

Harmonising and Integrating Two Domain Models Topography

Jantien Stoter¹, Wilko Quak² and Arjen Hofman³

¹ITC, the Netherlands, stoter@itc.nl

²TU Delft, the Netherlands, c.w.quak@tudelft.nl

³Logica, the Netherlands, arjen.hofman@logica.nl

Abstract

This article presents a case study on harmonising and integrating two domain models topography into a global model that have been established for different purposes; IM-Geo defines topography to serve municipalities in maintaining public and built-up area; TOP10NL defines topography for visualisation at 1:10k scale. The study identifies problems and proposes solutions to accomplish integration, which is required to provide the datasets within the principles of the national Spatial Data Infrastructure. At first the types of differences between the current models are analysed. Secondly, the article formulates recommendations to harmonise the differences which may be random (i.e. easy to solve) or fundamental (to be addressed in the integration). Finally, the article presents modelling principles for an integrated model topography based on two conclusions of the comparison study: two domain models are necessary to meet the specific demands of the two domains and secondly, TOP10NL cannot be derived from IMGeo because differences in perspective proved to be more dominant than scale differences did. Both the recommendations for harmonisation and the modelling principles are illustrated with prototypes which show the problems and potentials of harmonising and integrating different local (national) data models into global models.

Keywords: Information modelling, data harmonisation, integrating domain models, information model topography.

1. INTRODUCTION

An important objective of the INSPIRE Directive is to reduce duplicated data collection (INSPIRE, 2007). An absolute necessity for ‘collecting data once, and use it many times’ is harmonising specifications of datasets to fully integrate data from various sources. This is both valid for different datasets covering one state, but also for datasets of different states that touch at borders in order to “ensure that spatial data relating to a spatial feature the location of which spans the frontier between two Member States are coherent. Member States shall, where appropriate, decide by mutual consent on the depiction and position of such common features” (INSPIRE, 2007).

EuroGeographics has responded to the INSPIRE Directive by launching the EuroSpec project (EuroSpec, 2009). This project is the collective contribution of the National Mapping Agencies (NMAs) to build the European Spatial Data Infrastructure (ESDI), in line with the concepts of INSPIRE. An important activity of the project focuses on harmonised pan-European and cross-border specifications for large scale topographic data that goes beyond the successes of harmonised pan-European small scale products, such as EuroGlobalMap (scale 1:1million) and EuroRegionalMap (scale 1:250k) for topography and EuroBoundaryMap (scale 1:100k) for administrative units (EuroGeographics, 2009).

Although it may seem straightforward that topographic datasets within and between countries will contain similar types of objects, many (small) differences occur between

these datasets. Afflerbach et al. (2004) studied the differences between four national topographic datasets within the context of the GiMoDig project (Geospatial Info-Mobility Service by Real-Time Data-Integration and Generalisation (GiMoDig, 2009): Germany, Finland, Sweden and Denmark and found many differences. For example, different geometries for similar concepts (e.g., road centre lines representing individual lanes or road centre lines representing the whole road construction); differences between classifications for water (different types of water), for roads (using either widths or official administrative categories as classification criteria) and for recreational areas (which may have been further classified into 'amusement park', 'campground', 'parks', but not in all datasets); different minimum size criteria to collect area objects such as parks and forests; different collection criteria for hydrographical networks resulting in different densities for same types of hydrography.

Because of these differences in the definition of concepts several problems occur. Firstly, national, local data models are not easy to map to global (e.g., European) data models that realise the harmonisation without information loss. Secondly, a question such as "what is the total forest coverage of Europe?" is not easy to answer, because different conditions are used to identify an area with trees as forest. For instance what minimum density of trees is required to identify forest? What is the minimum dimension of the area to identify it as forest? What are, apart from presence of trees, criteria to identify forest, i.e. the function of the area (recreation or hunting), the maintenance characteristics of the area or the type of land-cover?

Besides cross-border differences between topographic datasets which is caused by lacking agreement of the national topographic data producers, differences may occur between topographic datasets covering the same country. The reason for these differences is also originating from different data organisations being responsible for producing and maintaining the different datasets with their own goals in mind. For example, municipalities collect and maintain large scale topographic data in support of the management of public and built-up area, while National Mapping and Cadastral Agencies (NMCAs) collect and maintain topographic data for the same area to represent maps at different scales (also at large scale). The last few decades these datasets are being translated into object oriented datasets to support database applications and GIS analysis. Key question for providing these different datasets within an Spatial Data Infrastructure (SDI) how they relate to each other to enable collecting data of same objects once in the future.

This article studies the feasibility of harmonising and integrating two independently established information models topography, expressed in UML (Unified Modelling Language) class diagrams. UML diagrams are often used to describe the content and meaning of datasets. The term 'harmonising' is used in this article as 'agreeing on thematic concepts' and 'integrating' as 'defining how objects in one dataset can be derived from objects in another dataset'. The article will provide insight into generic problems and solutions to accomplish such harmonisation and integration.

The case study contains two datasets representing topography at different scales for different purposes in the Netherlands. For both datasets information models have been established that describe the content and meaning of the data. The first dataset is the object oriented Large-scale Base Map of The Netherlands (*Grootschalige Basiskaart Nederland*: GBKN; LSV GBKN, 2007) defined in the Information Model Geography (IMGeo, 2007). Municipalities are the main providers (and users) of this dataset. The second dataset is the topographic dataset at scale 1:10k provided by the Netherlands' Kadaster and defined in the TOP10NL information model (TOP10NL, 2005). Harmonis-

ing and integrating these two domain models have become an important issue now 'key registers' are being established to support the Dutch SDI (VROM, 2009a). Legally established key registers contain authentic base data and their use is mandatory for all public organisations. As long as harmonisation and integration is not realised, municipalities and the Kadaster need to collect data of the same objects in parallel to meet the requirements of the specific domain in which they operate. This requires two key registers topography:

- *Basisregistratie Grootchalige Topografie* (BGT), 'key register large scale topography'. The feasibility of BGT as key register defined in IMGeo is being studied (VROM, 2009b);
- *Basisregistratie Topografie* (BRT), 'key register topography', in force since 2008. Currently BRT only contains topographic data at scale 1:10k. From 2010 the smaller scales will be added to this register (VROM, 2009c).

This article aims at formulating recommendations and proposing modelling principles, illustrated with UML examples, for an integrated information model topography that serves both domains. This integration at conceptual level can be used to move towards 'collect once, use many times' in the future. Although the case study is limited to the Netherlands, it presents common needs and problems for harmonising core national topographic databases within and across countries as well as solutions to establish global (e.g. European) data models.

Section 2 identifies the differences that need to be addressed in the integrated model topography. These differences are categorised based on the results of Hofman et al. (2008) and Stoter (2009), who studied the differences and commonalities between the two information models in detail. Section 3 proposes the integrated model topography based on results and conclusions of the comparison study. The proposal consists of two parts: a) recommendations to harmonise differences as much as possible, b) modelling principles for the integrated model topography. For both parts representative solutions are presented not limited to the case study. Consequently the proposed solutions can serve as recommendations and modelling solutions for harmonising and integrating information models established for different purposes. The article ends with conclusions in section 4.

2. DIFFERENCES TO BE ADDRESSED

Both IMGeo and TOP10NL are domain models extending the abstract data model NEN3610. The ISO compliant version of NEN3610 (Basismodel Geoinformation) was established in 2005 (NEN3610, 2005). This data model provides the concepts, definitions, and relations for objects which are related to the earth surface in the Netherlands. Domain models extend NEN3610 by defining their classes as subclasses of the NEN3610 GeoObject. Therefore these classes inherit all properties of the NEN3610 GeoObject. Examples of domain models are 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). ISO19109 defines such a domain model as "application schema" (ISO, 2005): 'a conceptual schema for data required by one or more applications'.

The idea behind NEN3610 and the extended domain models was that inheritance of the same NEN3610 GeoObject would assure harmonisation of the domain models. However, when comparing the information models TOP10NL and IMGeo, which both extend NEN3610, we observe that many differences need to be addressed before inte-

gration can be realised. This section lists the types of differences to be addressed which is a result of analysing the studies of Hofman et al. (2008) and Stoter (2009). The types of differences that we will present in this section are:

- Differences in perspective (section 2.1);
- Differences in main classes (section 2.2);
- Differences in object demarcation (section 2.3);
- Differences in attribute values (section 2.4);
- Different classes for same concepts (section 2.5);
- Same attribute name for different concepts (section 2.6);
- Differences in amount of information (section 2.7), and
- Differences in class definitions (section 2.8).

2.1 Difference in perspective

IMGeo and TOP10NL model the same geographic extent and same types of objects from a different perspective. The difference in perspective is due to differences in objectives, source data, scale, application domain, providers, acquisition method and rules, see Table 1. These differences have resulted in different contents of the datasets.

An example is how topology is implemented in the datasets (see Table 1). Terrain, water and road objects in TOP10NL that are visible from above form a planar partition (i.e. no overlap or gaps); whereas IMGeo models the planar partition at ground level. Consequently in IMGeo objects can be located above the planar partition (indicated with `relativeHeight > 0`). In contrast, in TOP10NL no objects can be located above the planar partition: objects with `heightlevel=0` are located at ground level or on top of a stack, for example in case of infrastructural objects at crossings, *and* they are part of the planar partition.

2.2 Differences in main classes

Table 2 lists the main, non-abstract classes that occur in either IMGeo, TOP10NL or in both models. Also the corresponding NEN3610 classes are listed. 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.

A comparison of the main classes provides the following insights. Six classes occur in both models. In addition, Geographical Area (used to link toponyms to objects), Functional Area (used to group objects of different classes) and Relief are only modelled in TOP10NL (Relief not available in NEN3610). Furthermore, IMGeo and NEN3610 distinguish Engineering Structure which is not available as separate class in TOP10NL. Another observation concerns the classes related to buildings. NEN3610 models Building Complex (*Gebouw*), Building (*Pand*) and Living Unit (*Verblijfsobject*). IMGeo only models Building and Living Unit in accordance with the Building and Address Register (BAG, 2006). TOP10NL only models Building Complex, which also includes single buildings when they are larger than a minimum size.

A final observation of comparing the main classes, is a similar granularity of NEN3610 classification at the one hand and IMGeo and TOP10NL at the other hand, i.e. they contain more or less the same number of classes. However, since IMGeo and TOP10NL extend the abstract data model NEN3610 to define content of specific datasets, one would expect refinement of the classes, i.e. more classes in the domain

Table 1: Differences in background between IMGeo and TOP10NL.

	IMGeo	TOP10NL
Objectives	enabling and standardising exchange of 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 semantic description of the terrain for TOP10vector, according to requirements of internal and external users of the TOP10vector dataset
Source data	Object oriented GBKN	TOP10vector
Scale	1:1k in urban area; 1:2k in rural area	1:10k
Application domain	management of public and built-up area	- visualising objects in map at scale 1:10k. - Large scale GIS analyses
Providers	municipalities, water boards, provinces, manager of Dutch railway infrastructure (Prorail), Department of Public Water (Rijkswaterstaat), Kadaster	Kadaster
Acquisition method	terrestrial measurements	aerial photographs completed with terrain acquisition
Acquisition rules	no generalisation is applied	little generalisation 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 smaller than 2 meters are represented as lines
Topology	all objects of any class with polygon geometry and relativeHeight '0' divide the terrain into objects without any gaps or overlaps - '0' means 'part of the terrain' - possible values are ..., -1, 0, 1 etc - all objects at ground level form planar partition	all objects of classes PartOfWater, PartOfRoad and Terrain and heightlevel '0' form a complete partition without any gaps or overlap. - '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 terrain	buildings are located on top of terrain

models. Compared to NEN3610, TOP10NL only further specialises the Relief class in five subtypes; IMGeo only further specialises Layout Element in eleven subclasses and Registration Area in nine subclasses. A major consequence of the limited number of classes in both models is heterogenous classes which are hard to harmonise: it is easier to agree on the definition of a lamp post or a recreational area than on the definition of layout element respectively terrain.

2.3 Differences in demarcation of objects

IMGeo and TOP10NL differ in how they demarcate objects during acquisition. The demarcation of objects is only limitedly defined in the models, but becomes clear when comparing the underlying datasets.

We use the example of road to illustrate two main differences in the demarcation of area objects. Because of a minimum width of 2 meters, many TOP10NL road areas are assigned to neighbouring objects. Examples are parallel roads (cycle paths and foot-paths) and parking areas and verges along a road. Sometimes these areas are assigned to the neighbouring terrain and sometimes to the neighbouring roads. The small

**Table 2: Main classes in NEN3610, IMGeo and TOP10NL.
Dutch translations added in italics.**

Class	NEN3610	IMGeo	TOP10NL
(PartOf)Road (<i>Wegdeel</i>)	Yes	Yes	Yes
Terrain (<i>Terrein</i>)	Yes	Yes	Yes
(part of)Water (<i>Waterdeel</i>)	Yes	Yes	Yes
(PartOf)Railway (<i>Spoorbaanddeel</i>)	Yes	Yes	Yes
Layout Element (<i>Inrichtingselement</i>)	Yes	Yes	Yes
Registration Area (<i>Registratief Gebied</i>)	Yes	Yes	Yes
Building (<i>Pand</i>)	Yes	Yes	No
Living Unit (<i>Verblijfsobject</i>)	Yes	Yes	No
Engineering Structure (<i>Kunstwerk</i>)	Yes	Yes	No
Building Complex (<i>Gebouw</i>)	Yes	No	Yes
Geographical Area (<i>Geografisch gebied</i>)	Yes	No	Yes
Functional Area (<i>Functioneel gebied</i>)	Yes	No	Yes
Relief (<i>Reliëf</i>)	No	No	Yes

roads themselves are represented with line geometry. In contrast, IMGeo does model these small area objects as specific types of roads. These differences become clear in Figure 1. In this figure TOP10NL road objects only cover the roadways (thus adding parking areas and footpath areas to terrain class), while IMGeo road objects cover the full construction of roads.

The second type of differences in the demarcation of area objects is the division of objects into 'part of' objects. This is well defined in TOP10NL, i.e. Road is divided into PartOfRoads at crossings. IMGeo will most probably follow the division as applied in GBKN. In this dataset roads are divided into parts based on maintenance characteristics such as paving type and administrative boundary. The difference can be seen in Figure 1, where the division of IMGeo roads is different than the division of TOP10NL roads.

Besides differences in demarcation of area objects, also linear objects may contain different geometries in both models as for rails. The line geometry for rails assigned to class Rails in IMGeo represents the middle of the rails (specialisation of Layout Element). In contrast TOP10NL models the geometry of rails as centre lines representing the whole railway body, assigned to class Railway.

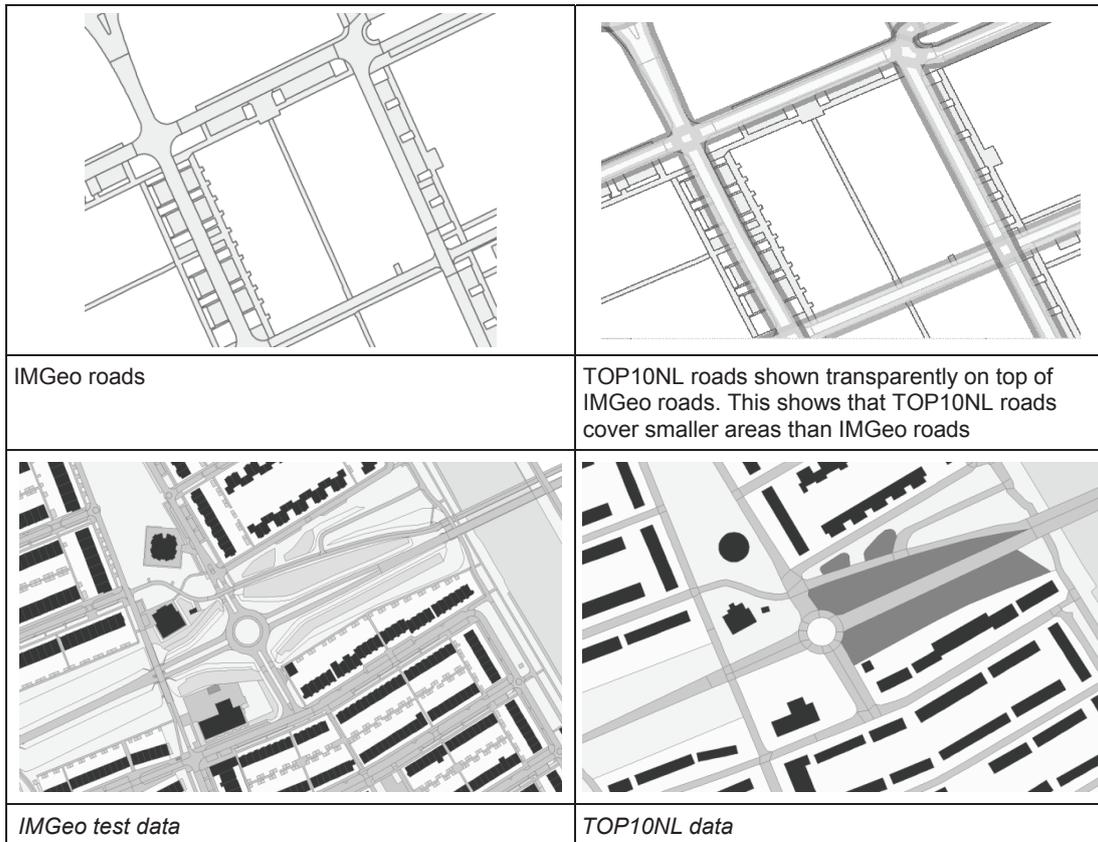
A last example of differences in demarcation of objects concerns whether buildings are included in the dataset or not. IMGeo models all buildings independent of their size. TOP10NL only models buildings that meet a minimal area (3x3 m).

The way objects are demarcated is mostly only available in acquisition rules and not in the models. Consequently for harmonising differences in demarcations, it is most important to formalise this information in the domain models.

2.4 Differences in attribute values

The attribute values in both models differ slightly in many cases. The differences between the paving types and railway types, as shown in Table 3, are representative for such differences. The differences are small and may not be significant, such as open pavement (IMGeo) and partially paved (TOP10NL).

Figure 1: Examples of IMGeo data (courtesy of municipality of Almere) and TOP10NL data (courtesy of Kadaster).



Another important difference to be solved for integration are the different values for terrain type (also shown in Table 3). None of the twelve IMGeo types has exactly the same label as one of the nineteen TOP10NL types. A few types are presumably the same (grass and grass-land; 'nature and landscape' and heather). In addition IMGeo contains coarser classifications for two types of terrain: forest and green object (see Table 3). A last difference in terrain types worth mentioning is 'built-up' area in TOP10NL to identify terrain on which buildings are located. Since buildings cause a gap in the terrain (see Table 1), IMGeo does not have 'built-up' terrain.

A last noticeable example of slightly different attribute values (not shown in Table 3) are Layout Element types. Of the eighty identified types in both models, only nine have exactly the same name, examples are tree, hedge, wall, and sign post. Ten types that differ in name presumably model the same concept, examples are road closing (TOP10NL) and barrier (IMGeo); hectometer stone (IMGeo) and milestone (TOP10NL). All other 60 types cannot be mapped. The IMGeo types are mainly originating from the utility sector or required for the management of public area. The TOP10NL types are needed for orientation or are related to Defense (the original application domain of TOP10vector).

Table 3: Slightly different attributes and attribute values pointing at same concepts.

Concept	IMGeo			TOP10NL		
	Class	Attribute	Value	Class	Attribute	Value
Paving	Part Of Road Terrain	typeOfPaving	closed pavement open pavement unpaved	Part Of Road	pavingType	paved unpaved partially paved unknown
Railway types	Rail	typeofRail	crane, metro, tram, train, fast tram/lightrail			
	Railway	typeofRailway	crane, metro, tram, train metro, fast tram/lightrail, railway verge, to be defined	Railway	typeofRailway	mixed metro tram train
Terrain type	Terrain	typePartOfTerrain	forest grass nature and landscape culture land other green object industrial terrain uncultivated terrain courtyard area with plants recreational area sport terrain embankment	Terrain	landUseType	mixed forest brushwood deciduous wood coniferous wood grassy area heather arable land orchard tree cultivation poplar graveyard graveyard in forest jetty sloped stones built-up area fruit cultivation loading bay area for railway sand

2.5 Different classes for same concepts

Several concepts are modelled with different classes, as shown in Table 4. An example is the concept 'verge' which is terrain of type 'grass' in TOP10NL and a specific type of road in IMGeo. Another example are sport area, recreational area and industrial area. These are considered Terrain in IMGeo and Functional Area, i.e. a collection of objects of different classes, in TOP10NL. Lastly, Engineering Structure is a specific class to model infrastructural structures such as viaducts, bridges, locks and dams in IMGeo. In TOP10NL these objects are modelled as a specific type of infrastructural objects (PartOfWater, PartOfRailway or PartOfRoad) or as a Layout Element.

Table 4: Same concept, differently modelled.

Concept	IMGeo	TOP10NL
	Modelled with class	
Verge	PartOfRoad	Terrain
Industrial Area	Terrain	Functional Area
Recreational Area	Terrain	Functional Area
Sport Area	Terrain	Functional Area
Engineering structure	Engineering Structure	Layout element PartOfRoad/PartOfWater/PartOfRailway

2.6 Same attribute name for different concepts

Table 5 shows how the attribute `typeOfRoad` assigned to `PartOfRoad` is used in a different manner in the two models. IMGeo uses the attribute to distinguish different parts of a road; TOP10NL to define a hierarchy of roads required for visualisation. As shown in Table 5, NEN3610 even uses the attribute in a third way.

Table 5: Different use of attribute `typeOfRoad`.

	IMGeo	TOP10NL	NEN3610
Used for	identifying different parts of a road	defining hierarchy for roads for visualisation	defining hierarchy for function of roads
Example values (not complete)	parking area public transport footpath verge roadway cycle path pedestrian area residential area	highway main road regional road local road street	continuous road access road access road to residential areas other roads facilities

2.7 Differences in amount of information

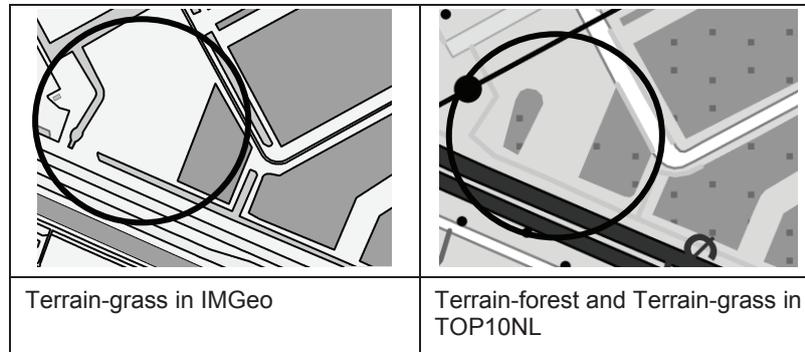
In general, TOP10NL models more information than IMGeo. For example TOP10NL contains more attributes for its classes than IMGeo. Most likely this is because GBKN, the underlying dataset of IMGeo, contains less attributes than TOP10NL data. Another example of less information in IMGeo compared to TOP10NL is that IMGeo identifies 'forest' versus four types of forest in TOP10NL and one 'green object' versus four types in TOP10NL (see Table 3). A last example of less information in IMGeo refers to sport area and recreational area. IMGeo classifies sport and recreational areas as single objects where TOP10NL identifies different types of objects (roads, buildings, terrain) that constitute the areas (see also section 2.5).

2.8 Differences in class definitions

Differences in application domains have led to different classifications for the same objects, for example when is an object 'forest', 'grass', 'recreational area' and/or 'area with plants'? We can observe these differences when comparing the datasets, i.e the instances of the classes, but they are not apparent from the models. In TOP10NL a wooded area may be split in two areas: deciduous wood and coniferous wood to identify the type of land-cover. The same area can be split in a different way in IMGeo depending on whether the wood is maintained (area with plants) or not (forest). Figure 2 (Almere) shows another example of such differences. A forest object is identified within

grassy area in TOP10NL data (right). IMGeo data does not identify this forest area (left). From the information models the reason for this difference does not become clear.

Figure 2: Differences in instances of same class (i.e. Terrain).



3. INTEGRATED INFORMATION MODEL TOPOGRAPHY

Although TOP10NL models topography at scale 1:10k and IMGeo at scale 1:1k in urban area and 1:2k in rural area, we can conclude from the previous section that TOP10NL cannot be considered as a derivation of IMGeo. This is not surprising since none of the models used the other model as starting point. Also the differences in history, objectives, providers, source data and stakeholders explain the differences of two domain models topography: one supports management of public and built-up area and one visualises topography at scale 1:10k.

Because of these differences, topographic data of the same objects is currently collected twice to serve two application domains. To collect topographic data that meets the requirements of both domains in the future, the domain models need to be integrated. Such integrated model will assure consistency when users (or applications) move from one dataset to another. Based on the conclusions of the comparison study, this section presents recommendations to accomplish the integration.

Starting from the current differences, two main steps are required to build the integrated information model topography. Firstly harmonisation, i.e. agreeing on definitions of concepts, should be accomplished as much as possible. Section 3.1 describes recommendations for harmonising the differences identified in section 2. The result will be two better aligned domain models topography. Section 3.2 presents and motivates the modelling principles for the integrated model topography. This model defines how topographic data on real-world objects can be collected once and used in both the IMGeo and TOP10NL domain, starting from the harmonised versions of IMGeo and TOP10NL.

3.1 Recommendations for harmonising

The first main step for harmonising is to study whether the differences between the models are principal or random: which differences in modelling can be harmonised based on agreement of concepts without having significant consequences for one of the models? To support this harmonisation, this section formulates recommendations for harmonising the differences identified in section 2.

Some differences are not clear from the models, but were identified by comparing the datasets. Therefore our first recommendation, before harmonising the models, is to model any information about the content and meaning of datasets that is currently not included in the domain models, e.g. acquisition rules. This allows harmonising and integrating this information as well. For example, the current TOP10NL acquisition rules state that railway banks are not measured, despite the presence of the class Railway in the model.

Many differences between the information models are caused by *difference in perspective* (section 2.1). These different perspectives cannot easily be harmonised because they are justified by differences in objectives, source data, application domain, providers, and acquisition methods. However, the law on key registers provides potentials for harmonising some parts of the perspective. Specifically two aspects of the law enforce municipalities to inform the Kadaster about updates for TOP10NL data. The first aspect is that any user of the key register must inform the provider when (s)he notices an error. Secondly, municipalities are obliged to use the Kadaster's TOP10NL data, updated every 2 years, instead of their self-produced 1:10k datasets, updated more frequently. To effectuate updates in TOP10NL data as soon as possible, some large municipalities will send TOP10NL updates based on their 1:10k datasets. For this purpose they are currently converting their 1:10k datasets into TOP10NL compliant datasets. This practice makes the integration issue of the two topographic datasets, i.e. IMGeo and TOP10NL, relevant within municipalities.

Because not all information on the data is laid down in the TOP10NL model, TOP10NL data can be generated from a municipal perspective without violating the model. An example is that the Kadaster often assigns road areas that are too small to be area objects (smaller than 2 meters) to terrain. However, assigning these areas to neighbouring roads better supports the municipal maintenance task of public area and fits better with the definition of PartOfRoad in the TOP10NL model. Consequently municipal TOP10NL roads may cover the full construction of all IMGeo road objects which solves the *differences in object demarcation* of roads (section 2.3). To illustrate this, two TOP10NL road implementations, one generated by Kadaster and one generated by municipality of Rotterdam, are compared with IMGeo roads in Figure 3.

Obviously this poses new research questions, since now the differences in perspective do not occur between IMGeo and TOP10NL, but within one dataset, i.e. TOP10NL. In conclusion, to solve the *differences in object demarcation* it is most important to make these demarcations unambiguously explicit in the models. In a next step it can be studied whether differences can be aligned and how.

A first step in harmonising *differences in main classes* (section 2.2) is to model more specialisations (i.e. subclasses). The result will be more homogenous classes on which it is easier to agree. Figure 4 shows an example. The left part of the figure shows the current Terrain class in TOP10NL with its different attribute values for different types of terrain. Integrating IMGeo and TOP10NL requires agreeing on the concept of Terrain. The alternative modelling with subclasses for different types of terrain (Figure 4, right) requires only agreeing on the definition of types of terrain, for example, Farmland or Forest.

Figure 3: IMGeo roads (a) and TOP10NL implementations of roads by Kadaster (b) and municipality of Rotterdam (c).

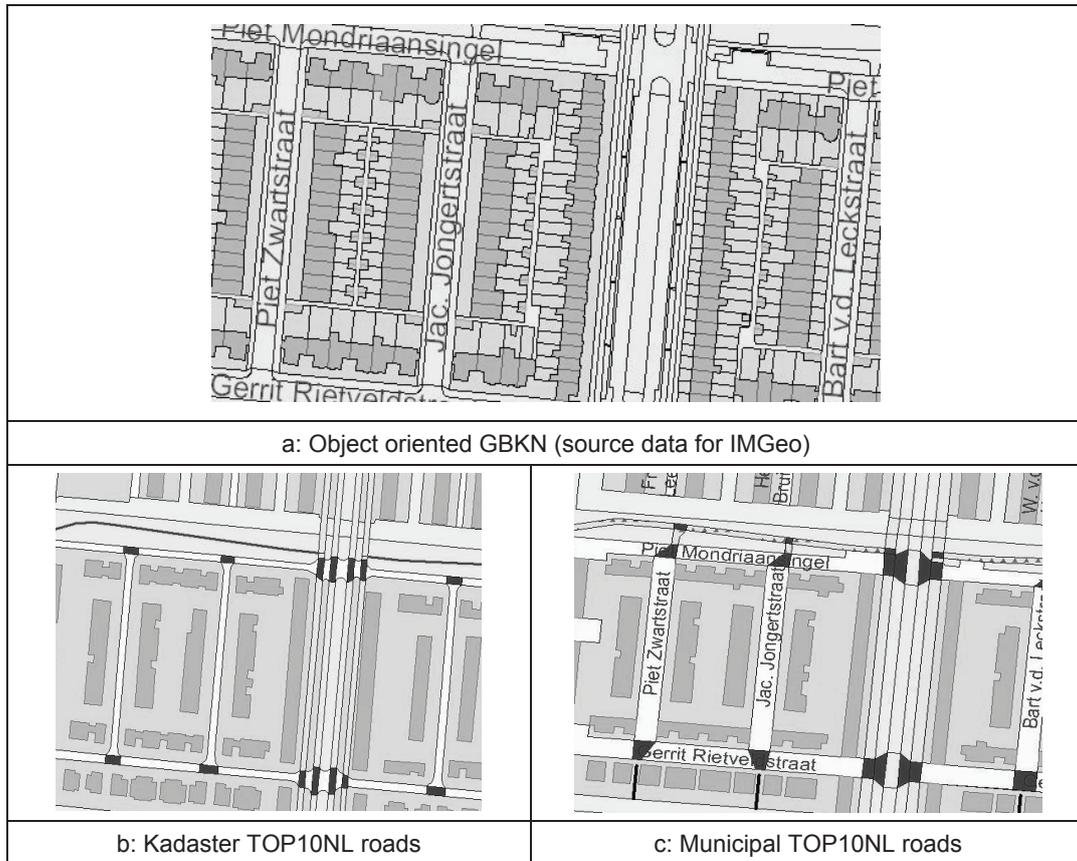
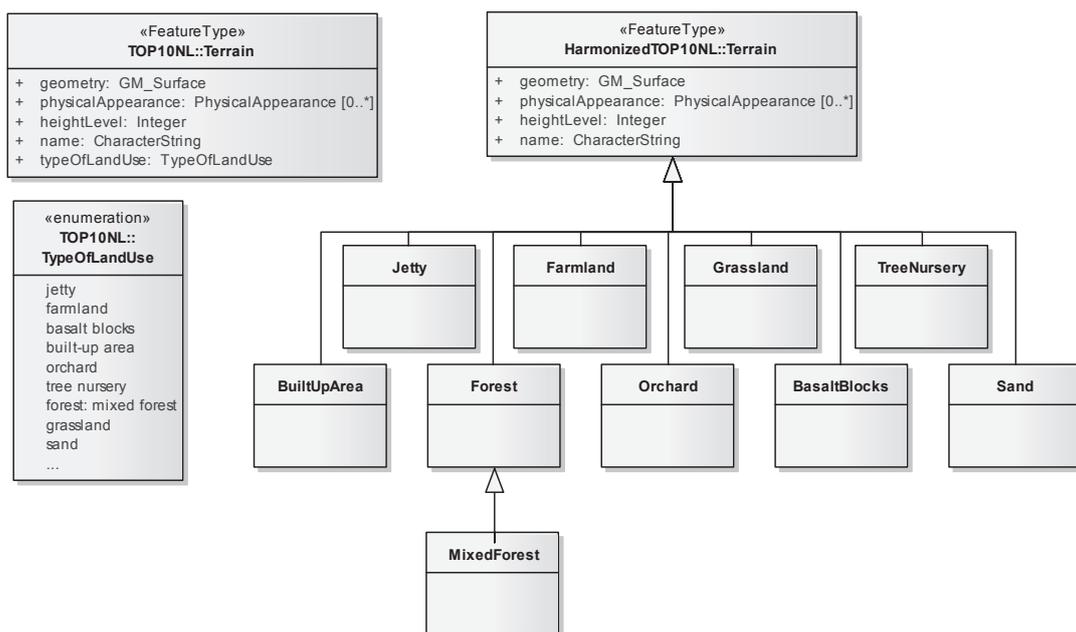


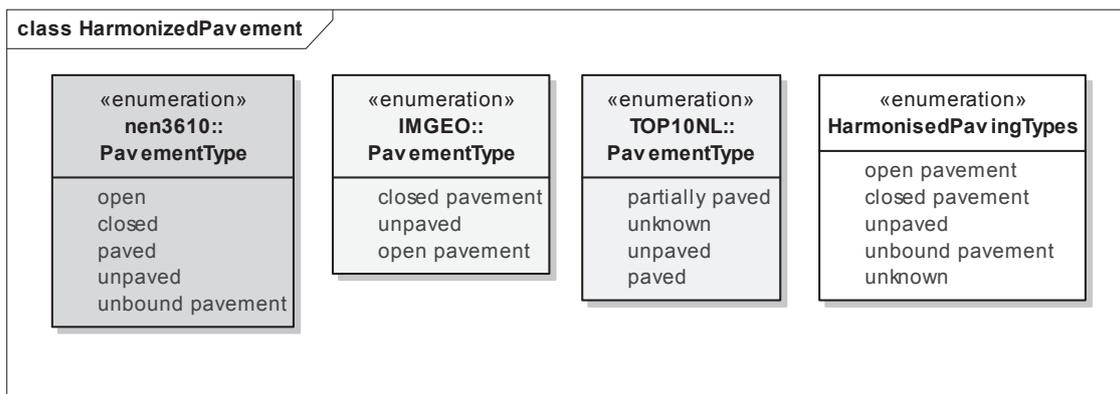
Figure 4: Left: Terrain class in current TOP10NL model. Right: Subclasses for different types of Terrain result in homogenised classes.



Such homogenous classes will also avoid that *different classes are used for the same concepts* (section 2.5).

Differences in attribute values (section 2.4) can be harmonised through lists of common types completed with harmonised values in case of non-significant differences, for example for the paving types (Figure 5) and railway types. Some values may remain information model specific, example is the built-up area for TOP10NL terrain types which is lacking in IMGeo because buildings cause gaps in the terrain.

Figure 5: Harmonised values for paving types.



The *same attribute name for different concepts* (section 2.6) can only be harmonised by agreeing on common use of attributes. To avoid such differences in the future, attribute names should be used that have less ambiguous semantics.

To solve *differences in amount of information* (section 2.7), information required at the smallest scale (TOP10NL), but not available in the largest scale (IMGeo) can be either moved down to the largest scale or be removed from the smallest scale. Moving down information to IMGeo is only of interest for municipalities when it is relevant for their application domain.

To solve *differences in class definitions* (section 2.8) new objects at cross sections of classifications could be generated. However, a class for every possible combination makes the models more complex. An example are the four possible combinations for *area with plants/forest* (IMGeo) and *deciduous/coniferous wood* (TOP10NL): *DeciduousWoodAreawithPlants* etc. A better option is therefore to keep the classes from the original models. This will result in overlapping polygons in the datasets, but since two different concepts are registered for the same area (maintenance and land-cover) this reflects the real-world situation. In any case the exact definitions of classes should be unambiguously defined in the information models. In the current situation such differences only become clear when comparing the data (i.e. instances of classes).

Harmonising the information models using these recommendations will result in better aligned IMGeo and TOP10NL models. The more harmonisation can be achieved, the more straightforward the integration of the two domain models will be.

3.2 Recommendations for integrating: Base Model Topography

We propose an information model topography that integrates the two information models. The modelling principles that we present here are motivated by two important conclusions of the comparison study (section 2).

Firstly, two datasets defined in two information models topography are necessary to meet the specific demands of the two domains, i.e. IMGeo for maintenance of public area and TOP10NL for visualisation at scale 1:10k. Secondly TOP10NL cannot be derived from IMGeo, because the application domain proves to be more dominant on the perception of topography than scale is.

Starting from these two conclusions, we propose an intermediate domain model between NEN3610 at the one side and IMGeo and TOP10NL at the other side: Base Model Topography (BMT). The motivation for this intermediate layer instead of solving the integration within NEN3610 is that NEN3610 is meant to integrate at a higher level of abstraction. The two conclusions that direct the modelling principles of BMT are invalid for all domain models under NEN3610. Consequently it is better to solve the integration of these two topographic domains outside NEN3610.

BMT is an information model defining scale-independent topographic classes where both IMGeo and TOP10NL can be derived from. The BMT classes respect both the IMGeo and TOP10NL perspectives on topography. However, they do not necessarily have the same label (see further). For the moment BMT defines how concepts in IMGeo are related to concepts in TOP10NL. This provides consistency for users (and applications) when moving from one dataset to the other. However, the data is still separately collected until an organisation has interest to collect data to serve both domains. In that case the 'collect once, use many times' principle will be realised through collecting data on BMT classes. Therefore they contain all information that becomes relevant in any dataset that needs to be derived from BMT.

The modelling principles of our approach are based on the multi-scale Information Model TOPography (IMTOP, see Stoter et al., 2008). IMTOP, which integrates topographic data at scales 1:10k to 1:1000k for the Netherlands' Kadaster, proposes an abstract super class for every topographic class. These super classes have subclasses at all scales and only contain attributes and attribute values valid for all scales. The super classes are abstract and the data is collected for the largest scale dataset, while smaller scale datasets are derived from the next larger scale dataset.

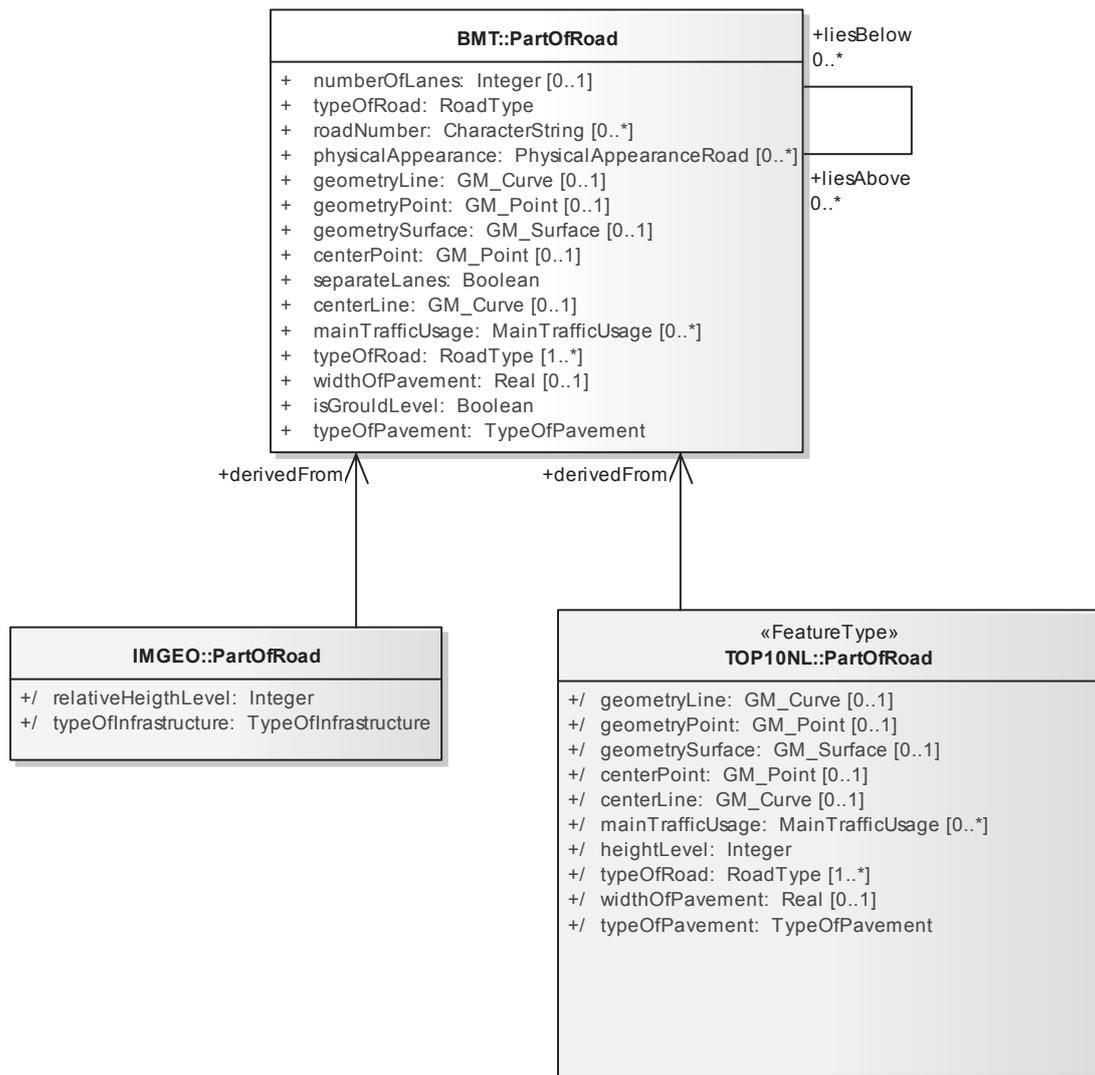
Similar to IMTOP, we define IMGeo and TOP10NL classes that are derived classes from BMT classes. An example is shown in Figure 6 where we model the derivation of the PartOfRoad object. Constraints defined in Object Constraint Language (OCL) can define how objects in IMGeo and TOP10NL can be derived from BMT.

Although we follow the main principles of IMTOP, the proposed BMT differs on a few fundamental aspects.

Firstly, the name of the BMT classes and the derived classes can be different because of different perspectives on concepts (see Figure 7). In contrast, every IMTOP super class occurs as subclass with the same name in each scale. For example the properties of an IMTOP road super class are inherited by road class at 1:10k scale, by road class at 1:50k scale, by road class at scale 1:100k. Secondly, we define association relationships between BMT classes and the derived classes, instead of a generalisation/specialisation relationship as in IMTOP. The reason is that BMT classes and the derived classes do not necessarily represent the same concepts. Thirdly, the BMT classes (comparable to super classes in IMTOP) are non-abstract. The reason for this is that objects in both domains (IMGeo and TOP10NL) are derived from instances of

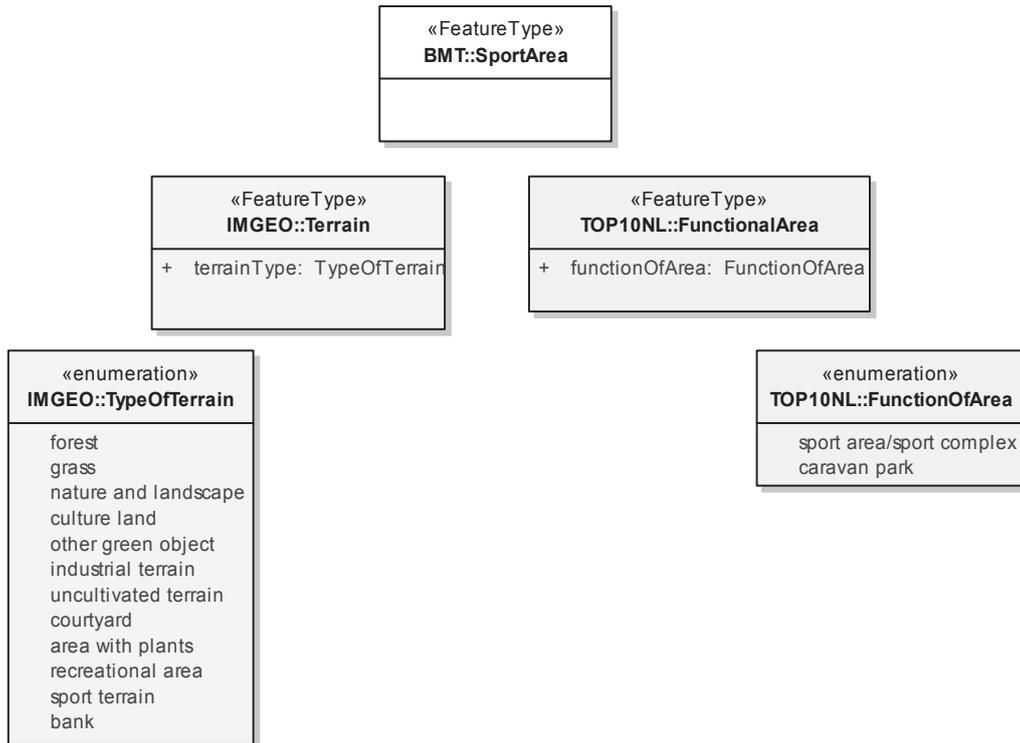
BMT classes, which contain all information required to derive both IMGeo and TOP10NL data. Fourthly, we recommend moving all information down to BMT to avoid extra data acquisition for derived datasets. This implies that all attributes of the IMGeo and TOP10NL objects are derived, except the Identifier and other system attributes not shown in Figure 6. In IMTOP the classes at smaller scales can have extra attributes which are only valid (and collected) for the specific scale. Our BMT approach also slightly differs with the i(integration)-classes as identified in the multi-representation approach of Friis-Christensen and Jensen (2003). The i-classes only contain attributes that are valid in the corresponding classes, as the super classes for IMTOP.

Figure 6: Concept of PartofRoad, modelled in integrated model topography.



Finally, we recommend relationships (liesAbove and liesBelow) and the Boolean attribute IsGroundLevel to every BMT class (as shown in Figure 6) to derive both the relativeHeight and heightlevel attributes required for IMGeo respectively TOP10NL. Consequently both the IMGeo and the TOP10NL implementation of topology can be derived from BMT.

Figure 7: Modelling sport area as BMT class and derived classes in IMGeo and TOP10NL.



4. CONCLUSIONS

In this article we studied requirements and possibilities to harmonise and integrate two independently established information models topography. The harmonisation and integration consists of several steps.

At first we identified types of differences that have to be addressed. Apart from these differences, four other conclusions from this comparison study are important for harmonising and integrating the two models. Firstly, for many differences it is not clear whether they are random (i.e. easy to harmonise) or fundamental (i.e. to be addressed in the integration). This requires further study. Secondly, because not all information is defined in the models, datasets compliant to the models may be implemented with different ‘flavours’. These ambiguities are unwanted when reusing data of other domains in SDIs. Therefore an important recommendation is to make information on the content and meaning of data as much as possible explicit in the information models. Thirdly, two information models topography are necessary to meet the specific demands of the two domains, i.e. maintenance of public area and visualisation at scale 1:10k. Finally TOP10NL cannot be derived from IMGeo, because the application domain of these two large scale data sets determines the different perspectives on topography rather than scale does. At large scales (also valid for scale 1:10k) objects can be represented with their true geometries, and therefore harmonisation and integration is mainly a schema matching problem. At the smaller scales, symbolisation causes objects to be altered with respect to reality. Consequently at smaller scales harmonisation and integration becomes merely a multi-scale problem, i.e. how can a dataset be converted into a dataset with fewer details.

Based on the conclusions, the article formulated recommendations to harmonise the differences and presented modelling principles to define an integrated model topogra-

phy, both illustrated with UML examples. These recommendations, principles and illustrations show the problems and potentials of harmonising and integrating different data models into global data models to enable data provision within national and international SDIs. First the proposed integrated model formally defines how concepts in one dataset relate to concepts in another dataset. In a future step the results of this study can be further developed to move towards 'collect data once, maintain it at several domain databases, and use it multiple times'. Comparing similar developments in other countries, for example aligning Teknisk Korte and TOP10 for the Danish SDI, can be very useful here.

REFERENCES

- Afflerbach, S., Illert, A. and T. Sarjakoski (2004). "The Harmonisation Challenge of Core National Topographic Databases in the EU-Project GiMoDig". *Proceedings of the XXth ISPRS Congress*, July 12-23, 2004, Istanbul, Turkey, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXV (B4:IV): 129-134.
- BAG (2006). Catalogus Basis Gebouwen Registratie (2006), versie 4.0, Ministerie van VROM, Projectbureau Basisregistraties voor Adressen en Gebouwen (BAG), at: http://bag.vrom.nl/ufc/file/bag_sites/a9c332aa4bb9765d34c1be8581949dd1/pu/BGR_cataloguc_2006_def.pdf. In Dutch.
- EuroGeographics (2009). http://www.eurogeographics.org/eng/00_home.asp.
- EuroSpec (2009). http://www.eurogeographics.org/eng/01_EuroSpec_goals.asp.
- Geonovum (2008). Standaarden voor Geo-informatie, the National Spatial Data Infrastructure (NSDI) executive committee in the Netherlands, at: <http://www.geonovum.nl/standaarden.html>. In Dutch.
- Friis-Christensen A. and Jensen C.S. (2003). "Object-relational management of multiply represented geographic entities". *Proceedings of the Fifteenth International Conference on Scientific and Statistical Database Management*. Cambridge, MA, USA, July 9–11, pp 183–192.
- GiMoDig (2009). <http://gimodig.fgi.fi/>.
- 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, at: <http://www.geonovum.nl/informatiemodellen/imgeo/>, 2007. In Dutch.
- INSPIRE (2007). <http://inspire.jrc.ec.europa.eu/>.
- ISO (2005). ISO 19109:2005 Geographic information - Rules for application schema.
- LSV GBKN (2007). GBKN Handboek, versie 2.1, Stichting LSV GBKN, Document number 07.05/052, at: <http://www.gbkn.nl/nieuweweb/downloads/07.05.065%20GBKN%20handboek%20VIPU2.1.pdf>. In Dutch.
- 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 surface. Normcommissie 351 240 "Geo-informatie". In Dutch.

- 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 topographic data", *Lecture Notes in Geoinformation and Cartography*.
- Stoter, J.E. (2009). "Towards one domain model and one key register topography", in: P. van Oosterom (ed.). *Core Spatial Data*, Delft: Netherlands Geodetic Commission.
- TOP10NL (2005). TOP10NL versie 2.3, februari 2005, http://www.kadaster.nl/top10nl/gegevensmodel_top10nl_2%203.pdf. In Dutch.
- VROM, (2009a). Ministry of Housing, Spatial Planning and the Environment <http://www.vrom.nl/pagina.html?id=12050>. In Dutch.
- VROM (2009b). Ministry of Housing, Spatial Planning and the Environment <http://www.vrom.nl/pagina.html?id=36586>. In Dutch.
- VROM (2009c). Ministry of Housing, Spatial Planning and the Environment <http://www.vrom.nl/pagina.html?id=36688>. In Dutch.