possible). Some NMAs have made more fundamental steps towards automated generalisation than others. This section describes the status of automated generalisation within NMAs by addressing the four main steps introducing automated generalisation:

1. Renewing data models
2. Designing the conceptual architecture
3. Implementing generalisation processes
4. Managing relationships between different scales

The status of every step for the individual NMAs are summarized in Table 1 and will be further explained in this section using representative examples.

Renewing data models

As mentioned in the introduction, incompatible data models cause difficulties with respect to automated generalisation. Therefore, the first step towards automated generalisation is restructuring existing datasets into datasets compliant to data models that meet today's requirements of base datasets. Example requirements are data delivery within a spatial data infrastructure, history management, unique IDs and object-oriented datasets. This step has been taken by all participating NMAs. For the base datasets, new data models have been designed. For the smaller scales, data models are being restructured to make them compatible with the base dataset. Examples are the Dutch multi-scale information model topography (IMTOP) covering scales from 1 : 10k to 1 : 1000k (Stoter *et al.*, 2008) and the Danish multi-scale GeoDB data model that contains renewed data models for scales 1 : 50k and 1 : 100k. Another example is Ireland who developed new conceptual model covering all scales. A prototype dataset was reengineered to evaluate this model. The next step in Ireland is to reengineer all data to the new model, and to create a new production flow line for all large and small scale products (paper and digital, vector and raster). This will result in a single object-oriented database, which will be the source for all map products, irrespective of scale.

Some NMAs maintain a base dataset with varying scale. An example is the OS MasterMap product of Great Britain, a seamless topographic database for which the data have been collected at 1 : 1.25k in urban areas, 1 : 2.5k in rural areas and 1 : 10k in mountain and moorland areas.

Designing the conceptual architecture

The second step to enable automated generalisation is to design the conceptual architecture of the (automated) generalisation process. The main question for the conceptual architecture are: what is the intended source and target datasets of the generalisation processes?

The approach that is followed by all NMAs is to convert available datasets into datasets compliant with the new data models. After this step, the smaller scales are updated by generalizing the updates from the base dataset. Consequently, generalisation within NMAs focuses on generalisation of updates and data matching to make updating more efficient.

In some NMAs, the smaller scale datasets are newer than the base dataset due to different update cycles. In such cases, the smaller scales are updated independently of the base dataset until the update cycles are harmonized. Examples are 1 : 100k dataset in Denmark, 1 : 50k dataset in Belgium until 2005 (Fechir and Waele, 2005) and 1 : 100k dataset in The Netherlands.

An important decision concerning the conceptual architecture is whether to follow the ladder approach or star approach (EuroGeographics, 2005). In the ladder approach (followed by Denmark, Belgium, Germany, Sweden and The Netherlands), the (updates of) smaller scales are derived from (the updates from) a large scale dataset in steps (scale by scale). For example, Denmark applies the ladder approach for generalizing 1 : 50k and 1 : 100k datasets. The 1 : 200k and 1 : 250k datasets of Denmark are still independently processed. The alternative is the star approach in which every small scale dataset is generalized from the same base dataset. France, Switzerland and Catalonia have chosen for a mixture of both. In the mixed approach, the large to mid-scale datasets are derived from the base dataset while smaller scales are derived from one mid-scale dataset. Ireland considers the ladder model for defined products (both large- and small-scale) and the star model for a Web service-oriented dissemination approach. Great Britain has still to decide on which approach to follow. An important observation is that the scales that are produced and the scale transitions, at which generalisation is applied, differ among all NMAs (see the section on 'Method to quantify problems of automated generalisation of NMAs').

Another important decision for the conceptual architecture is whether to distinguish between model generalisation and cartographic generalisation. NMAs such as Denmark, Lower Saxony in Germany, Sweden and The Netherlands argue that if users are interested in geometries close to reality, they should use a large scale dataset with topographical precision instead of geometries that have been modified (e.g. displaced) based on their symbolisation in the map. Therefore, they do not distinguish between model and cartographic generalisation, but apply both types of generalisation in one process. The other NMAs consider some operators such as displacing of objects only appropriate when producing a readable map and therefore, they do distinguish between model and cartographic generalisation. However, within this last group, there is no consensus on which operator belongs to which process.

Implementing generalisation processes

The third step for introducing automated generalisation in production lines is the implementation. In all participating NMAs, specific generalisation operations have been automated. However, full automated generalisation processes

Table 1. Analysis of the four main steps to introduce automated generalisation (for the definition of the star and ladder approach please refer to the following paragraph)

	Data models			Conceptual architecture		Implementation					Relationships between different scales
NMA	Renewed data model base dataset	Renewed data model smaller scales	Varying scale for base dataset	Star (S) or ladder (L) approach	Distinguishing between model and cartographic generalisation	Update process indepen-dent of larger scale dataset	Generalisation of whole dataset once, then generalisa-tion of updates	Considerable effort has been put in automated generalisation	Commercial software used	Self-developed algorithms used	Maintenance of relationships between scales
Belgium	Yes	Yes		L	No	1 : 50k			ArcGIS and Lamps2/Agent	Yes	Is being studied
Catalonia	Yes	Yes		Mix	Mix[†]	1 : 50k	1 : 25k db	Yes	CHANGE	Yes	
Denmark	Yes	Yes		Mix	No	1 : 100k	1 : 50k	Yes	Lamps2/Agent	Yes	
France	Yes	Yes		Mix	Yes		1 : 100k	Yes	Data Draw and Lamps2/Agent	Yes	Smaller scales
Germany-North Rhine Westphalia and Baden-Württemberg	Yes	Yes		L	Yes		1 : 50k	Yes	Lamps2/Agent	Yes	Yes
Germany-Lower Saxony	Yes	Yes		L	No		1 : 50k	Yes	CHANGE/ PUSH and SICAD/Open	Yes	
Great Britain	Yes	No	Yes	*	*			Yes		Yes	
The Netherlands	Yes	Yes		L	No	1 : 100k		No	MicroStation		
Ireland	Yes	Yes	Yes	*	Yes	No	*		Mercator (Star-apic)		Is being studied
Sweden	Yes	No	Yes	Mix	No				ESRI and FME	Yes	
Switzerland	Yes	No	Yes[‡]	Mix	*	No	Yes	Yes	ESRI, FME and Axpand		Yes

*Not yet decided.
[†]Not for 1 : 5k, 1 : 10k and 1 : 25k; Yes for 1 : 50k and 1 : 250k.
[‡]Base dataset is three-dimensional with target precision of 1 m (but not for all layers of information).

do not exist. The NMAs of Catalonia, Denmark, Germany, France and Great Britain have made major steps towards automated generalisation by adjusting available software or developing their own algorithms. In Great Britain, the results have only been implemented prototypically in a research environment.

An important conclusion of all NMAs is that human interaction will always be required to improve the automated results (see the section on 'Necessary research to improve current generalisation practice'). Therefore, on-the-fly generalized datasets are not considered to be realistic. Worth mentioning are the achievements of Denmark to automatically generalize 1 : 50k dataset from 1 : 10k dataset with only minimum human interaction. They can produce a complete new series of 1 : 50k map sheets in less than 18 months. Denmark is also preparing a new 1 : 100k dataset generalized from the 1 : 50k dataset.

If the specific dataset did not yet exist [1 : 50k dataset in Germany and Denmark; 1 : 25k database in Catalonia (Baella and Pla, 2005); 1 : 100k dataset in France], a first edition was generalized as automated as possible with manual improvements of the automated results. After the new dataset has been generated, generalisation focuses on the updates, with the exception of Denmark. Denmark still generalizes the entire dataset 1 : 50k every time to update this dataset. The reason is that the base dataset is renewed and changed constantly, because of a major change in topographic data collection. Therefore, Denmark produces a new complete, generalized dataset once a year.

Managing relationships between different scales

The last step for automated generalisation within NMAs is establishing links between objects at different scales. The AdV Project (Arbeitsgemeinschaft der Vermessungsverwaltungen) in Germany (AdV, 2007) builds and maintains references between different datasets. Catalonia has adjusted its data models at different scales in order to keep the semantic coherence between different scales. In France, relationships are maintained between BDCarto (about 1 : 50k) on the one side and the 1 : 100k and 1 : 120k datasets on the other side (Lecordix *et al.*, 2007). The other NMAs maintain no or little information on links between the datasets at different scales.

Necessary research to improve current generalisation practice

The second question of the workshop was: what research is needed to improve current generalisation practice? Before agreeing on topics for further research, the participants concluded that results of previous studies on automated generalisation have not always found their way to practice. The directions for future generalisation research as stated by Mueller *et al.* (1995) seem still to be valid:

> *Research cooperation between NMAs and academic research should be intensified. NMAs should state their requirements with respect to generalisation functions more clearly and academic research should take up these issues. Likewise, the third player in R&D, software vendors, should be in close contact with developments taking place at NMAs and sponsor research at academic institutions.*

The participants identified three reasons for the difficulty to implement research results into practice. First, results have to be implemented in commercial software to become available for NMAs, but generalisation requirements are very diverse and NMA-specific, depending on data models, software, source and target scales. It is hard for software vendors to provide a general solution while taking individual NMA demands into account. Generic requirements may be specifically suitable to be addressed by vendors. However, customisable software is more appropriate to meet NMA-specific requirements. This also implies that NMAs need to invest in building expertise and skills to customize generalisation software. The second reason for the difficulty of introducing research results into practice is that generalisation research is often limited to specific themes or selections from datasets. In practice, generalisation is applied to existing datasets that may contain errors or have limitations with respect to generalisation (e.g. lack of object orientation, missing semantic, geometric and topological relationships between objects).

The last reason for the difficult introduction of research results into practice is the subjectivity of generalisation. When two cartographers are given the same generalisation rules for the same area, they will come to different results (Rieger and Coulson, 1993). Exceptions are common in the generalisation process, and there may be more than one ideal generalisation solution. This is not easy to automate.

Nonetheless, the participants identified topics for further research. First, formalizing generalisation requirements are of uttermost importance to automate the process and to unambiguously understand the requirements of NMAs. This includes the possibility of automatically evaluating the requirements after or as part of the process. Second, a system is required that understands the problem of generalisation laid down by the formal requirements. Such a system should implement generalisation functionality that takes the global context (e.g. mountains, rural and urban) and local context (e.g. neighbouring objects) into account. The system should support databases, which are enriched with semantics for generalisation. Examples of such additional information are object density and distribution, relative importance of objects, and semantic and topological relationships between objects (Weibel and Dutton, 1999). A third need for NMAs is generic generalisation functionality that is adaptable to different data models. This requires compatible data models that support multi-scale databases. In addition, support for multi-representation databases (i.e. maintenance of links between derived and original dataset, automated updating of derived datasets and relevance check during updates) is hardly available in mainstream DBMSs but important when maintaining multi-scale data. Finally, some participants in the workshop would like to see major progress in automated generalisation of contour lines, place names, buildings in the urban context and pruning of artificial networks.

QUANTITATIVE ANALYSIS OF RELEVANT GENERALISATION OPERATORS FOR MAP PRODUCTION

One of the topics identified for further research in the previous section is generically applicable generalisation functionality. To get more insights into what kind of functionality is lacking at NMAs and how important this functionality is for NMAs, a survey was carried out. The aim of this survey was to extend the findings of the workshop and to provide a quantitative view of the NMAs on missing generalisation functionality. This analysis enables formulating more specific recommendations for NMAs, software suppliers and the research community for developing generalisation solutions.

The title of the survey was 'the current problems of automated generalisation of topographic data at NMAs' and was carried out at the end of 2007. It was completed by 11 NMAs from eight countries and three German states. The structure of the survey was two-fold. The first part addressed the kind of implementation of the generalisation process to derive topographic products at the specific NMA (model versus cartographic generalisation) and their degree of automation. This part of the survey was used to update the workshop findings and to outline generalisation practice at the NMAs as background to the second part of the survey.

The second part aimed at analysing in more detail and in a quantifiable way the missing generalisation functionality within NMAs. For an indicator of the missing generalisation functionality, we used the importance and problematic (i.e. lacking) characteristics of generalisation operators. The motivation for using operators as an indicator is that operators are one of the main building blocks for generalisation processes. In addition, they are suitable to quantify problems of automated generalisation. In the survey, the operators were analysed with respect to the importance of feature types to which they are applied and considered for each scale transition separately.

This section presents a quantitative analysis of the second part of the survey (on generalisation operators) and introduces a 'relevance' factor for the operators, which integrates the importance and problematic characteristics of operators. The analysis shows how generalisation operators are used in practice. It also shows the obstacles and requirements of operators for generalisation processes within NMAs. This analysis thereby goes beyond the preliminary results as presented in Foerster and Stoter (2008).

The major observations of Foerster and Stoter (2008) were that the model generalisation process at NMAs is far more automated and advanced than the cartographic generalisation process. In addition, lots of different operators are required for a successful generalisation process as identified by the importance of operators. Finally, the most problematic operators during the generalisation process are displacement and typification.

The section on 'Method to quantify problems of automated generalisation of NMAs' explains the method that quantifies the current problems of automated generalisation at NMAs through generalisation operators. It also explains the 'relevance' measure used in our method and describes the three steps of the method. The results of these three steps are presented in the sections on 'Important and problematic operators for model and cartographic generalisation', 'Relevant operators for model and cartographic generalisation' and 'Relevance of operators weighted by importance value of feature types'.

Table 2. Classification of generalisation operators applied in the survey based on Foerster *et al.* (2007)

Model generalisation	Class selection
	Reclassification
	Collapse
	Combine
	Simplification
	Amalgamation
Cartographic Generalisation	Enhancement
	Displacement
	Elimination
	Typification
	Enlargement
	Amalgamation

Method to quantify problems of automated generalisation of NMAs

Model and cartographic generalisation operators

To learn more about the specific problems of operators, it is useful to distinguish between operators for model generalisation and operators for cartographic generalisation. Since no consensus exists on a distinction between model and cartographic generalisation, as mentioned in the section on 'Necessary research to improve current generalisation practice', the survey followed the approach of Foerster *et al.* (2007), as shown in Table 2. Their classification is based on models of ISO and OGC. It has also been formalized using the object constraint language (OCL) as described in Foerster *et al.* (2008).

Topographic feature types

Generalisation operators are always applied to a specific feature type (or group of feature types). To include the aspect of feature types in our analysis, the survey studied the operators regarding the topographic feature types they are applied to. The set of topographic feature types that was used is adopted from the EuroRegional Map Project (Delattre, 2004) and is depicted in Table 3.

It is important to note that the generalisation operators are rated regarding the specific feature type they are applied to. However, they may take other features or feature types into account. Thus, generalisation operators are considered to be contextual, if appropriate.

Analysis steps

The analysis steps of the method to measure the relevance of operators are depicted in Figure 2 and can be summarized as follows.

Table 3. Classification of topographic feature types applied in the survey

Administration
Buildings
Railways
Roads
Relief
Lake
River
Coastal feature
Landcover

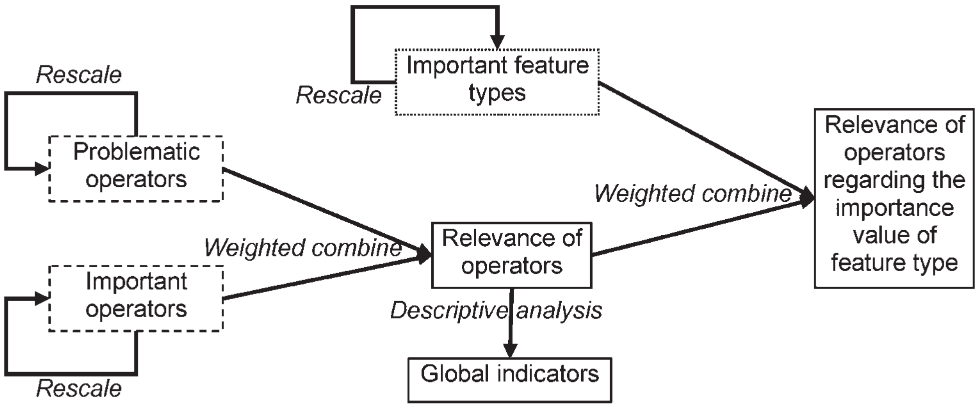

Figure 2. Quantitative analysis process of measuring the relevance of operators

Step 1. Rescaling important and problematic operators

The survey separated between 'important' and 'problematic' operators. Important means that an operator is often applied and plays a dominant role in the specific generalisation process (applied on a specific scale transition and on a specific feature type). Whereas problematic means that a specific operator is lacking and it therefore exposes problems to the generalisation process. Both measures address an important and specific aspect. The results of these two separate measures have been reported in Foerster and Stoter (2008).

In this paper, we combine the two measures in an aggregated value. Therefore, the values (C) for the important and problematic generalisation operators are rescaled to their local minimum and maximum using equation (1). Originally, the participants were asked to rate the different variables using a value range from 0 (low) to 5 (high). After rescaling, all values are between 0 and 1 which allows us to compare and combine results of the different measures

$$\forall c \in C,\ c = \frac{c - \min(C)}{\max(C) - \min(C)} \tag{1}$$

The resulting values are standardized on the local maximum $\max(C)$ and the local minimum $\min(C)$ of C. The rescaled values for the important and problematic operators are presented in the section on 'Important and problematic operators for model and cartographic generalisation'.

Step 2. Calculating the relevance of operators

To get a complete picture of the operators, this paper introduces an integrated measure, termed as the relevance of a specific generalisation operator. The relevance measure combines the (rescaled) important and problematic values of operators using equation (2).

$$\forall g \in G,\ \exists f \in F,\ c = 0{\cdot}5g + 0{\cdot}5f \tag{2}$$

Equation (2) weights the values of a set (g of G) by a corresponding measure (f) of another set (F). Equation (2) applies a linear factor of 0.5, which weights both aspects equally.

The results of this analysis separated for model and cartographic generalisation operators are presented in the section on 'Relevant operators for model and cartographic generalisation'. The relevance measure is further compiled to global indicators by descriptive statistics which are visualized using boxplot diagrams in the section on 'Relevant operators for model and cartographic generalisation'. The global indicators represent first quartile, third quartile, arithmetic mean and median for each of the scale transitions. The global indicators give additional information about the outcomes of the relevance measures for model and cartographic generalisation operators at specific scales. Any variance indicator would also have been an interesting global indicator. However, they have not been calculated as the number of collected survey answers per scale was too small.

Step 3. Weighting the relevance of the operators by the importance of feature types

In a next step, the relevance of the operators is weighted by the rescaled importance values of the feature type. The results are presented in the section on 'Relevance of operators weighted by importance value of feature types'. The relevance of operators already implicitly incorporates a certain degree of importance of the specific feature types. However, combining relevance with importance of feature types will both filter and exaggerate the relevant operators with respect to the most important feature types in the current products of NMAs. This new indicator better exposes the requirements for map production, since it provides not only insight into missing functionality, but also insight into which operators might be relevant in the future, i.e. how bad it is that they are missing?

The relevance of operators and the importance of feature type are weighted 0.5 and 0.5. Consequently, the importance and problematic characteristics of operators only influence this second measure by 0.25 each, whereas the importance of the feature type is 0.5 of the complete measure. It may have been possible to weight the values by one-third each. However, in order to stress the role of the feature type within the generalisation process and its importance regarding the operator, we equally weighted the relevance values of operators and the importance value of feature types.

Scale transitions considered in the analysis

Apart from specific feature types at which operators are applied, scale transitions at which the operators are applied are important to identify missing generalisation functionality. Therefore, the survey distinguished between scale

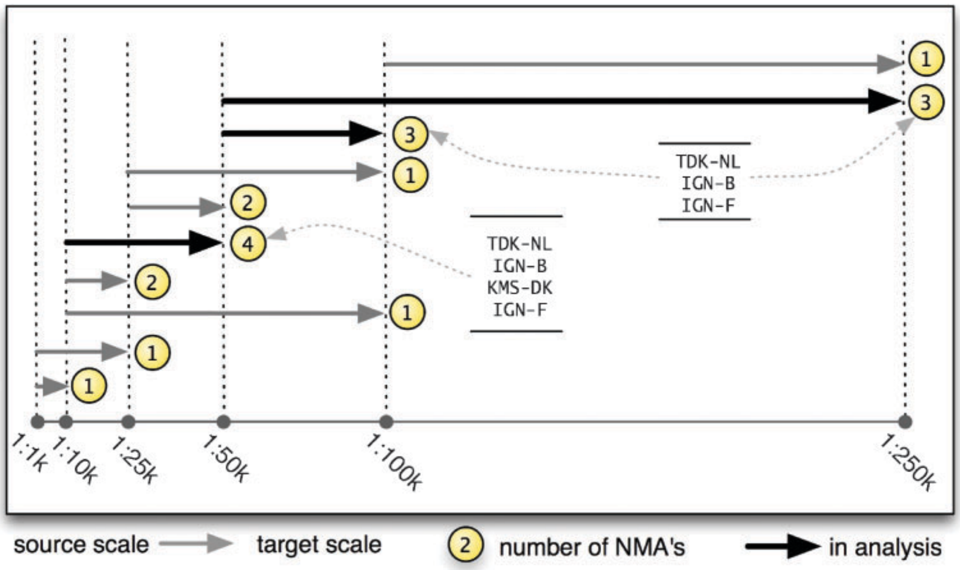

Figure 3. The scale transitions collected in the survey – the highlighted ones, will be considered in the presented analysis

transitions as they are carried out at the NMAs. To conduct representative results, the analysis focused only on scale transitions that are applied by more than three participants, i.e. 1 : 10k–1 : 50k, 1 : 50k–1 : 100k and 1 : 50–1 : 250k (Figure 3). All results in the remainder of this section are analysed for these three scale transitions separately.

Important and problematic operators for model and cartographic generalisation

This section introduces the rescaled values for important and problematic generalisation operators. The original values were collected from 0 to 5 and can be found in Foerster and Stoter (2008).

Important generalisation operators

The rescaled values representing the importance of operators in relation to the different feature types are presented in Figure 4 for model generalisation operators and in Figure 5 for cartographic generalisation operators. The importance values of these two types of operators differ when considering the specific scale transition. The importance of model generalisation is significantly higher at scale transition at smaller scales (1 : 50k–1 : 250k). However, the importance of cartographic generalisation operators is higher at larger scales (1 : 10k–1 : 50k). NMAs consider simplification, amalgamation (model generalisation) and displacement (cartographic generalisation) as most important operators.

Problematic generalisation operators

The lack of specific generalisation operators in relation to a specific feature type and scale are depicted in Figure 6 (model generalisation) and Figure 7 (cartographic generalisation). Figure 6 shows that model generalisation operators are not considered as problematic. On the contrary, the

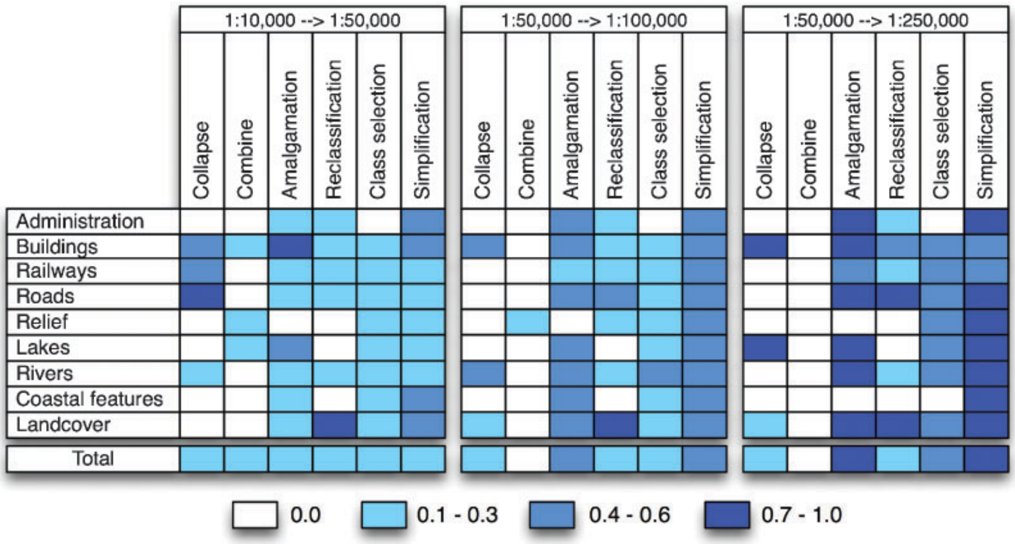

Figure 4. Importance of model generalisation operators versus feature types related to scale

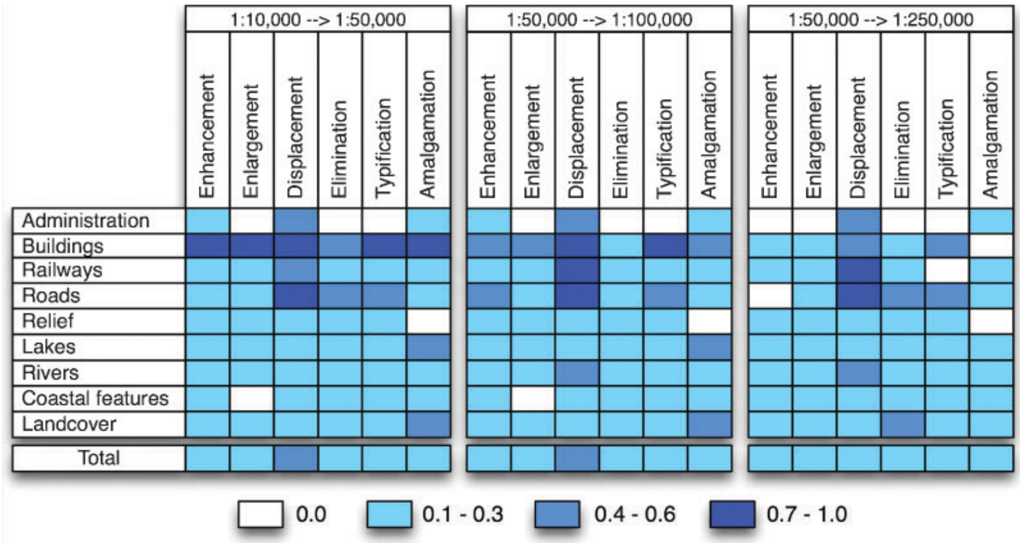

Figure 5. Importance of cartographic generalisation operators versus feature types related to scale

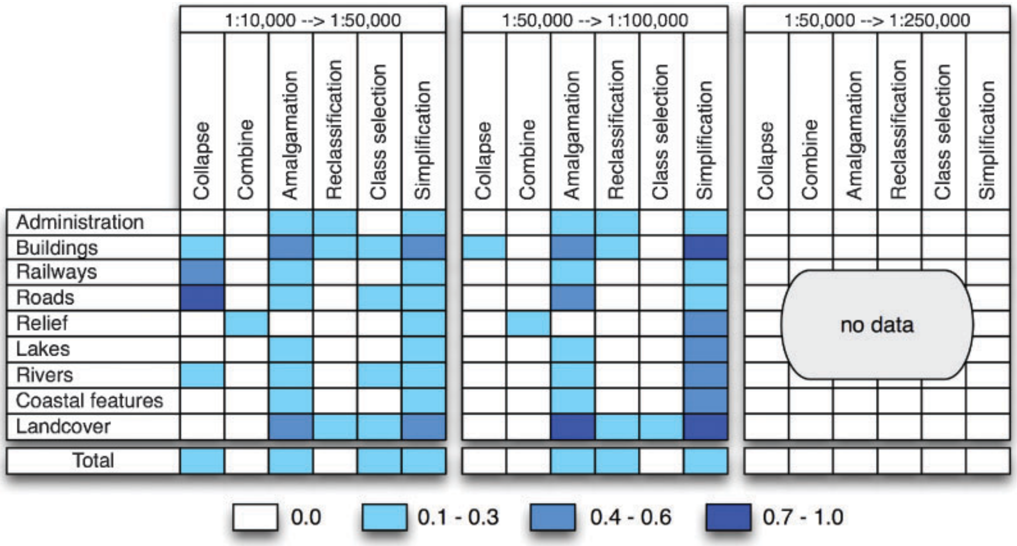

Figure 6. Problematic model generalisation operators (no answers for 1 : 50k–1 : 250k available)

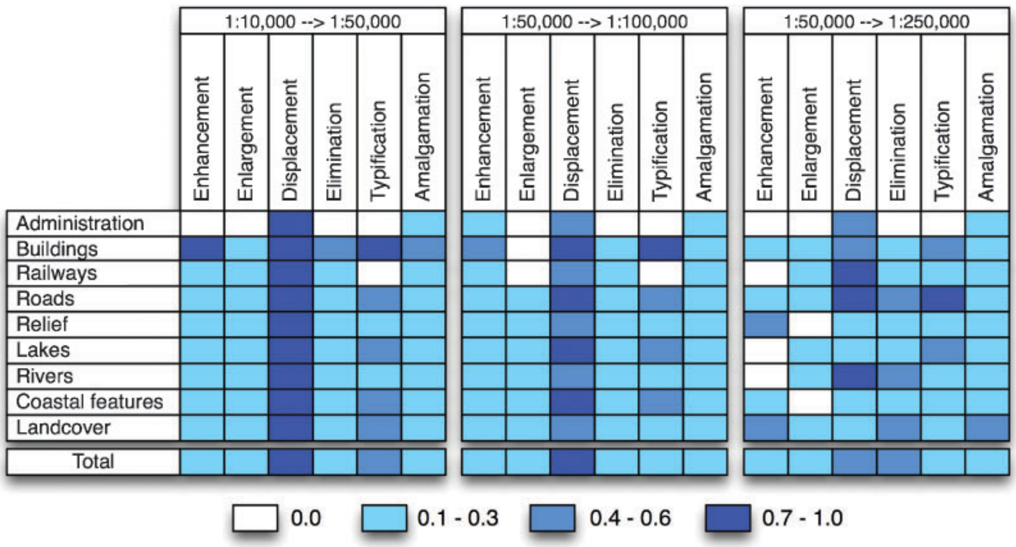

Figure 7. Problematic cartographic generalisation operators

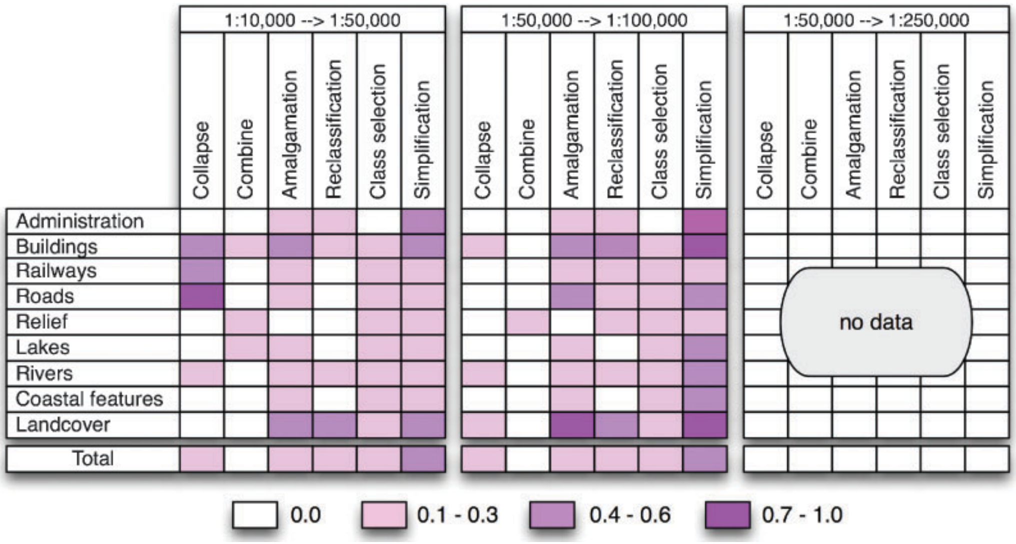

Figure 8. Calculated relevance values of model generalisation operators separated for the major scale transitions. The survey did not result in sufficient information regarding model generalisation operators at the highest scale transition (1 : 50k–1 : 250k). Thus, values for this scale transition have been excluded

cartographic generalisation operators (Figure 7) are more problematic for current production lines. The most problematic operators are displacement and typification.

Relevant operators for model and cartographic generalisation

The results of the relevance measure, combining the importance and lacking characteristics of operators, are presented in Figure 8 (model generalisation) and Figure 9 (cartographic generalisation). All values are calculated based on the rescaled measures presented in the section on 'Important and problematic operators for model and cartographic generalisation'.

We can draw the following conclusions from these tables. Simplification, collapse and amalgamation are the most relevant model generalisation operators. Collapse is relevant at lower scale transitions (1 : 10k–1 : 50k), especially for roads, buildings and railways but not at the higher scale transition (1 : 50k–1 : 100k). This can be explained because already collapsed roads are reused at higher scales.

Figure 9 shows that the most relevant generalisation operators for cartographic generalisation are displacement and typification. Additionally, any operator applied to feature type buildings is highly relevant.

To compare the overall relevance of operators at certain scale transitions and between model generalisation and cartographic generalisation, Figures 10 and 11 presents the results of the global indicators (boxplot diagram). The rescaled values are the basis for those diagrams. Thus the value range is always between 0 and 1.

Several conclusions can be drawn from these global indicators. First, the relevance of model generalisation operators increases with decreasing scales (from 1 : 10k–1 : 50k to 1 : 50k–1 : 100k), whereas the relevance of

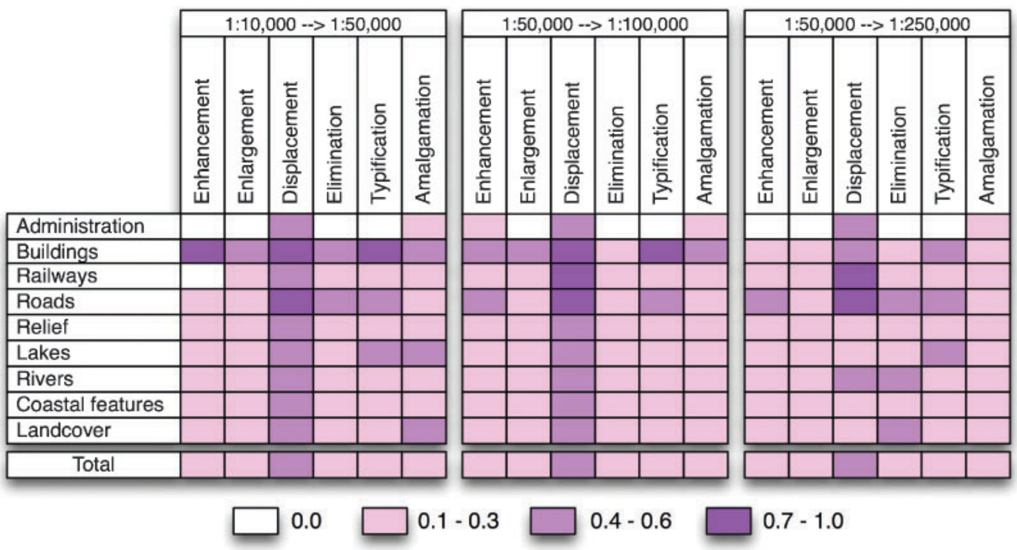

Figure 9. Calculated relevance values of cartographic generalisation operators separated for the major scale transitions

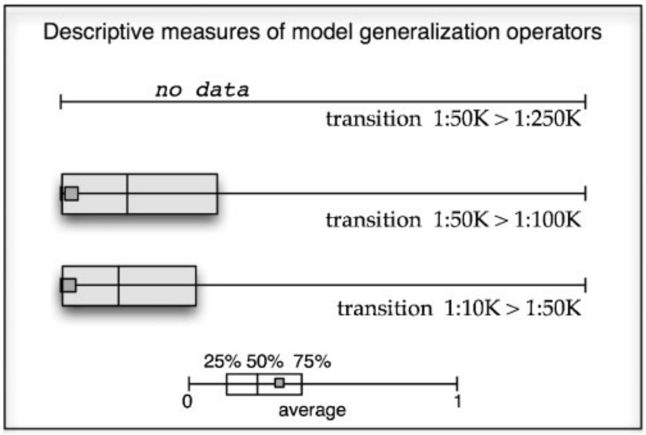

Figure 10. Boxplot diagram of the model generalisation operator measures (min.=0, max.=1) as presented in Figure 8

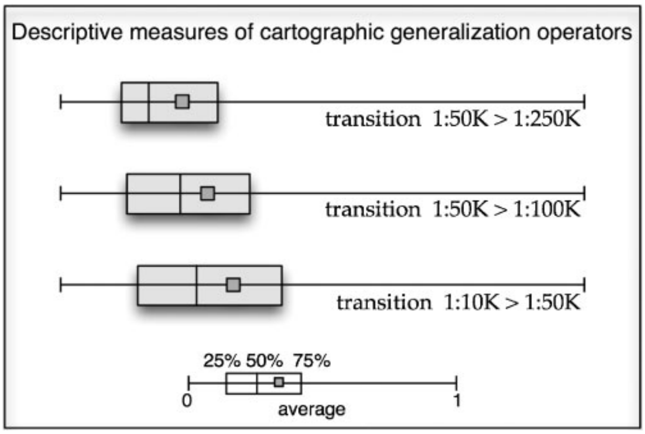

Figure 11. Boxplot diagram of the cartographic generalisation measures (min.=0, max.=1) as presented in Figure 9

cartographic operators decreases with decreasing scale. A second conclusion is that cartographic generalisation operators are overall more relevant than model generalisation operators. This is in line with the workshop conclusions that especially contextual operators (mostly cartographic generalisation operators) are considered as problematic. In addition, the numbers support the initial findings of the survey reported in Foerster and Stoter (2008). Another observation from Figures 10 and 11 is that the distribution of the values is different, as the median is above the average mean for model generalisation operators. This can be explained by low relevance values for model generalisation operators as shown in Figure 8. In the case of cartographic generalisation, it is slightly different. Some operators seem to be more relevant, as the mean is higher than the median, which is an indicator for statistical outliers.

Relevance of operators weighted by importance value of feature types

Figure 12 shows the rescaled importance values of the different feature types regarding the specific scale transitions, which were originally collected from 0 (low) to 5 (high). The table shows that rivers and roads are the most dominant feature types for all scale transitions. However, the building feature type becomes less important over decreasing scale. In addition, networks become more important at smaller scales.

In the second step, the relevance of generalisation operators (the section on 'Relevant operators for model and cartographic generalisation') are weighted by the rescaled importance values of the feature types. This indicator combines the importance values of the feature type (Figure 12) according to equation (2) with the relevance values of the model generalisation and cartographic generalisation operators (Figures 8 and 9). The results are depicted in Figures 13 and 14 for respectively model generalisation operators and cartographic generalisation operators.

The following observations can be made from these tables. The generalisation of buildings and roads appear to be the most relevant for model generalisation (Figure 13). Especially amalgamation of buildings seems to be highly relevant for map production at 1 : 10k–1 : 50k. In line with Figure 8, Figure 13 shows that amalgamation is of major concern at the investigated scales. In contrast to some of the extremes that disappeared compared to Figure 8. For example, simplification turns out to be not that relevant overall for model generalisation.

Also for cartographic generalisation (Figure 14), weighting the relevance measures by importance values of feature types causes some extreme values to disappear. For instance, displacement got a lower relevance, due to the lower

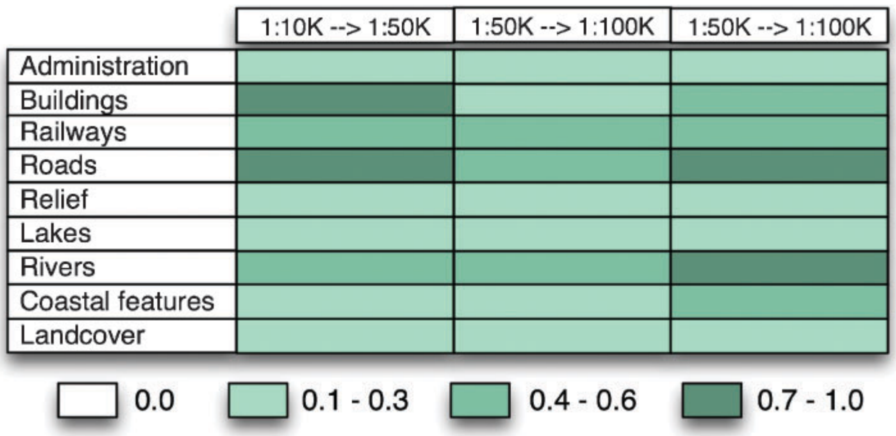

Figure 12. Importance values of feature types at certain scale transitions. The values are scaled regarding the local minimum and maximum

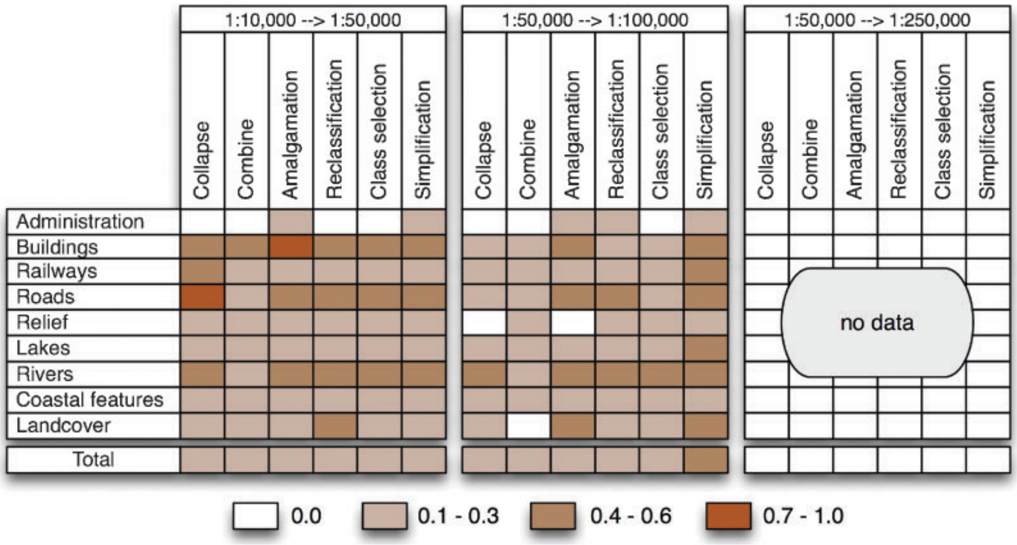

Figure 13. Relevance of model generalisation operators weighted by the importance of feature types

importance values of the combined feature types. However, as rivers are highly relevant in map production, all the related operators (i.e. enhancement and elimination of rivers) become more relevant. The same conclusion applies to roads (i.e. enlargement and elimination) and also to railways (i.e. elimination and enhancement).

CONCLUSIONS

The aim of the research presented in this paper was to analyse the challenges of automated generalisation as experienced by NMAs as well as to elaborate on the consequences for research, NMAs and software vendors.

First, a qualitative analysis was carried out about the trends and policies on automated generalisation within NMAs (the section on 'QUALITATIVE ANALYSIS OF CHALLENGES FOR AUTOMATED GENERALISATION'). The analysis is based on a workshop held in 2005 and attended by 12 NMAs. Recently, findings have been updated by the participants. From this analysis, it can be concluded that full automation is not implemented at any NMA, although some NMAs have made large investments and achieved major steps, a representative example being Denmark. Another important conclusion from the workshop is that there is no single approach for the adoption of automated generalisation within NMAs. It heavily depends on NMA-specific factors such as the level of detail of initial datasets, supported scales, applied scale transitions, specific configuration of the landscape, variance in information at the largest scale according to type of area, specific data content per scale, geometry types of features, distinction between model and cartographic generalisation, and organisational aspects such as the availability of special resources for strategic research, type of customers to serve, business model, etc.

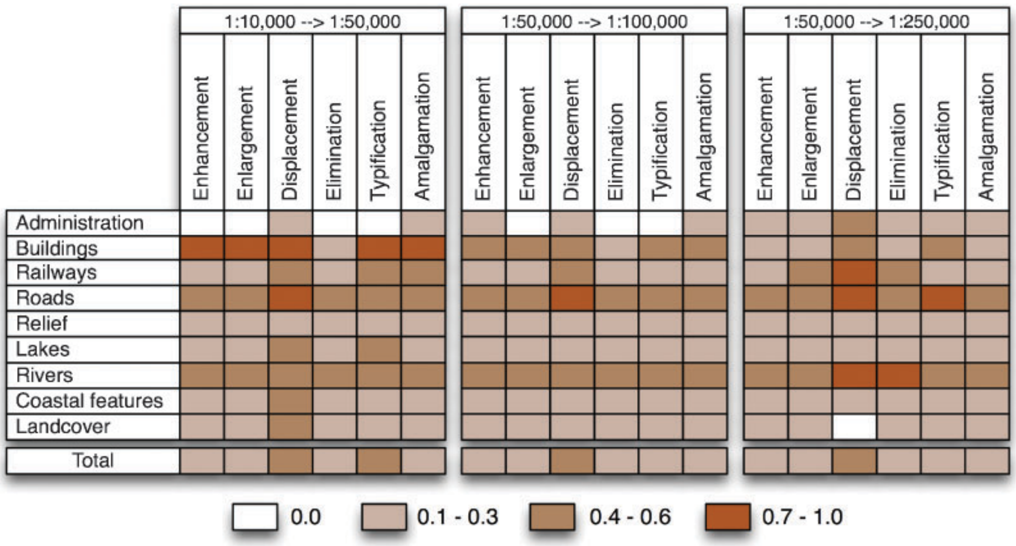

Figure 14. Relevance of cartographic generalisation operators weighted by the importance of feature types

Ready to use software for automated generalisation is therefore not considered as appropriate for automated generalisation. Instead, it will require implementation as well as remodelling efforts of NMAs to introduce automated generalisation into own production lines. To support these NMA-specific processes, NMAs need adjustable systems as well as generic applicable generalisation functionality. Providing a common view on such functionality, reflecting NMA requirements, may support researchers and software vendors to develop automated generalisation solutions for NMAs.

This motivates the quantitative analysis on missing generalisation functionality as described in the section on 'QUANTITATIVE ANALYSIS OF RELEVANT GENERALISATION OPERATORS FOR MAP PRODUCTION'. This analysis provides detailed insights into currently applied strategies towards generalisation operators and current problems of generalisation operators at NMAs.

The analysis demonstrates the relevance of specific generalisation operators by combining the importance and problematic (i.e. lacking) aspects of operators. This shows that the relevance of model generalisation operators increases with decreasing scales, but never reaches the relevance level of cartographic generalisation operators.

Weighting the relevance measures by importance values of feature types results in another valuable conclusion. Especially network-based feature types such as rivers, railways and roads are relevant for NMAs in combination with the operators enhancement, typification and elimination. Overall, contextual operators and operators that create generalized features that inherit a network-based structure are the main challenges for cartographic generalisation. This underlines the workshop findings.

The presented results of both the (updated findings of the) workshop as well as the survey describe the long term challenges for NMAs. They may therefore serve as a guideline for NMAs, researchers and software suppliers to better align their activities. The presented work also extends the findings of the OEEPE project (Ruas, 2001) and the EuroSDR project as it studies generalisation operators not limited to specific generalisation solutions or test cases but as applied and required in NMA production lines.

Mackaness *et al.* (2007a) state that research on automated generalisation should 'connect' to practice in order to better meet their requirements and to streamline research activities. This study is an example of obtaining better understanding of NMA requirements for automated generalisation and of identifying topics for further research starting from a requirement analysis at NMAs. In addition, exchanging knowledge about generalisation operators, the main building blocks of automated generalisation processes, sharpens the terminology within NMAs and research groups and thereby improves the interoperability of concepts. This will enable more flexible and effective solutions both in databases as well as on the Web. In the future, the presented criteria could be reassessed to identify the success and remaining problems of NMAs of automated generalisation. The resulting index could then be used to assess the undertaken effort of the generalisation community.

BIOGRAPHICAL NOTES

Theodor Förster (1980) is a PhD candidate at ITC's Department of Geo-Information Processing (GIP) since the end of 2005. His PhD research covers the technical generalisation aspects within the framework of the Ruimte voor Geo-Informatie (RGI) 002 research project 'Generation and use of base maps for integrated querying of digital physical development plans' (http://www.durpondergronden.nl). In this context, he is mostly interested into Web service standards for geo-processes and geo-data and the formalisation of generalisation operators. The PhD study is supervised by Jantien Stoter and Menno-Jan Kraak.

In 2005, he received his Diploma in Geoinformatics at the University of Muenster, Germany (http://www.uni-muenster.de). Before and after graduation, he worked at the Institute for Geoinformatics in Muenster (http://ifgi.uni-muenster.de) for different projects related to geo-information Web-application development.

He is an active member of the 52north open source initiative (http://www.52north.org) and is leading the development of the 52north Web Processing Service (52n WPS). Besides that, he is participating in the standardisation of geo-processing services at OGC.

ACKNOWLEDGEMENTS

We highly appreciate the individual contributions of members of (*participating in the survey; †participating in the workshop): LGN Lower Saxony*† (German state), LVG Bavaria*† (German state), LVA Baden-Wuerttemberg*† (German state), LVA North Rhine-Westphalia† (German state), BKG* (Germany), NLS* (Finland), IGN*† (Belgium), IGN*† (France), OS*† (Ireland), OS† (UK), KMS*† (Denmark), Swiss Topo*† (Switzerland), TDK*† (The Netherlands), ICC† (Spain, Catalonia) and Lantmäteriet† (Sweden).

We also kindly appreciate the feedback of Nicolas Regnauld (Ordnance Survey, UK) and Harry Uitermark (Dutch Kadaster) who helped us improving the survey.

REFERENCES

AdV. (2007). Produktblaetter Arbeitskreis Geotopographie. Arbeitsgemeinschaft der Vermessungsverwaltungen der Laender der Bundesrepublik Deutschland, http://www.adv-online.de/extdeu/binarywriterservlet?imgUid=cc550b56-9e72-f711-a3b2-1718a438ad1b&uBasVariant=22222222-2222-2222-2222-22222222&isDownload=true (accessed 27 October 2008).

Baella, B. and Pla, M. (2005). 'Reorganizing the Topographic Databases of the Institut Cartographic de Catalunya Applying Generalization', in **8th ICA Workshop on Generalisation and**

Multiple Representation, A Coruña, Jul 7–8, http://ica.ign.fr/Acoruna/Papers/Baella_Pla.pdf (accessed 9 October 2008).

Brewer, C. A. and Buttenfield, B. P. (2007). Framing guidelines for multi-scale map design using databases at multiple resolutions. **Cartography and Geographic Information Science**, 34, pp. 3–15.

Buttenfield, B. P. (1991). 'A rule for describing line feature geometry', in **Map Generalization: Making Rules for Knowledge Representation**, ed. by Buttenfield, B. P. and McMaster, R. B., pp. 150–171, Longman, London.

Delattre, N. (2004). 'Euroregionalmap, A Pragmatic Solution to Create an ESDI at Medium Scale through the Harmonization of National Topographic Data Collections', in **10th EC-GI&GIS Workshop**, Jun 23–25.

EuroGeographics. (2005). 'Generalisation processes: a benchmark study of the expert group on quality', **EuroGeographics**, http://www.eurogeographics.org/eng/05_quality_reports.asp

Fechir, A. and Waele, J. D. (2005). 'The Future of Production of Generalised Maps at IGN Belgium', in **8th ICA Workshop on Generalization and Multiple Representation**, A Coruña, Jul 7–8.

Foerster, T., Morales, J. and Stoter, J. E. (2008). 'A classification of generalization operators formalised in OCL', in **Proceedings of GI-days 2008**, ed. by Bishr, M., Pebezma, E. and Bartoschek, T., pp. 141–156, Institute for Geoinformatics, Muenster.

Foerster, T. and Stoter, J. E. (2008). 'Generalization Operators for Practice – A Survey at NMAs', in **11th ICA Workshop on Generalization and Multiple Representation**, Montpellier, Jun 20–21.

Foerster, T., Stoter, J. E. and Kobben, B. (2007). 'Towards a Formal Classification of Generalization Operators', in **23rd International Cartographic Conference (ICC 2007)**, Moscow, Aug 4–10.

Gruenreich, D. (1992). 'ATKIS – a topographic information system as a basis for GIS and digital cartography in West Germany', **Geologisches Jahrbuch Reihe A**, 122A, pp. 207–216.

Hubert, F. and Ruas, A. (2003). 'A Method Based on Samples to Capture User Needs for Generalisation', in **5th Workshop on Progress in Automated Map Generalization**, Paris, Apr 28–30.

Kilpelaïnen, T. (2000). 'Knowledge acquisition for generalization rules', **Cartography and Geographic Information Science**, 27, pp. 41–50.

Lecordix, F., Gallic, J. L., Gondol, L. and Braun, A. (2007). 'Development of a New Generalization Flowline for Topographic Maps', in **10th ICA Workshop on Generalisation and Multiple Representation**, Moscow, Aug 2–3, http://aci.ign.fr/BDpubli/moscow2007/Lecordix_ICAWorkshop.pdf (accessed 19 September 2008).

Leitner, M. and Buttenfield, B. (1995). 'Acquisition of procedural cartographic knowledge by reverse engineering', **Cartography and Geographic Information Systems**, 22, pp. 232–241.

Mackaness, W. A., Ruas, A. and Sarjakoski, L. T. (Eds.). (2007). **Generalisation of Geographic Information: Cartographic Modelling and Applications**, Elsevier, Oxford.

Mackaness, W. A., Ruas, A. and Sarjakoski, L. T. (2007). 'Observations and research challenges in map generalization and multiple representation', **Generalization of Geographic Information: Cartographic Modelling and Applications**, ed. by Mackaness, W. A., Ruas, A. and Sarjakoski L. T., pp. 315–323, Elsevier, London.

McGraw, K. L. and Harbison-Briggs, K. (1989). **Knowledge Acquisition: Principles and Guidelines**, Prentice Hall, Upper Saddle River, NJ.

McMaster, R. B. (1995). 'Knowledge acquisition for cartographic generalization', in **GIS and Generalization: Methodology and Practice**, ed. by Mueller, J. C., Lagrange, J. P. and Weibel, R., pp. 161–179, Taylor & Francis, London.

McMaster, R. B. and Shea, K. S. (1992). **Generalization in Digital Cartography**, American Association of Geographers, Washington, DC.

Mueller, J. C., Weibel, R., Lagrange, J. P. and Salge, F. (1995). 'Generalization: state of the art and issues', in **GIS and Generalization, Methodology and Practice**, ed. by Mueller, J. C., Lagrange, J. P. and Weibel, R., pp. 3–18, Taylor & Francis, London.

Muller, J. and Mouwes, P. (1990). 'Knowledge acquisition and representation for rule based map generalization: an example from The Netherlands', in: **GIS/LIS '90**, pp. 58–67, Anaheim, CA, Nov 14–17.

Mustiere, S. (2001). 'Apprentissage supervisé pour la généralisation cartographique', PhD thesis, Université Paris VI, Paris, France.

Mustiere, S. (2005). 'Cartographic generalization of roads in a local and adaptive approach: a knowledge acquisition problem', **International Journal of Geographical Information Science**, 19, pp. 937–955.

Nickerson, B. G. (1991). 'Knowledge engineering for generalization', **Map Generalization: Making Rules for Knowledge Representation**, ed. by Buttenfield, B. and McMaster, R. B., pp. 40–55, Longman, London.

Plazanet, C., Bigolin, N. and Ruas, A. (1998). 'Experiments with learning techniques for spatial model enrichment and line generalization', **Geoinformatica**, 2, pp. 315–333.

Regnauld, N. and Revell, P. (2007). 'Automatic amalgamation of buildings for producing Ordnance Survey 1 : 50 000 scale maps', **The Cartographic Journal**, 44, pp. 239–250.

Reichenbacher, T. (1995). 'Knowledge acquisition in map generalization using interactive systems and machine learning', in **17th International Cartographic Conference**, pp. 2221–2230, Barcelona, Sep 3–9.

Rieger, M. K. and Coulson, M. R. C. (1993). 'Consensus or confusion: cartographers' knowledge of generalization', **Cartographica**, 30, pp. 69–80.

Ruas, A. (2001). **Automatic Generalisation Project: Learning Process from Interactive Generalisation**, OEEPE, Frankfurt-am-Main.

Stoter, J. E. (2005a). 'Generalisation within NMAs in the 21st century', in **22nd International Cartographic Conference**, A Coruña, Jul 9–16.

Stoter, J. E. (2005b). 'Generalization: the gap between research and practice', in **8th ICA Workshop on Generalization and Multiple Representation**, A Coruña, Jul 7–8.

Stoter, J. E., Burghardt, D., Duchene, C., Baella, B., Bakker, N., Blok, C., Pla, M., Regnauld, N. and Touya, G. (2009) 'Methodology for evaluating automated map generalization in commercial software', **Computers, Environment and Urban Systems**, 33, pp. 311–324.

Stoter, J. E., Morales, J., Lemmens, R., Meijers, M., van Oosterom, P., Quak, W. and Uitermark, H. (2008). 'A data model for multi-scale topographical data', in **Headway in Spatial Data Handling, Lecture Notes in Geoinformation and Cartography**, ed. by Ruas, A. and Gold, C., pp. 233–254, Springer-Verlag, Berlin.

Weibel, R. (1991). 'Amplified intelligence and rule-based systems', in **Map Generalization: Making Rules for Knowledge Representation**, ed. Buttenfield, B. and McMaster, R. B., pp.172–186, Longman, New York.

Weibel, R. (1995). 'Three essential building blocks for automated generalization', in **GIS and Generalization: Methodology and Practice**, ed. by Mueller, J., Lagrange, J. P. and Weibel, R., pp. 56–70, Taylor & Francis, London.

Weibel, R. and Dutton, G. (1999). 'Generalising spatial data and dealing with multiple representations', in **Geographic Information Systems – Principles and Technical Issues**, ed. by Longley, P., Goodchild, M., Maguire, D. and Rhind, D., pp. 125–155, John Wiley & Sons, New York.

Weibel, R., Keller, S. and Reichenbacher, T. (1995). 'Overcoming the knowledge acquisition bottleneck in map generalization: the role of interactive systems and computational intelligence', in **Spatial Information Theory – A Theoretical Basis for GIS**, ed. by Frank, A. U. and Kuhn, W., pp. 139–156, Berlin, Springer-Verlag.