

Automatische generalisatie voor nieuwe productielijn Kadaster

Van links naar rechts:
Jantien Stoter, Kadaster en TU Delft
Ron Nijhuis, Kadaster
Jan Bulder, Kadaster
Ben Bruns, Kadaster



Op dit moment worden de topografische producten van het Kadaster op schaal 1:10.000, 1:50.000, 1:100.000 en 1:250.000 - die te samen de Basisregistratie Topografie (BRT) vormen - apart van elkaar geproduceerd en beheerd. Kartografen werken de producten op schaal 1:50.000 en kleiner bij op basis van het 1:10.000 product (TOP10NL) volgens generalisatierichtlijnen. Dit vergt relatief veel tijd, te meer omdat de kleinere schalen stapsgewijs van elkaar worden afgeleid (iedere schaal van de schaal eronder). Om de productielijn efficiënter maar ook consistentere te maken, is een onderzoeksteam van het Kadaster sinds 2011 bezig met een studie naar de implementatie van automatische generalisatie. In eerste instantie voor de productie van kaarten op schaal 1:50.000 en 1:100.000. (NB: De 1:25.000 kaart is een verrasterde versie van TOP10NL en valt daarmee buiten het generalisatieproces).

Het actueel houden van de BRT kost gemiddeld 500-2000 uur voor een 1:50.000 kaartblad van 500 vierkante kilometer (in totaal zijn er 93 kaartbladen). De actualiteit en de kostenbeheersing van de BRT zijn voor het Kadaster dan ook belangrijke motivaties geweest om automatische generalisatie te bestuderen als alternatief voor de huidige productielijn. Het vraagstuk van automatische generalisatie is al jaren een populair wetenschappelijk onderzoeksonderwerp, maar er zijn ook externe ontwikkelingen die automatische generalisatie tot een realistische optie maken. Ten eerste is recentelijk gebruikersvriendelijke, kartografische generalisatiefunctie beschikbaar gekomen in commerciële software. Dit maakt het interessant om te kijken naar de haalbaarheid hiervan voor de productielijn van het Kadaster. Ten tweede is het gebruik van informatie in het algemeen en geoinformatie in het bijzonder de laatste jaren sterk gewijzigd. Topografische informatie

wordt door veel meer mensen en voor veel meer en bredere toepassingen gebruikt dan tien jaar geleden. Bovendien is de behoefte aan actuele en consistente informatie groter dan ooit. Deze behoefte zal in de toekomst alleen maar toenemen.

Niet nabootsen van bestaande kaarten

Om de potenties van automatische generalisatie toekomstgericht te bestuderen, beperkt het Kadaster zich niet tot het namaken van de bestaande topografische producten. Ten eerste omdat de specificaties van de bestaande kaarten vaak meer dan 60 jaar oud zijn en dus zijn vastgesteld op basis van gebruikerseisen van destijds. Nu het Kadaster een nieuwe productielijn ontwikkelt, benut het de kans om de bestaande specificaties te herzien vanuit nieuwe gebruikerscontexten. Ten tweede is het erg ingewikkeld om een product

dat tot nu toe op basis van interpretatie door een kartograaf wordt gemaakt, volledig identiek te genereren door een geautomatiseerd proces. Een laatste reden om de bestaande specificaties te herzien, is dat het geautomatiseerde proces parametergestuurd is en daardoor (secundair) gebruikt kan worden om producten 'op maat' (on-demand) te produceren voor verschillende toepassingen. De eis 'one product fits all uses' omdat er veel interactie aan te pas zou moeten komen is daarmee niet meer valide.

Principes voor automatische generalisatie

Omdat al vroeg duidelijk was, dat het Kadaster zich niet uitsluitend zou richten op het namaken van de huidige kaarten, was het erg belangrijk om expliciet te maken, wat het dan wél wilde bereiken. Er werd daarom als leidraad een lijst met principes opgesteld:

- De behoefte aan actuele informatie kan voor gebruikers in sommige situaties belangrijker zijn dan een product dat voldoet aan alle kartografische principes. In de praktijk betekent de introductie van automatische generalisatie dat een 1:50.000 kaart twee weken na de inwinning van TOP10NL beschikbaar zal zijn i.p.v. de huidige 4, 6 of 8 jaar. De huidige generalisatie-richtlijnen worden wel gebruikt om de belangrijke kartografische principes toe te passen, maar tegelijkertijd wordt de vrijheid genomen om van deze regels af te wijken als de experimenten daar reden toe geven. Bijvoorbeeld als het te moeilijk is een bepaald principe te automatiseren. Uiteraard is het wel belangrijk dat de resultaten nog steeds van acceptabele kwaliteit zijn.
- Niet het Kadaster (als producent), maar de gebruiker moet bepalen wat een acceptabel product is. Daarom is de gebruiker vroeg in het ontwikkelproces betrokken.
- Generalisatie zonder enige interactie is de enige optie voor de realisatie van efficiënte, consistente en on-demand producten. Daarom is de ambitie 100% automatische generalisatie.
- Landsdekkende, topografische informatie die alle klassen integreert in één bestand is op kleinere schalen alleen nodig voor visualisatie. Tegelijkertijd kunnen TOP10NL gegevens voldoen aan de behoefte aan landsdekkende, topografische informatie nodig voor GIS-bewerkingen. Voor specifieke thema's (klassen) kan er wel behoefte zijn aan kleinschalige objectgegevens zoals bijvoorbeeld gegeneraliseerde water- en wegennetwerken. Voor die thema's zouden afzonderlijke, objectgerichte producten kunnen worden gemaakt.
- De focus van het gegeneraliseerde product ligt dus op visualisatie die 1 op 1 samenhangt met de gebruikte symbolisatie. Dat betekent dat de gebruikers zich bewust moeten zijn op welke wijze zij de punt-, lijn- en vlakgeometrie kunnen gebruiken. Gebruikers zouden de kaartgebaseerde

Door automatische generalisatie is kleinschalige kaart al beschikbaar 2 weken na herziening van grootschalige kaart

gegevens immers onbedoeld op een verkeerde manier kunnen gebruiken. Bijvoorbeeld voor ruimtelijke analyses of voor een kaart met andere symbolisatie of tekenvolgorde van kaartlagen. Dit kan resulteren in onverwachte (incorrecte) visualisaties.

Deze beperkingen van de geometrie zullen duidelijk worden vastgelegd in de nieuwe productspecificaties.

- In eerste instantie wordt gekeken naar de schalen 1:50.000 en 1:100.000 omdat automatisering hier de meeste besparing oplevert ten opzichte van het interactieve proces. Bovendien worden momenteel de kaartspecificaties voor de schalen 1:250.000 en 1:1 miljoen geharmoniseerd op Europees niveau (EuroRegional Map respectievelijk EuroGlobal Map) waardoor het onpraktisch is om op dit moment deze producten te herzien.
- De meest eenvoudige manier voor wijzigingen is het vervangen van de complete oude versie met een complete nieuwe versie van het betreffende gebied. Het is daarom niet nodig om relaties tussen brongegevens en gegeneraliseerde gegevens te behouden.

Omdat het Kadaster snel met resultaat wil komen, werd beschikbare commerciële software gebruikt en ook voortgebouwd op resultaten uit eerder onderzoek (zie Geo-Info 5 (2008|2)). Deze resultaten zijn inmiddels geïmplementeerd in ArcGIS 10 (uitgebracht in 2010) dat daarom is gebruikt voor de implementatie.

Aanpak

Vanwege de complexiteit en het onverwachte karakter van het generalisatieproces (als je ergens in het proces iets verandert, weet je niet precies waar dat effect zal hebben) is het basisprincipe van de aanpak heel

simpel: 'al doende kijken wat er goed en fout gaat; het goede behouden en het foute verbeteren'. De huidige generalisatie- en symbolisatie-richtlijnen (in eerste instantie van 1:50.000) dienden daarbij als uitgangspunt. Bij dit proces heeft een senior kartograaf van het Kadaster geholpen om de generalisatie-richtlijnen te verduidelijken en alternatieven te vinden als bleek dat een bepaalde richtlijn moeilijk te implementeren was. Door het implementeren, het iteratief testen van deelstappen en het zorgvuldig evalueren van tussentijdse resultaten is de optimale implementatie van één geïntegreerd werkproces bepaald. Dat wil zeggen: de beste volgorde van stappen als ook de meest geschikte algoritmes en parameterwaarden voor elke stap. Uitgangspunt was steeds: het bereiken van 100% automatische generalisatie. Dat kon enerzijds door het proces te verbeteren. Maar vaak bleek ook dat de basis (TOP10NL) verrijkt moest worden om betere resultaten te krijgen, bijvoorbeeld door het toevoegen van op- en afrit-informatie over wegen. Uiteindelijk leverde deze aanpak een werkproces op waarmee op volledig automatische wijze een 1:50.000 kaart uit TOP10NL wordt gegeneraliseerd. Het uiteindelijke werkproces bestaat uit drie onderdelen:

1. Gegevensgeneralisatie gericht op het versimpelen en uitdunnen van de objecten die moeten worden gevisualiseerd
 2. Symbolisatie
 3. Kartografische generalisatie om de kartografische conflicten van gesymboliseerde objecten op te lossen
- Het uitdunnen van de gegevens bestaat met name uit de volgende stappen:
- TOP10NL hartlijnen (een extra laag die beschikbaar is naast de TOP10NL wegpolygonen) samenvoegen tot TOP50NL hartlijnen. In TOP10NL is iedere rijstrook een hartlijn, terwijl in TOP50NL een volledige weg één hartlijn is. Het gebruikte algoritme werkt alleen goed als de input een correct wegennetwerk is.



De afgeleide hartlijnen van de wegpolygonen in TOP10NL.



De gegeneraliseerde hartlijnen van de wegobjecten in TOP50NL.

Bij de omzetting van Top10Vector naar TOP10NL data (in 2005) is het wegennet gebouwd door het creëren van afzonderlijke wegsegmenten op kruispunten. Echter, veel van deze segmenten lijken te zijn gecodeerd met 'type = overig' waardoor het wegennet niet goed kan worden gebouwd. We hebben dit verbeterd. Ook is het kenmerk van op- en afritten toegevoegd aan TOP10NL om het resultaat van het algoritme te verbeteren.

- Terreingrenzen worden doorgetrokken naar de TOP50-hartlijnen. Als gevolg van het vervangen van de TOP10NL wegpolygonen door de hartlijnen, vallen er namelijk gaten die moeten worden toegewezen aan naastgelegen terrein.
- Converteren van gebieden met veel gebouwen (dat wil zeggen bedekking > 17%;

experimenteel vastgesteld) binnen stedelijk gebied en

buiten bedrijventerrein in 'bebouwd gebied'. De losse gebouwen worden hier verwijderd. De geïsoleerde gebouwen worden hier niet langer afzonderlijk afgebeeld. Belangrijke gebouwen, zoals

scholen, ziekenhuizen en kerken worden behouden als aparte puntobjecten om vervolgens als symbool gevisualiseerd te kunnen worden.

- Uitdunnen van het wegennetwerk, omdat er onvoldoende ruimte is om het TOP10NL wegennet op kleinere schalen af te beelden. Het gebruikte algoritme houdt rekening met connectiviteit, algemene karakter en een hiërarchie gebaseerd op wegtypen. Hierbij worden vrijliggende fietspaden behouden en alle fietspaden parallel aan wegen verwijderd.
- Waterwegennetwerk wordt uitgedund met hetzelfde algoritme als voor wegen, omdat ons kunstmatige waternetwerk meer kenmerken heeft van een wegennetwerk dan van een natuurlijk waternetwerk (waarvoor andere uitdunningalgoritmes beschikbaar zijn).
- Het hoofdspoor netwerk is bepaald en op rangeerterreinen tevens de buitenste sporen. Deze informatie wordt gebruikt om de spoorlijnen uit te dunnen.
- Geïsoleerde, kleine terreinvlakken worden verwijderd en groepjes kleine terreinvlakken worden samengevoegd (maar niet over harde terreingrenzen heen).

Nadat de uitgedunde geometrieën zijn gesymboliseerd, start het kartografische generalisatieproces.

Dit proces is erop gericht om kartografische conflicten - zoals te veel geometrische details en overlappings - op te lossen. Overlap ontstaat, doordat een aantal gesymboliseerde objecten op schaal 1:50.000 en 1:100.000 op de kaart (disproportioneel) groter zijn dan in werkelijkheid.

Het kartografische generalisatieproces bestaat uit de volgende stappen:

- De gesymboliseerde lijnobjecten (wegen en spoor) en de grenzen (lijnen) van de vlakobjecten (terrein en water) worden ten opzichte van

elkaar verplaatst. Het algoritme maakt gebruik van een hiërarchie van klassen. Sommigen klassen mogen verplaatst worden (terreingrenzen, water en onbelangrijke wegen) en anderen niet

of minder (snelwegen, spoorwegen, belangrijke wegen, belangrijke dijken).

- De overgebleven gebouwen worden versimpeld, vergroot en ten opzichte van elkaar en andere objecten verplaatst. Het toegepaste algoritme probeert hierbij het oorspronkelijke distributiepatroon te behouden.
- De terreinvlakken worden opnieuw 'gebouwd'. Met behulp van de verplaatste vlakgrenzen en met behulp van links/rechts coderingen worden de oorspronkelijke attributen weer aan de vlakken gehangen.

Resultaten en bevindingen

De input in het uiteindelijke werkproces bestaat uit de verschillende TOP10NL klassen en het resultaat wordt met één 'druk op de knop' gegenereerd. Voor een gebied van 55 vierkante meter rond Amersfoort duurt het ongeveer 15 minuten om het model te draaien. Het gegeneraliseerde resultaat ziet er veelbelovend uit. Ook de eerste gebruikersonderzoeken bevestigen dit. Het model wordt inmiddels op andere testgebieden toegepast. Voor ieder nieuw, gegeneraliseerd testgebied wordt gekeken of er nieuwe situaties ontstaan die niet voldoende worden opgelost door het huidige model en dus om verfijning vragen. Er wordt bijvoorbeeld gekeken naar gebieden met reliëf, dichte wegenpatronen, gecompliceerde kruisingen, veel terreingrenzen, dicht waternetwerk, stedelijke en industriële gebieden.

Openstaande knelpunten

Op dit moment is het Kadaster bezig de openstaande knelpunten op te lossen. Dit zijn punten waarvan bekend is dat die nog niet in het model worden meegenomen. Maar ook punten die naar voren (zullen) komen in de tussenevaluaties door gebruikers en de interne kwaliteitscontroles.

Deze interne kwaliteitscontroles lopen parallel aan het onderzoek. Voorbeelden van onderwerpen voor verder onderzoek zijn:

- De generalisatie van objectklassen als geïsoleerde bomen en inrichtingselementen (zoals hoogspanningsmasten en bomenrijen),

Gebruikers betrokken bij acceptatie van gegeneraliseerde eindresultaten

laag- en hoog-waterlijnen, administratieve gebieden en annotatie moet nog worden aangepakt.

- Sommige generalisaties verlopen nog niet goed zoals complexe fly-over kruisingen en aan elkaar grenzende wegen (deze laatste worden te ver van elkaar weggeduwd in het verplaatsingsproces).
- Soms is het beter om watervlakken om te zetten naar lijnen wanneer de watervlakten te smal worden (smaller dan 40 meter). Maar wat te doen als de waterloop slechts voor een klein stukje (een beetje) te smal is: de waterloop voor dit gedeelte met een lijngeometrie representeren of lokaal verbreden tot de minimale grootte?
- Op dit moment worden de gegevens alleen uitgedund ten behoeve van visualisatie. De vraag is of de geometrie ook uitgedund dient te worden om de producten bruikbaar te maken. Dit uitdunnen (filteren van vertices) zal niet altijd direct zichtbaar zijn in de kaart, maar kan er wel voor zorgen dat onjuiste geometrieën (zoals het over elkaar schuiven van datapunten) wordt tegengegaan.

Opdelen van Nederland

De uiteindelijke implementatie zal in deelgebieden gebeuren. Dat maakt zowel het werkproces als het uitleveren van wijzigingen (vervangen van gehele gebieden) haalbaar. Nederland is hiervoor opgedeeld volgens grenzen (lineaire objecten) die nooit verplaatst mogen



Nederland opgedeeld in 580 gebieden ten behoeve van de automatische generalisatie.

worden, zoals snelwegen. Aan de kust zijn ook andere objecten gebruikt om deelgebieden te maken. Dit resulteerde in de 580 gebieden. Deze gebieden worden momenteel een voor een gegeneraliseerd waarbij in eerste instantie wordt geprobeerd zo veel mogelijk verschillende soorten gebieden te processen.

Betrokkenheid van gebruikers

De nieuwe producten zullen een balans zijn tussen gebruikerswensen en technische mogelijkheden. Het gebruikersoverleg BRT (de formele vertegenwoordigers van BRT-gebruikers) is cruciaal bij het definiëren van de specificaties voor de nieuwe producten. Deze groep wordt periodiek gevraagd tussenresultaten te evalueren en te commentariëren. De deelnemers van het gebruikersoverleg proberen een veel bredere groep te bereiken door bijvoorbeeld enquêtes te verspreiden binnen GeoBusiness Nederland en de OGT (Overlegplatform Gebruikers Topproducten). Tevens wordt er in 2012 gestart met een BRT LinkedIn groep. Iedereen die dat wil, kan via deze media input leveren. Een belangrijk punt bij de gebruikersacceptatie is de nadruk op de visualisatie voor kleinschalige topografische producten. Daarnaast wil het Kadaster de nog openstaande knelpunten in overleg met de gebruikers oplossen: hoe erg is het als een specifiek knelpunt niet wordt opgelost en is er een goed alternatief voorhanden, dat makkelijker te implementeren is?

Kwaliteitscontroles

Afgezien van de externe, periodieke gebruikersconsultaties zijn ook interne kwaliteitscontroles ingebouwd. Immers, de nieuwe kaarten zullen nog steeds aan bepaalde basiscondities moeten voldoen om een 'echte' en 'leesbare' weergave van de werkelijkheid te zijn. Dit valt niet geheel door gebruikers te controleren. Bovendien ligt de verantwoordelijkheid hiervan bij de producent. Het Kadaster heeft twee typen kwaliteitscontroles opgezet. De eerste is een kwalitatieve controle waar kartografische experts het resultaat evalueren. Zij wijzen op kartografische fouten en op situaties die zij anders zouden hebben gegeneraliseerd.

Dat laatste hoeft niet per se erg te zijn, maar kan wel inzicht geven in de werking van het model. Daarnaast zijn kwantitatieve controles geïmplementeerd die statistieken berekenen na verschillende stappen in het proces.

Deze statistieken laten zien hoe de input en het resultaat na een stap van elkaar verschillen wat betreft het totale aantal objecten of oppervlakte van een bepaalde klasse. Momenteel worden metingen geïmplementeerd die iets zeggen over het gehele proces, zoals metingen die de reductiefactor (=mate van generalisatie) uitdrukken.

Hoe nu verder

Op basis van deze studie heeft het Kadaster besloten, dat een volledig automatische generalisatie de beste en waarschijnlijk enige mogelijkheid is om in de toekomst zowel een snelle wijzigingscyclus te realiseren (bijna tegelijkertijd met wijzigingen in TOP10NL) als een meer flexibele productielijn die producten 'on-demand' mogelijk maakt. Er zijn echter nog wel openstaande punten te onderzoeken. Het moment waarop de nieuwe productielijn zal worden ingevoerd is daarom nog niet duidelijk. Parallel aan dit onderzoek en op basis van de bevindingen is - in overleg met het Ministerie van Infrastructuur en Milieu als opdrachtgever voor de BRT - wel al besloten om in 2012 de kleinschalige kaartproducten zoals TOP50NL en TOP100NL niet te gaan leveren, maar in plaats daarvan de bestaande vectorproducten tijdelijk te handhaven en te herzien, totdat de nieuwe productielijn praktijk is. Naast het oppakken van de openstaande onderzoekspunten voor 1:50.000 worden de onderzoeksresultaten uitgebreid naar schaal 1:100.000. Veel resultaten van 1:50.000 (zoals het gegenereerde wegennetwerk) kunnen hiervoor worden hergebruikt. Hoewel het invoeren van automatische generalisatie resulteert in een revolutionaire verandering in de producten en de productielijn van het Kadaster, heeft dit onderzoek laten zien dat dit de enige manier is om de topografische informatie voorziening voor te bereiden op de toekomst. 📍

Referenties

Van Smaalen, J.W.N, 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.

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

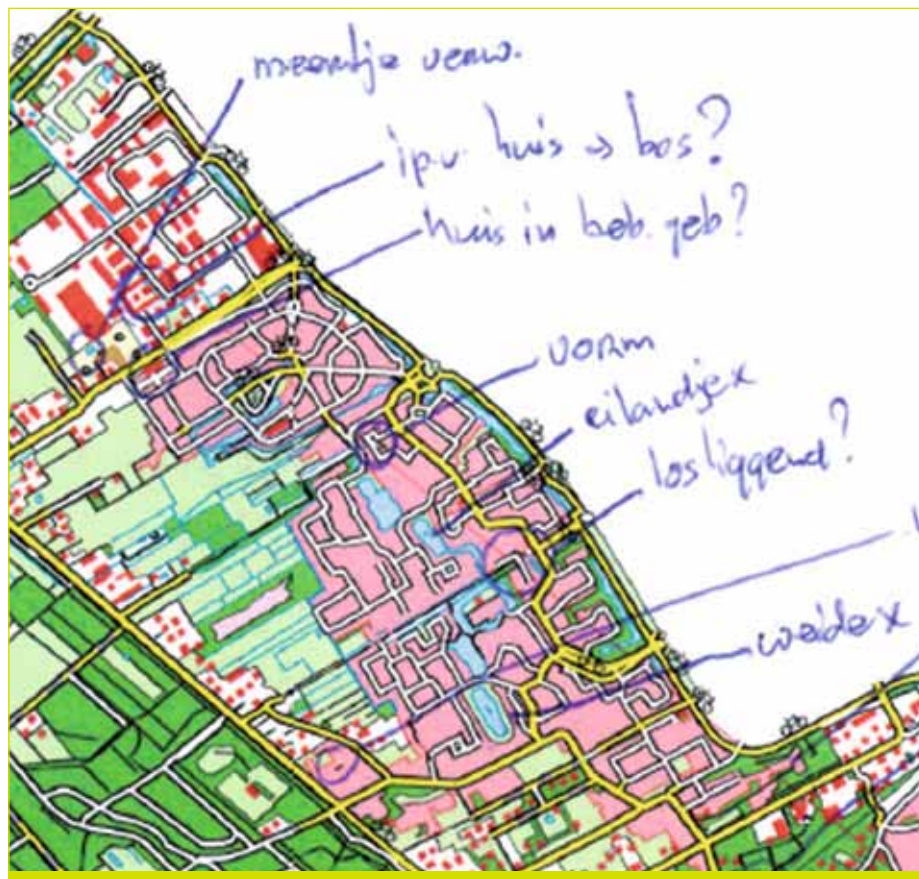
Stoter, J.E. van Smaalen, N. Bakker, P. Hardy, Specifying map requirements for automated generalisation of topographic data, The Cartographic Journal Vol. 46 No. 3 pp. 214-227 August 2009.

Samenvatting

Sinds begin 2011 onderzoekt het Kadaster de haalbaarheid van automatische generalisatie voor de productie van haar topografische kaartseries. Dit artikel gaat in op de implementatie van volledig geautomatiseerd werkproces voor het generaliseren van een kaart op schaal 1:50.000 uit TOP10NL, de 1:10.000 kaartserie. Door het iteratief testen van deelstappen en het evalueren van tussentijdse resultaten in nauw overleg met gebruikers is de optimale implementatie bepaald. Op basis van deze studie heeft het Kadaster besloten, dat een volledig automatische generalisatie werkproces de beste en waarschijnlijk enige mogelijkheid is om zowel een korte wijzigingscyclus te realiseren als een meer flexibele productielijn die 'on-demand' producten mogelijk maakt.

Summary

Since beginning of 2011 the Dutch Land Registry Office (Kadaster) is studying the feasibility of automated generalization for the generation of its topographic map series. This article addresses the implementation of a fully automated workflow to generalise a map at scale 1:50.000 from TOP10NL, the 1:10.000 map series.



Kwalitatieve controle door kartografische expert.



During this study, the optimal implementation and parameter values have been iteratively determined liaising closely with end users. Based on this study, the Dutch Kadaster has decided that a fully automated generalisation workflow is not only the best and only viable option to accommodate short update cycles, but also to produce flexible and on-demand products.

Het onderzoeksteam naar automatische generalisatie bestaat naast de auteurs uit Marc Post, Vincent van Altena, Hedwig Kupers, Henrik Oomes, Peter Lentjes (Kadaster) en John van Smaalen (Esri).