

# Hapklare basisregistratiedata via GeoToko.nl

Door Just van den Broecke

De Basisregistraties: velen van ons werken er dagelijks mee. Het Kadaster en andere instellingen, zoals het CBS en Rijkswaterstaat, stellen Basisregistraties en andere geodata in toenemende mate beschikbaar als open data. Via bijvoorbeeld PDK of data.overheid.nl kunnen datasets voor bijvoorbeeld de BAG, BGT, BRT en BRK 'gedownload' worden en naar eigen believen toegepast.

De meeste Basisregistraties zijn gestructureerd volgens uitgebreide, object-gerichte, 'IM'-datamodellen zoals IMBAG en IMGeo (voor de BGT). Bestands-uitlevering is meestal in een uitgebreid GML-formaat, gestructureerd volgens de conventies van een 'GML Application Schema', zoals CityGML voor de BGT. De BAG kent daarentegen een eigen uitleverformaat: XML met GML-fragmenten ('Namespaces'), al wordt dit vaak (onterecht) GML genoemd.

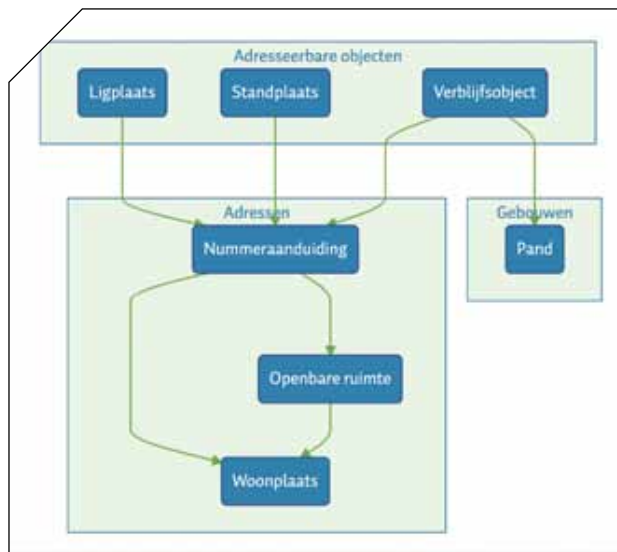
Om deze (GML en XML voor BAG) data te gebruiken binnen een toepassing zijn conversies nodig.

Een dergelijke data-omzetting wordt ETL, 'Extract, Transform and Load' genoemd, omdat vaak drie stappen te onderscheiden zijn: brondata worden uitgelezen oftewel 'geëxtraheerd', vervolgens getransformeerd en ten slotte ingeladen in een doelbestand. Dit laatste meestal in een database.

ETL voor de Basisregistraties kent een drievoudige uitdaging: uitgebreidheid van de datamodellen, complexiteit van het (GML)uitleverformaat en grootte van de uitleverbestanden (bijvoorbeeld zo'n 220 GB ongecomprimeerd voor de BGT GML van februari 2021). Daarnaast zijn de modellen ook nog aan verandering onderhevig, bijvoorbeeld BAG 1.0 naar 2.0 in 2021.

## NLEextract

ETL voor de Basisregistraties wordt door organisaties op velerlei manieren aangepakt: met



Het BAG-datamodel van het Kadaster. Bron: Kadaster.



BAG Pand XML-fragment. Bron: Kadaster.

gespecialiseerde ETL-pakketten zoals FME en/of door expliciete programmatuur te ontwikkelen. Steeds vaker wordt hierbij open source gebruikt, waarbij NLEextract het leidende product is – daarover later meer.

Al meer dan tien jaar werken ontwikkelaars

samen aan ETL voor Basisregistraties binnen het open-sourceproject NLEextract. Dit betreft vooral het inlezen, 'extraheren', van GML/XML-brondata en de transformatie naar de open source ruimtelijke database PostGIS. Momenteel ondersteunt NLEextract onder andere ETL voor de BAG, BGT, BRK (-DKK, de Digitale Kadastrale Kaart), de BRT TOP10NL plus overige BRT-Vector-producten zoals TOP25NL en TOP50NL. Ook worden in NLEextract componenten voor data-verrijkingen ontwikkeld plus omzetting naar additionele formaten zoals CSV, en binnenkort ook GeoPackage. Een voorbeeld is het verrijken van ingelezen BAG-data met CBS-data voor wijken, buurten, gemeenten en provincies. Hiermee worden bijvoorbeeld verschillende 'platte' adressen-sets voor heel Nederland (ongeveer 9,5 miljoen adressen) afgeleid. Er wordt daarbij ook automatisch rekening gehouden met de tijdlijn van gemeentelijke (her)indelingen. Voor de meest uitgebreide adressen-set, 'Adres Plus', wordt ook Kadaster Woningtypering (geometrisch) bepaald.

NLEextract is de 'stille motor' achter letterlijk zichtbare toepassingen die velen mogelijk kennen, zoals de OpenTopo-kaarten van Jan-Willem van Aalst, 3D-BAG van TUDelft en toepassingen zoals Waag Society CitySDK.

## GeoToko

Het installeren, onderhouden en gebruiken van NLEextract vereist kennis, IT-infrastructuur, en voor- al tijd. Daarom is vanuit de NLEextract-community



OpenTopo-kaart. Bron: J.W. van Aalst via map5.nl – opentopo.png.

het initiatief GeoToko.nl voortgekomen. 'GeoToko' is in feite een catalogus en webshop waar met NLEExtract geconverteerde Basisregistraties tegen vergoeding worden aangeboden, dus in bovengenoemde formaten zoals PostGIS en CSV, en straks ook GeoPackage. Via de GeoCatalogus kunnen GeoToko-datasets gezocht, bekeken en proefbestanden worden uitgeprobeerd. Datasets kunnen eenmalig of als abonnement worden afgenomen. Daarbij is er voor (potentiële) afnemers ook een helpdesk en een Kennisbank. Regelmatig verschijnt ook een nieuwsbrief. Instellingen of individuen die de data aanwenden voor open-source- en open-data-toepassingen kunnen korting krijgen. Bijvoorbeeld academische en onderwijsinstellingen, OpenStreetMap 'mappers' en publieke kaartenmakers. Met de opbrengsten van GeoToko.nl worden ook open-sourceprojecten, zoals NLEExtract, 'gevoed'.

### GeoFabriek

De meeste producten op GeoToko worden maandelijks ververst. Om hiervoor continui-

teit en datakwaliteit te garanderen is het interne product GeoFabriek ontwikkeld. Deze is gebouwd op NLEExtract en is vooral gericht op de interne proces-ketens die nodig zijn om vanuit basisregistratie-brondata de producten en hun metadata voor de GeoToko-afnemers beschikbaar te maken.

De GeoFabriek draait continu en autonoom/onbemenst in 'de cloud'. Een typische procesketen voor bijvoorbeeld de BAG kent de volgende stappen: controleer of er nieuwe maand-data zijn bij Kadaster/PDOK, zo ja, download deze, dan verwerken en verrijken (Adressen!) met NLEExtract, bestanden maken ('Dumpen'), de GeoCatalogus updaten en voor aankoop/download beschikbaar maken in GeoToko. Elk van deze stappen kent controlemomenten zoals datakwaliteitscontroles. Voor de verwerking zelf wordt binnen de GeoFabriek de standaardversie van NLEExtract gebruikt. Er is dus geen interne, 'betere' versie van NLEExtract binnen de GeoFabriek of GeoToko. Dit maakt het voor ieder moge-

lijk om te kiezen: zelf NLEExtract gebruiken of kant-en-klare, 'hapklare' Basisregistraties van GeoToko afnemen. Dit is een kosten-batenafweging voor eindgebruikers.

### Toekomst

GeoToko past in de stijgende trend om web- of clouddiensten, gebouwd met open-source-componenten, aan te bieden. Ook in de gewereld gebeurt dit steeds vaker. Bijvoorbeeld QGIS Cloud en dichterbij huis 'GeoCat Live' en Map5.nl. Deze ontwikkeling biedt niet alleen voordelen voor de eindgebruikers: de aanbieders zijn vaak zelf hoofdontwikkelaars bij de betreffende open-sourceprojecten. Hierdoor ook ontstaat kwaliteit en continuïteit in de evolutie van het onderliggende open-sourceproject.

Toekomstplannen zijn er genoeg voor GeoToko/GeoFabriek. Veel wordt door afnemers aangedragen via de GeoToko Helpdesk, maar ook het NLEExtract-project staat niet stil. Als voornaamste ontwikkelingen zijn hier te noemen: migratie van BAG 1.0 naar 2.0, met daarbij levering van BAG Compact (gediscontinueerd door Kadaster in BAG 2.0) en levering in het GeoPackage-formaat.

Ook wordt verwerking middels 'computing-on-demand' ontwikkeld, zodat capaciteit alleen voor de duur van een ETL-verwerking wordt aangewend. Dit bespaart zowel kosten als energie. ■

*Just van den Broecke is zelfstandig open source geo-professional. Voor meer informatie, zie <https://justobjects.nl>, <https://geotoko.nl> en <https://geocatalogus.nl>.*



Een 3D BAG-screenshot van Haarlem.