Automated generalisation in production at Kadaster NL

Vincent van Altena[1], Ron Nijhuis[1], Marc Post[1], Ben Bruns[1], Jantien Stoter[1,2]

[1] Kadaster, The Netherlands, email: firstname.secondname@kadaster.nl
[1,2] OTB, Delft University of Technology, The Netherlands, j.e.stoter@tudelft.nl

Abstract. This paper presents the implementation of a fully automated production workflow to generalise a 1:50k map from 1:10k data. The feasibility study for this workflow started in 2010 and has led to a production of a countrywide 1:50k map in 2013. From that moment on, the automatically generalised 1:50k map will replace the existing one. Because of the limited time needed to generalise the 1:50k map series from 1:10k data, a new 1:50k map update is foreseen with every new release of 1:10k data, i.e. five times a year.

Keywords: automated generalisation, cartography, multiscale topographic data

1. Introduction

In 2010 the Netherlands' Kadaster, who also holds the national mapping agency, started a feasibility study to introduce automated generalisation in its map production line. The study was motivated by the encountered problems to meet the legal obligations of Kadaster within available budgets. The Kadaster is legally obliged to produce 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 (or shorter). To meet this obligation Kadaster has converted its vectorised maps into object-oriented databases since 2007. The interactively update (i.e. generalisation) of these products by cartographers takes too much time (and consequently costs) to meet the obliged update cycle within available budgets. For small-scale maps this problem is even bigger since maps at every scale are not generalised directly from the 1:10k source, but in steps from the next larger scale map in a ladder approach. Therefore small-scale maps require even longer time before they are updated. The time-consuming update process also has limitations to produce on-demand products (i.e. different products for different demands).

The feasibility study to meet these problems has led to a fully automated workflow to produce 1:50k maps, which will be practiced from 2013. The following sections describe the Kadaster approach on introducing generalisation in production. Section 2 describes the generic characteristics of the automated generalisation approach, Section 3 describes implementation details and Section 4 closes with results, findings and future plans.

2. The Kadaster automated generalisation approach

The feasibility study on automated generalisation firstly focused on the workflow from object oriented 1:10k data (called TOP10NL) to a 1:50k map. Both source and target data cover the complete skin of the earth (no gaps or overlap). The source data is shown in Figure 1. Now the feasibility study has been developed into a production line, the generalisation from 1:100k map from 1:10k data is under development.

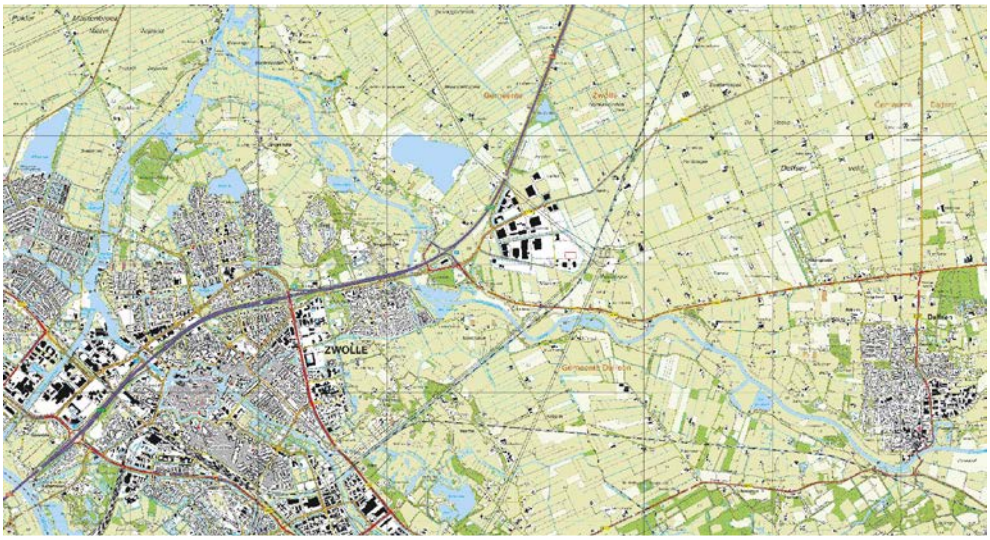

Figure 1 Source data (1:10k), displayed at smaller scale

From the beginning it was clear, that the aim of the automated generalisation workflow should not be replicating the existing map. This had several reasons.

Firstly, legacy topographic products – with an origin of often more than 60 years ago - may overemphasis past (cartographic) requirements and may ignore new requirements of multiscale topographic information, e.g. topographic information is used in much more applications by a wider public than ever before and the user may prefer up-to-date maps over maps that meet all traditional cartographic principles (although the results should still be of acceptable quality). In addition automating a previously interactive process, which was designed for a past technical and organisational context, has appeared to be very complicated (Foerster et al, 2010; Stoter et al, 2009b). Another reason to reconsider existing map requirements is that the time-saving aspect of automation makes the process well suited to produce various products for different demands. Therefore the requirements of automatically generated multi-products may differ from the requirements of the existing single product that should fit all uses.

A few other aspects further refined the scope of the Kadaster approach:

- The focus is on producing a map. Therefore disruptions of the geometry to meet cartographic requirements do not have to be controlled apart from assuring their consistency in the resulting map, i.e. roads and water should still form complete networks after generalisation. To accomplish this, the workflow performs generalisation using a ‘smart‘ partitioning, instead of the existing approach of generalising within map sheets (see Section 3.4).
- Generalisation without any interaction is the best guarantee for efficiency and consistency and the only way to produce multiple on-demand products. Therefore we do not allow that the automatically achieved results are interactively improved afterwards.
- The most straightforward way for updates in an 100% automated generalisation workflow is to completely replace the old version map. This is in line with Regnauld (2011). Therefore at this moment, we do not maintain links between the objects at the different scale levels. If these links are required, this will be part of a subsequent study.

Since we reconsider existing map specifications based on the potentials of technologies, one of the main challenges was how to redefine specifications for automated generalisation taking existing guidelines as starting point while assuring that users requirements are met. We have accomplished this in the following way.
Firstly, we generated an initial 1:50k map in a semi-automatic manner by extending the work of Stoter et al (2009a; 2012). The aim of this first step was to see how much automation we can achieve with currently available tooling and some self-developed algorithms. This work implemented existing generalisation guidelines for interactive generalisation in an automated process and improved the implementation by evaluating intermediate results.

The initial map was sent to a selection of main customers of the current 1:50k map, who are formally organised in a users group, to test the main principles and assumptions. Based on obtained insights the process was improved and refined and implemented as one integrated workflow. In a next stage the evaluation and improvement process was repeated, by asking more costumers to assess the resulting map in more detail and for different types of areas, i.e. 'Relief and dense road pattern'; 'Complicated crossings and dense parcel boundaries'; 'Dense water network' and 'Urban and industrial area'.
Based on those evaluations and iterative testing, the optimal sequence of steps was determined as well as the most appropriate algorithms and parameter values for each step, integrated in one workflow.
Interestingly, the evaluations showed that the customers appreciated the "same appearance of the map" less than "more frequent update cycles". Indeed, they confirmed to be very pleased that updated 1:50k maps will appear two years earlier than currently is the case as well as that the 1:50k maps will be 100% consistent with the 1:10k source data because of the synchronised releases. Another interesting observation was that some results of the automated generalisation were appreciated higher than the results of interactive generalisation. For example the automatically thinned road network appeared to be more appropriate for navigation than the interactively thinned road network. Finally several respondents were satisfied by improved uniformity of the whole map.

3. Implementation details

This section describes the used software (section 3.1), pre-processing of the data (section 3.2), the implemented automated generalisation workflow (section 3.3) and finally the approach that was applied to automatically generalise the whole country (section 3.4).

3.1. Used software and technology

For the implementation we use a mixture of standard ArcGIS tools, self-developed tools within Python and a series of FME tools. ArcGIS contains some specialised generalization tools for collapsing two lanes of a road into a single road line, displacing symbolised geometries, simplifying of (symbolised) buildings and thinning of networks, see Punt and Watkins (2010). The complete generalization workflow is implemented within the Model builder tool of ArcGIS. The workflow consists of three main models, consisting of about 200 sub models that are responsible for each specific generalization problem that we need to solve in the process.

3.2. Pre-processing the data

Since the aim is 100% automation, the process should cover as many generalisation aspects as possible. This is accomplished by either improving the process step-by-step, or - if that did not work - by improving and enriching the source data. Besides correcting (hidden) errors (resulting in TOP10NL basis), the enrichment of the source data (resulting in TOP10Extra) is done in two ways. Either external data sources are used or the required knowledge is made explicit by computation. Examples of enrichments of the input data are determining urban extents by defining areas with high density of buildings (i.e. higher than 10%) and attributing TOP10NL road segments with information on exits to better control the process that generalises the road network from the TOP10NL roads.

3.3. Implemented workflow

The implemented automated generalisation workflow consists of the following steps:

(1) Model generalisation aiming at reducing the data that has to be visualised. This is the largest part of the process. Model generalisation is not only conversion of geometric objects to the lower density and structure of the TOP50 model, but also translation and reclassification of attributes to the TOP50 model. The main operations are replacing road polygons with road centrelines; merging individual road lanes into single line geometry; pruning the road and water network; and, generalisation of small land use areas.

(2) Symbolisation of the data. The symbolisation process assigns symbols to all geometries, as they should appear on the map. In this process basis symbols are used which exactly correspond to shape and outline of portrayed features, but which lack all cartographic refinement. Sophisticated symbolisation is postponed to a later stage in the process. The symbolisation may result in objects that appear larger on the map than they are in reality and in overlapping objects. This is solved in the next step.

Graphic generalisation to solve cartographic conflicts of symbolised objects. The graphic generalisation process consists of simplifying, typifying and displacing buildings and displacing the linear objects (roads, water) and boundaries of symbolised water and terrain objects as well as all other point and linear objects (i.e. administrative boundaries, height contours, engineering constructs). At the end of the process polygon-objects are rebuilt from the displaced boundaries and former codes are assigned to the new areas by using left/right information of the boundaries.

3.4. Countrywide coverage

To be able to generalise a map for the whole of The Netherlands, the workflow is applied on about 400 generated partitions. These partitions were generated using linear objects that must never be displaced, which are highways and main roads. In contrast to map sheet boundaries, these boundaries also appear in the real world and they hardly clip any objects. Near the coast, where these linear road objects are missing, artificial partitions have been made. Besides some global operations that are applied for the whole country (such as creating and simplifying the power line network), the workflow is applied per partition and partitions are connected afterwards. Because vertices of objects at and near partition boundaries are prohibited to move in the displacement process, the objects at neighbouring partitions still fit after generalisation.
The generalization process of the 1:50k map from 1:10k source data for the whole country can be achieved in 50 hours. This is realised through the multiprocessing capabilities of ython, which allows parallel processing of six partitions on each of the six available systems.

4. Results, findings and future plans

Figure 4a shows the 1:50k map that is generalised fully automatically with the resulting workflow from the data shown in Figure 1. Figure4b shows the interactively generalised version for a global comparison.

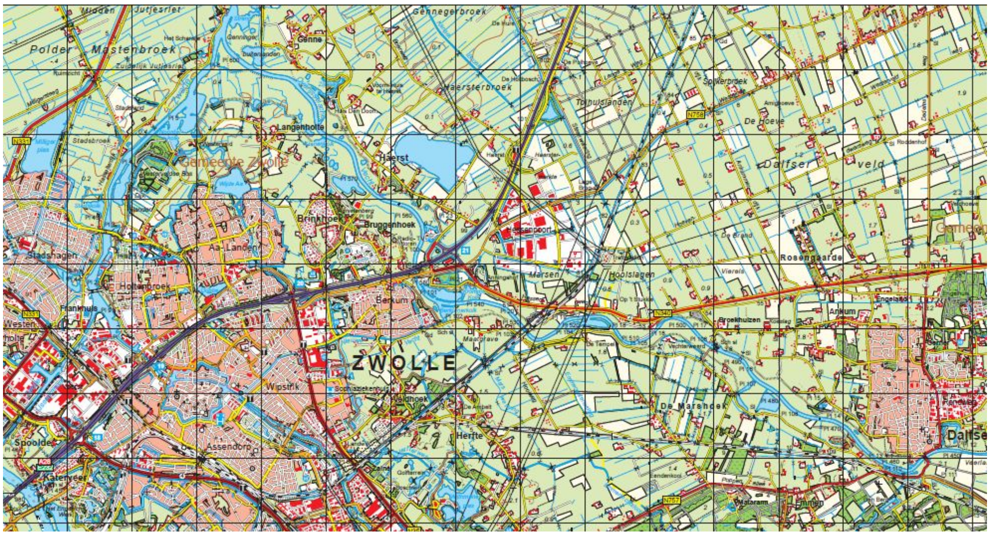

Figure 4a: 1:50k map, obtained fully automatically; displayed at smaller scale

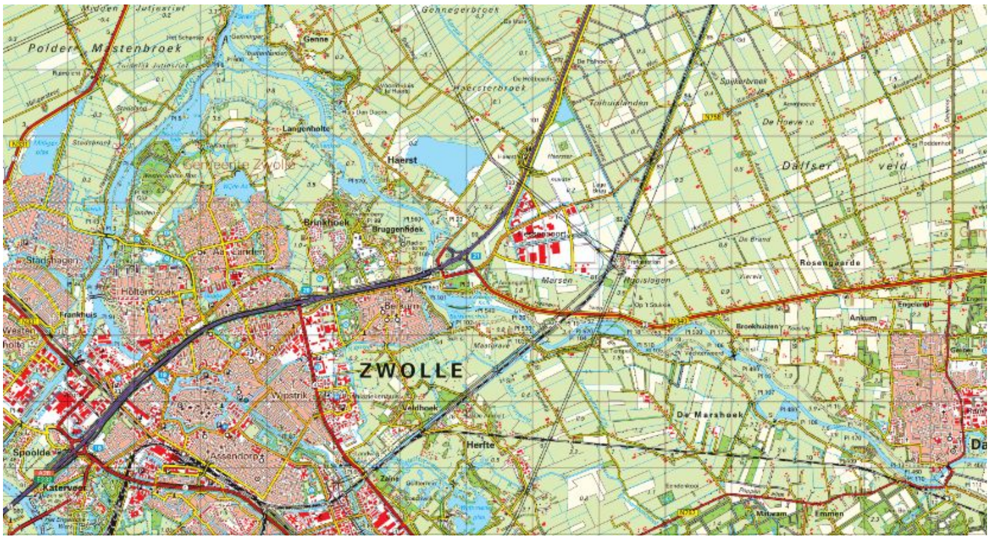

Figure 4b: 1:50k map, interactively generalised; displayed at smaller scale

Based on the results and very good users' evaluations the Kadaster decided that a fully automated generalisation workflow is the most sustainable workflow for the future as well as they only way to produce products on demand.
As mentioned above, the thirty-six parallel Python processes perform the core generalising process of the complete Netherlands within approximately 50 hours. Including pre-processing, generalisation, visualisation and printing the whole turnaround is 3 weeks (at most) for the whole country. Therefore a 1:50k update is foreseen with every new delivery of TOP10NL (five times a year).

Based on the experiences with the new 1:50k product, the automated generalisation approach is currently being extended to the 1:100k map and to on-demand products, such as the backdrop map at multiple (between 7 and 16) scales for the national geo-portal.

5. References

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

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

Regnauld, N. (2011). OS Vectormap district: automated generalisation, text placement and conflation in support of making pubic data public, presented at the 25th International Cartographic Conference, July, 2011, Paris. Available online:

http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/D3-Generalisation/CO-358.pdf

Stoter, JE, Smaalen, J van, Nijhuis, R, Dortland, A, Bulder J & Bruns, B (2012). Fully autmated generalisation of topographic data in current geo-information environments. In S Zlatanova, H Ledoux, EM Fendel & M Rumor (Eds.), Urban and Regional Data Management - UDMS Annual 2011 (pp. 109-120). Leiden: CRC Press - Taylor & Francis Group.

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

Stoter, J.E., D. Burghardt, C. Duchêne, B. Baella, N. Bakker, C. Blok, M. Pla, N. Regnauld, G. Touya, S. Schmid (2009b). Methodology for evaluating automated map generalization in commercial software, Pages 311-324 In: Computers, Environment and Urban Systems Volume 33, Issue 5, September 2009.