Automated generalisation of 1:10k topographic data from municipal data

Dogan Altundag[1] and Jantien Stoter[1, 2, 3]

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

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

[3] *Geonovum, Amersfoort, The Netherlands, j.stoter@geonovum.nl*

The research in this paper is a collaboration between Kadaster and four large Municipalities: The Hague, Amsterdam, Utrecht and Rotterdam

1. INTRODUCTION

This paper presents the ongoing research on automatically deriving a topographic dataset at scale 1:10k from large scale municipal topographic data (scale 1:1k).
The motivation of the research are the currently two independent 'key-registers' of topography in the Netherlands: one for municipal topographic data (not yet practiced) and one for topographic data at scale 1:10k and smaller. These legal key-register, established to support the national Spatial Data Infrastructure (SDI), contain authentic base data and their use is mandatory for all public organisations. The two key-registers on topography, both covering the whole of the Netherlands, are:

1. *Basisregistratie Grootschalige Topografie (BGT)*, (Geonovum, 2010)'key-register large scale topography', expected to become an operational key-register in 2015. The information model that defines the content of the register is currently being established. The BGT data will be the object oriented version of the Large scale Base Map of The Netherlands at scale 1:1k (*Grootschalige Basiskaart Nederland: GBKN*). Providers of the GBKN are mainly municipalities as well as water boards, provinces, ProRail (the manager of Dutch railway network infrastructure) and Rijkswaterstaat
2. *Basisregistratie Topografie (BRT)*, (Kadaster, 2008), 'key-register topography' in force as key register since 2008. The BRT consists of the separate object oriented topographic vector datasets at scale 1:10k, 1:100k, 1:250k, 1:500k, and 1:1000k. These datasets are provided by one organisation, namely the Kadaster who also holds the national mapping agency.

The current situation of separate key-registers is the consequence of history: traditionally municipalities collect large sale topographic data to maintain public and built-up area and Kadaster collects data to produce topographic maps at scale 1:10k and smaller.
The situation of two registers topography does not fulfil the SDI principle of collecting data once and use it many times. Instead the optimal situation would be to collect data for the most detailed information (i.e. within the municipal application domain) and automatically derive topographical data at scale 1:10k and smaller from this dataset. To obtain more knowledge on this optimal situation including its potentials, limitations and consequences, a study has been started on the automated generalisation of TOP10NL data (the object oriented database containing topography at scale 1:10k) from BGT data. The main research question is whether a 1:10k dataset can be automatically generated that serve the needs of a 1:10k data set in the new situation that BGT is practice (from 2015). Consequently some present TOP10NL users may shift to BGT data instead which may change the needs for 1:10k data.
Based on results of dedicated generalisation tests, the research aims at formulating recommendations for a closer link between BGT and BRT, ultimately resulting in one integrated key-register topography.

The study is carried out in collaboration with four large municipalities, i.e. Amsterdam, Rotterdam, Utrecht and The Hague. These municipalities maintain an own 1:10k dataset to serve their municipal tasks, which is updated in an interactive manner from the municipal map, see Table 1. The four municipalities are currently converting their 1:10k dataset into TOP10NL data because of the new law on key-registers topography, i.e. only one 1:10k dataset is allowed. These municipal TOP10NL data will replace the TOP10NL data from Kadaster. Since the four municipalities will become producers of TOP10NL data, they also have significant interest in generalising TOP10NL data from municipal large scale topographical data in an automated manner.

Table 1: Links between municipal dataset at scale 1:1k and dataset at scale 1:10k in four municipalities

Municipality	*Objects in 1:1k data?*	*Acquisition of 1:10k*	*TOP10NL*	*Use of 1:10k dataset*
Amsterdam	Polygons with topology, but no classes	Interactive generalisation of 1:1k data and aerial photo's	Conversion of municipal 1:10k data into TOP10NL information model	Visualisation and network analysis
Utrecht	No objects; objects are generated from geo data by maintainers of green areas and roads	Interactive generalisation of 1:1k data and aerial photo's	Conversion of municipal 1:10k data into TOP10NL information model	Mainly as visualisation
The Hague	No objects	Interactive generalisation of 1:1k data and aerial photo's	Throw away own 1:10k data; insert TOP10NL data in own database and enrich the data for municipal applications	Mainly as visualisation
Rotterdam	Yes, object oriented data	Interactive generalisation of 1:1k data and aerial photo's	Conversion of municipal 1:10k data	Mainly as visualisation

First this paper details the scope of the research in Section 2. Section 3 describes the methodology applied for this research. Section 4 presents the results and Section 5 ends with conclusions and outlook.

2. SCOPE OF THE RESEARCH

The aim of the research defined in consultation with the above mentioned municipalities is:
To derive a ~ 1:10k dataset from a BGT dataset in a fully automated manner to get insight into the feasibility of closer integration of the two key registers on topography, to identify issues for further research and to provide insights into the consequences of automatic derivation. The last aspect is important since the product of automated generalisation will be different from the current TOP10NL. The question is if these differences are acceptable when considering the significant advantages of automated derivation above separate maintenance (i.e. cost reduction, improved consistency, better up-to-date data). In addition 1:10k data may serve another purpose once BGT is operational (expected in 2015) and current TOP10NL data users may be well served with BGT data. Therefore the tests should also provide insight into the relationship between BGT and BRT once they are both practice.
The main purpose of the 1:10k data is to have a (visual) representation of topographic data at that scale. In addition the municipalities also use their current 1:10k data set for network analyses. Therefore a correct road- en water network is identified as important for the target data. Follow up research is required if current TOP10NL customizsrs can work with the new situation: i.e. either use the new 1:10k product or use BGT data in situations where they used to use TOP10NL data.

3. METHODOLOGY

The methodology that deploys the generalisation process of this study is depicted in Figure 1. The methodology contains five steps. At first the BGT data is translated into the TOP10NL data model. Secondlygeneralisation operators are applied according to the specific conditions of each class in the target dataset. The main operators are elimination, simplification and aggregation. Although data is generalised in the dataset at scale 1:10k (i.e. areas narrower than 2 meters are collapsed to lines), conflicts because of symbolisation hardly play a role. Consequently the focus of geometric generalisation is the reduction of data rather than solving cartographic conflicts. The third step is combining the generalised objects in one dataset to generate a topologically correct dataset. Finally, the last step assesses the quality of the generalised set by comparing original and generalised features. For the research a BGT compliant test dataset is available from Rotterdam. It should be noted that the information model for BGT is still in consultation and is therefore not yet approved. Consequently the official BGT information model (expected in September 2011) may differ from the data model used for this paper. Until now ArcGIS 10 and FME have been used for executing the tests.

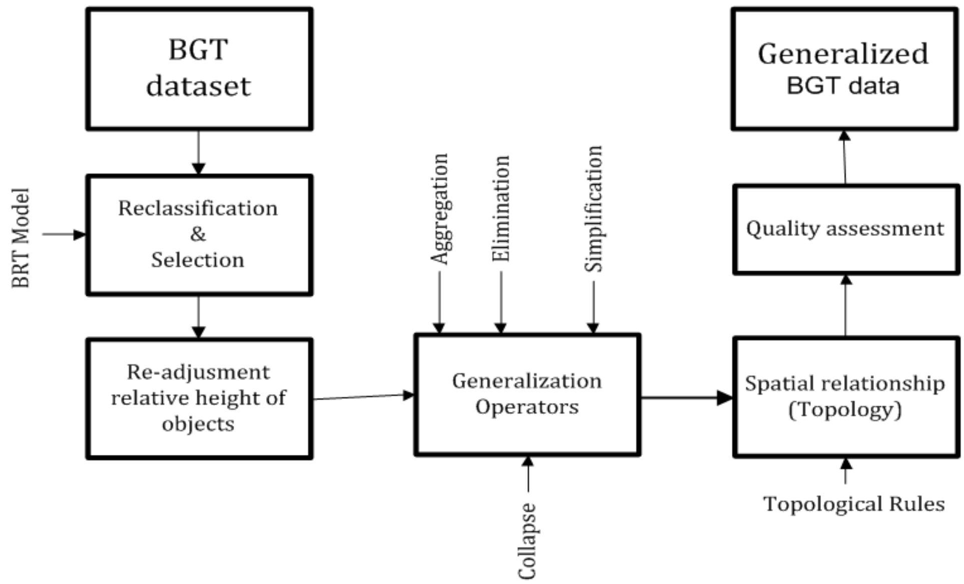


Figure 1: Automatic generalisation procedure as studied in this research

4. RESULTS

This section presents the results of each step:
1. Reclassification
2. Applying generalisation operators
3. Repair topology
4. Quality check

Step 1. Reclassification of BGT data into TOP10NL model

First step of the automated derivation of 1:10k data from BGT data, is reclassification. To know how BGT data should be translated into TOP10NL data model, a comparison was carried out between the two data models, i.e. which classes, attribute and attribute values represent more or less the same phenomenon (see also Stoter (2009) and Stoter et al (2009)). This information was used to translate the BGT data into the TOP10NL model in a next step.
To illustrate which differences had to be addressed in the model translation, Table 2 shows the main classes in both models. The class names are translated into English; the original Dutch names are added in italics and in brackets. From this table the following differences and similarities can be identified.
Classes that occur in both models are (Part of) Terrain, Part of Road, Part of Water, Part of Railway, Layout Element and Registration Area (for non-physical objects such as province, municipality and quarter). The 'part of' concept is to model the division of whole objects into several geometries in an object oriented approach.

Table 2: Main classes in BGT and TOP10NL. Dutch translations added in italics.

Class	BGT	TOP10NL
(PartOfRoad (*Wegdeel*)	Yes	Yes
Terrain (*Terrein*)	Yes	Yes
(part of)Water (*Waterdeel*)	Yes	Yes
(PartOf)Railway (*Spoorbaandeel*)	Yes	Yes
Layout Element (*Inrichtingselement*)	Yes	Yes
Registration Area (*Registratief Gebied*)	Yes	Yes
Building (*Pand*)	Yes	No
Engineering Structure (*Kunstwerk*)	Yes	No
Building Complex (*Gebouw*)	No	Yes
Geographical Area (*Geografisch gebied*)	No	Yes
Functional Area (*Functioneel gebied*)	No	Yes
Relief (*Reliëf*)	No	Yes

The concept 'Building' is modelled differently in both models. BGT contains Building and TOP10NL models Building Complex. Geographical Area, Functional Area and Relief are only modelled in TOP10NL. Geographical Area is used to link annotations in TOP10NL to geographical objects. Functional Area is used to group several objects into one object, for example a sport-area consisting of roads, building complexes and grass. Relief is used for topographical objects such as quays, peaks, isotopes and height differences. This information is less important for management of public and built-up areas and therefore missing in BGT.
BGT distinguishes Engineering Structure for infrastructural engineering structures such as bridges, viaducts, locks and dams, represented with polygon geometry. In TOP10NL these classes are modelled as a specific type of infrastructural objects (Part of Water, Railway or Road) or as a Layout Element. TOP10NL models much more attributes for its classes. The reason is firstly because these attributes are needed to visually distinguish different objects within one class. Secondly, BGT does not define more attributes than available in the underlying GBKN data and required for the municipal application domain.
To actually convert the municipal BGT data into the TOP10NL information model, the test dataset of Rotterdam municipality was studied and a conversion table was defined for each class-attribute-attribute value combination. Sometimes that conversion was straightforward; sometimes the conversion needed further interpretation because different terms are used for the same concepts, for example bicycle path (BGT) and cyclist (TOP10NL).

The translations between BGT data model and TOP10NL data model were done via SQL queries on the BGT data according to the translations rules as determined in the comparison study. For example:
Class road: ([KLASSE] = 'Wegberm' OR [KLASSE] = 'Voetpad' OR [KLASSE] = 'Rijwielpad' OR [KLASSE] = 'Rijbaan' OR [KLASSE] = 'Parkeerplaats' OR [KLASSE] = 'Overige Verharding')).
Further study will identify the missing information in the target data set compared to TOP10NL and if these TOP10NL concepts can be inferred from other BGT information. If not, than automated generalisation of 1:10k data from BGT data results in a loss of these concepts.

Step 2: Geometric generalisation for specific classes
The second step of the generalisation process is applying generalisation operators according to the specific conditions for each class. For the different classes, the next operators have been applied:

Buildings (Figure 2)
- The selected buildings of step 1 that are closer than 3 meters are amalgamated while keeping the orthogonal shape of the input features.
- Buildings smaller than 25m² are removed.

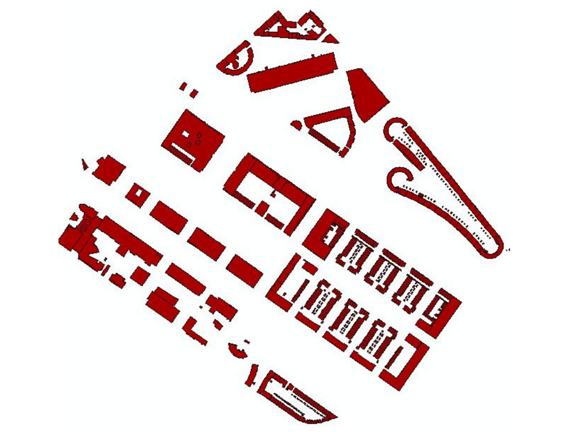
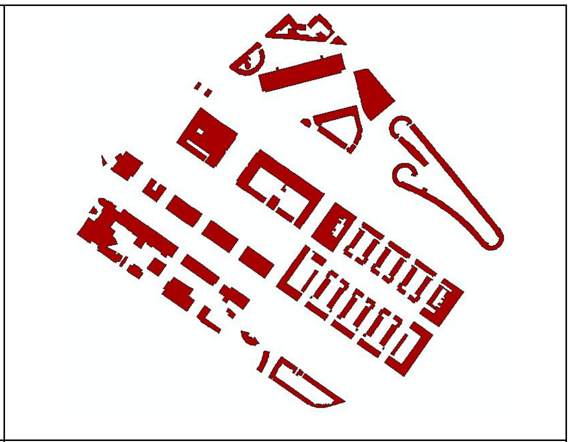
	
Before generalisation	After generalisation

Figure 2: Generalisation of buildings

Water bodies

- Water features that have the same attributes after reclassification (step 1) are aggregated and amalgamated if distance is < 3 meters.

Terrain (Figure 3)

- Terrain is aggregated (if they have the same attributes after step 1) and simplified.
- Polygons smaller than 100m² are removed, as well as holes <100m².
- Boundaries are simplified.

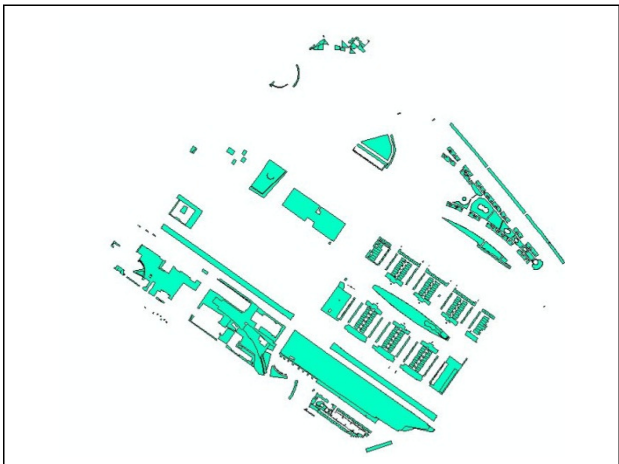
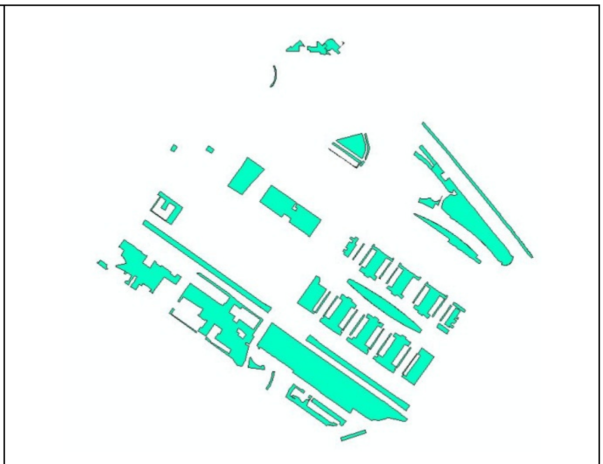
	
Before generalisation	After generalisation

Figure 3: Generalisation of terrain

Roads

- Roads narrower than 2 meters are eliminated and assigned to neighboring areas. The width of many road areas change across their geometry, because road geometries continue at crossing if attributes do not change. Consequently it is not straightforward to measure the width of one road geometry. Therefore a specific method was developed consisting of the following steps:
  1. The centre line of the road polygons are generated with the Medial Axis or Straight Skeleton algorithm of Fenkel and Obdrzalek (1998).
  2. Vertices are generated on polygon boundaries and its centre lines at every 25cm interval.
  3. For each vertex on the center line the closest vertex on the surrounding polygon boundary is identified and the distance to this vertex is calculated.
  4. Based on the threshold of 2 meters the road geometry is divided when the width of one of its part is narrower than 2 meters.

5. Roads narrower than 2 meters are eliminated. If roads remain without any connection, as in Figure 4, it might be better to eliminate these as well.
6. The tiny areas mistakenly generated by medial axes algorithm are eliminated.

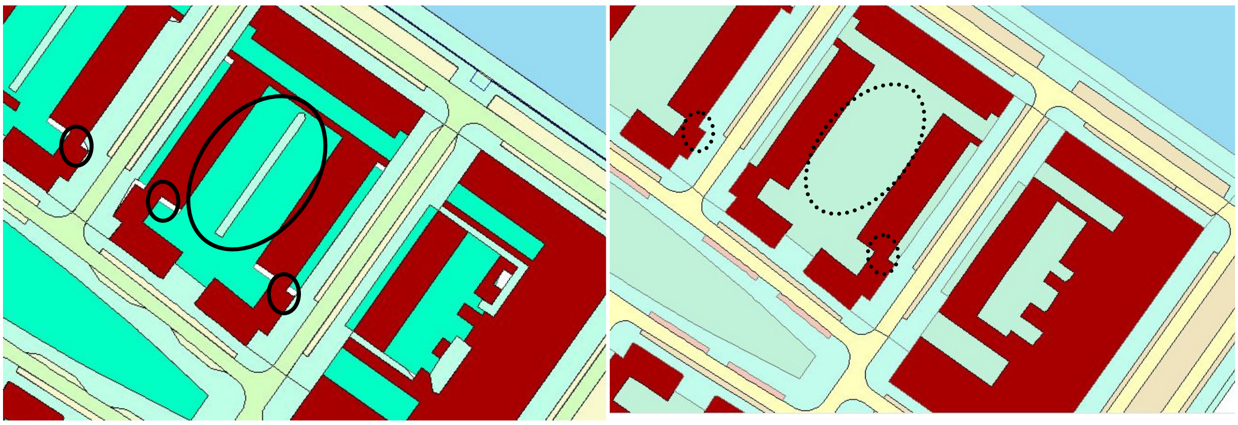
- Road width is added as attribute (as in TOP10NL information model) so that it can be used for symbolisation.

Step 3: Repairing the topology

Step 2 results in a set of individually generalised objects. In order to resolve the errors on the shared geometries, a set of topological rules, including cluster tolerances, have been set up and processed. The defined cluster tolerance allows the movement of the edge vertices in the specified distance range to dissolve topological errors. The following topological rules have been applied (and the optimal cluster tolerance of 0.9 m, empirically identified, has been used):

- *Terrain : 'Must not overlap with'; Building, Water and Road features*
- *Building: 'Must not overlap with'; Terrain, Water and Road features*
- *Water: 'Must not overlap with'; Terrain, Building and Road features*
- *Road (side walk, Highway, cycle path, parking lots): 'Must not overlap with' each sub-road parts*

In addition to the defined rule above, for each feature class a rule *'must not have gaps'* was also applied shared geometries. After the process, gaps and overlaps have been removed by moving vertices. As can be seen in Figure 4 non-connected road parts were assigned to the largest neighboring feature classes.



Before topological rules applied After topological rules applied

Figure 4: The result of applied topological rules.

Step 4: Quality assessment

The geometrical distortions caused by the generalisation process is important knowledge because at this scale range topographical data is often used in spatial analysis. To estimate the geometric distortions caused by the applied generalisation operators, a series of estimation indicators is applied to assess the rate in the changes between before and after generalisation. As it is very difficult to estimate the changes of the object similarity according to a single parameter, different criteria were studied such as the area distribution of an object in space and the area difference (Podolskaya et. al., 2007). The applied quality assessment procedure is based on the changes in the positional accuracy by comparing the deviation of the polygon centroids between initial and generalised polygons.

To calculate the change in the area of an object, the intersections of the same objects were taken into account before and after generalisation. The difference between the intersected area and total area of the original polygon (O) indicate the removed parts and 'intrusions' (e.q 1). On the contrary, the difference between the intersected area and total area of the generalised polygon (G) indicate the added parts and 'extrusions' (eq.2). Both of the quantity parameters represent the symmetric differences. (Filippovska et. al., 2008).

$$R_{intrusion} = \frac{Area\ (O \cap G)}{Area\ (O)} \quad (1)$$

$$R_{Extrusion} = \frac{Area\ (O \cap G)}{Area\ (G)} \quad (2)$$

$$R_{Area} = \frac{Area\ (G)}{Area\ (O)} \quad (3)$$

To estimate the differences of the polygon boundaries for each feature, the weighted mean center of each polygon is calculated before and after the generalisation procedure (e.q 4 and 5). Then, the positional deviation of the same feature is estimated by calculating the Euclidean distance between the centroids obtained before and after generalisation.

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad (4)$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \quad (5)$$

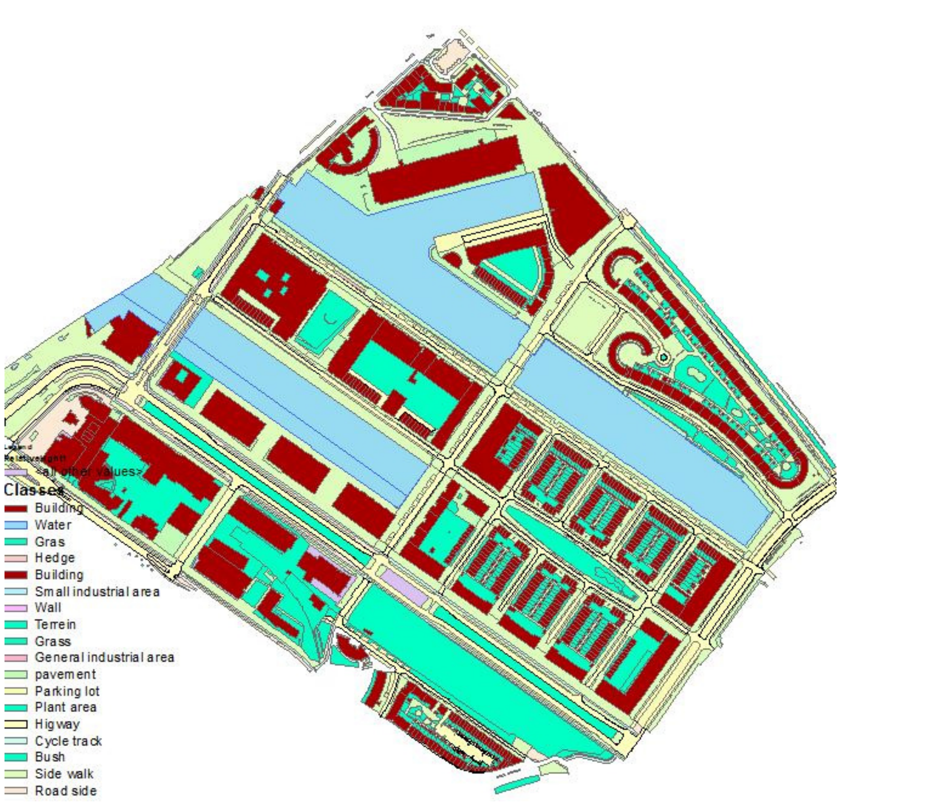
Where, w_i is the calculated perimeter of each feature before and after generalisation. x_i is the X coordinates of the polygon vertices and y_i is the Y coordinates of the polygon vertices of a feature. The results of the positional accuracy and the area distribution of an object in space are given in table 3.

Table 3: Quality assessment results according to object classes

Characteristic	***Area (m^2)***		***Perimeter(m)***		***Intersection rate***		***Ratio***	***Average Euclidean distance(m)***
Object Class	*(O)*	*(G)*	*(O)*	*(G)*	$R_{Intrusion}$	R_{Ext}	R_{Area}	*Centroids deviation*
Building	131712,3	133989,5	29326,72	16912,91	0,98	0,97	1,01	0,42
Terrain	103954,8	112302,8	21972,48	15128,18	0,98	0,90	1.1	0,90
Sidewalk	299573,1	295646,9	97947,3	76427,82	0,46	0,49	0,94	0,62
Parking lot	23357,06	25139,47	19085,19	18291,08	0,91	0,48	6,47	0,77

The results show that the positional deviation between original and generalised features is in the range of defined cluster tolerance. The total positional deviation is found as an average 0,67 m. The results of the rates of intrusion and extrusion indicate that added and the removed parts are compensated by the generalisation operators. The positional accuracies calculated for TOP10NL is around 4 m. Consequently the generalised BGT data provides better positional accuracy.

The final results are shown in Figure 5.



Before generalisation

After generalisation

Figure 5: Results of automated generalisation of 1:10k dataset (below) from municipal data at scale 1:1k (left) in test area of Rotterdam.

4. CONCLUSIONS AND OUTLOOK

4.1 Conclusions

The generalisation tests as described in this paper aim at automatically generalising a 1:10k topographical dataset from large scale municipal topographic data to formulate recommendations for a sustainable topographic information structure serving the national SDI.

Currently two authentic key-registers on topography exist. The different backgrounds of the two registers have resulted in datasets that differ with respect to source, provider, customers, responsible organisation, objectives and collection method (i.e. terrestrial measurements versus aerial photographs).

Because an object oriented version of 1:10k dataset existed already in 2005, the key-register on topography at scale 1:10k and smaller (BRT) was earlier in force than the key register on large scale topography (BGT) (Bakker, 2005). That is why many customers use the TOP10NL data as database. Now the key-register on large scale topography is in reach (expected to become practice in 2015), it is relevant to reconsider BRT as source for GIS and database applications, i.e. TOP10NL data users may be well served by BGT data. In principle two alternatives are possible:

Keep the current situation of two registers topography justified by the different application domains and different interests of the involved organisations. The advantages of this solution that it does not take much effort for the short term (no change required). In addition harmonising information models from two different domains can be avoided (which can be quiet complicated because of different interests). The main disadvantage is that separate registrations are kept (per definition not consistent) and that automated derivation is not practiced. Therefore this is not a sustainable option.

The second more advanced solution is one register topography that integrates topography at scale 1:1k to 1:1000k supported by harmonised information models and implemented generalisation processes. The most detailed base data within this register is BGT data, enriched if necessary. The consequence of this solution is significant, since current TOP10NL data will be replaced by the automated derived product. Furthermore it is not clear yet whether all TOP10NL concepts can be inferred from BGT information. But most probably the information at scale 1:10k and smaller will become poorer, since BGT contains less attributes. Besides these consequences, the disadvantages are throwing away the efforts of the past (new products will be generated) and the uncertainties about how users will perceive the new situations. However the realisation of an integrated key-register topography will be optimal for automated update propagation, for consistency, and consequently for a sustainable topographic information provision supporting the SDI. This will therefore be the long term goal envisaged in this research.

4.2 Outlook

This paper has indicated the importance of an integrated register topography. It requires further investigation to determine what small products can and should be generalised from the largest scale data. This is an interaction between the available technology, an enriched BGT data set, (contemporary) needs for multiscale topography (which scale levels; which objects), willingness to compromise in favor of frequent update cycles and finally the role /function and thus demands for a 1:10k data set once BGT is practiced. This research will also pay attention to generalisation to the smaller scales, including the question what scales are needed instead of using the current scale range as a starting point (see Stoter et al, 2010; Stoter et al, 2009b). For example the integrated register for topography could contain a series consisting of large scale local level, regional level at high resolution, regional level at medium resolution and national level. Moreover, the option of the vario-scale should be considered (Meijers, 2009), a data structure specifically designed to generate representations from a large-scale topographic base.

The ongoing research presented in this paper will show the technical feasibility and consequences of one integrated register topography starting from BGT. In addition it will provide the policy makers and involved organisations with concrete example to decide on the best (i.e. cost-effective and optimally serving user needs) situation.

Finally, the progresses in the Netherlands on integrated registers topography will also learn from experiences abroad, as in Denmark where an information model is established that integrates large scale data obtained by the municipalities and the 1:10k data set of the Danish Land Registry.

REFERENCES

Bakker, N., 2005, Developing a new geographical object database. Experiences from idea to delivering datasets. In: Proceedings of the 22nd International Cartographic Conference (ICC 2005), A Coruña, Spain, 9 - 16 July 2005

Fenkel P. and Obdrzalek S., Straight Skeleton Implementation, 1998, "SCCG 98: Proceedings of the 14th Spring Conference on Computer Graphics," pages 210-218, ISBN 80-223-0837-4 (Publisher: Comenius University, Bratislava).

Filippovska Y., Walter V., and Fritsch D (2008) Quality evaluation of generalisation algorithms, ISPRS.Vol.XXXVII. Part B2. Beijing.

Geonovum, 2008, Catalogus Basisregistratie Topografie (BRT), 2008, versie 2, Kadaster.

Kadaster, 2010, Informatiemodel BGT, 2010, gegevenscatalogus, versie concept 0.95

Meijers, M., Peter van Oosterom and Wilko Quak, 2009, A storage and transfer efficient data structure for variable scale vector data, In: Sester, Bernard, Paelke (Eds.); Advances in GIScience, Springer, 2009, pp. 345-367

Podolskaya E.S., Anders K.H., Haunert JH., Sester M. (2007). Qualiy assesment for polygon generalisation, ISPRS, Vol.34, Part XXX, Commission 2,WG7.

Stoter, J.E., C.W. Quak, and A. Hofman (2009). Harmonisation and integration of topographical concepts for a multi-scale base register topography, In: SDI convergence, research emerging trends and critical assessment / ed by. B. van Loenen, J.W.J. Besemer, and J. Zevenbergen. Delft: Netherlands Geodetic Commission, 2009. ISBN 978-90-6132-310-5. (Netherland Geodetic Commission: green series, 48) pp. 89-106.

Stoter, JE (2009). Towards one domain model and one key register topography. In Core Spatial Data (pp. 1-23). Delft: KNAW / NGC Netherlands Geodetic Commission.

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

Stoter, J.E., T. Visser, P. van Oosterom, W. Quak and N.Bakker (2010), A semantic-rich multi-scale information model for topography In: International Journal of Geographical Information Science, iFirst, 2010, 25 p.