Considerations for the design of a semantic data model for a multi-representation topographical database

Jantien Stoter[1], Wilko Quak[2], Peter van Oosterom[2], Martijn Meijers[2], Rob Lemmens[1], Harry Uitermark[3]

[1]Institute ITC, International Institute for geo-information science and earth observation
P.O. Box 6, 7500 AA Enschede, The Netherlands
{stoter|lemmens}@itc.nl

[2]Delft University of Technology, OTB, GIS technology
Jaffalaan 9, 2628 BX Delft, The Netherlands
{B.M.Meijers|C.W.Quak|P.J.M.vanOosterom}@tudelft.nl

[3]Kadaster
Hofstraat 110, 7311 KZ, Apeldoorn, The Netherlands
harry.uitermark@kadaster.nl

Abstract. This paper reports on the design of a semantic data model that integrates topographical databases at different map-scales at machine level in order to facilitate the (semi-)automated generalisation process (generation of small scale databases from large scale databases). Apart from the data content to produce coherent topographical databases at the different scales (object classes, attributes, attribute values, relationships and constraints), the data model should contain information to produce coherent topographical maps at different scales. The context, requirements, possible alternatives and criteria to be fulfilled by the semantic data model are presented. From the first experiments it can be concluded that for modelling data content covering all scales, a UML class diagram that starts with topographical object classes that occur at any scale looks most promising. These object classes are defined as superclasses and subclasses are created at the moment specific attributes or attribute values, relationships or constraints are only applicable for a specific scale. Future research will focus on how the consistency of geometry, topology and generalisation related spatial structures can be defined in the UML class diagram. In addition it will be explored how cartographic constraints can be defined in the UML <<interface>> object which is created for every map product that has to be produced.

Keywords. generalisation, multi-representation, multi-scale, ontology, semantic data modelling, topographical databases

1 Introduction

Since several decades National Mapping Agencies (NMAs) are maintaining vector databases at different map-scales to support map production processes. To make the production line more efficient and to assure consistency of the databases - and consequently maps - machine based integration of the databases at the different scales is required (note that in this paper 'scale' refers to the scale of a map). The integration of topographical databases at different scales should be accomplished in such a way that it supports the automatic generation of (updates in) a smaller scale database from (updates in) a larger scale database. The integration, resulting in a multi-representation topographical database, starts with the design of an integrated data model in which all relevant information and knowledge is formalised. This information comprises more than object classes defined with their geometry, topology, attributes, attribute values, constraints and relationships. It also comprises knowledge on the geographical meaning (semantics) of object classes in order to produce meaningful presentations at different scales: what is the spatial and semantic context of an object class, what are the cartographic characteristics (such as graphical conditions to adhere) of an object and of a collection of objects for different scales?

The problems of data integration of multi-scale and multi-representation databases have been addressed extensively in previous research, for example on database integrity with multiple representations and on automatic retrieval of spatial information from a multiple representation database (Devogele et al., 1996; Jones et al. 1996; Friis-Christensen and Jensen, 2003; Parent, 2006). In our research we build on these findings. However the main objective of the integrated model in the research presented in this paper is to support automatic generation of small scale databases from large scales databases rather than to support the management of an existing multi-representation database. Therefore information related to generalisation and scale (transitions) should be explicitly modeled. At first the integrated model is meant for internal processes. However having a formal description addressing meaning of geo objects (an ontology) it is easier to exchange information with other domains. The Ordnance Survey (UK) published several articles on research to use an ontology for exchanging spatial data with other domains (Regnauld, 2007; Goodwin, 2005a; Goodwin, 2005b; Hart et al, 2004; Schwering and Hart, 2004; Greenwood and Hart, 2003; Hart and Greenwood, 2003).

This paper presents the approach to design an integrated semantic data model for a multi-representation topographical database for the Dutch NMA (TD Kadaster) in order to facilitate (semi-)automated generalisation between topographical databases and maps at different scales. The integrated model will be the result of collaboration between the Department of Geo Information Processing at ITC, the Department of GIS technology at TU Delft and TD Kadaster. The design of the model is currently in progress and is expected to be finished by the end of 2007. Consequently this paper will not present the model itself. Instead the paper defines the scope and rationale for the model: what should it be able to do, how to model databases at different scales while also addressing the transitions between them, what modelling directions are possible and what are the consequences of these directions, how much formalisation is required and possible. The paper is organised as follows. Section 2 describes the context of the model. Requirements that are laid down by the need to use the integrated model in automated generalisation processes are described in section 3. Section 3 also describes different modelling approaches for a multi-scale data model and shows their possibilities by presenting some modelling experiments. Section 4 describes criteria to test the quality and functionality of the integrated data model. The paper ends with a discussion (section 5).

2 The context of the semantic data model

The context of the integrated semantic data model is outlined by several developments that will be explained in this section. These are:

- object oriented databases at TD Kadaster (section 2.1)
- legally identified key-registers (section 2.2)
- standard data model for geo-applications in the Netherlands (section 2.3)
- compliance to standards (section 2.4).

2.1 Object oriented databases at TD Kadaster

The integrated model (working title: *TOPNL*) will be the follow-up of the data model that has been designed in 2005 for the largest scale that is produced by TD Kadaster, which is 1:10,000. This data model is called TOP10NL (Bakker, 2005). The object oriented TOP10NL data model is based on user's surveys and is therefore tailored to user requirements. Currently TD Kadaster is in the process of finalising the conversion of the 'old' TOP10VEC data sets into a database according to the object oriented TOP10NL data model. TOPNL should cover the object oriented versions of all TOPxxVEC databases (TOP50VEC, TOP100VEC, TOP250VEC, TOP500VEC and TOPmillionVEC). The first step was therefore to analyse the current product specifications in order to identify which object classes, attributes and attribute values should be covered in TOPNL. This analysis was carried out by TD Kadaster (see figure 1). The result is a multi-dimensional table with the following entrances: for every object class, for every attribute, for every attribute value, for every scale whether the combination (object class, attribute, attribute value) exists or not. In the analysis also the application schema for large scale topographical data (IMGEO) has been taken into account (see section 2.3).
It can occur that certain attributes or attribute values are available on a smaller scale but not in TOP10NL, e.g. attribute value 'roundabout' (*rotonde*), for attribute 'type infrastructure' defined for object class 'road' (*wegdeel*) is available in TOP50VEC but not in TOP10NL.
The table presented in figure 1 contains all information that should be contained in TOPNL: object classes to be represented at the different scales as well as their attributes and attribute values. The yes-no in the 'scale' columns (*ja,nee)* indicate relationships of the attributes and attribute values at the different scales: which attributes and attribute values are applicable to all scales or are altered (e.g. combined) at a certain scale. An attribute value (or object class or attribute) that does not change at a scale transition, does not necessarily mean that the attribute value has exactly the same meaning throughout all scales. For examples it is questionable if 'forest' at scale 1:10,000 has exactly the same meaning at scale 1:500,000. This aspect needs more attention. Currently the table is being extended with geometry types, e.g. if a road is

represented by a polygon or a line at a certain scale. Figure 1 is one representation of TOPNL. The next step is to translate this information into a semantic machine interpretable data model (see section 3.4). For generalisation it might be needed to add extra spatial information to the model. For example in order to treat buildings within and outside urban extents differently these extents should be known.

Feature class	Attribute	Attribute value	IMGEO	10	50	100	250	500	1000
geo-object	locatie	Aanduiding van locatie door adresgegevens	ja	nee	nee	nee	nee	nee	nee
geo-object	naam	Benaming van het geo-object	ja	nee	nee	nee	nee	nee	nee
geo-object	brontype	luchtfoto	nee	ja	ja	ja	ja	ja	ja
geo-object	brontype	kaart	nee	ja	ja	ja	ja	ja	ja
geo-object	brontype	RD	nee	ja	ja	ja	ja	ja	ja
geo-object	brontype	GBKN	nee	ja	ja	ja	ja	ja	ja
geo-object	brontype	top10vector	nee	ja	ja	ja	ja	ja	ja
geo-object	brontype	overig	nee	ja	ja	ja	ja	ja	ja
geo-object	bronbeschrijving	< tekst >	nee	ja	ja	ja	ja	ja	ja
geo-object	bronactualiteit	< datum >	nee	ja	ja	ja	ja	ja	ja
geo-object	bronnauwkeurigheid	< getal >	nee	ja	ja	ja	ja	ja	ja
geo-object	dimensie	2D	nee	ja	ja	ja	ja	ja	ja
geo-object	dimensie	3D	nee	ja	ja	ja	ja	ja	ja
wegdeel	type infrastructuur	verbinding	ja	ja	ja	nee	nee	nee	nee
wegdeel	type infrastructuur	kruising	ja	ja	ja	nee	nee	nee	nee
wegdeel	type infrastructuur	vlakte	ja	nee	nee	nee	nee	nee	nee
wegdeel	type infrastructuur	rotonde	nee	nee	ja	nee	nee	nee	nee
wegdeel	type infrastructuur	overig verkeersgebied	nee	ja	ja	nee	nee	nee	nee
wegdeel	type weg	OV-baan	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	overweg	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	pad	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	parkeervlak	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	perron (voor tramverkeer)	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	rijbaan	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	rijwielpad	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	vluchtheuvel	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	voetgangersgebied	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	voetpad	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	wegberm	ja	nee	nee	nee	nee	nee	nee
wegdeel	type weg	woonerf	ja	nee	nee	nee	nee	nee	nee

Fig 1: Relationships between data content at different scales

2.2 Legally identified key-registers

Important drive for the integration of databases at different scales is the new Dutch law on key-registers. This law, which will come into force January 2008, has identified TOP10NL as a key-register (VROM, 2007). The law states that smaller scales should follow as key-register from January 2010. Key-registers contain authentic data and are mandatory (and free) to be used by all public organisations. Two of the currently specified ten key-registers contain topographical data: the cadastral map and TOP10NL. Key-registers should be up to date, object oriented and produced with an update cycle of two years. In what way the Large Scale Map of The Netherlands (GBKN) will become a key-register has not been decided yet (an object oriented database does for example not yet exist). The law on key-registers as well as considerations to reduce production costs have pushed TD Kadaster to consider improving generalisation by automation. The first step in the automation is the design of TOPNL.

2.3 Standard data model for geo-applications in The Netherlands

TOPNL will be modelled within the generic basic scheme for geo-information in The Netherlands. A general data model for geo-objects in the Netherlands (Base model for Geo Information) was completed in 2005 (NEN3610, 2005). The aim of this model is to have common definitions for object classes in the geo-information domain. The objects involved are objects that occur in the terrain of which the location and geometry are fixed within a certain time period. Also objects which are not visible in the terrain are included such as cadastral parcels and administrative boundaries. The model was designed by RAVI (Dutch organisation for geo information in the public domain) in collaboration with ICTU (Dutch organisation for information and communication technology in the public sector), Wageningen University and Research Centre, Technical University of Delft and the Netherlands' Kadaster. To achieve commitment many users were consulted. The model defines object classes with identifiers, and descriptive, geometrical, temporal and meta data attributes. NEN3610:2005 describes these object classes at an abstract level. Geo-application domains are building/have built their specific domain models on this generic model, such as data model for spatial planning (IMRO), data model for cables and pipelines (IMKL) and data model for large scale geo-information (IMGEO). 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'. In the application schemas attributes of the NEN3610:2005 model can be refined.
The application schema TOP10NL has been embedded in NEN3610:2005. This was achieved by defining every object class of TOP10NL as subclass of a NEN3610-class: every TOP10NL object is a NEN3610 GeoObject. Consequently all TOP10NL objects inherit the properties of the NEN3610:2005 GeoObject (see also Quak et al., 2005). Also the coding lists for TOP10NL attribute values are integrated with the NEN3610:2005 coding lists. The general coding lists of NEN3610:2005 can be extended with TOP10NL specific codes. For example the TOP10NL model defines more types of forests than the NEN3610 data model. Similar as TOP10NL, TOPNL will also be embedded in NEN3610:2005.

2.4 Compliance to standards

As the generic NEN3610:2005 is modelled using the ISO TC/211 standards also TOPNL will comply to these standards. Most notably the model complies with ISO19109 'Rules for applications schema'. ISO19109 formalises the rules for building spatial data application schemas. The standard distinguishes between the conventional data interchange by transfer and the increasingly popular data interchange by transaction. The first implies that the application schema holds information about data structure and data content of the data set. Examples are SDTS (USA), Interlis (Switzerland), ATLIS (Germany) and NEN1878/NEN3610 (The Netherlands). In the second case, application schemas also hold information on a common interface to be used by the interchanging applications. The development of Spatial Data Infrastructures, such as the European INSPIRE (INSPIRE, 2007), is necessitating existing

spatial data transfer standards to be harmonised or revised, based on modern international standards, developed by ISO.
NEN3610:2005 implements ISO 19109 by making use of a general feature model (GFM) and a feature catalogue model. It also adopts OGC's Geography Markup Language (GML) as interchange standard. A feature catalogue defines object classes and their attributes. ISO has standardised the method of feature cataloguing in ISO 19110 'Methodology for feature cataloguing'. The feature catalogue of NEN3610:2005 defines a feature on a high abstraction level by describing generic properties such as location and geometry. In this way NEN3610 forms the common basis for more specific geo-information models used in topographic mapping, land-use planning, water management, etc. Feature catalogues are built from domain terminology classifications.

3 Integrated model for supporting automated generalisation

TOPNL should ultimately support automated generalisation to produce coherent topographic databases and topographical presentations at different scales (maps). This lays down requirements for the information captured in the model as well as its functionality. This section will elaborate on the information requirements for the model that are dictated by the (semi-)automated generalisation process as well as on possible modelling approaches for a multi-scale data model illustrated with some examples. The issues addressed in this section are:

- Three approaches for spatial data models for single applications (section 3.1)
- Define coherent topographical structures at different scales (section 3.2)
- Define coherent topographical presentations at different scales (section 3.3)
- Modelling approaches for a multi-scale data model (section 3.4)
- Define relationships between object classes at different scales (section 3.5)

3.1. Three approaches for spatial data models for single applications

When looking at spatial modelling in the past, three main approaches can be distinguished: 1) geometry/topology-first approach, 2) object-first approach, and 3) a hybrid approach. Because the related models have quite a different starting point, there is sometimes confusion between modellers. In the geometry-first approach, the models start from the geometry (topology). Attributes are added to these geometries in order to classify the objects. The result is typically a set of tables in the database such as point/symbol table, text/label table, line table and area table. Within a table all objects (records) have the same set of attributes. For example in the area table there may be houses and roads, all having the same attributes. In this approach, it is also possible to explicitly model the topological structure (e.g. linear network, or partition of space) with well-known advantages (explicitly connectively, avoiding redundancy, better guarantees for quality under updates). The Dutch cadastral map in *LKI* is a typical example of this geometry-first approach (Lemmen et al., 1998). In this solution objects may share, via topology, their geometry with other objects. It could be argued that map representations (on paper or screen) themselves, i.e. the visualisation of the spatial data, is also a geometry-first type of model as all objects are considered together in one geometry model.

The object-first approach models the object classes first with added geometry. Every object class can have its own set of thematic attributes which may vary for the different object classes. Also every object class has its own geometric description independent of any other object. TOP10NL is an example of this approach. Typically the result is a set of tables in the database such as houses, roads, waterways, which have among others their own simple object geometry type attribute. Sometimes additional rules (constraints) are added in order to avoid unwanted situations (often topology based), e.g. a house polygon should not overlap with a road polygon at same level. Drawback is that all constraints have to be explicitly stated (and checked when updates are performed) and are not embedded in the main structure of the model. Also the model does not explicitly contain the topological relationships, which may support various types of analysis (e.g. quality control of updates). It must be noted that topological relationships are very important for map generalisation; e.g. what are the neighbours of this object (candidates for aggregation), is the network connectivity damaged when this road segment is removed, etc.

The third approach, the hybrid approach, treats the geometry and object class equally. It combines the strengths of both approaches: the thematic attributes are specifically designed for every object class, but the model also enables shared geometry and use of embedded structure. The spatial domain is partitioned and the result is described using tables for nodes, edges, and faces (and solids in 3D). The objects are modelled in the same way as in the object-first approach with the exception that the objects do not have their own independent geometry-attributes, but refer to primitives in the geometry/topology part of the model (node, edge, face,…). This is the approach as described in the 'formal data structure' (FDS) theory of Molenaar (1989) and quite recently implemented in products such as 1Spatial (LaserScan) Radius Topology and Oracle spatial topology (first introduced in version 10g).

De Hoop et al. (1993) discuss the different modelling approaches and the consequences for realisation and use. It cannot be claimed that one model is 'better' than another model. This depends on the application context and use. If one specifies a number of important characteristic of the application domain and typical use, then it is possible to state which approach is preferred. Considerations could be: 1. allow exceptional overlapping of objects in certain cases (e.g. bridge over water), 2. allow modelling of systematically overlapping sets of object classes (e.g. topographic objects at one hand and administrative units at the other hand), 3. enable multiple geometry representations of single objects (e.g. road area polygon and road centre line, or building footprint polygon, building rooftop polygon, and building centroid), 4. support consistent updating/maintenance, 5. support efficient querying, analysis and viewing of data, 6. avoid storage space consuming representations (might also be expensive for data transfer), 7. support easy delivery for customers (simple objects might be easier to receive in another system than topology structure), etc. All three approaches can be extended in one way or another to add multi- or vario-scale aspects to the model. For example, it can be claimed that the tGAP structure (Van Oosterom,

2005; Van Oosterom et al., 2006) is a typical example of a vario-scale model based on the geometry-first principle.
NEN3610 implements both the object-first and the hybrid approach. Since TOPNL is the follow-up of TOP10NL, first focus will be on the object-first approach. Future research will focus on how geometry, topology and generalisation related spatial structures can be embedded in the model (hybrid approach).

3.2 Define coherent topographical structures at different scales

To assure coherent topographical structures at different scales the model should contain information on data content (object classes), real world abstraction and data capture rules. In the geo-domain the conceptual model covering data content is called a Digital Landscape Model (DLM). For a topographical database this also includes the geometry types of object classes at all scales (point, line, polygon) and the geometrical, topological and semantic relationship between object classes. It should be noted that the INSPIRE directive does not address the coherence of object classes at one scale (INSPIRE, 2007); it only defines separated topographical themes that need to be provided by national governments.As explained in section 3.1 starting from the object-first approach information addressing coherence of topographical structure can be specified in two ways: 1) defining constraints on and between object classes, or 2) defining geometry and topology properties of object classes as primitives in the geometry/topology part of the model (for example tGAP structure, Minimum Spanning Tree, Delauny Triangulation, Planar Graph).

3.3 Define coherent topographical presentations at different scales

Complementary to the DLM supporting a multi-representation topographical database a second type of model should be covered by TOPNL which assures coherent topographical presentations at different scales, called Digital Cartographic Models (DCMs). The DCM should define precisely how object classes should appear in a presentation (map) also with respect to their surroundings. This can be done using cartographic constraints. The concept of cartographic constraints was introduced in the field of automated generalisation late eighties to provide a framework for the flexible description of cartographic requirements in a formalised way. Weibel and Dutton, 1998 regard cartographic constraints as design specifications to which generalisation solutions should adhere. There are many publications on the successful application of cartographic constraints in the automated generalisation process either to trigger the generalisation process or to use the constraints in the evaluation of generalisation output. Examples are Mackaness et al., 1986; Nickerson, 1988; Brassel and Weibel, 1988; Beard, 1991, Schylberg, 1993; Mackaness, 1995; Ruas and Plazanet, 1996; Weibel and Dutton, 1998; Ruas, 1999; Barrault et al., 2001; Regnauld, 2001; Ware et al., 2003; Bard, 2004; Burghardt and Neun, 2006. We can distinguish between constraints addressing the preservation of shape and constraints addressing legibility aspects (Burghardt et al., 2007). At scale transitions cartographic constraints addressing preservation conditions are completely satisfied. These are constraints prescribing topology, position/orientation, shape, pattern and distribution/statistics. Preservation constraints can become violated when operations are applied in order to adhere legibility constraints (minimal dimensions and emphasis of objects). Cartographic constraints to be modelled in TOPNL can be divided into:

- geometric constraints for one object (e.g. minimal dimensions)
- geometric constraints for a meaningful group of objects (e.g. proximity constraints for a group of buildings bounded by a road network)
- geometric constraints between two objects (e.g. minimal distance between two parallel roads)
- topological constraints between two objects (e.g. adjacency between road and built-up area)
- semantic constraints between two objects (e.g. buildings should be accessible by a road; buildings cannot be surrounded by a lake)

According to (Ruas, 1998) cartographic constraints should be accompanied by information on the constraints in order to design the best generalisation process. This information covers:

- attributes that describe the characteristics of the constraint (initial and ongoing values), e.g. value of building density in an area
- threshold value under or over which a conflict occurs
- severity of conflict used for (intermediate) evaluation
- priority according to constraint type and severity of conflict
- list of advised generalisation operations to solve the conflict
- list of generalisation methods that should be avoided to maintain a specific situation (e.g. high filtering of a very sinuous road)

3.4 Modelling approaches for a multi-scale data model

TOPNL is not only meant to communicate the data content at the different scales to humans. Instead the main aim of the model is to implement machine based data integration in order to guide the (semi-)automated generalisation process. Therefore the model should have a high level of formalisation. The first approach will be to use UML to formalise the model. This model will be evaluated against the criteria defined for the model (see section 4), whereupon it will be decided whether a technology should be applied enabling a higher level of formalisation of knowledge (e.g. OWL; Web Ontology Language). The UML modelling covers two types of knowledge: a) knowledge that assures coherent topographical structures at different scales and b) knowledge that assures coherent topographical presentations at different scales. UML class diagrams are used to model the first type of knowledge. UML, with its focus on object classes, is most suited for the object-first strategy as described in section 3.1. For the second type of knowledge (related to cartographic presentations) it will be studied whether it is possible to define a UML <<interface>> object to represent different outputs with their own cartographic constraints attached. Note that these classes themselves do not really contain data, but can be seen as selections of data from the objects in the DLM-part of the model. In (Lemmen and Van Oosterom, 2006) an example of modelling such an interface object is given in the context of the cadastral domain model: OwnershipFolio (collection of facts from Person, RightRestrictionResponsibility and RegisterObject) and CadastralMap (collection of Parcel and Building objects).

For the purpose of TOPNL a difference can be made between data model constraints (used to define data content) and cartographic constraints (rules related to the topographical presentation at a certain scale as described in section 3.3). The use of constraints in data modelling has recently received more attention. Traditionally models (e.g. UML class diagrams) consist of object classes (with class name, attributes and operations) and relationships (inheritance, composition, association). Recently the role of constraints to specify valid contents of a model has been used more and more to include additional semantics within the model (Louwsma et al, 2006; Van Oosterom, 2006). TOP10NL is an example of adding constraints to the model to specify valid content. The same constraints are used in different parts of the system; e.g. enforced within the database (at check-in time), but also during interactive editing. For the UML class diagram covering data content and topographical structure at the different scales, there are several alternatives:

1. For every map-scale a UML class diagram is designed separately. In a second step all scale models are integrated into one supermodel in which relationships between classes at the different scales are defined.
2. The object classes that occur at any scale are the entities in the UML diagram. The specific appearance of the class at a specific scale is defined by constraints, e.g. if TOP50NL then geometry type of ‘secondary roads’ is line.
3. One UML class diagram is designed covering all possible object classes. For every class a “super” class exists containing properties that are valid at all scales. Attributes and attribute values that are scale dependent are modelled with subclasses (as specialisations), e.g. geometry type.

The three approaches are tested with road parts (in Dutch: *wegdelen*) which are modelled at scales 1:10,000 and 1:50,000 in order to show pros and cons of the different modelling approaches.

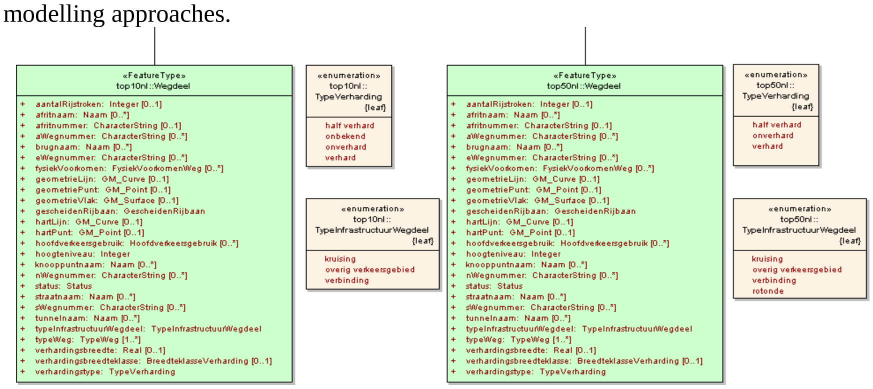

Fig 2: Separate UML class diagram per scale (inheritance from GeoObject is not shown)

Figure 2 shows the first modelling approach. This approach gives good insight into which object classes, attributes and attribute values (and relationships between object classes if the example was a bit more extended) are available at each scale. However, it is not immediately clear what the differences are between different scales. Also, because each scale has its own model, a lot of redundancy exists between the models (attributes are repeated for every scale), or at least, it can be questioned whether the meaning of similar attributes throughout the different diagrams remains the same. Finally it is questionable how flexible the model is when other outputs are required in the future such as a TOP50NL map with only water themes represented according to TOP10NL or if an output is required at a vario-scale (between two predefined map scales).

In figure 3 constraints are used to model the different scales. Note that there is no formal syntax used for the constraints; they are only meant as an example (for formal syntax OCL could be used). Attributes and attribute values that are disallowed at certain scales are restricted with constraints.

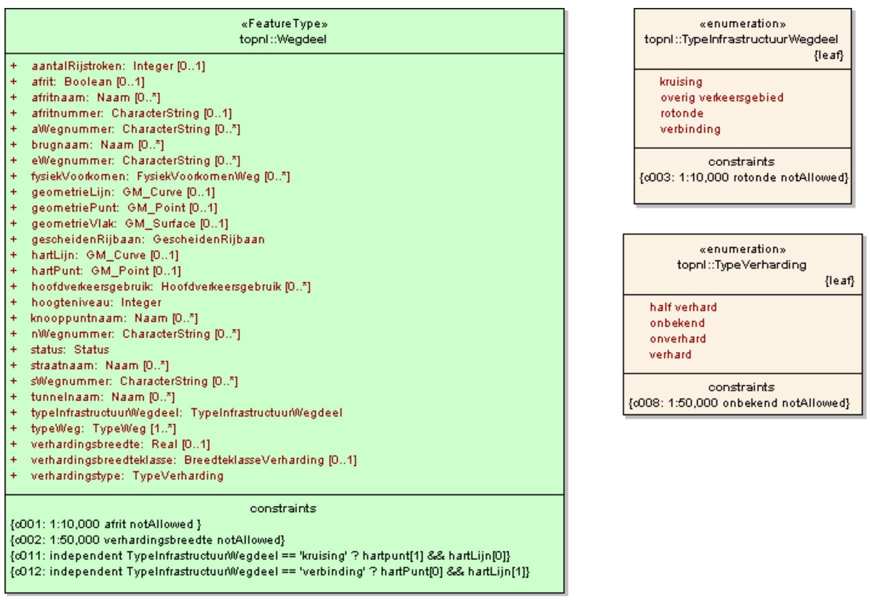

Fig 3: Using constraints to model scale dependent information (inheritance from GeoObject is not shown)

One of the disadvantages of this approach is that the model can become quite complex due to all the constraints necessary to define the appearances at every scale. Also because geometry can be dependent on multiple attributes. For example "if a road part is a crossing, then a point geometry should be stored, but no line geometry is allowed". It is also not clear how to automatically derive a model per scale with this approach. Another disadvantage of using constraints for scale-dependent information is that the data model can become non-transparent since constraints are also used to define data content as was mentioned before. Constraints that were not yet taken into account in the example are 'cross-class constraints', i.e. constraints that define data capture rules between classes (e.g. house and shed as separate objects must be located x meter apart from each other.).

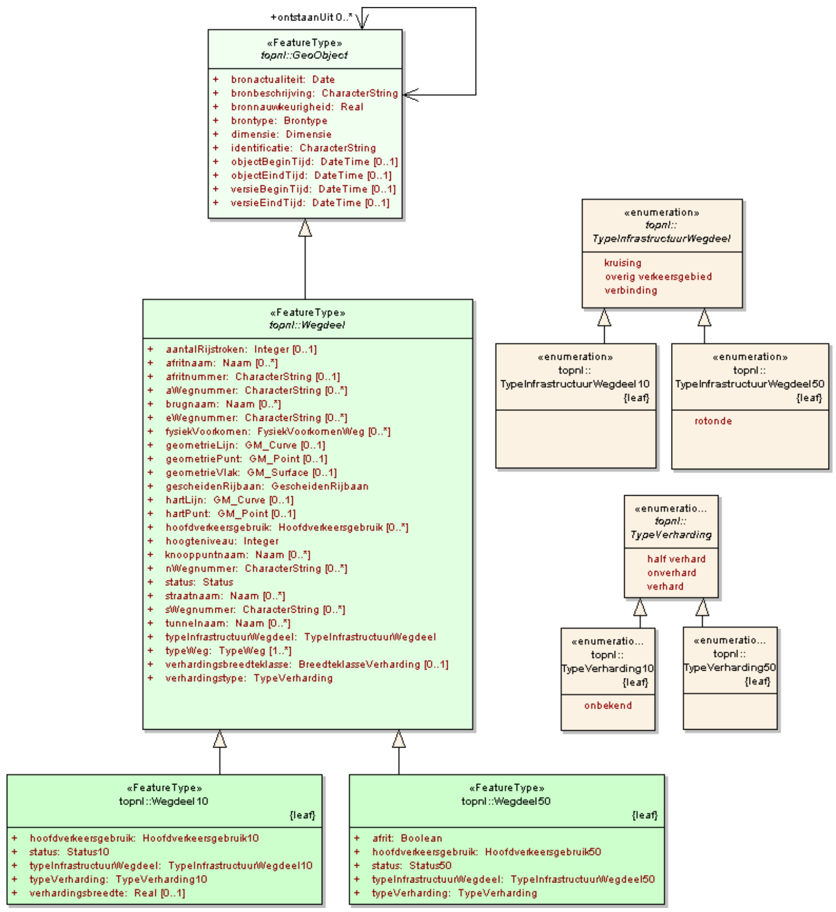

Fig 4: Using inheritance to define what is valid on every scale

Inheritance is used in the third approach to express what is applicable for all scales (figure 4). This approach is similar to the multi-representation strategy of the MADS model (Parent et al., 2006) and the approach to manage multiply represented geographic entities in a DBMS by Friis-Christensen and Jensen (2003), although these researches do not focus on deriving multi-representations at different scales but on keeping consistency of different existing representations of the same real world entity. As can be seen from figure 4 the disadvantage of this approach is that a lot of duplicate subclasses are modelled if one attribute is missing or present in only one of the scales. For example the '*rotonde*' attribute value (i.e. roundabout), is only present in the 1:50,000 scale, yet TypeInfrastructuurWegdeel has to be subclassed for each scale, due to its existence on the 1:50,000 scale only.

Advantages of this approach are that it is quite easy to get back to a model per scale by just showing only the relevant classes for that scale and that it is possible to work with scale ranges.

In order to make UML more suited for geometry (and topology) modeling we can extend the UML model with concepts for geometry. There is a formal way of extending the functionality of UML by extending the underlying meta-model of UML. In the example of figure 5 we use stereotypes and tagged values to extend the UML model: the tag <<MultiScale>> in the figure is used to indicate that the given class will have different representations at different scales. The stereotype is an indication for the user of the model that the stereotyped class has additional semantics. For MultiScale objects one of the most important semantics is that they are valid at specific scales. We use second extension mechanism of UML 'tagged values' to tag all MultiScale objects with a 'minScale' and 'maxScale' tag to indicate on which scales the object occurs. The same extension mechanism can be used to indicate that a specific attribute is only valid at specific scales. Of course the scale interval of an attribute should never exceed the scale interval of its class. We have stereotyped all spatial attributes with the <<MultiScale>> stereotype and tagged the geometries with the correct minScale and maxScale (unfortunately only the tags are visible in the graphic representation of the class).

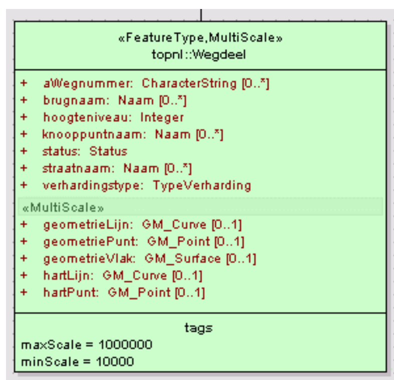

Fig 5: Example of using a MultiScale stereotype to model scale dependent information

This modelling exercise shows that the third approach (using superclasses for all object classes and subclasses to model information only related to specific scales) has the best potentials. Constraints are used to address the data content whereas information related to generalisation and scale issues are modelled with other means, which makes the model most transparent. This approach also links to previous approaches to model multi-representation databases and it shows good potentials to extract a data model for one specific scale.

The exercise has also raised some more questions. It is not clear how explicit the notion of scale is embedded in each of the modelling approaches. The experiments with the meta-model of UML look promising to make this more explicit. As was mentioned before it is also not clear whether the meaning of object classes (or attributes, attribute values) changes (slightly) from one scale to another so whether they should become new object classes at a certain scale.

3.5 Define relationships between object classes at different scales

In order to generalise a database at one scale from a database at a larger scale, it is important to lay down relationships between object classes at the different scales. Mostly this will be hierarchical relationships in the sense that object classes at larger scales have ‘specialisation-generalisations’ (taxonomic) or ‘composite-aggregation’ (partonomic) relationships with object classes at smaller scales. These relationships can be based on attribute values (“churches” and “mosques” are combined into “religious buildings”) or be a consequence of a generalisation operation that was triggered because a constraint was violated (buildings are aggregated into built-up areas when a threshold building density has been reached). At scale transitions first transformations are applied for a whole data set to convert geometry types (collapse or combine) and to take care of partonomic and taxonomic relationships based on attribute values. These transformations are based on transition rules embedded in the data model and are applied to all instances of an object class. Secondly cartographic constraints as described in section 3.3 are tested in order to solve graphical conflicts. These are only applied to specific instances and situations. It should be noted that relationships between object classes at different scales can also be non-hierarchical such as semantic relationships (e.g. between road and building). Besides relationships between different object classes relationships between instances should be maintained in order to appropriately manage the multi-representation topographical database.

4. Criteria to test the semantic data model

To evaluate the quality and functionality of TOPNL it must be clear and testable what the model should facilitate when it is finished. The criteria to test TOPNL are related to supporting the semi-automated generalisation process, also taking into account to which extent it is possible to formalise required information. These criteria comprise the possibilities:

1. to extract a UML class diagram for a specific scale
2. to produce a GML application schema (.xsd) for a specific scale from the UML diagrams generated in step 1
3. to generate the database structure for a specific scale from every schema generated in step 2
4. to generate the database structure for multi-representation topographical database by linking schemas at the different scales. This includes information on which object classes correspond to each other and how, also derivation and transition rules should be embedded in the database structure
5. to link instances at different scales and to attribute this linkages

6. to use all information (including supporting structures, scale and transition information) directly in the generalisation process, without (or with minor) human intervention
7. to store, manage and query a multi-representation topographical database
8. to support vario-scale presentations (between fixed scales)
9. to support progressive transfer (see Van Oosterom, 2005).

5. Discussion

In this paper context and requirements for an integrated semantic data model for a multi-representation topographical database are described. Key-issue for the data model is that all related information and knowledge should have a high level of formalisation since it should support the (semi-)automated generalisation process. The semantic data model contains two types of knowledge: firstly knowledge on the data content at the different scales including relationships and derivation and transition rules at the different scales; secondly knowledge on how objects should appear on a map also in relation to the surrounding objects. The first type of knowledge is modelled using UML class diagrams of which several alternatives were presented in this paper. The approach that defines superclasses for all object classes and subclasses to model information only related to specific scales looks most promising. Next step is to also address geometry, topology and generalisation related spatial information (e.g. proximity) in the UML class diagram. For the second type of knowledge (related to map presentations) the potentials to use a UML <<interface>> object will be studied. The considerations and experiments presented in this paper will guide the design of the model, which is expected to be finished end 2007.
At this moment it is not known to what extent UML is suitable for automatic knowledge extraction needed for the generalisation process. A UML approach might require implicit assumptions with respect to semantics that have to become explicit in the implementation (the software code executing the generalisation). Therefore the UML will be tested on the requirements described in section 4 and at the same time other alternatives will be investigated such as OWL to create an intermediate model (as interpreter between the UML models and the generalisation implementation). Such a semantic intermediate model has the advantage of making the semantics more explicit and more transferable between heterogeneous software systems (see Lemmens, 2006).

Acknowledgements

We would like to thank our colleagues who contributed to this research. They are: Theo Tijssen (TU Delft), Richard Knippers and Barend Köbben (ITC), Maarten Storm and Nico Bakker (TD Kadaster).

References

Bakker, 2005, Developing a new geographical object database. Experiences from idea to delivering datasets. In: Proceedings of the 22nd International Cartographic Conference (ICC 2005), A Coruña, Spain, 9 - 16 July 2005

Bard, S., 2004. Quality Assessment of Cartographic Generalisation. Transaction in GIS, Vol. 8, No. 1, pp. 63-81.

Barrault, M., Regnauld, N., Duchene, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W., Ruas, A. and R. Weibel, 2001. Integrating multi-agent, object-oriented, and algorithmic techniques for improved automated map generalization. Proceedings 20th International Cartographic Conference, Beijing, China, 6-10 August, pp. 2210-2216.

Beard, M. K., 1991. Constraints on Rule Formation. In Buttenfield, B.P., and R.B. McMaster (eds), Map Generalization: Making Rules for Knowledge Representation, Longman Group, pp. 121-135.

Burghardt, D. and M. Neun, 2006. Automated sequencing of generalisation services based on collaborative filtering. In: M. Raubal, H.J. Miller, A.U. Frank and M. Goodchild (eds): Geographic Information Science. 4th Int. Conf., GIScience 2006, IfGIprints 28, pp. 41-46.

Brassel, K. E., and R. Weibel, 1988. A Review and Conceptual Framework of Automated Map Generalization. International Journal of Geographical Information Systems, Vol. 2, No. 3, pp. 229-244.

Burghardt, D., S. Schmid and J. Stoter, 2007, Investigations on cartographic constraints formalisation, ICA workshop on generalisation and Multiple Representation 2-3 August 2007, Moscow

Friis-Christensen, A. and C. S. Jensen. 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, 183-192, 2003.

Goodwin, J., 2005a, Experiences of using OWL at the Ordnance Survey, presented at the workshop: OWL: experiences and directions, Galway, Ireland, November 11-12 2005.

Goodwin, J., 2005b, What Have Ontologies Ever Done For Us – PotentialApplications at a National Mapping Agency, presented at the workshop: OWL: experiences and directions, Galway, Ireland, November 11-12 2005.

Greenwood, J. and Hart, G., 2003. Sharing Object Based Geographic Information - A Data Model Perspective. Geocomputation.

Hart, G. and J. Greenwood., 2003. A component based approach to geo-ontologies and geodata modelling to enable data sharing. Proceedings of the 6th AGILE, April 24–26 2003, Lyon, France.

Hart, G., H. Mizen and S. Temple. 2004. Tales of the River Bank: An overview of the First Stages in the Development of a Topographic Ontology. 12th Annual Conference on GIS Research UK (GISRUK) pp 179-186.

Hoop, S. de, P. van Oosterom and M. Molenaar, 1993. Topological Querying of Multiple Map Layers. COSIT'93, European Conference on Spatial Information Theory, Italy, September 19-22, (LNCS nr. 716), pages 139-157, September 1993.

INSPIRE, 2007, DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007, establishing an Infrastructure for Spatial Information in the European Community, www.inspire.jrc.it/directive

ISO, 2005, ISO 19109:2005, IDT, Geographic Information – rules for application schema

Jones, C.B., D.B. Kidner, L.Q. Luo, G.L. Bundy, J.M. Ware. Database design for a multi-scale spatial information system. International Journal of Geographical Information Systems, 10(8): 901-920, 1996.

Lemmen, C.H.J., E.P. Oosterbroek, and P.J.M. van Oosterom, 1998, Spatial data management in the Netherlands Cadastre. In Proceedings of the FIG XXI International Congress, Commission 3, Land Information Systems, pages 398–409, Brighton, United Kingdom, July 1998.

Lemmen, C. and P. van Oosterom, 2006. Version 1.0 of the FIG Core Cadastral Domain Model, XXIII International FIG congress, Munich, October 2006, 18 p.

Lemmens, R., 2006, Semantic Interoperability of distributed geo-services, 2006, PhD Dissertation, ITC, Enschede.

Louwsma, J., S. Zlatanova, R. Lammeren and P. van Oosterom, 2006. Specifying and Implementing Constraints in GIS - with Examples from a Geo-Virtual Reality System. In: GeoInformatica, Volume 10, 4, pp. 531-550

Mackaness, W. A., Fisher, P. F. and G. G. Wilkinson, 1986. Towards a cartographic expert system. Proceedings Auto-Carto London, Vol. 1, 578-587.

Mackaness, W. A., 1995. A Constraint Based Approach to Human Computer Interaction in Automated Cartography. Proceedings of the 17th International Cartographic Conference, Barcelona, pp. 1423-1432.

Molenaar, 1989, Single valued vector maps: a concept in geographic information systems, in: International Journal of GIS, Vol.2, No.1, pp.18-27

NEN3610, 2005: Basic scheme for geo-information – Terms, definitions, relations and general rules for the interchange pf information of spatial objects related to the earth’s surface. In Dutch, Nederlands Normalisatie-instituut, Delft, The Netherlands.

Nickerson, B. G., 1988. Automated Cartographic Generalization for Linear Objects. Cartographica, Vol. 25, No. 3, pp.15-66.

Oosterom, P. van, 2005. Variable-scale Topological Data Structures Suitable for Progressive Data Transfer: The GAPface Tree and GAP-edge Forest. In: Cartography and Geographic Information Science, Vol. 32, No. 4, pp. 331-346.

Oosterom, P.J.M., van, M. de Vries and M. Meijers, 2006. Vario-scale data server in a web service context. Workshop ICA Commission on Map Generalisation and Multiple Representation, Vancouver USA, June 2006, 14 p.

Oosterom, P. van 2006. Constraints in Spatial Data Models, in a Dynamic Context. In: J. Drummond, R. Billen, E. Joao and D. Forrest (Eds.); Dynamic and Mobile GIS: Investigating Changes in Space and Time, pp. 104-137.

Parent, C., S. Spaccapietra, and E. Zimányi, 2006, Conceptual Modelling for Traditional and Spatio-Temporal Applications, The MADS Approach, ISBN: 3-540-30153-4

Quak, C.W., P.J.M. van Oosterom, M.E. de Vries, B. Bruns and N. Bakker, 2005, TOP10NL in GML3, Geo Info, Volume 2005, Number 11. In Dutch.

Regnauld N., 2001. Constraint based mechanism to achieve automatic generalization using agent model, Proceedings GIS Research UK GISRUK, University of Glamorgan, pp. 329-332.

Regnauld, N., 2007, Evolving from automating existing map production systems to producing maps on demand automatically, ICA workshop on generalisation and Multiple Representation 2-3 August 2007, Moscow

Ruas, A. and C. Plazanet, 1996. Strategies for Automated Generalization. Proceedings of the 7th Spatial Data Handling Symposium, Delft, the Netherlands, pp. 319-336.

Ruas, A., 1998, OO-constraint modelling to automate urban generalisation process, Proceedings of the 8th Spatial Data Handling Symposium, Vancouver, pp/ 225-236

Ruas, A., 1999. Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. Doctoral Thesis, Université de Marne-la-Vallée.

Schylberg, L., 1993. Computational Methods for Generalization of Cartographic Data in a Raster Environment, Doctoral Thesis, Department of Geodesy and Photogrammetry, Royal Institute of Technology, Stockholm, TRITA-FMI Report 1993:7.

Schwering, A. and G. Hart., 2004. A Case Study for Semantic Translation of the Water Framework Directive and a Topographic Database. Crete University Press. 7th Conference on Geographic Information Science (AGILE). Heraklion, Greece. Seite(n) 503-510.

VROM, 2007, Ministry of Housing, Spatial Planning and the Environment, http://www.e-overheid.nl/thema/basisvoorzieningen/basisregistraties/

Ware, J. M., C. B. Jones and N. Thomas, 2003. Automated Map Generalization with Multiple Operators: A Simulated Annealing Approach. International Journal of Geographical Information Science, Vol. 17, No. 8, pp. 743-769.

Weibel, R., and G. Dutton, 1998. Constraint-Based Automated Map Generalization. Proceedings of the 8th Spatial Data Handling Symposium, Vancouver, pp. 214-224.