Feasibility study on an automated generalisation production line for multiscale topographic products

Jantien Stoter[1,2], Ron Nijhuis[1], Marc Post[1], Vincent van Altena[1], Jan Bulder[1], Ben Bruns[1], John van Smaalen[3,2]

[1]*Kadaster, Apeldoorn, The Netherlands, firstname.lastname@kadaster.nl*

[2]*GISt, OTB, TU Delft, The Netherlands, j.e.stoter@tudelft.nl*

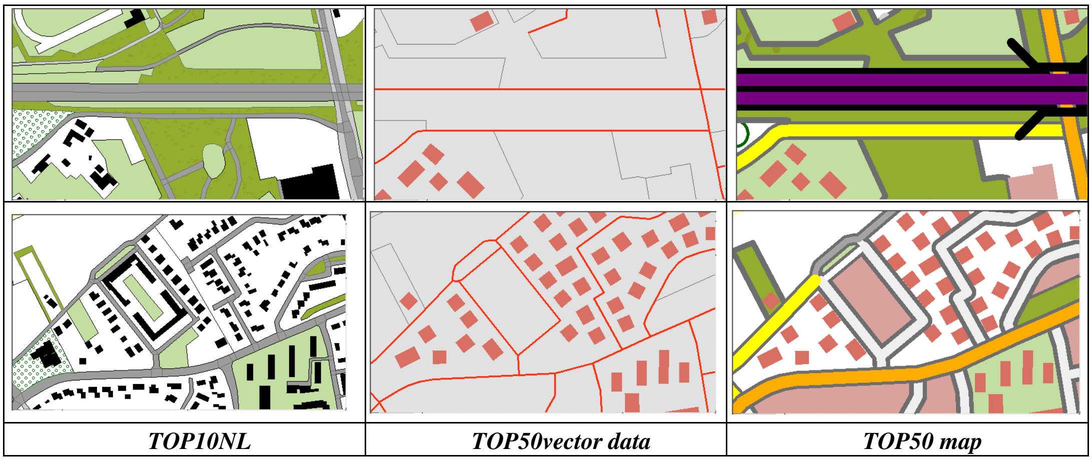
[3]*Esri NL, Rotterdam, The Netherlands, jvansmaalen@esri.nl*

1 INTRODUCTION

This paper presents the ongoing feasibility study on a new production line by the Kadaster (the Dutch Cadastral and National Mapping Agency). This production line should enable fully automated generalisation of smaller scale maps from one detailed base data set.

The reason to reconsider the present production line is that limitations are met to fulfil the legal task of producing topographic vector data and raster maps at scale 1:10k, 1:50k, 1:100k, 1:250k, 1:500k and 1:1.000k in an update cycle of two years. To meet this legal obligation Kadaster has converted its vectorised maps into object oriented databases since 2007. These products (called TOP10NL, TOP50NL etc) are still updated interactively by cartographers following generalisation guidelines. Consequently updates in the current multiscale data require considerable time, i.e. 500-2000 hour on average to update a 1:50k map sheet of 500 square kilometres. Specifically since the step-wise process (i.e. every scale is generalised from the next larger scale in a ladder approach) causes updates to be propagated even later in smaller scale maps. Besides the costs involved, this lengthy process does no longer fulfil current requirements with respect to up-to-dateness of data and maps. Therefore automated generalisation is being considered for a new production line. (To understand the generalisation challenge of a small country as the Netherlands: the whole country is covered by 93 map sheets at scale 1:50k, 20x25km each.)

The feasibility study reassesses the current multiscale products, because it is expected that the maps of which the origin is often more than 60 years old, cause limitations for automated generalisation for several reasons. At first taking the legacy topographic databases as starting point may overemphasis past requirements and may ignore new requirements of multiscale topographic data. In addition automating a previously interactive process, which was defined for a past technical and organisational context, may be too complex. Another limitations of the current products is that the databases containing geometry of vectorised maps, mix cartographic and database representations of topography. The objects may have been enlarged or displaced for the sake of readability of the symbolised objects (see Figure 1).



TOP10NL | ***TOP50vector data*** | ***TOP50 map***

Figure 1 Displacement of objects in TOP50vector due to graphical conflicts of symbolised features

These geometrical adjustments have not been controlled to meet the accuracy of that scale. Instead of this mixed approach, distinguishing between (requirements for) multiscale data and multiscale visualisations may offer better potentials to move forward in the domain of automated generalisation. The ongoing research on a new production line builds on the findings of the internal proof-of-concept project that was conducted in 2010 (Stoter et al, 2011; Stoter et al, 2009; Van Smaalen and Stoter, 2008). In that project a fully automated generalisation workflow was set up for deriving a 1:50k map from TOP10NL database taking current requirements and new technologies into account.

The Dutch Kadaster has no history in investing in automated generalisation in contrast to other NMAs such as IGN, France (Lecordix et al. 2007), KMS, Denmark (Foerster et al, 2010), ICC, Catalonia (Baella & Pla 2005) and Germany (AdV, 2007). Consequently the proof-of-concept pilot provided important insights for the new production line. For this reason the main findings of this pilot are reported in Section 2. Section 3 elaborates on the issues for the new production line that are currently being studied. Section 4 ends with work in progress.

2 MAIN FINDINGS OF THE GENERALISATION PILOT

The aim of the generalisation pilot which set up a fully automated workflow for generalising 1:50k map from TOP10NL data was to show the potentials of automated generalisation for Dutch landscape and data. In addition the pilot identified topics for further research such as enhancement of TOP10NL database to better support automated generalisation (see Section 3). Finally the pilot provided insights into how the products of fully automated generalisation differ from current products and the consequences this has, also considering the gain in consistency and efficiency.

The generalisation pilot built on the research of Stoter et al (2009) and Van Smaalen and Stoter (2008). These studies defined and implemented an action plan for automated generalisation of 1:50k map from TOP10NL data based on generalisation guidelines as used by cartographers (Kadaster, 2005). The internal generalisation pilot optimised the action plan by experimentally determining the optimal tools, algorithms and parameter values for each step in the action plan. The different steps were prototyped with different alternatives and steps were implemented in different orders to identify the optimal implementation. The pilot made use of generalisation techniques that recently have become available in Esri ArcGIS software (version 10), such as displacement of infrastructure and buildings (Punt and Watkinson, 2010).

At the end of the five weeks pilot, the optimised action plan was implemented as one automated workflow in ArcGIS model builder and run on the test area. The implemented workflow is summarised in Section 2.1. Section 2.2 describes the main conclusions.

2.1 Implemented workflow

The workflow consists of three parts:

i. Data generalisation aiming at reducing the data that has to be visualised, most important steps:
    - TOP10NL codes are recoded into TOP50 codes
    - Urban areas are detected based on building density to distinguish between urban and rural area.
    - Areas with many buildings (i.e. coverage > 17%) are converted to built-up areas. Symbols for important buildings (schools, hospitals, churches) are kept.
    - TOP10NL road polygons are collapsed to road centrelines by converting TOP10NL road centrelines (additional information in TOP10NL for representing every road lane) into TOP50 centrelines (where every road has one centreline)
    - The road network is pruned with a thinning road algorithm. Before the pruning, cycle paths parallel to roads and non-paved dead-ends shorter than 1 km in rural areas are automatically detected and deleted. In addition free cycle paths and access roads to buildings are detected and kept.
    - Land use parcels are extended to the new road centrelines to fill up the area originally covered by a road polygon.
    - Water network is pruned by removing small water parallel to roads and by selecting the remaining free lying water.

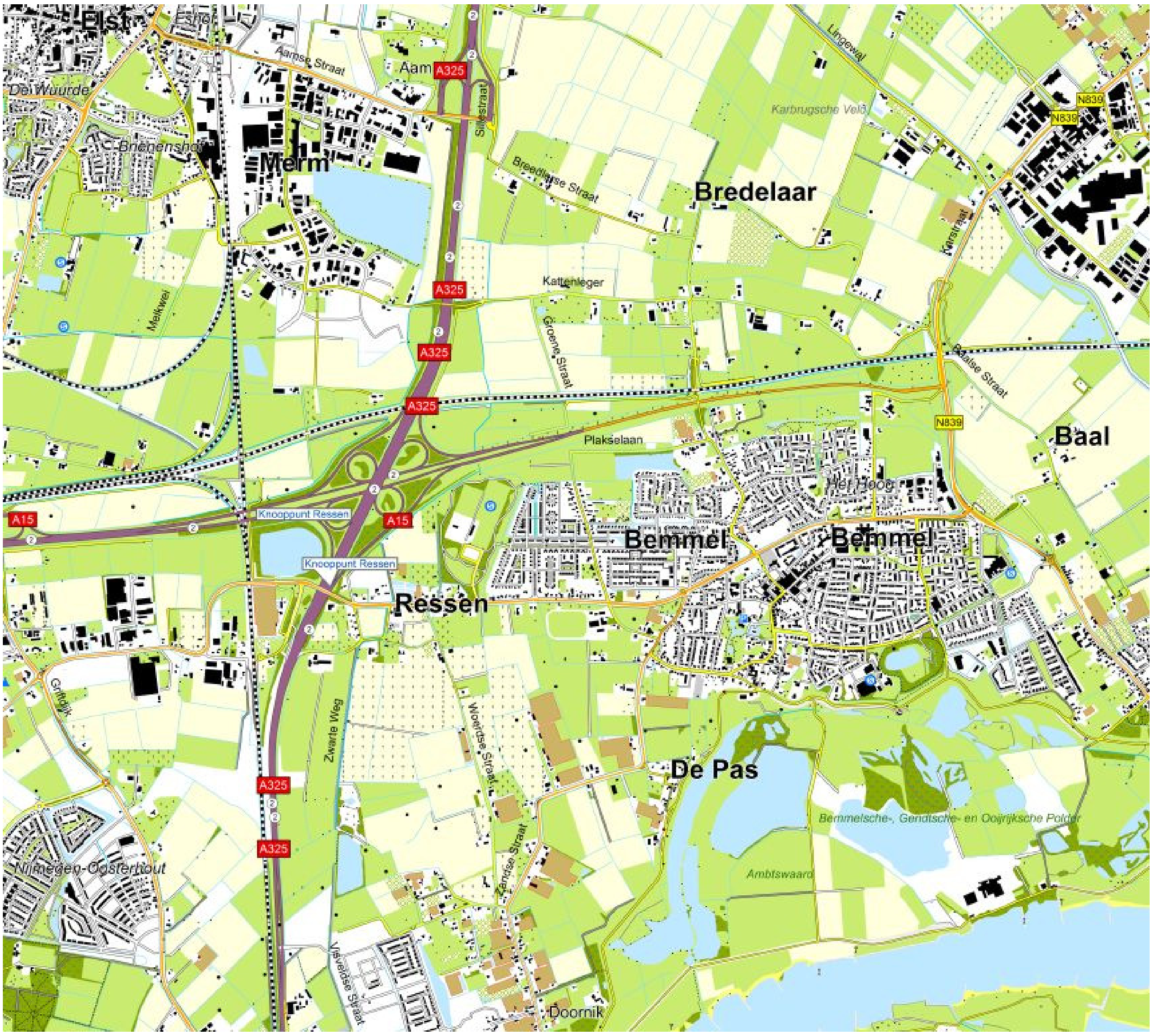
- The centreline of railways is detected as well as the outer railways and this information is used to calculate a thinned railway network .
- Too small water and land use polygons are deleted if they are isolated otherwise they are amalgamated.

ii. Symbolisation of the data

iii. Cartographic generalisation to solve cartographic conflicts of symbolised objects, most important steps:

- Symbolised dikes, railways, roads, water, terrain (both linear objects and boundaries of areas) are displaced.
- Remaining buildings (not converted into built-up areas) are simplified and displaced to avoid overlap and to meet a minimum building size.
- Land use polygons are recreated from the displaced linear object and the former terrain codes are assigned to the new terrain areas by an overlay with the old terrain areas.

The source data of the pilot (characterised by dense road network and relief), is shown in Figure 2.



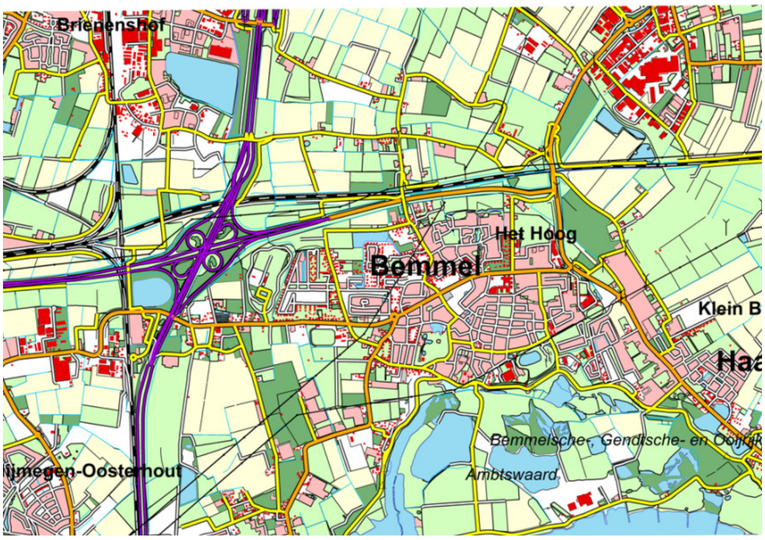
Source data (TOP10NL) in generalisation pilot

Figure 2 Source data of test area in generalisation pilot

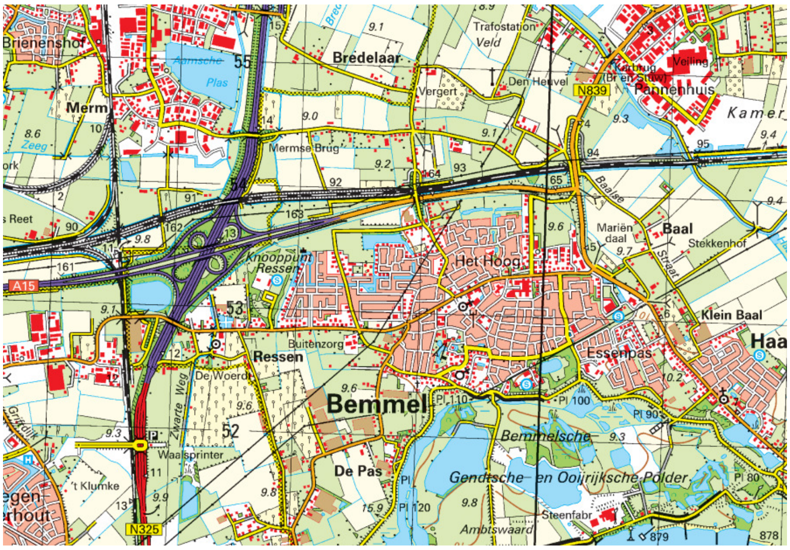
2.2 Results of the pilot

Figure 3 shows the outcomes of the pilot. It took 15 minutes to run this model for an area of 55 km$^{2.}$ The pilot results are further detailed in Stoter et al (2011). The results show that automated generalisation of smaller scale maps from TOP10NL data is feasible with the use of new techniques. However an important aspect of the success is to evaluate the result of the automated process within nowadays context which is the focus area of the pilot. A main requirement of today's use of topographic data is highly up-to-date maps at different scales. Therefore the requirement of up-to-dateness may get priority over some generalisation guidelines, although the results should still be of acceptable quality.

In line with this finding and based on the intermediate results, the implemented workflow ignored guidelines which appeared to be complicated to implement such as 'keep detached houses as single buildings in densely built areas' (instead these houses are converted into built-up areas as other buildings), 'treat ditches as separate objects' (instead they were kept as ordinary terrain boundaries, because of their limited importance and because their visualisation is identical to terrain boundaries); 'keep water smaller than 2 meters as linear objects' (instead small water bodies parallel to roads are eliminated to save space because we regarded them as unimportant) and 'generate built-up area only in that part of the land use polygon that is covered with buildings (instead a complete land use polygon was converted into built-up area when it is covered for more than 17% with buildings).



a: 1:50k map, obtained fully automatically; displayed at smaller scale



b: 1:50k map, interactively generalised;, displayed at smaller scale

Figure 3 Results of automated generalised map (a) and interactively generalised map of the same area (b)

The results from this pilot also identified issues that require further actions. Many of these issues are related to enhancing and enriching TOP10NL database as base data set for the automated generalisation workflow. These issues are:

- The collapse road dual to centrelines algorithm works well, but only if the road network can be built. In the conversion of Top10Vec to TOP10NL data (in 2005), the road network was built by creating road segments at crossings. However many of these segments appear to be encoded with "type=other" and therefore the network could be built over these segments. To solve this issues these errors have been identified and repaired in TOP10NL. Another improvement for this algorithm is to keep exits out of this operation. Therefore TOP10NL road segments have been enriched with this information.
- Pruning of the (artificial) water networks can be improved on two aspects which both require enrichment of the source data: pruning based on drainage pattern and pruning with a road thinning algorithm. For the first aspect drainage information has been added to TOP10NL (obtained from water boards) which showed promising results. For the second the source data need to form a complete water network, which is currently not the case for all water objects. Pruning water based on drainage pattern seems sufficient for 1:50k. However for 1:100k, further pruning of water is definitely required. In addition at scale 1:100k and smaller, collapsing narrow waterways (under 40m) to lines brings an extra challenge, specifically when the width of a water course is (a little) too small for only a part of its course (what to do here: collapse for a small part or not?).
- Damming dikes are not available as single objects in TOP10NL, although important for generalisation (i.e. they may never be displaced). Instead these can only be recognized by hatches on the map. Because the main dikes are predominant objects, these type of dikes have been identified in TOP10NL data.
- The proper placement of buildings in a limited amount of space can be improved. For example by assigning lower hierarchy to small (can be calculated) and unimportant buildings, for example sheds versus houses (to be retrieved from external sources).
- The generalisation of small areas of forest and water which are surrounded by roads (or water) needs improvement. Since these areas are often important for orientation on the road, they cannot always simply be eliminated.
- The generalisation of additional objects as isolated forest (tree lines) and power pylons needs specific attention to make sure they still fit with the displaced objects.
- Also generalisation of annotation requires specific attention because of its complexity. The solution is sought in a separate database with topographical names that also contains scale-related information, e.g. information on every name on if and how it is portrayed at each scale.
- Generalisation of fly-over crossings could be improved by taking the (relative) heights into account.

The results of the internal generalisation pilot look promising and therefore a follow up study has started on the feasibility of a fully automated production line. This follow up study further evaluates the pilot results. It also scales and validates the results by applying the workflow on other test areas and other scales (at first 1:100k). Experiences so far have identified issues to be considered for the new production line. The next section will elaborate on these issues.

3 CONSIDERATIONS FOR THE NEW PRODUCTION LINE

Based on experiences of the current production line and the results of the generalisation pilot, the Kadaster defined the following starting points for the new production line that is currently investigated (the study will reveal the feasibility of these points):

- Only at the most detailed level both a database and map will be produced, where as topographic information at smaller scales is considered for visual applications only (for orientation and human navigation). The consequence is that the smaller scale data are only cartographic

representations (on paper or screen) and therefore disruptions of the geometry for meeting cartographic requirements do not have to be controlled. Model generalisation (i.e. generalising the data without considering the symbolisation) will be applied, but only to prepare the data for the optimal visualisation. In addition a study with the customers is being set up to identify data that is required at smaller scales, such as a road network. This can be an additional product.
- For keeping consistency and enabling automated update propagation, as well as for correct visualisation, we do consider well structured data underlying the maps an important prerequisite. Therefore obtaining a planar partition is currently being studied (this was not yet achieved in the pilot).
- The aim of the automated process is not to replicate the small scale maps and therefore we allow ourselves some freedom to diverge from existing cartographic generalisation guidelines. We do use these guidelines to copy the important principles, but at the same time we may reconsider the guidelines in case our experiments show reasons for this.
- At first, the scope of the multiscale products will be limited to the static visualisations on screen or paper. Dynamic generalisation for Internet use (real time and not per se related to the predefined scales) has to meet different requirements but has also different solutions and is therefore outside the scope of this research.
- Generalisation without any interaction is the best guarantee for efficiency and consistency. Therefore the new multiscale products at predefined scales should be generalised fully automatically.

With tests that extend the 1:50k generalisation pilot to other test areas (such as an area with complicated crossings and dense parcel boundaries; an area with dense water network and an urban and industrial area) as well as to other scales the feasibility of a fully automated production line is currently being tested. Issues that we have identified until now are:
- The base data set, built on TOP10NL, should be enhanced and enriched with aspects as identified in Section 2.
- At smaller scales information should be added that is only available in smaller scales (exits, roundabouts) .
- A correct result of the collapse dual to centrelines can be reused at smaller scales and should therefore be part of the base data set.
- An automated generalisation process for the whole of The Netherlands at once is not feasible. Therefore the workflow will implement fully automated generalisation of parts of The Netherlands. An update will cause a full replacement of such a part. Currently we have generated a partition based on features that must never be displaced (mainly large waterways, roads and railways), resulting in about 400 areas. Other objects such as secondary roads in different partitions should still fit after the generalisation. This requires further investigation. A simpler to implement but less advanced solution is to divide the country into tiles while cutting objects at tile boundaries. Although this would mean a decline of the seamless TOP10NL database, it will provide major potentials for automated processing of the whole country. See also Chaudry and Mackaness (2008) and Touya (2010) for relevant partitioning to support automated generalisation.
- Further study should identify whether the optimal way for deriving smaller scale products is a ladder, a star or a mixed approach (Eurogeographics, 2005).

The conceptual schema in Figure 4 shows the role of the base data set in the automated generalisation production line.

The workflow starts with base data set (i.e. TOP10Basis) that is the result of translated TO10NL data. In a next step this input data is validated on attribute, geometry and topology. Problems are resolved either in the input-datasets (before it is inserted as in the case of roadsegment preparation mentioned before) or as generic action performed on the inserted data. After the geometry, topology and geometrical errors have been solved, the database is enriched with information as described above (e.g. urban/rural extent (calculated); collapsed road centrelines (calculated); richer attributes for roads and water (obtained from external sources)). The object lifecycle identification completes the preparation of a generic base dataset. This includes a scheme conversion from the base set to an

information model that supports all map scales and adding attributes to be able to follow the lifecycle of objects throughout the map scales.
The resulting dataset that is the input for automated generalisation workflow is called *TopActueelbasis* (*Actueel* means up-todate), because it is prepared to serve automated generalisation of all map scales to quickly obtain small scale maps of updated 1:10k data. This conceptual workflow is currently prototyped, tested and refined.

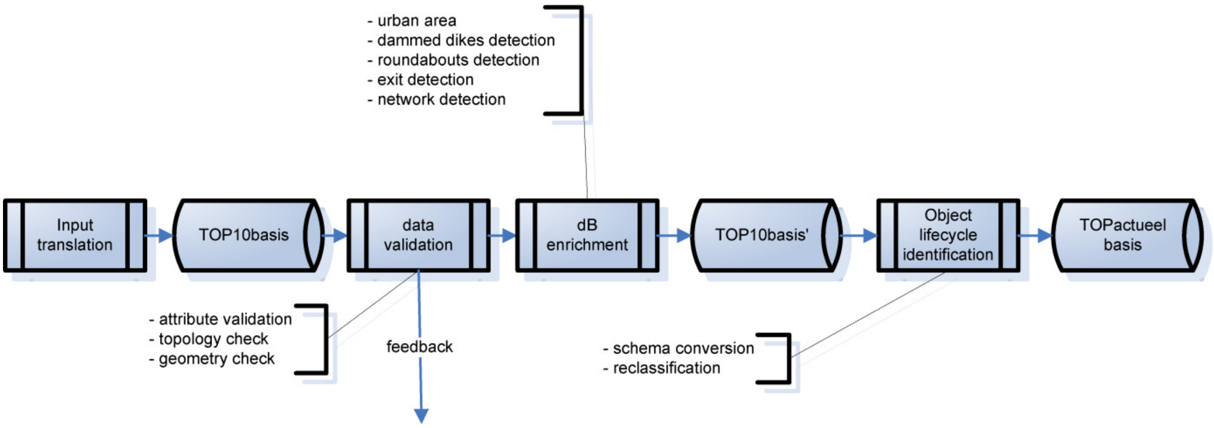


Figure 4: Conceptual workflow of automated generalisation production line (under study).

4 WORK IN PROGRESS

This paper describes the ongoing study of Kadaster to reconsider the production line of multiscale products. Main principle of the new production line is ***fully*** automated generalisation to meet (at least) the 2- year update cycle in an cost-effective manner. Another principle is the reconsideration of current products to meet new requirements with new technologies. Therefore customizers of multiscale data and maps are closely involved in formulating acceptance criteria for the new products.
The research on the new production line gets high priority within the Kadaster and therefore a research team is dedicated to this study to realise the new production line mid 2012.
One of the next steps of the study is the implementation of an automated generalisation workflow to obtain a 1:100k map for the whole country, which will also address the remaining issues from the 1:50k pilot. Because of its actuality (i.e. 2004) and because of the promising results of the generalisation pilot on 1:50k, automated generalisation is considered as the only option to obtain an updated 1:100k map within a limited amount of time. The implementation of the workflow for 1:100k generalisation into a real production line will yield valuable knowledge for implementing an automated generalisation production line from start to finish.

REFERENCES

AdV. 2007. *Produktblaetter Arbeitskreis Geotopographie*, Arbeitsgemeinschaft der Vermessungsverwaltungen der Laender der Bundesrepublik Deutschland.

Baella, B. & Pla, M.. 2005. Reorganizing the topographic databases of the Institut Cartographic de Catalunya applying generalization. In Proceedings of the 8th ICA workshop on Generalisation and Multiple Representation. A Coruña, Spain.

Chaudry, O.Z. and W.A. Mackaness. 2008. Partitioning Techniques to Make Manageable the Generalisation of National Spatial Datasets, In: Proceedings of the 11th ICA workshop on generalisation and multiple representation, 20-21 June 2008, Montpellier

EuroGeographics. 2005. Generalisation Processes: a benchmark study of the expert group on quality, EuroGeographics. Available at: http://www.eurogeographics.org/eng/05_quality_reports.asp.

Foerster,T., J.E. Stoter and M Kraak. 2010. Challenges for Automated Generalisation at European Mapping Agencies: A Qualitative and Quantitative Analysis, In: The Cartographic Journal, Volume 47, 1, pp. 41-54

Kadaster. 2005. Generalisatievoorschriften TOP50vector, (Generalisation Regulations), Topografische Dienst, Emmen.

Lecordix, F., Gallic, J.L., Gondol, L. & Braun, A.. 2007. Development of a new generalization flowline for topographic maps. In 10th ICA workshop on Generalisation and Multiple Representation. Moscow.

Punt, E. and D. Watkins. 2010. user-directed generalization of roads and buildings for multi-scale topography, 13th ICA Workshop on Generalisation and Multiple Representation, 2010 Zurich

Stoter, J., R. Nijhuis1, J. van Smaalen, A. Dortland, J. Bulder, B. Bruns (2011), Fully automated generalisation of topographic data in current geo-information environments, paper accepted for Urban Data Management Symposium, 28-30 September, Delft, The Netherlands, 2011.

Stoter, J.E., J. van Smaalen, N. Bakker, P. Hardy. 2009. Specifying map requirements for automated generalisation of topographic data, The Cartographic Journal Vol. 46 No. 3 pp. 214–227 August 2009

Touya, G.. 2010. Relevant Space Partitioning for Collaborative Generalisation, 13th Workshop of the ICA commission on Generalisation and Multiple Representation, 2010 Zurich

Van Smaalen, J.W.N. and J.E. Stoter. 2008. Automatische generalisatie van TOP10NL naar TOP50NL haalbaar?. In: Geo-info: tijdschrift voor geo - informatie Nederland (GIN), 5 (2008)2, pp. 48-53