Aligning heterogenous topographical data to derive multi-scale content for the Dutch nationwide Spatial Data Infrastructure

Keywords: automated generalisation, harmonisation, alignment of data models

Abstract:
To adhere to the Spatial Data Infrastructure principle "collect once, use many times" automated generalisation processes are needed to efficiently derive data and maps at lower levels of detail (i.e. small-scale) from already collected, large-scale, spatial data. Therefore, in recent years a fully automated generalisation process has been developed by the Dutch Kadaster that automatically generalizes the smaller scale products from the base data at scale 1:10K, called TOP10NL (Stoter et al, 2014a). The TOP10NL ((i.e. topographic data at scale 1:10k) is acquired by topographers from aerial images and panorama images. Together, the multi-scale products (both data products and maps) starting from 1:10k form the Base Register Topography (BRT), see Figure 1. Kadaster is responsible for the production and maintenance of the BRT.

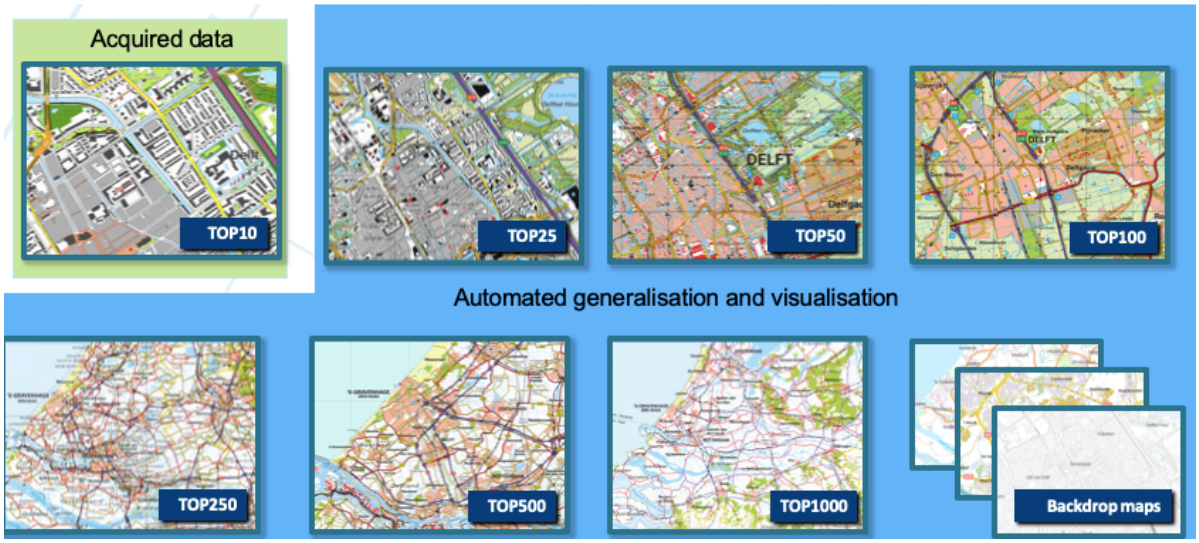

Figure 1: the multi-scale products of Kadaster

The BRT is part of the framework of base registers that has been established in the Netherlands (also called "key registers"). In this framework, specific governments are responsible to collect specific data and other governments are obliged to use the data collected by other governmental organisations. At this moment, there are ten Base Registers (Digital Government 2020). Two of these, the base register large-scale topography (*Basisregistratie Grootschalige Topografie, BGT*) and BRT (for scales 1:10k and smaller), define multi-scale topographical data about roads, water, land use, bridges, buildings, etc. In addition, information (including geometry) about buildings is not only stored in the BRT and BGT, but also part of the Base Register Addresses and Buildings (BAG). The acquisition of BGT (and BAG data) is the responsibility of many organisations, mostly municipalities.

The BRT, BAG and BGT have been developed apart from each other and for different purposes: the BRT has its origin in multi-scale map production and therefore maintains many attributes to support symbolisation on a map; BAG and BGT data are meant to support the maintenance of public space (by e.g. municipalities, provinces, waterboards, etc.) and contain attributes needed for this purpose. The differences in the purposes of the data has led to differences in the datasets. In the extended paper we will describe in more detail the input data of the automated generalisation (BAG/BGT), the current BRT-TOP10NL data, as well as an overview of differences between the two.

Nowadays, these base registers have matured and the question is how a BRT, containing multi-scale products, can be derived automatically from BGT and BAG, starting from TOP10NL. The interest to automatically derive BRT from BGT has two reasons. The first reason is consistency of topographical data provided by different governmental organisations and this requires the harmonization and alignment of BRT, BAG and BGT. The second reason is cost reduction which demands the implementation of the "collect once and use multiple times" principle in order to increase the reuse of once collected data.

For this reason, Kadaster has started a project to define a new product family: BRT.Next. The BRT.Next will be automatically generated from BGT/BAG data, and consequently, will be different than the existing BRT acquired by topographers. These changes are not per se problematic because the BRT specifications are based on map specifications that sometimes were defined decades ago with military use in mind. For example, the width classes of water are based on how far a soldier could jump (3m) or what width could still be taken by a tank (6m). Requirements that are not necessarily valid nowadays. In addition, the application of current BRT data - countrywide available since 2008; and before 2008 as vector data underlying the maps - has broadened to more and diverse disciplines. These new users might be served evenly well by BGT data, which has recently become available countrywide (since 2016).

The aim of the project is to reconsider the content of the BRT (starting from TOP10NL) and to define new content for a future proof BRT (called BRT.Next) that has a different context than current BRT: BRT.Next will be automatically derived from BGT and BAG instead of acquired from areal images by topographers; BGT data has become countrywide available and could replace current use of BRT-data which will change the user requirements of BRT-products; and, some BRT-specifications might no longer be valid.

Our previous studies on this topic (Stoter, Altena, et al. 2014b; Stoter et al, 2009; Stoter, 2009) were mainly based on a comparison of *data models*, because of a lack of countrywide BGT data. Our current project also investigates harmonisation and generalisation of countrywide BGT *data* that has become available.

Our methodology to define the new specifications of BRT.Next is not straightforward since the end result of BRT.Next is still open and depends on many interrelated considerations: what BRT data can be automatically derived from BAG and BGT, and how? What are the users' expectations? Which current usages of BRT could use BGT instead and what are the remaining use cases for BRT, both for the data and for the maps? What additional information is required to be able to continue the production of small-scale products at Kadaster? What are the minimal requirements of BRT.Next that we must fulfil? What impact has the quality of BGT-data - collected by hundreds of organisations and therefore still rather heterogenous - on the results that we can achieve with automated generalisation?

To address this multi-aspect context, the definition of BRT.Next follows the following principles:

(1) imitating the current TOP10NL (and the other small-scale products) is not the aim. However, to have a target to work to (and compare with) we do use TOP10NL as target data set. This has another advantage: the closer the data is to the current multi-scale data, the more we can reuse the automated generalisation process in place to automatically derive the small-scale products. Any significant change in this process will result in an increase of the costs instead of a reduction of the costs, as is the aim of our project.

(2) The production should be based on as much as automation as possible in order to realise the foreseen cost reduction. Only in exceptional cases acquisition (and maintenance) of additional data should be done by Kadaster

(3) We will use the semantics (i.e. attribute names and values) of BGT as much as possible in order to harmonise both data sets and to provide the user with consistent governmental topographical data.

(4) The BRT.Next data should meet the main and basic user needs. Needs that are specific for one user could eventually be met in an additional product.

(5) BRT.Next should be able to support the current automated production of the small-scale data products. Data requirements of this production line should therefore also be met as much as possible , additional to users' needs.

We are applying the following approach to define the content of BRT.Next balancing between a range of aspects:

1. Automated derivation possibilities: for different themes (water, roads, buildings, administrative areas, nature), a small study is carried out to identify what TOP10NL data (geometries and attributes) can automatically be derived from BGT/BAG.
2. Harmonisation with BGT and BAG data models: the semantics of BRT is as much as possible expressed according to the BGT data model (class names, attribute names, attribute values, definition)
3. Step 1 and 2 lead to an initial data model of BRT.Next that is presented to the users for evaluation in a survey. For each original TOP10NL concept the TOP10NL.Next equivalent is mentioned as well as if it could atomically be derived or otherwise kept and maintained in an additional process. The users are asked for each concept if the new proposal will affect current processes. The users' survey is supported by use case studies to understand in depth the needs of the users and the impact of BRT.Next on these needs.
4. We check whether there are other conditions that we need to fulfil, such as data requirements enforced by law and data requirements needed to produce small-scale data and maps.

5. Parallel to this, a study is carried out for each theme on how it is currently available in BGT and how this will impact BRT.Next. The BGT is acquired by hundred organisations. We know from previous studies that this results in highly heterogenous data, even if they are all adhering to the same acquisition guidelines. This is also due to the fact that the BGT is very new and therefore there has been no time to resolve different interpretations of the guidelines. However, such differences will affect the results we can achieve. The study is meant to provide insights on the impact of current BGT data on the result of BRT.Next.

The above steps are being carried out for individual themes where also the connection between themes will be addressed.

In the proposed presentation/paper we will present the results of each step for a few themes. These results will serve as input for the definition of BRT.Next in a later stage.