Towards one domain model and one key register topography

Dr. J.E. Stoter
TU Delft (research carried out at ITC), the Netherlands

Abstract
Two domain models for topography have been independently established in the Netherlands: Information Model Geography (for large scale topography) and TOP10NL (for small scale topography). The two domain models IMGeo and TOP10NL model the content and meaning of existing datasets which will be legally established as key registers for the national Spatial Data Infrastructure (SDI). Since both domain models and corresponding datasets represent the same types of object and cover the same geographical extent, the question is if one domain model and one register topography will be feasible to serve the notion of 'collect once, use many times' within the Dutch SDI. This paper contains a thorough comparison of how similar concepts in the two domain models are defined. The conclusion is that two key registers topography need to be kept for the moment. The main reason is the differences in content due to differences in data source, providers, objectives and stake holders. However since many differences seem random and easy to solve, harmonising of concepts is recommended before integrating the models. For the integration the paper proposes a Base Model Topography that models how TOP10NL data can be derived from IMGeo data to serve the goal: collect data once, maintain it at two key registers to use it many times.

1. Introduction
A main drive for establishing Spatial Data Infrastructures (SDIs) in general (and INSPIRE in particular) is to collect spatial data once and use it many times. To be able to reuse already collected data, it is most important to understand the content of the data. Revealing the content and meaning of data to outside (either human beings or other applications) is accomplished by means of data models, today often expressed as UML (Unified Modelling Language) class diagrams. The data model defines the concepts of concern as a collection of object classes, the hierarchical classification of the concepts, the mutual association between the concepts and their cardinality. It also contains the definition of the attributes (names and types) and the constraints associated with the data.

For reusing data from another application, the next challenge, after having specified the data in data models, is to agree on similar concepts defined in different data models. Agreeing on spatial concepts is the first step. Open Geospatial Consortium (OGC) and ISO/TC211 have developed a rich set of standards for spatial features such as point, line, and polygons, independent of specific themes or domains. This alone is not sufficient to understand each other's information. For combining data meaningfully, agreement is required on thematic concepts defined in different domain models.

It can be expected that it is difficult to achieve such agreement in different domains. For example the concept of 'water' is perceived differently in the tourist domain (recreation), in farming domain (critical factor for harvest), in domain topography (object for orientation), in domain of water management (source of flooding), in water sport domain (information for navigation) etc. However also within similar domains it might not be easy to agree on concepts. This was the motivation for this paper to study feasibility of integrating two domain models both dealing with topography.

For this study two datasets were selected representing topography at different scales for different purposes. The first dataset is the Large scale Base Map of the Netherlands (*Grootschalige Basiskaart Nederland*: GBKN). For the object oriented version of GBKN an information model in UML was established in 2007, called IMGeo (Information Model Geography) (IMGeo, 2007). Providers (and users) of the GBKN are municipalities, water boards, provinces, ProRail (the manager of Dutch railway network infrastructure), Rijkswaterstaat and Kadaster. The second dataset is the topographical dataset at scale 1:10k provided by the Netherlands' Kadaster of which the content is defined in the TOP10NL information model, established in 2005 (TOP10NL, 2005).

The harmonisation of these two domain models as well as the integration of the two datasets have become an important issue now 'key registers' are being established to support the Dutch SDI. Legally established key registers contain authentic base data and their use is mandatory for all public organisations. For topography two key registers have been identified, both covering the whole of the Netherlands:

- *Basisregistratie Grootschalige Topografie (BGT)*, 'key register large scale topography', expected to become a key register the coming years. IMGeo models how to exchange BGT-data. A complete IMGeo-compliant dataset is not yet available.
- *Basisregistratie Topografie (BRT)*, 'key register topography', in force since 2008. BRT currently consists of topographical data at scale 1:10k. From 2010 the smaller scales will be added to this register. TOP10NL, which is currently being extended to the smaller scales in the Information Model TOPography (IMTOP), models the data content of BRT.

Apart from key registers, the INSPIRE directive lays down requirements for harmonisation and exchange of topographic data. Although INSPIRE does not explicitly name topography as theme, it does address topography-related themes (see Table 1).

The most optimal situation for key registers serving the SDI would be to have one key register topography containing most detailed information from which the topographical datasets at smaller scales are derived automatically when required. This should be supported by one information model for multi-scale topography, specifying data content and meaning at the largest scale and describing how classes change at scale transitions. This covers both generalisation possibilities to derive TOP10NL-data from GBKN as well as to derive 1:50k, 1:100k, 1:250k etc from TOP10NL-data. Harmonisation of concepts currently modelled in IMGeo and TOP10NL is a key requirement for this approach. Hofman et al. (2008) studied the geometrical integration possibilities between IMGeo and TOP10NL. This paper will studies the feasibility of one domain model and one key register topogra-

Annex I Themes	Annex II Themes	Annex III Themes
Coordinate reference systems Geographical grid systems Geographical names. Administrative units Addresses. Cadastral parcels Transport networks Hydrography Protected sites	Elevation Land cover Orthoimagery Geology	Statistical units Buildings Soil Land use Human health and safety Utility and Government services Environmental monitoring facilities Production and industrial facilities Agricultural and aquaculture facilities Population distribution – demography Area management/restriction/regulation zones and reporting units Natural risk zones Atmospheric conditions Meteorological geographical features Oceanographic geographical features Sea regions Bio-geographical regions Habitats and biotopes Species distribution Energy resources Mineral resources

Table 1. Spatial themes of INPIRE (INSPIRE, 2009).

phy from a data model perspective. Similarities and differences between the two domain models have been analysed to show what is needed to harmonise concepts and to design an integrated model topography.

Section 2 describes the case study of this paper in more detail. Section 3 compares the model-ling approaches in the two domain models for a selected number of concepts. Section 4 analyses the findings of Section 3 and elaborates on the requirements for one domain model and one key register topography. The paper ends with conclusions in Section 5.

2. Background

In this section the domain models of this study are described in more detail: IMGeo (Section 2.2) and TOP10NL (Section 2.3). IMGeo and TOP10NL are both extensions of the Base Model Geo-information (NEN3610). This model is first introduced in Section 2.1.

2.1 NEN3610: Base Model Geo-information

The information model NEN3610 (NEN3610, 2005) of which the OGC compliant version was established in 2005 provides the concepts, definitions, relations and general rules for exchanging information on objects which are related to the earth surface in the Netherlands. The aim of this model is to have common definitions for object classes in the geo-information domain required for interoperability. NEN3610 describes geo-classes at an abstract level. Geo-application domains have built and are building their specific domain models on this generic model. Exam-ples are the information model for physical planning (IMRO), information model for cables and pipelines (IMKL), information model for soil and subsurface (IMBOD), and information model for water (IMWA) (Geonovum, 2008). Also the two information models that are studied in this paper are domain models that extend NEN3610. In the domain models the classes are subclasses of the NEN3610 GeoObject (root object class) and therefore they inherit all properties of the NEN3610 GeoObject. ISO19109 defines such a domain model as 'application schema' (ISO, 2005). An application schema is a 'conceptual schema for data required by one or more applications'. Figure 1 shows the relationship between an abstract class (example of PartOfRoad (*Wegdeel*)) in NEN3610 and the same class in a domain model (IMGeo in this case).

2.2 GBKN, IMGeo and key register large scale topography (BGT)

The UML class diagram of IMGeo is shown in Figure 2. The Large scale Base Map of the Netherlands (GBKN) will be the main source for IMGeo data. The most recent specifications of GBKN have taken into account the required conversion of the GBKN lines (often terrain boundaries) into polygonal objects (LSV GBKN, 2007). Although many providers have gener-ated an object oriented GBKN, there is yet no IMGeo compliant dataset available, except for some test datasets created by municipalities such as Den Haag and Almere and the province Noord-Brabant. It is expected that the IMGeo compliant GBKN covering the whole country will become a key register for large scale topography (BGT) within several years. GBKN (and there-fore IMGeo) is acquired for presentations at scale 1:1k in built-up area and 1:2k in rural area. The aim of IMGeo is "enabling and standardising exchange of object oriented geographical in-formation, IMGeo should be a framework of concepts for all organisations that collect, maintain and disseminate large scale

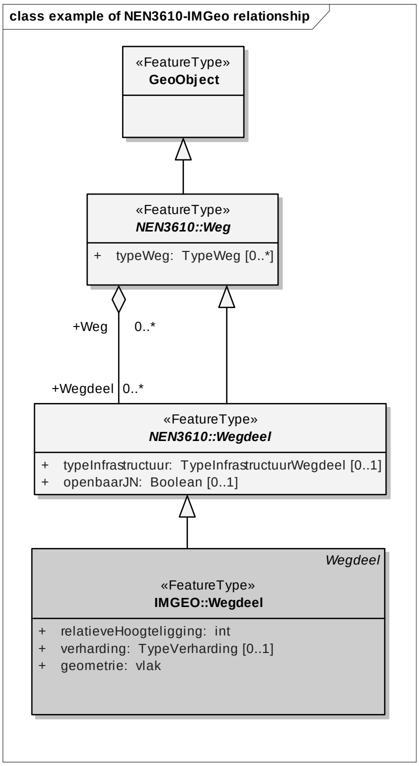


Figure 1. Relationship between abstract class 'PartOfRoad' (Wegdeel) in NEN3610 and the same class in the domain model IMGeo.

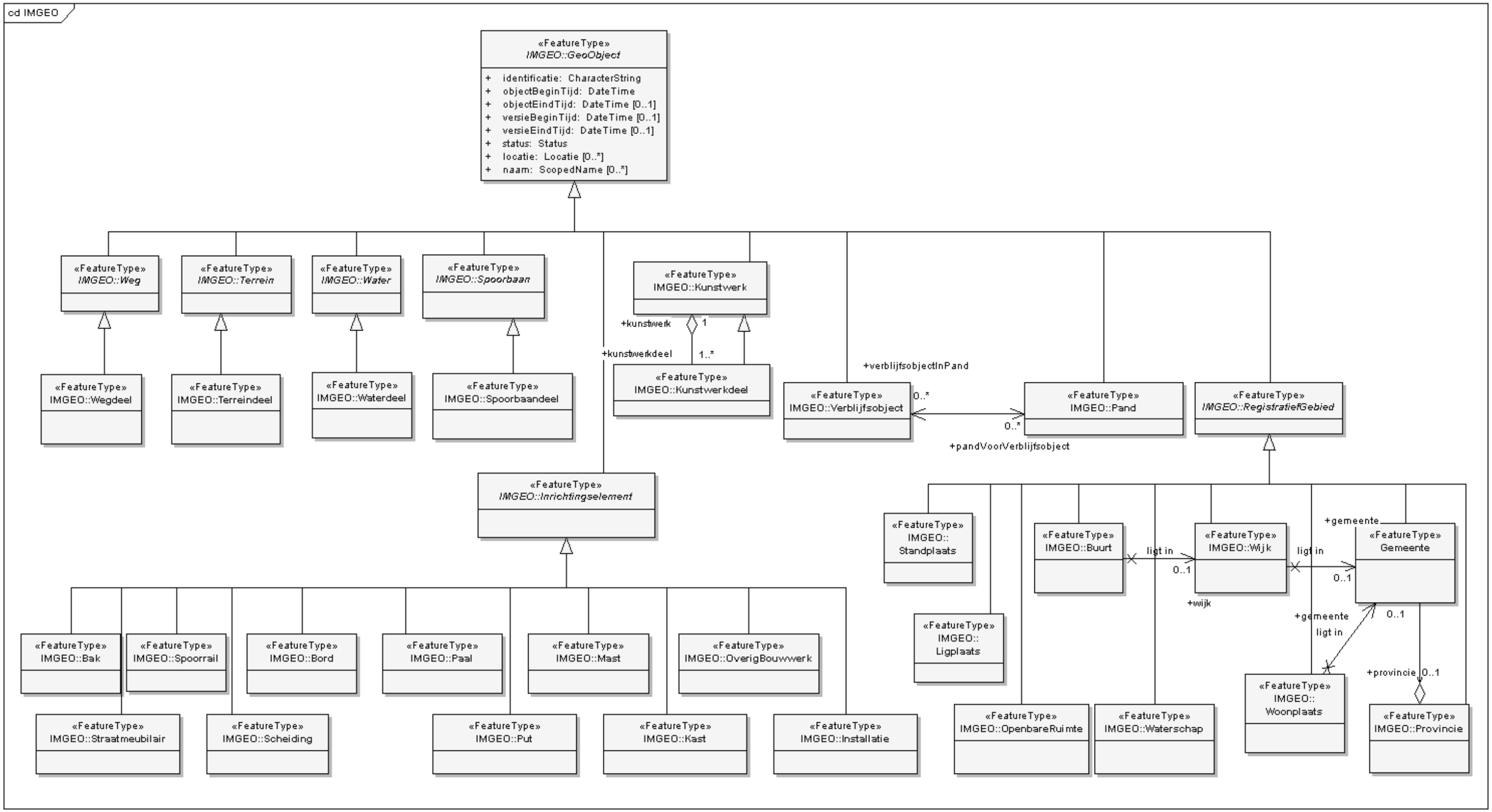


Figure 2. UML class diagram of IMGeo (IMGeo, 2007).

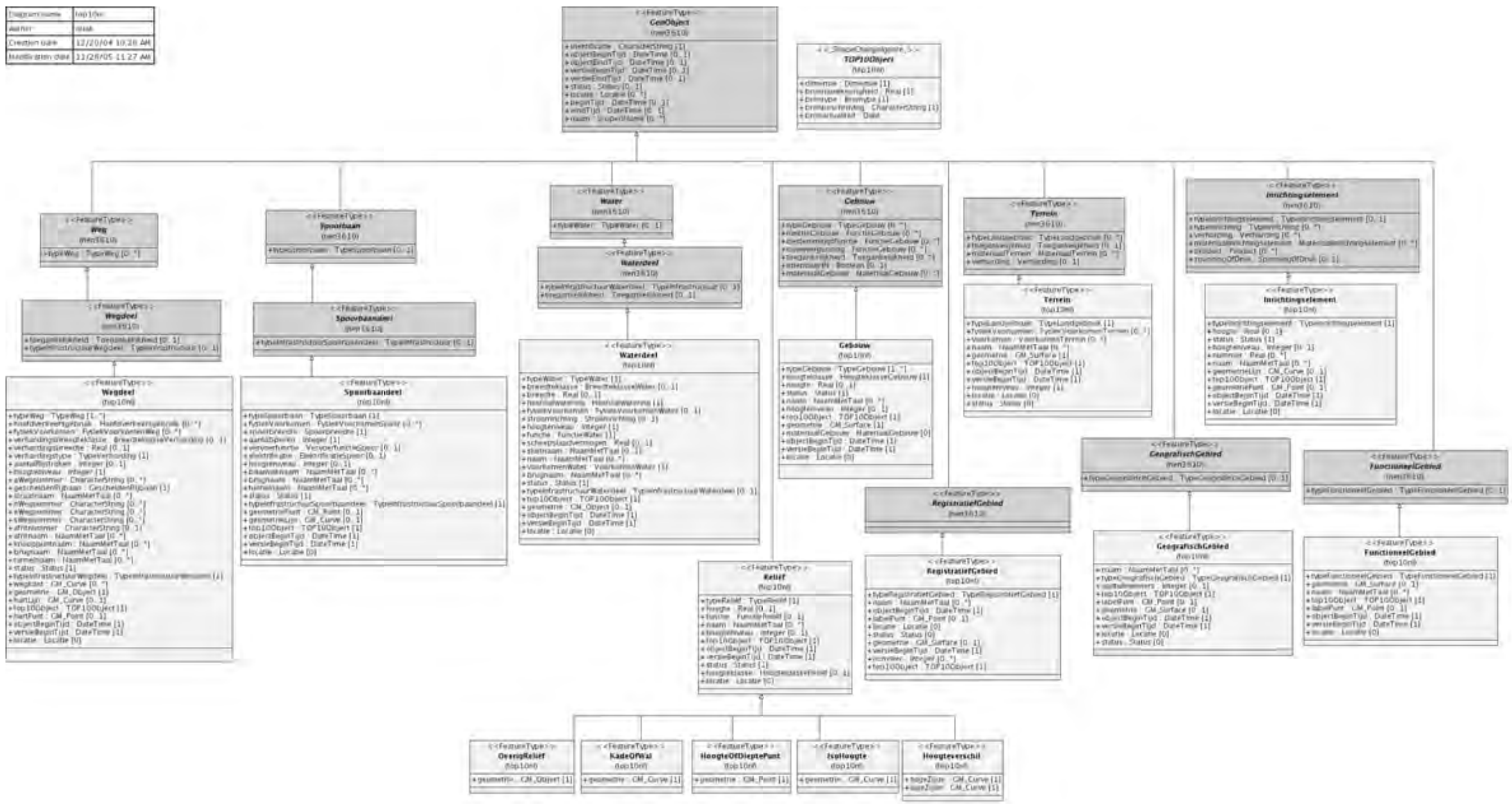


Figure 3. UML class diagram of TOP10NL (TOP10NL, 2005).

geographical information" (IMGeo, 2007; translated from Dutch by the author). The data that is modelled in IMGeo is not only meant to produce maps but mainly to support management of public and built-up area. IMGeo was established a few years after TOP10NL. However TOP10NL concepts were not used as starting point (IMGeo, 2007).

2.3 TOP10NL, IMTOP and key register topography (BRT)

The UML class diagram of TOP10NL is shown in Figure 3. Since January 2008 a TOP10NL dataset covering the whole country is available as key register topography (BRT). Currently TOP50vector, TOP100vector, TOP250vector etc are being converted into object oriented datasets. These vector datasets were created in the eighties to support the map production process. The object oriented versions (TOP50NL, TOP100NL, TOP250NL, TOP500NL and TOP1000NL) will be added to the BRT from 2010. A multi-scale information model is being defined to support this multi-scale key register. This information model is called IMTOP (In-formation Model TOPography). A detailed description of IMTOP can be found in Stoter et al. (2008). The aim of TOP10NL is "an object oriented, semantic description of the terrain for TOP10vector, according to requirements of internal and external users of the TOP10vector dataset" (TOP10NL, 2005; translated from Dutch by the author). Because TOP10NL has its origin in TOP10vector, the objective is tightly linked with visualising objects for a map at scale 1:10k. Nowadays TOP10NL data is also frequently used in GIS analyses.

3. IMGeo and TOP10NL: differences and simularities

This section compares IMGeo and TOP10NL models in order to answer the question how different the models are and to see if concepts defined in the different models can be harmonised. Section 3.1 compares the two models globally. Section 3.2 focuses on a selection of classes.

3.1 General comparison

Although IMGeo and TOP10NL model the same geographical extent and similar types of objects, it is important to realise that they differ with respect to source, provider, objectives and collection method. These differences resulted in differences in content of the datasets defined in the two information models. IMGeo data is mainly acquired using terrestrial measurements; TOP10NL data by means of aerial photographs supported by terrestrial measurements. IMGeo data is meant to support management of public and built-up areas and visualise the geometry of these objects at a scale of 1:1k and 1:2k; TOP10NL is meant to model objects for an acceptable visual presentation at scale 1:10k. Table 2 shows the comparison of the non-abstract classes in both models. The class names are translated into English; the original Dutch names are added in italics and between brackets. Also the corresponding NEN3610 classes are shown. NEN3610 contains more classes than displayed in Table 2.

As can be seen in the table a few classes start with 'part of'. This is to model the division of whole objects into several geometries in an object oriented approach. Classes that occur in both models are PartOfRoad (Section 3.2.1), PartOfWater, PartOfRailway (Section 3.2.2) and Layout Element (Section 3.2.5). Terrain (Section 3.2.4) exists in both

NEN3610	IMGeo	TOP10NL
PartOfRoad (*Wegdeel*)	PartOfRoad	PartOfRoad
Terrain (*Terrein*)	PartOfTerrain	Terrain
PartOfWater (*Waterdeel*)	PartOfWater	PartOfWater
PartOfRailway (*Spoorbaandeel*)	PartOfRailway	PartOfRailway
Layout Element (*Inrichtingselement*)	Layout Element	Layout Element
Building Complex (*Gebouw*)		Building Complex
Building (*Pand*)	Building	
Living unit (*Verblijfsobject*)	Living unit	
Engineering Structure (*Kunstwerk*)	Engineering Structure	
Registration Area (*Registratief Gebied*)	Registration Area	Registration Area
Geographical Area (*Geografisch gebied*)		Geographical Area
Functional Area (*Functioneel gebied*)		Functional Area
		Relief (*Reliëf*)

Table 2: Comparison of main classes in NEN3610, IMGeo and TOP10NL.

models, but IMGeo also distinguishes PartOfTerrain. Registration Area is defined in both models and is related to non-physical objects such as province, municipality and quarter. For building related objects NEN3610 models Building Complex (*Gebouw*), Building (*Pand*) and Living Unit (*Verblijfsobject*). IMGeo only models Building and Living Unit (in accordance with the Base register Addresses and Buildings: BAG (2006)) and TOP10NL only models Building Complex, which also includes single buildings. More details on building related objects in the three models are described in Section 3.2.3.

Geographical Area, Functional Area and Relief are only modelled in TOP10NL (Relief not available in NEN3610). Geographical Area is used to link toponyms 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.
IMGeo distinguishes Engineering Structure for infrastructural engineering structures such as bridges, viaducts, locks and dams, represented with polygon geometry (also available in NEN3610). In TOP10NL these classes are modelled as a specific type of infrastructural objects (PartOfWater, PartOfRailway or PartOfRoad) 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. IMGeo is mainly an exchange model and therefore does not need these kinds of attributes. Secondly, IMGeo does not define more attributes than available in the underlying GBKN data.

None of the two models defines topology, e.g. by the use of topological primitives. However TOP10NL (2005) describes that the classes Part of Water, PartOfRoad and Terrain form a complete partition of the country, without any gaps or overlap. Consequently building complexes, and also functional and geographical areas overlap with other objects. In addition different infrastructural objects can cross (i.e. overlap in space). This is modelled using two attributes assigned to infrastructural classes with polygon geometry

(PartOfWater or PartOfRoad): typeOfInfrastructure attribute, which models whether the infrastructure object is a connection or a crossing (see Figure 5a) and the heightLevel attribute. This last attribute models the relative order of objects where a value of '0' indicates that the object is on top of a stack of two or more objects. All other objects of the planar partition are located at heightLevel '0'.

The object catalogue of IMGeo (IMGeo, 2007) indicates that all objects with polygon geometry and relativeHeight '0' divide the terrain into objects that do not overlap. IMGeo does not specify which classes do (or which classes do not) contribute to the partition as TOP10NL does. For some classes with polygonal geometry it is obvious that they are not part of the terrain because they may overlap with other objects, for example Registration Area. But in principle all objects at level '0' contribute and therefore a building does cause a 'gap' in the underlying terrain. The TOP10NL attribute heightLevel and the IMGeo attribute relativeHeight assigned to different classes represent the same concept. However in TOP10NL value '0' means 'on top' and 'part of the planar partition', whereas in IMGeo the same value means 'at ground level' and 'part of the planar partition'. Consequently TOP10NL forms a planar partition of objects seen from above; whereas IMGeo models the planar partition on ground level.
The fact that IMGeo data should contain no gaps is not a requirement since it should be possible to exchange a limited number of themes via IMGeo. However generating full partitions for IMGeo data is a strong advice to data producers to guarantee consistency.

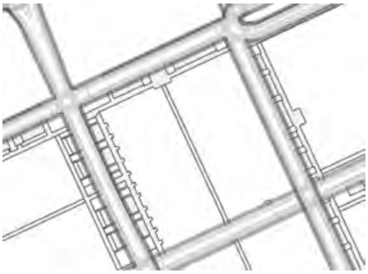
IMGeo and TOP10NL both model all their classes as children of the NEN3610 GeoObject, although explicitly in TOP10NL (leading to a formal relationship) and implicitly in IMGeo (the root class is not a NEN3610 class), as can be seen from Figure 2 and 3. TOP10NL has defined some additional attributes for all its classes, namely *dimensie* (dimension), *bronnauwkeurigheid* (precision of source), *brontype* (type of source), *bronbeschrijving* (source description) and *bronactualiteit* (source uptodateness). It should be noted that neither IMGeo nor TOP10NL contain composite relationships with the NEN3610 root object as proposed in NEN3610 (see Figure 1). Another important remark is that relationships between different object classes are rare in both models, for example to avoid overlap.

3.2 Comparison in detail

TOP10NL was established before IMGeo. However TOP10NL was not used as starting point for IMGeo. Consequently there are no relationships formulated between the two models to show which classes, attributes and attribute values model the same concepts. This section studies for a selection of concepts how these are modelled in both IMGeo and TOP10NL:

- Road (Section 3.2.1);
- Railway (Section 3.2.2);
- Building (Section 3.2.3);
- Terrain (Section 3.2.4);
- Layout Element (Section 3.2.5).

IMGeo roads.



TOP10NL roads transparently projected on IMGeo roads.

IMGeo data.

TOP10NL data.

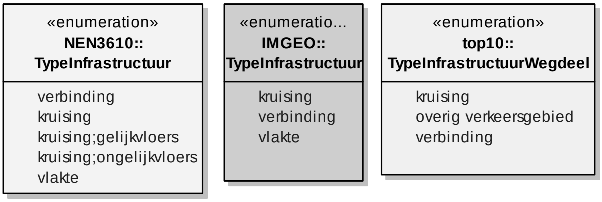
Figure 4. Visualisation of IMGeo data (courtesy of Municipality of Almere) and TOP10NL data.

The examples have been chosen to highlight some typical differences and similarities. It is not the intention to be complete here.

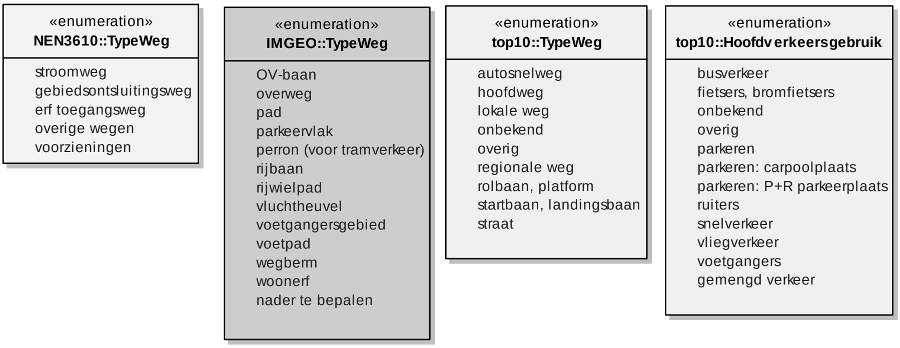
3.2.1 Road

In IMGeo all parts of roads are polygons. In TOP10NL the parts of roads contain multi-geometry: polygons and (centre) lines, unless a road is smaller than two meters. For those narrow roads only line geometry is maintained. For harmonising and integrating the two information models, it is important to note that the road concept is differently implemented in the datasets. These differences do not all become clear when comparing the models but when comparing the datasets: apparently some implicit information, for instance written down in acquisition rules, is not made explicit in the models.
TOP10NL data only contains one object per road namely the area that covers the roadway. In contrast IMGeo data identifies different objects for a road, for example foothpath (*voetpad*), cyclepath (*fietspad*), roadway (*rijbaan*), parking areas (*parkeervlakken*) and verge (*wegberm*). These differences can clearly be seen in Figure 4 where TOP10NL roads are simplified compared to IMGeo roads and where TOP10NL roads cover a smaller area.

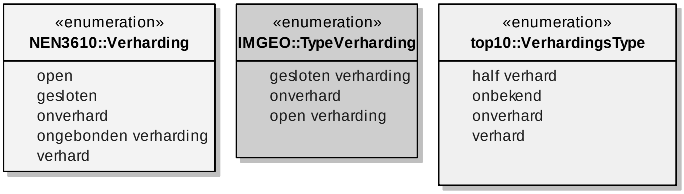
Another significant difference is that verge is considered PartofRoad in IMGeo but in TOP10NL verge is identified as Terrain, often with land use 'gras'. It would be possible to define this difference in a derivation rule, i.e. IMGeo PartOfRoad-verge is converted into TOP10NL Terrain-'gras'. However in this derivation, information on the function of the grassy area is lost. Consequently, if it is required to enlarge TOP10NL road to make it sufficiently visible in TOP50NL in a future process, information is lacking to assign the grassy



a. Attribute values for typeInfrastructuur (typeOfInfrastructure).



b. Attribute values for attribute typeWeg (typeOfRoad).



c. Attribute values for type of pavement assigned to PartOfRoad in TOP10NL, to PartOfTerrain and PartOfRoad in IMGeo, and to Terrain in NEN3610.

Figure 5. Attribute values for attributes related to PartOfRoad in NEN3610, IMGeo and TOP10NL.

areas, which are original IMGeo verges, to the roads in TOP50NL. To solve this TOP10NL objects should be enriched with an attribute describing the function of the objects.
The division of roads in parts is well defined in TOP10NL by use of the attribute typeOfInfrastructure as mentioned earlier, i.e. roads are divided into PartOfRoads at crossings. In contrast, IMGeo does not have clear rules to divide roads in PartofRoads, but most likely it will follow the division as applied in GBKN which is different than TOP10NL division. The division in GBKN is based on maintenance characteristics (e.g. pavement type). The differences in division of roads can also be seen in Figure 4.
Despite these differences, NEN3610, IMGeo as well as TOP10NL all contain the attribute typeOfInfrastructure (see Figure 5a) with values that seem easy to harmonise.

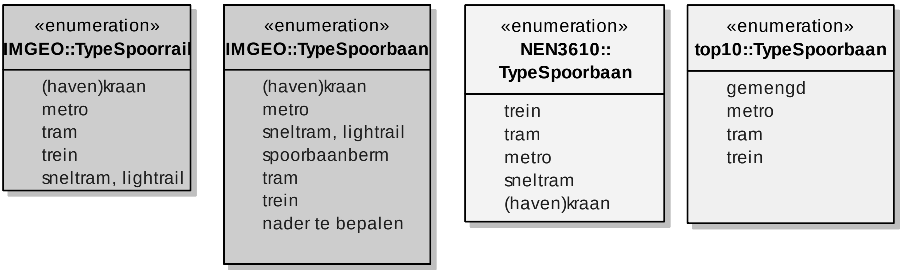
Another difference in road definition is the attribute typeOfRoad (*typeWeg*) used in a different way in both models, see Figure 5b. IMGeo uses the attribute to indicate different objects contributing to a road for example parking area, public transport-lane, footpath, verge, roadway, cycle path, pedestrian area, or residential area. The attribute value roadway identifies here all roads for motorists. In contrast TOP10NL needs to distinguish between different types of roads, also for motorists, to be able to visualise them differently. Therefore TOP10NL uses the attribute typeOfRoad to identify either a motorway, a main road, a regional road, or street. The attribute mainRoadUse (*hoofdVerkeersgebruik*) defines in a next step the main user of the road (not available in IMGeo). This can be cyclists, pedestrians, fast traffic, bus traffic etc, also shown in Figure 5b. TOP10NL therefore does not contain an equivalent concept for cycle path, pedestrian area, footpath or public transport-lane. By approximation these types of objects can be found via the attribute mainRoadUse. NEN3610 also models the attribute typeOfRoad with yet other values (also shown in Figure 5b): continuous road, access road, access road to residential areas, other roads, facilities.

Another interesting difference between IMGeo and TOP10NL is the attribute pavingType with values paved, unpaved, half paved and unknown assigned to TOP10NL PartOfRoads and attribute typeOfPaving with slightly different values as closed paving, open paving and not paved assigned to IMGeo PartOfTerrain and PartOfRoad. NEN3610 models even a different attribute Paving, with values open, closed, not paved, paved, assigned to Terrain. See Figure 5c. Although a human being can understand that most probably the same concepts are meant, for use in computers additional information is required to harmonise or map the concepts, i.e. explain that different terms are used for the same concept. Many similar examples with more or less same attribute names and more or less same attribute values exist. Important question is what the reason for the differences is: is it due to a lack of cooperation or are these differences fundamental?
As mentioned before TOP10NL contains more attributes for all classes. Examples of such extra attributes for PartOfRoads are physical occurrence, pavementWidthClass, pavementWidth, yes/noSeparationOfLanes, numberOfLanes, streetName, exit, crossway, bridge, tunnel.

3.2.2 Railway

As for PartOfRoad, IMGeo only allows polygon geometry for PartOfRailway. This geometry represents the whole area covered by the tracks. Information on the tracks are stored in attributes typeOfRailtrack (*typeSpoorbaan*) and typeOfInfrastrucutureRailway (*typeInfrastructuurSpoorbaandeel*) assigned to Railway. The middle of the rails is modelled with line geometry assigned to class Rail (*Spoorrail*) which is a specialization of Layout Element. This class has an attribute typeOfRail (*typeSpoorrail*) with mainly the same values as the attribute typeOfRailway assigned to Railway. This last attribute has two extra possible values, namely to be determined and railway-verge (see Figure 6). There is no explicit relationship between IMGeo Railway and IMGeo Rail. Consequently it is not clear whether it deals here with the same object (i.e. if it is a 1 to 1 relationship).

TOP10NL models all information on the railway as attributes of PartOfRailway represented by lines and points (for crossings). The lines are the centre lines of the railway and are therefore different than the rail-lines in IMGeo. The polygon geometries of the railways can also be represented in TOP10NL, but as Terrain, land use 'railway body'. The reason to model area covered by railways as Terrain is that TOP10vector (main source



Attribute values for typeOfRail, class Rail (left) and typeOfRailway, class Railway (right) both IMGeo.

Attribute values for typeOfRailway assigned to Railway in NEN3610 (left) and in TOP10NL (right).

Figure 6. Attribute values related to Railway (Spoorbaan) and Rail (Spoorrail) in NEN3610, IMGeo and TOP10NL.

of TOP10NL) contains a lot of land use of type other. To be able to specify these types in the future, more types are distinguished in TOP10NL which do not yet exist in TOP10NL data. Also here TOP10NL models more attributes, namely physicalOccurrence, railway-Width, numberOfLanes, transportFunction, electrification, and names of bridges, tunnels and railways.

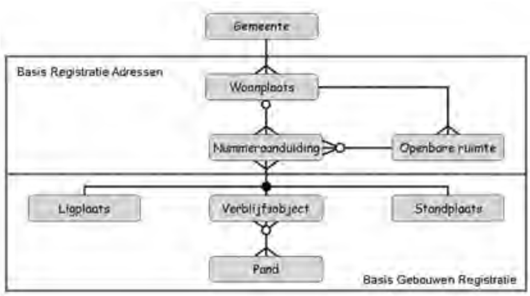
As for the road concept, we can conclude that the railway concept is differently defined in IMGeo and TOP10NL. In addition the defined types of railway differ (even between the two classes Rail and Railway both defined in IMGeo). As in the case of pavement type and type of infrastructure for roads, it seems not difficult to harmonise the types.

3.2.3 Building

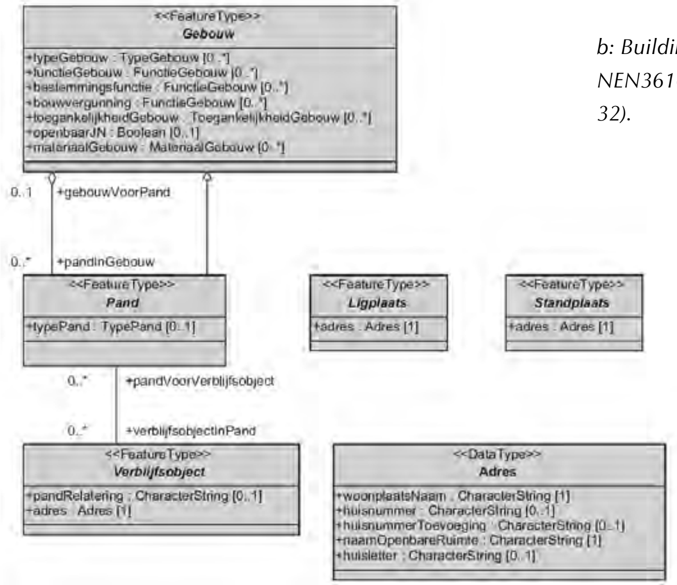
There are three building related classes defined in the models: Building Complex (*Gebouw*), Building (*Pand*) and Living Unit (*Verblijfsobject*). The last two are prescribed by BAG. BAG does not contain a class for Building Complex (*Gebouw*; see Figure 7a), since the designers of BAG could not find an unambiguous definition (BAG, 2006). NEN3610 models all three concepts (see Figure 7b). IMGeo follows BAG and only models Building (*Pand*) and Living Unit (*Verblijfsobject*) as main classes. The buildings are represented by the geometry seen from above (as prescribed by BAG) as well as by the extent of the building at surface level (as used in GBKN). Other BAG classes in IMGeo are Location for Mobile Homes, Location for Living Boats and Public Area. These three classes are modelled as subclasses of Registration Area, as prescribed by BAG.

TOP10NL only contains the class Building Complex (*Gebouw*), which is also used for single buildings. The class contains the orthogonal projection of the complex. Attributes are typeOfBuildingcomplex, name, height, heightClass.

IMGeo models all buildings, i.e. with and without addresses. TOP10NL models only a selection of building complexes, i.e. those meeting a minimal size condition of 3x3 meter. TOP10NL also merges buildings into one building complex in case of direct neighbours and when the distance is smaller than 2 meters.



a: Building related classes in BAG (BAG, 2006; pp. 12).



b: Building related classes in NEN3610 (NEN3610, 2005; p. 32).

Figure 7. Gebouw, Pand and Verblijfsobject in NEN3610 and BAG.

As can be seen in Figure 4, IMGeo buildings have a higher precision than TOP10NL buildings. Since TOP10NL will use GBKN buildings in the future, the differences will largely disappear (Kadaster, 2005).
We can conclude from above that the classes Building and Building Complex have a different meaning in NEN3610, IMGeo/BAG and TOP10NL.

3.2.4 Terrain

For the class Terrain (IMGeo also models PartOfTerrain) the attributes typePartOfTerrain (IMGeo) and typeOfLandUse (TOP10NL, and also NEN3610) model the same concept. Table 3 compares the possible terrain types in both models; also all NEN3610 types are shown. An important observation is that none of the types mentioned in IMGeo has exactly the same name as a TOP10NL type. A few types are presumably the same (gras and gras-land; 'nature and landscape' and heather).
Another observation is that IMGeo models one type of forest where TOP10NL models four types of forest. Also for IMGeo green object we can observe four possible values in TOP10NL. Apparently higher level of detail is required here for the smaller scale dataset.

NEN3610 typeOfLandUse	IMGeo typePartOfTerrain	TOP10NL typeOfLandUse	Translation
bos	bos	bos: gemengd bos	forest
bos: gemengd bos			mixed forest
		bos: griend	brush forest
bos: loofbos		bos: loofbos	deciduous forest
bos: naaldbos		bos: naaldbos	coniferous forest
grasland	gras	grasland	grassy area
natuur	natuur en landschap		nature
hoogveen			peat
moeras			swamp
heide		heide	heather
	cultuurgrond		culture land
akkerland		akkerland	arable land
agrarisch			agriculture
	overig groenobject		other green object
		boomgaard	orchard
		boomkwekerij	tree cultivation
		populieren	poplar
		dodenakker met bos	graveyard with forest
	bedrijfsterrein		industrial terrain
	braakliggend terrein		uncultivated terrain
	erf		courtyard
	plantvak		area with plants
	recreatieterrein		recreational area
	sportterrein		sport area
	talud		embankment
		aanlegsteiger	jetty
		basaltblokken/steenglooiing	sloped stones
bebouwd gebied		bebouwd gebied	built-up area
		dodenakker	graveyard
		fruitkwekerij	fruit cultivation
		laadperron	loading platform
		spoorbaanlichaam	area for railway
		zand	sand
		overig	other
		onbekend	unknown

Table 3. Comparison of types of terrain in NEN3610, IMGeo and TOP10NL.

For IMGeo the choice was made to make industrial area (*bedrijfsterrein*), recreational area (*recreatieterrein*) and sport area (*sportterrein*) as complete objects part of the terrain. In TOP10NL these are modelled as Functional Areas, i.e. as a collection of objects and thus with more detail. The first two types are modelled with slightly different values in TOP10NL: *bedrijventerein* and *recreatiegebied*. It is not clear what the motivation is behind these differences.

IMGeo has no equivalent for TOP10NL land use 'graveyard' (Functional Area in NEN3610). Also built-up area is not available in IMGeo because all buildings contribute to the planar partition, in contrast to TOP10NL buildings, and therefore built-up area is exactly the same area as the area of buildings.

TOP10NL Terrain has also more attributes than IMGeo Terrain: heightLevel, physical occurrence, name. As mentioned before IMGeo contains the extra attributes typeOfPaving and relativeHeight.

Another difference between definition of IMGeo Terrain and TOP10NL Terrain is caused by differences in acquisition rules which do not become clear from the models. Small terrain objects (width smaller than 6 meters) cannot be objects on their own in TOP10NL. Therefore they are assigned to their neighbours in the data acquisition process. Consequently objects such as verges are sometimes assigned to neighbouring road and sometimes to neighbouring terrain. Since these narrow objects can exist in IMGeo these are identified as road objects (see Figure 4).

3.2.5 Layout Element

In IMGeo the class Layout Element is divided into eleven subclasses, such as Street Furniture, Traffic sign, Pole, Installation and Well. All these classes have an own 'typeOfxx' attribute resulting in about 80 possible types of Layout Element. TOP10NL also identifies about 80 types of Layout Element by means of the typeOfLayoutElement attribute. Of these 80 identified types of layout elements in both models, nine have exactly the same label. These are: tree, hedge, high-tension pole, wall, pole, crane, sign pots, wind turbine and mast. In addition there are ten types which are presumably modelling the same concept, for example road closing (TOP10NL) and barrier (IMGeo); hectometer stone (IMGeo) and milestone (TOP10NL). All other types (about 60) cannot be mapped.
The types in IMGeo are mainly from the utility sector or required for the management of public area. The TOP10NL elements are needed for orientation. Other differences are that TOP10NL, in contrast to IMGeo, has many elements required for water navigation. In addition TOP10NL identifies a few elements originating form the military history of TOP10NL.

4. Towards one domain model and one key register topography

Based on the findings of Section 3, this section discusses the feasibility of one domain model topography (Section 4.1) and of one key register topography (Section 4.2).

4.1 Towards one domain model topography

The first question for 'collect once, use many times' is how feasible one domain model topography is using both the requirements for such a model as well as the two domain models TOP10NL and IMGeo as starting point.

Such an integrated model can be accomplished in several ways. In the most optimal way, that is when concepts are modelled in exactly the same way, it can be realised by modelling the concepts at the largest scale (= IMGeo) and model TOP10NL classes as derivation of IMGeo classes (and TOP50NL as derivation of TOP10NL etc). In this approach, information at smaller scales is usually reduced by applying coarser classification and generalisation operators such as merge, simplify etc. At the same time information that is only relevant at smaller scales can be introduced at these smaller scales, but should preferably be collected during the largest scale data collection process.

This optimal integration in one domain model topography seems to be obvious, giving the fact that IMGeo models reality at scale 1:1k and TOP10NL models the same reality at scale 1:10k. However in Section 3 many differences were identified in the definition of concepts. Consequently deriving current TOP10NL from current IMGeo is almost impossible. Examples are difference in definition and division of roads; lacking classes, attributes and attribute values compared to the other model; attributes with same name and different use; and, concepts modelled with different definitions, classes and/or attribute(value)s such as building, railway and terrain.

Two steps are required to integrate TOP10NL and IMGeo. The first step is harmonisation of the two information models, which also requires that information that is currently not defined in the models, but for example in acquisition rules, should be made explicit. Many differences seem to be random and easy to solve. Consequently the following questions need to be answered:

- Are there any errors (for example lack of classes, attributes or attribute values) in the models?
- Which differences in model approaches should persist since the underlying motivation justifies the differences?
- Which differences in modelling can be harmonised based on agreement of concepts without having significant consequences for one of the models?
- Which information only becomes relevant at smaller scale?
- Which classes, attributes and attribute values have different names but are defining the same concept?
- Which classes, attributes and attribute values have the same name but are used differently?

The second step for integrating TOP10NL and IMGeo is defining a set of rules that unambiguously define how TOP10NL objects can be derived from IMGeo objects. For example that IMGeo verges are converted into Terrain, land use 'gras' in TOP10NL. (Although one should realise that information is lost here that might be needed again at smaller scales where roads do cover a larger area including verges.) For defining derivation rules between the two information models, we propose a Base Model Topography (BMT) that maps IMGeo concepts to TOP10NL concepts and that contains clear derivation rules in UML in combination with Object Constraint Language (OCL). This model starts with modelling reality as a coherent, scale-independent collection of topographical classes where both IMGeo and TOP10NL can be derived from. The model can function as intermediate model between the abstract NEN3610 model at the one side and IMGeo and TOP10NL, TOP50NL, TOP100 etc at the other side. Most optimally the Base Model Topography will be incorporated in IMTOP, which integrates TOP10NL to TOP1000NL.
The approach of a Base model Topography is illustrated in Figure 8 for the concept PartOfRoad (*Wegdeel*).

The modelling principles for this example are based on the multi-scale information model IMTOP (see Stoter et al, 2008). For every concept a super class is modelled which con-

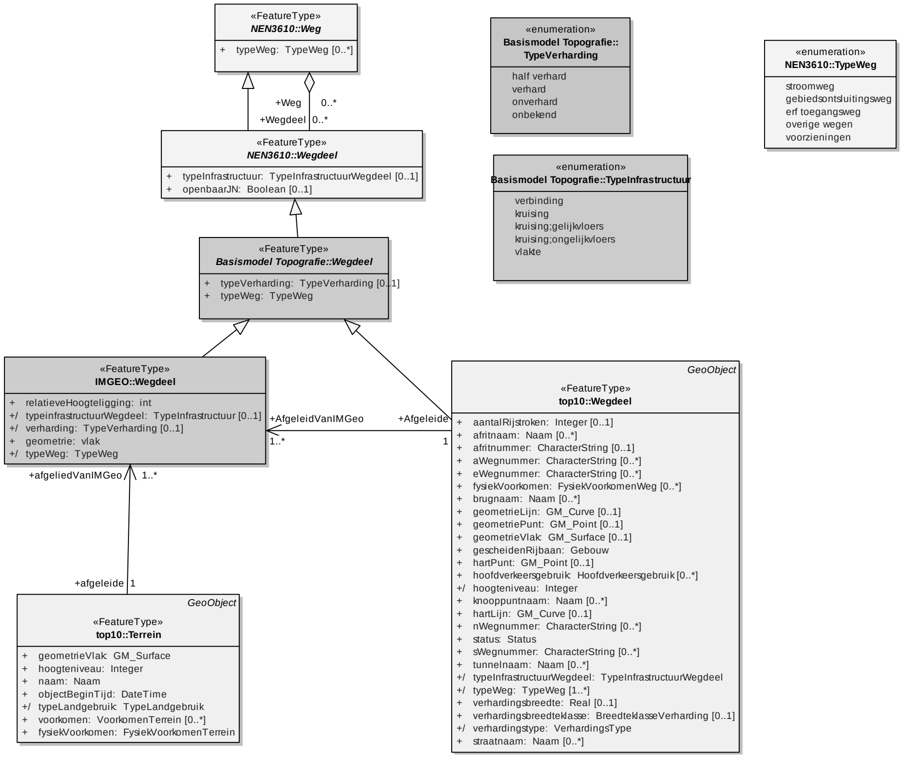


Figure 8. Concept PartOfRoad (Wegdeel), modelled in Base Model Topography.

tains attributes and attribute values which are valid for both domains. In a next step for both IMGeo and TOP10NL the same class is defined that inherits all properties form the super class. In addition 'derived from' (*afgeleid van*) relationships are defined to specify that TOP10NL classes are derived from IMGeo classes. With Object Constraint Language (OCL) it can be defined how the objects and attributes are derived (denoted with '/' in UML). The notion that PartOfRoad instances in IMGeo of type 'verge' may be converted in TOP10NL terrain instances (as currently is the case) is shown in Figure 8. The domain specific classes can have extra attributes which are only valid for the specific domain. It should be noted that the ideal solution for the 'verge' concept would be harmonisation, i.e. agree whether it belongs to terrain or road.

Although IMGeo is the largest scale model, new information is introduced in TOP10NL (see Figure 8). To enable collecting data for any scale and purpose during IMGeo data acquisition information that is required at any scale should be pushed down to the IMGeo model. However if the extra information is not relevant for IMGeo domain, one should consider modelling the BMT-classes as non-abstract classes and collecting information on the BMT classes. In a next step both IMGeo and TOP10NL can be derived from BMT.

4.2 Towards one key register topography

The second issue for 'collect once, use many times' is the feasibility to maintain one key register topography at the largest scale, from which topographical datasets at each predefined smaller scale can be derived. On the medium term this will not be possible, since no full automated solutions are available (Mackaness et al, 2007). Since this is partly due to incompatible data models and specifications, harmonisation of models will partly solve the generalisation problem.

Furthermore the question to automatically derive a topographical dataset at scale 1:10k from GBKN would only be relevant if TOP10NL data would not exist independently. In a few years there will be object oriented, well structured datasets at all required scales: IMGeo, TOP10NL, TOP50NL, TOP100NL etc. Collect once therefore mainly concerns the data acquisition for updates at the largest scale. Datasets at smaller scales should make use of these data in the re update processes, most optimally via update propagation, see for example (Uitermark, 2001). In this way the two key registers BGT (for large scale data) and BRT (covering datasets at several scales) can co-exist and co-function in the SDI according to the principle collect once, maintain multiple times at key registers at different scales and use many times, until the optimal situation will be achieved.

5. Conclusion

This paper reported about a research aiming at integrating two domain models that model topography at different scales and for different purposes. Integration is required to achieve one domain model and one key register topography to serve the national SDI in general and INSPIRE in particular. This integration is not straightforward as was shown in this paper. The proposed short to medium term approach is therefore to study which differences are random and can easily be harmonised. For the fundamental differences it is recommended to respect the two different points of view captured in the domain models. For integrating the IMGeo and TOP10ML models, after they have been better aligned, a Base Model Topography is proposed (most optimally extending IMTOP) that maps similar concepts in the two domain models. The consequence is the co-existence of topographical datasets at different scales and for different purposes. Condition for maintaining two key registers topography within an SDI is that these multi-scale representations should be accomplished in a smart manner so that different representations of the same real world object are aware of each other.

Research questions for this approach are: which updates at the largest scale are relevant for the smaller scales? In what way can database objects at different scales representing the same real world object be linked, which can be very complicated in case of n:m relationships, or when objects at smaller scales are deleted or when the definition of concepts change at scale transitions as in the case of IMGeo and TOP10NL? How can updates in a large scale dataset be generalised into updates in smaller scale datasets taking into account the complicated relationships between the different scales? How can the information related to scale, application and derivation as specified in the Base Model Topography be implemented in a DataBase Management System?

The starting point in the presented research are the already available datasets and domain models. Providing these datasets within the context of key registers available in an SDI requires harmonisation of concepts. Firstly to reuse the collected data in providing smaller scale datasets as was shown in this paper. However the effort that is required to make implicit information on meaning and content of data explicit will also be indispensable for reusing the data by applications from other domains.

Acknowledgements

I would like to express my sincere gratitude to Arjan Hofman (Logica), Dick Krijtenburg (Municipality of Amsterdam) and Nico Bakker (Kadaster) for their very important inputs and fruitful discussions. In addition I would like to thank Peter van Oosterom and Jaap Zevenbergen who reviewed an earlier version of this paper and gave insightful suggestions for improvement.

References

BAG, 2006, Catalogus Basis Gebouwen Registratie (2006), versie 4.0, Ministerie van VROM, Projectbureau Basisregistraties voor Adressen en Gebouwen (BAG), http://bag.vrom.nl/ufc/file/bag_sites/a9c332aa4bb9765d34c1be8581949dd1/pu/BGR_cataloguc_2006_def.pdf, 2006.

Geonovum, 2008, Standaarden voor Geo-informatie, the National Spatial Data Infrastructure (NSDI) executive committee in the Netherlands, http://www.geonovum.nl/standaarden.html.

Hofman, A.M., P. Dilo, P. van Oosterom and N. Borkens (2008). Developing a vario-scale IMGeo using the constrained tGAP structure. Using the constrained tGAP for generalisation of IMGeo to Top10NL model, Workshop on Generalisation and Multi-representation of the International Cartographic Association, Lyon, June, 2008. http://aci.ign.fr/montpellier2008/papers/27_Hofman_et_al.pdf.

IMGeo (2007). Informatiemodel Geografie (IMGeo); beschrijving van het model, IMGeo versie 1.0, Werkgroep IMGeo, http://www.geonovum.nl/informatiemodellen/imgeo/, 2007.

INSPIRE (2009). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 http://inspire.jrc.ec.europa.eu/.

ISO (2005). ISO 19109:2005 Geographic information – Rules for application schema.

Kadaster (2005). GBKN-gebouwen in TOP10NL. Een productieplanning op basis van de uitkomsten van de 'workloadtests' van de bladen 7 West en 22 West. Kenmerk: X6557. Kadaster. Apeldoorn, 2005.

LSV GBKN (2007). GBKN Handboek, versie 2.1, Stichting LSV GBKN, Documentnr. 07.05/052 2007. http://www.gbkn.nl/nieuwesite/downloads/07.05.065%20GBKN%20handboek%20VIP U2.1.pdf.

Mackaness, W.A., Ruas, A. and Sarjakoski, L.T. (2007). Generalization of Geographic Information: Cartographic Modelling and Applications. Elsevier, ISBN 9780080453743.

NEN3610 (2005). Basic scheme for geo-information – Terms, definitions, relations and general rules for interchange of information of spatial objects related to the earth's surfaceNormcommissie 351 240 "Geo-informatie".

Stoter, J.E., Morales, J.M., Lemmens, R.L.G., Meijers, B.M., van Oosterom, P.J.M., Quak, C.W., Uitermark, H.T. and van den Brink, L. (2008). A data model for multi – scale topographical data. In: Headway in spatial data handling: proceedings of the 13th international symposium on Spatial Data Handling, SDH 2008 / ed. by A. Ruas and C. Gold. Berlin: Springer, 2008. ISBN 978-3-540-68565-4 (Lecture Notes in Geoinformaton and Cartography) pp. 233-254.

TOP10NL (2005), TOP10NL versie 2.3, februari 2005, http://www.kadaster.nl/top10nl/gegevensmodel_top10nl_2%203.pdf.

Uitermark, H.T., 2001. Ontology-based geographic dataset integration. PhD Thesis. Computer Science Department, University of Twente, Enschede, the Netherlands (http://purl.org/utwente/36476).