Generalisation of a 1:10k map from municipal data

Vincent van Altena[1], Jan Bakermans[1], Peter Lentjes[1], Ron Nijhuis[1], Marc Post[1], Marcel Reuvers[1], Jantien Stoter[1,2,]

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

Abstract
This paper reports about the feasibility study carried out by the Dutch Kadaster to automatically generalise the largest scale topographical data set maintained by the Kadaster (i.e. TOP10NL) from the 1:1k topographical object oriented data set, which is currently being collected and structured by organisations that need to maintain public space such as municipalities, the railway company and provinces. The two data sets do not only differ in scale but also with respect to objectives, source data, application domain, providers, acquisition method and rules, and definition of topology. Therefore not only a scale step has to overcome.

Section 1 Introduction

In the Netherlands, the Kadaster is responsible for maintaining topographic data at scale 1:10k and smaller. Since 2005, the object oriented data at scale 1:10k called TOP10NL, covering the whole skin of the earth (no gaps or overlap) has been the most detailed, countrywide available data set in the Netherlands. This will change from 2016 onwards when large-scale topography, structured according the Dutch Information model Geography (IMGeo) will be served from a national portal. Figure 1 shows excerpts of both data sets.
IMGeo published in 2012, describes how object-based, large-scale (between scale 1:1000 and 1:2000) topographic features must be defined to make the national exchange of this information possible. From 2016, data providers such as municipalities, organisations responsible for the road, water and railway infrastructure etc. are required by law to provide their objects that fall under the definitions of IMGeo 2.0 to a national 'base registry' (*Basisregistratie Grootschalige Topografie, BGT*) where they are available for reuse as open data.
The mandatory core of IMGeo 2.0 contains object definitions for large-scale representations of roads, water, land use, land cover, bridges, tunnels etc. The optional part of IMGeo allows further classification of these objects into categories suitable for maintenance, and contains definitions for all kinds of city furniture and other non-mandatory classes (and also allows extension into 3D, see Stoter et al, 2013).

IMGeo data of an urban area	TOP10NL data of an urban area

Figure 1 Excerpts of IMGeo data (left) and TOP10NL data (right)

Until now the production and use of IMGeo data (and the predecessor *GBKN*: Large-scale basemap of The Netherlands) and TOP10NL data have been separated domains (different background; different stakeholders, different producers etc). With the principle of "collecting once, use many time", the

Kadaster has started a research how to generalise TOP10NL data from IMGeo data, from 2016 onwards. This research and preliminary results are described in this paper.
Examples of generalisation work of other National Mapping Agencies focusing on large-scale to mid-scale data are Ordnance Survey UK (Regnaud, 2011), ICC (Baella and Pla, 2005) and Swisstopo (Kauferle, 2013) (see also Duchene et al, 2014). The additional challenge that we have to face (besides bridging a scale step) is to bring together two data sets that have a long history in being produced, maintained and used separately.

This short paper is structured as follows. Section 2 summarises the main challenges to generalise TOP10NL data from IMGeo. Our approach is explained in Section 3. Section 4 presents generalisation experiments and preliminary results for water and Section 5 ends with conclusions.

Section 2 Generalization of IMGeo to TOP10NL: the challenge

Because of the different background, TOP10NL and IMGeo do not only differ in scale but also with respect to objectives, source data, application domain, providers, acquisition method and rules, and definition of topology, see Table 1. These differences have resulted in differences between the contents of the datasets.

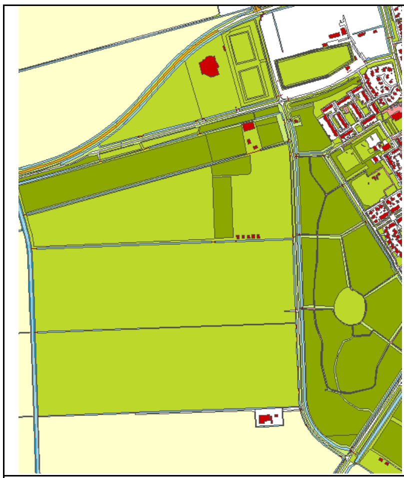
Table 1: Main differences between IMGeo and TOP10NL

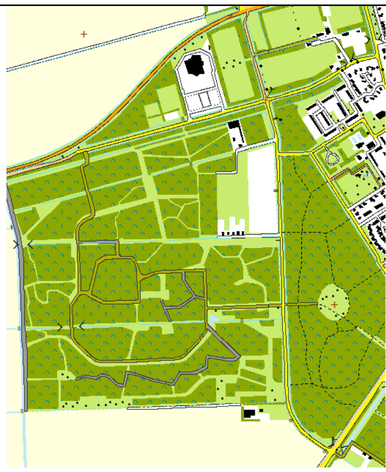
	IMGeo	TOP10NL
Objectives	enabling and standardising exchange of large-scale object oriented geographical information. IMGeo should be a framework of concepts for all organisations that collect, maintain and disseminate large scale geographical information	object oriented semantical description of the terrain for TOP10vector, according to requirements of internal and external users of the TOP10vector dataset
Source data	Object oriented large-scale municipal data	TOP10vector (originally digitized maps)
Application domain	management of public and built-up area	- visualising objects in map at scale 1:10k. - GIS analyses
Providers	municipalities, water boards, provinces, manager of Dutch railway infrastructure and Rijkswaterstaat	Kadaster
Acquisition method	terrestrial measurements	aerial photographs completed with terrain acquisition
Acquisition rules	(almost) no generalization applied	generalization is applied, e.g.: - only buildings with minimum area of 3x3 meter are acquired - buildings are merged when the distance is closer than 2 meters - roads and water smaller than 6m are represented as lines etc
Topology	all objects of any class (also buildings) with polygon geometry and height level '0' divide the terrain into objects that do not overlap	all objects of classes Part of Water, Part of Road and Terrain and height leve '0' form a complete partition without any gaps or overlap.
	- height level '0' means 'part of the terrain' - possible values are .., -1, 0 ,1 etc - all objects at ground level form planar partition	- height level '0' indicates that the object is on top of a stack of two or more objects - only values smaller than 0 are allowed (-1, -2 etc) - objects visible from above form a planar partition
	buildings are part of the planar partition	buildings are located on top of the planar partition

Besides those general differences, IMGeo and TOP10NL differ in how they model reality. A comparison of the information models shows a different approach for modelling same concepts. For example bridges and tunnels are separate classes in IMGeo but a specific type of road (and water) in TOP10NL. In addition, there are several differences in attribute names and attribute values. The differences are often small but still need to be solved when generalising TOP10NL form IMGeo. In general, TOP10NL models more information, i.e. more attributes are assigned to the classes, than IMGeo. Other prominent differences are:

- modelling of cross sections: IMGeo only models one object at a cross section for the "on going road" (which is not always unambiguously identifiable); TOP10NL models a stack of road objects (for every involved road) to keep continuity for all connecting roads
- geometry types being used: "small" IMGeo-areas are points or lines in TOP10NL.
- modelling of buildings: IMGeo defines footprints which are part of the planar partition; TOP10NL defines building geometries as seen from above in a separate layer as the planar partition.

Interestingly, because of the difference in purpose and background, the scale-difference is not evident everywhere. With the higher focus on urban areas, IMGeo is even less detailed that TOP10NL in rural areas, as can be seen in Figure 2.





IMGeo in rural area	TOP10NL in rural area

Figure 2 IMGeo and TOP10NL in rural area

Section 3 Methodology

The aim of our research is to automatically derive 1:10k data from IMGeo data, as alternative for TOP10NL data in 2016. The obtained target data should account for requirements of TOP10NL in 2016 that may have been changed since TOP10NL was established (2005) as well as the fact that from 2016 onwards a more detailed countrywide data set will be available than TOP10NL (which users can use instead of TOP10NL). The research will result in a workflow to automatically derive 1:10k data from IMGeo data, recommendations to acquire additional data (before or as part of the generalisation process) as well as in new specifications of these data, called TOP10NL 2.0.

Scope

Several starting points define the scope of our research:

- The mandatory part of IMGeo is the source data for generalization, since the optional part will not be filled in all areas (it depends on the data supplier)
- Our aim is not to imitate TOP10NL. To have a target to work to, we do use TOP10NL as reference target. The identified differences between the generalised 1:10k data and TOP10NL data are topic of further (iterative) study.
- The differences in semantics between both data sets will be harmonized and it will be studied if it is possible to use IMGeo semantics as much as possible.
- Because TOP10NL objects will be derived from IMGeo objects, the TOP10NL objects will change and will cause a deviation from the original product. This may cause problems for organisations that have assigned information to the TOP10NL objects.
- Other starting points are open and part of the study, such as support of:
    - A planar partition for the same objects as in the current version
    - ID's with history and update management
    - Incremental updates (the alternative is redo the generalisation for the complete map)
    - Meeting data for the smaller scales (i.e. 1:50k and smaller)
    - INSPIRE requirements
    - Other use of data, such as differentiation of the current "one product fits all uses" product
    - Road and water networks (not supported in both data sets; but may be required by nowadays users)
    - New information such as the identification of complete roads (now only part of roads are defined)

Approach

First a study has been carried out to identify the feasibility of deriving Kadaster TOP10NL data from municipal IMGeo data. Within a limited amount of time (five days) interactive generalisation experiments on the main themes (buildings, roads, water, terrain) showed that this generalisation is feasible (see Figure 14), but also that some issues need further research. In addition, the study showed that the derived TOP10NL product will differ from the existing TOP10NL product and therefore that users need to be involved in the follow-up of the study.
This follow-up study further develops the experiments to automatically generalise TOP10NL data from IMGeo data using available tools (i.e. mainly FME and ArcGIS) with self-developed extensions, making use of our results obtained for our study to generalise 1:50k map from 1:10k data (see Stoter et al, 2013). The intermediate generalisation results will show differences with TOP10NL 1.0. These

differences will be discussed with users and further studied. The themes and issues that we will study and will discuss in six sequential user-consultations are:
1. Water
2. Roads, Height levels and Bridges
3. Buildings
4. Engineering objects, Railway, Relief
5. Terrain, Land use
6. Topology, ID's, temporal aspects, semantic harmonization

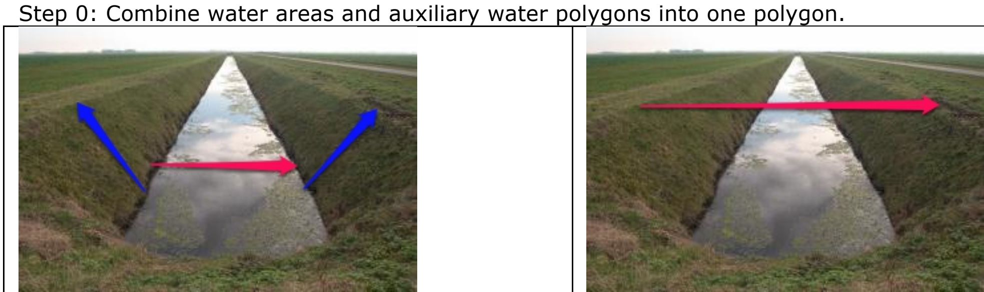
The iterative workflow that we will follow for each item is:
Step 1: Automatically generalise the specific item with a certain amount of effort for two or three test areas (appr 6 days).
Step 2: Identify the differences between the derived TOP10NL data and original TOP10NL data
Step 3: For each difference consult the users: do the users see a problem? If not, then the iteration stops here. If the users do see a problem, it will be studied whether the information can be extracted and generalised from another national source (for example the national road data set or the building and address register)? If not, than there are two main options:
   1. Iterate the above steps to see if more can be achieved with more effort.
   2. It is concluded that better generalisation result cannot be achieved.

The option of using an external data set is preferred to adhere to the principle "collect once, use many times" (for both efficiency and consistency reasons).
Step 4: After the above steps led to option 2, the next three alternatives are:
   1. It will be decided not to support the information any longer (for cost efficient reasons).
   2. The information will be manually added to the derivation process.
   3. The source data (IMGeo) will be enriched to make better derivation possible.

Section 4 Experiments and results for water

The experiments and user sessions for the six above topics are scheduled from May till December (2014). This short paper will report about the results for water (results for Roads, Height levels and Bridges as well as Buildings will be described in future publications).
The steps followed to generalise TOP10NL water objects from IMGeo water objects are:

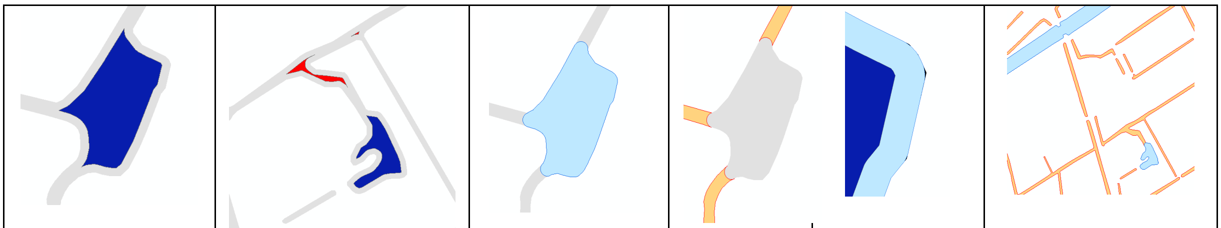
Step 0: Combine water areas and auxiliary water polygons into one polygon.



IMGeo distinguishes between water and auxiliary water objects (e.g. banks)	TOP10NL generalises banks and water into one water object

Figure 3 Modelling of water in IMGeo and TOP10NL

Step 1: identify which water polygons need to be kept as polygons (width >6 m) and which need to be collapsed (width < 6m)



Water polygons are buffered with a value of -3m	Water polygons < 250M^2 are selected	Water polygons > 250m^2 are buffered "back" (value of +3m)	Buffers Are removed	Slivers > 6x6 m are assigned to water polygons	Result: candidate objects for collapsing

Figure 4 Workflow to separate water objects wider than 6m from water objects smaller than 6m

The buffer value of 3m has been chosen to remove the objects that are wider than 6m from the collapsing process. To avoid that water polygons with widths close to 6m are kept, the remainders of the negative buffer are removed. 250M^2 has been experimentally identified, but may need adjustments for other test cases. The size of slivers to be taken into account (6x6m) is related to the buffer distance.

Step 2: Small water polygons are collapsed into lines
The ReplaceCenterLine-tool from FME has been used to generate the centre lines of the small water polygons.

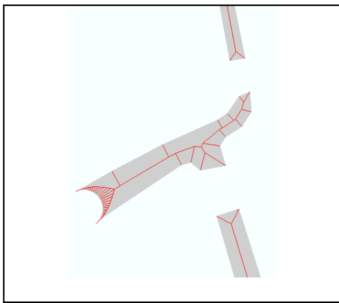
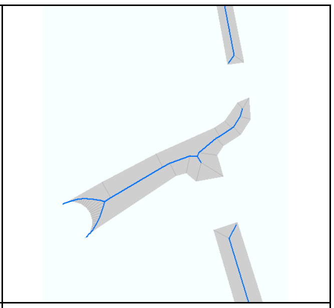
	
Straight skeleton	Medial axis

Figure 5 Result of ReplaceCenterLine-tool with different parameters

Side effects of this tool are the artificial branches near the end of the polygons and centrelines that stop before the end of the polygons. The steps that have been taken to remove the artificial branches and to extend the centrelines are as follows:

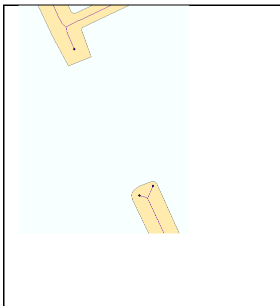
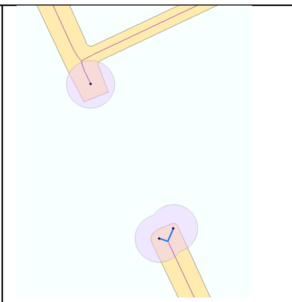
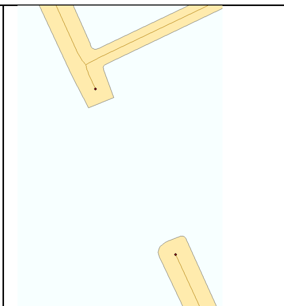
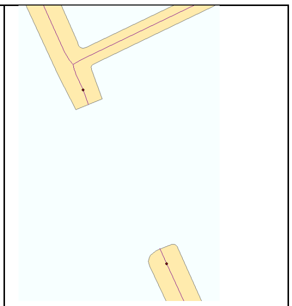
			
A node is attached to the end points of the centrelines	The buffer is perfectly round in case of one end node. If not, then the nodes and involved edges are removed.	Nodes are again attached to the end points	and extended to the end of the original water polygons

Figure 6 Workflow to remove side effects from the FME centreline generation tool

Step 3: Connect collapsed lines to remaining water polygons
The skeleton that was created for the centrelines is used to build a network between water lines and water areas. The polygons of this skeleton that are neighbouring water polygons are assigned to the water polygons, resulting in water lines connecting water polygons.

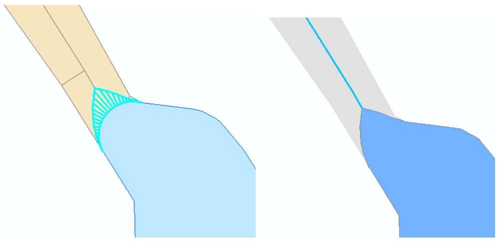


Figure 7 Building a network from water lines and water areas

Step 4: identify "width class"
The width-class in TOP10NL is used for symbolisation of water lines. To measure the width-class of water lines, these are split at each vertex and buffered with 3 meters ('end type' = FLAT to get rectangular buffers).

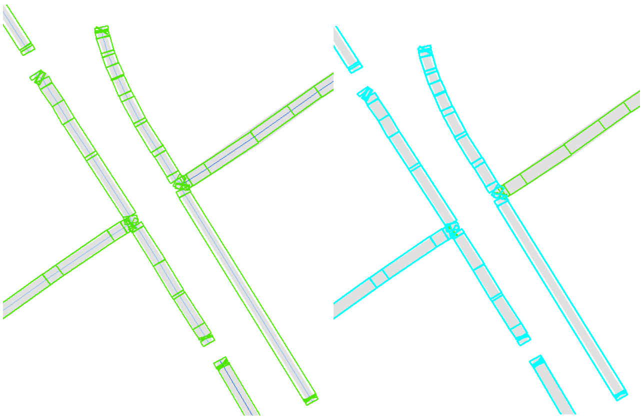
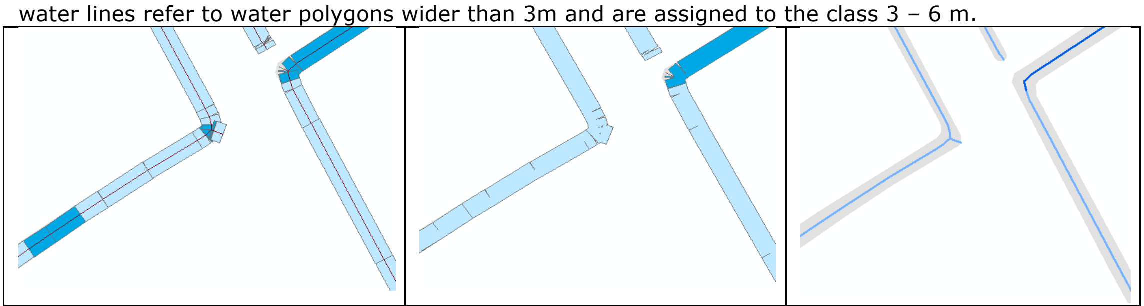


Figure 8 Determination of width classes

A spatial query determines whether the buffered polygons intersect with the original water polygons. If so, than the water polygon was smaller than 3m and falls within the class 0.5 – 3m. The other water lines refer to water polygons wider than 3m and are assigned to the class 3 – 6 m.



1	2	3

Figure 9 Generlisation and assignment of width classes

Using 3m as hard value and not taking any tolerance into account may result in fragmented width-class assignment. To solve this, small polygons are assigned to the width class of the (large) neighbouring polygon (1). In a next step all polygons with the same class vale are dissolved (2) and assigned to the water lines (3). NB: automatically calculated water widths may differ with TOP10NL-water width.

Step 5: Dams
When water goes underground, this is not modelled in IMGeo and therefore the network is broken (see Figure 10, left). To repair the network in TOP10NL at those locations, the water lines are extended to surrounding water (if the distance is smaller than 10 meters), see Figure 10 (right).

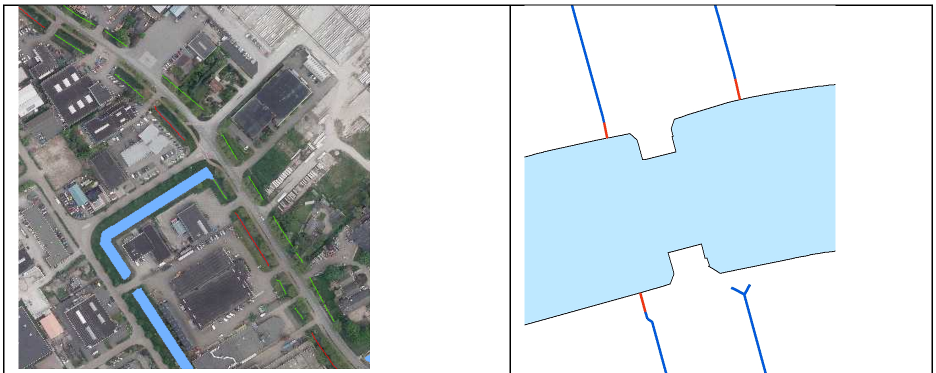


Figure 10 Water connections missing in IMGeo (left) are calculated for TOP10NL and identified as such (red lines) (right)

Step 6: Move water close to roads
If water polygons run very close to roads than these are moved adjacent to roads in the acquisition of TOP10NL data. This process is automated: water close to roads is identified and the water lines are located on top of the road boundaries. NB: The roads are yet un-generalised but will be simplified in a later stage of the research. This may alter the sequence of operations in the final workflow (first generalisation of roads; than moving water lines to roads).

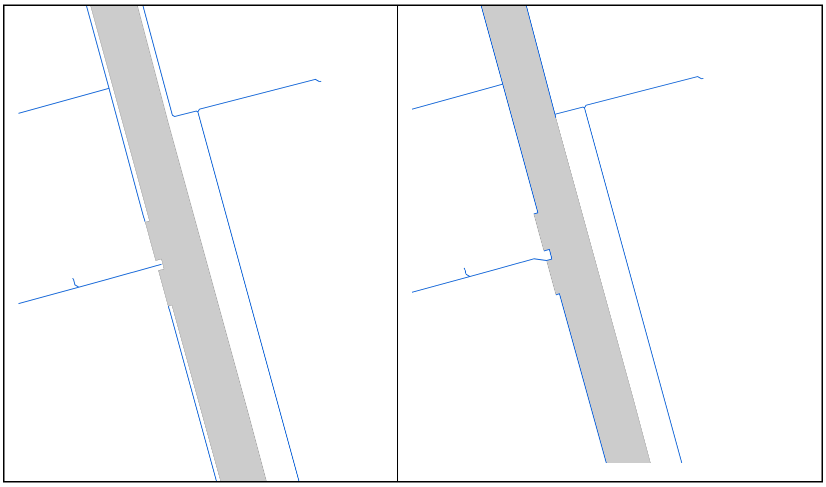


Figure 11 Water close to roads is moved adjacent to the roads (roads not generalised yet)

Step 7 Assign removed water polygons to surroundings
To keep the planar partition, the areas that are left because of collapsing of water polygons need to be assigned to the surroundings. We use the tools developed in earlier research t accomplish this.

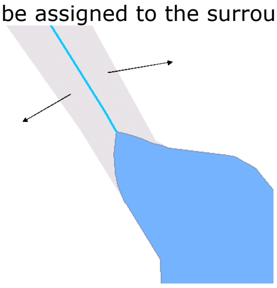


Figure 12 Removed water polygons need to be assigned to surroundings

Step 8: Identify semantic class of water polygons
There are some differences in water types between IMGeo and TOP10NL:

Table 1 Water types in IMGeo and TOP10NL

IMGeo	TOP10NL
Sea	Sea
Canal	Canal
Water area	Lake, fen, pond
Dry ditch	Dry ditch
x	Well
Bank	x
Mud flat	x

To convert IMGeo water types to TOP10NL water types, the following mappings have been applied:

- IMGeo dry ditches (polygon) becomes TOP10NL dry ditchs (line)
- All other water object polygons that have been collapsed into lines become "canal"
- Isolated water areas are identified and classified as "Lake, fen, pond"; Water polygons connected in a network are classified as "Canal" (see Figure 13).
- IMGeo Sea becomes TOP10NL sea
- IMGeo banks outside dikes are kept as banks (the only ones kept from IMGeo); the others are combined with water objects (see step 0).

Open issues in the mapping are TOP10NL wells and IMGeo mud flats.

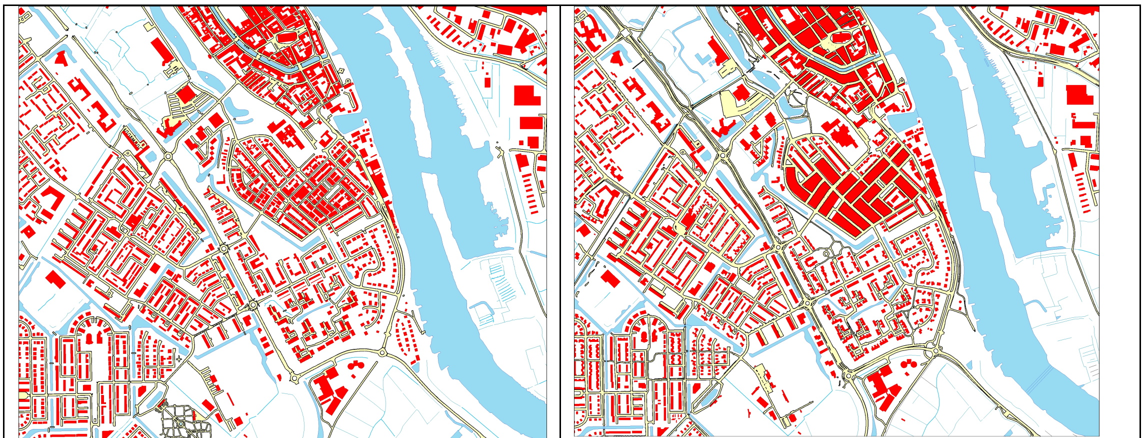
Figure 13 Determination of continuous and isolated water objects

Section 5 Preliminary Conclusions

This paper presents the research of Dutch Kadaster to generalise (a revised) version of TOP10NL data from large-scale, municipal data. This large-scale data will be countrywide available according to a common information model (i.e. IMGeo) from 2016 onwards. Initial results (see Figure 14) showed that it is feasible to derive TOP10NL data from IMGeo data ***if*** it is accepted that TOP10NL 2.0 differs from TOP10NL 1.0, for example because classification in both data sets are different or because some information cannot be derived or because automated generalisation from terrestrially acquired data results in different data than acquired by topographers from aerial images.
A follow up study is now focusing on specific issues and involving users in allowing differences between TOP10NL version 1.0 and version 2.0 or studying alternative solutions. The experiments and user consultations are planned for the coming months (until January 2015). The paper details our experiments on water. For water the following conclusions can be drawn:

- Water polygons and centre lines can successfully derived from IMGeo water polygons
- The connections between water lines and water polygons can be built as well as between water lines at the locations where water goes underground and disappears from IMGeo.
- The width-class can be determined.
- The attribute "width" (absolute value) is harder to determine. But because it is an optional value, the users were asked if they need this information. And because they do not; it will be removed from the TOP10NL 2.0 specifications.
- Water lines can be moved adjacent to roads if they are close to roads.
- The TOP10NL water types can be derived from IMGeo water types except from wells.

In addition the coming months, other issues will be studied and discussed with users to obtain both a generalisation workflow and new specifications for TOP10NL. This will integrate the large-scale data and mid-scale data - until now produced and maintained in different contexts and by different stakeholders – which will be an important milestone for adhering to the "collecting once, using many times" principle of our Spatial Data Infrastructure.



TOP10NL derived from IMGeo	Existing TOP10NL data

Figure 14 Results from feasibility study to generalise TOP10NL data from IMGeo

Acknowledgements

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs (Project code: 11300).

References

Cécile Duchêne, Blanca Baella, Cynthia A. Brewer, Dirk Burghardt, Barbara P. Buttenfield, Julien Gaffuri, Dominik Käuferle, François Lecordix, Emmanuel Maugeais, Ron Nijhuis, Maria Pla, Marc Post, Nicolas Regnauld, Lawrence V. Stanislawski, Jantien Stoter, Katalin Tóth, Sabine Urbanke, Vincent van Altena and Antje Wiedemann (2014) Generalisation in practice within National Mapping Agencies, Chapter 11 in Methodologies and Applications of Map Generalisation, Series: Publications of the International Cartographic Association (ICA), Editors: Burghardt, Dirk, Duchene, Cécile, Mackaness, William (Eds.) Published 2014, 407 p.

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.

D. Käuferle (2013) New National Maps of Switzerland, 26th International Cartographic Conference Dresden, August 2013

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

Jantien Stoter, Marc Post, Vincent van Altena, Ron Nijhuis & Ben Bruns (2014) Fully automated generalization of a 1:50k map from 1:10k data, Cartography and Geographic Information Science, 41:1, 1-13, DOI: 10.1080/15230406.2013.824637

Stoter, Jantien, Linda van den brink, Jakob Beetz, Hugo Ledoux, Marcel Reuvers, Rick Klooster, Paul Janssen, Friso Penninga, George Vosselman (2013) Establishing and implementing a national 3D standard, Journal for Photogrammetry and Geoinformation (PFG), 2013, Issue 4, DOI: 10.1127/1432-8364/2013/0184