Core Spatial Data

Core Spatial Data

NCG seminar on the occasion of the 25th year jubilee of
Mathias J.P.M. Lemmens at TU Delft, Delft, 18 December 2008

Peter J.M. van Oosterom (Editor)

NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 47

Delft, December 2009

Core Spatial Data
NCG seminar on the occasion of the 25th year jubilee of Mathias J.P.M. Lemmens at TU Delft, Delft,
18 December 2008
Peter J.M. van Oosterom (Editor)
Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 47, 2009
ISBN: 978 90 6132 317 4

Published by: NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission,
Delft, the Netherlands
Printed by: Optima Grafische Communicatie, Optima Graphic Communication, Rotterdam,
the Netherlands
Cover illustration: Ben Gorte, TU Delft

NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission
P.O. Box 5030, 2600 GA Delft, the Netherlands
T: +31 (0)15 278 28 19
F: +31 (0)15 278 17 75
E: info@ncg.knaw.nl
W: www.ncg.knaw.nl

The NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission is part of
the Royal Netherlands Academy of Arts and Sciences (KNAW)

Contents

Introduction Core Spatial Data

Prof.Dr.Ir. P.J.M. van Oosterom
TU Delft, the Netherlands

18 December 2008

This publication is a very special one for several reasons. First of all it is on the happy occasion of the 25th jubilee of Tjeu Lemmens at the TU Delft, which makes it a very pleasant setting. Second, this publication is the first publication of the new KNAW/NCG Subcommission 'Ruimtelijke Basisgegevens' (or in English 'Core Spatial Data'). Finally, the NCG research oriented approach to the topic of the Core Spatial Data has its counterpart in practice with the recent legislation on key registers of which several are containing spatial data. The topic of Core Spatial Data is very close to the interests of Tjeu Lemmens and it is therefore great that the authors in this publication were willing to contribute. Many of the authors are in one way or another (former) colleagues of Tjeu.

A short history

In 2004 an NCG task group, chaired by Martien Molenaar, started with the assignment to investigate the future needs and developments in the area of core spatial data. In March 2006 this group delivered the final version of the *Rapport Ruimtelijke Basisgegevens 2010* (Molenaar et al., 2006), which included among others the advise create new subcommission within the NCG on this topic. In June 2007 the Subcommission Core Spatial Data (chair Vosselman) was established together with another new Subcommission Geo-Information Infrastructure (chair Bregt). Both these subcommissions proceed from the Subcommission Geo-Information Models (GIM) which existed from November 1988 until June 2007. The creation of two subcommissions reflects the growing importance of this part of the research field within the NCG.

Previous NCG/GIM seminars

This first seminar of the Subcommission Core Spatial Data continues the good tradition of the Subcommission Geo-Information Models to organize seminars and publish the results. A small selection of some of the more recent NCG/GIM seminars (of which most resulted in an NCG publication; see Figure 1):

- Studiedag Sensor web enablement, Utrecht, 2007 (Grothe and Kooijmans, 2008).
- Studiedag Geo-information and computational geometry, Utrecht, 2005 (Van Oosterom and Van Kreveld, 2006).
- Seminar Standards in Action, Delft, 2004 (Van Oosterom, 2005).
- Studiedag GeoMetaMatica, Utrecht, 2004 (Heres, 2004).
- Themamiddag 3D Models and Applications, Delft, 2003.
- Studiedag Europese GIS-projecten with among other things INSPIRE, Utrecht, 2003.
- Geo-norm(ale) studiedag, Wageningen, 2002.

Figure 1. Some of the recent NCG GIM publications.

Relevance of a research agenda

Besides organizing seminars to exchange knowledge and discuss open problems, an important activity of an NCG subcommission is to specify a common research agenda for the field. The previous research agenda of the NCG Subcommission GIM formed the basis of scientific chapter of Bsik 'Ruimte voor Geo-Informatie' (RGI, Space for Geo-information) proposal. The RGI project lasted 4 years form 2004 – 2008 and did have an overall budget of 40 Meuro invested in geo-information research/innovation. The result was a boost in the Dutch geo-information developments. A sequel to Bsik RGI is to be expected in the context of the FES2009 theme 'Water, Klimaat & Ruimte' (WKR, Water, Climate and Space). It can only be hoped that the research agenda's of NCG Subcommissions GII+RB will be starting point for the Geo-Information aspects of the FES/WKR proposal and that they have the same impact as the GIM agenda did for the Bsik RGI project!

Contributions

The first contribution is by Jantien Stoter (TU Delft), who discusses the feasibility of a multi-scale information model and one key register 'Topography' instead of the current practice with separate information models for every scale. The growing importance of the third dimension is addressed by Stefan Flos (SJF Projects & Support) exploring the (im)possibilities to treat the more raw height data, as currently collected in the context of Actueel Hoogtebestand Nederland-2 (AHN-2, Actual Height model of the Netherlands) as core spatial data. Airborne laser scanning is a technique applied to collect these data. However, terrestrial laser scanning also has many interesting applications; see Figure 2. In the third contribution of this publication, the members of the NCG Subcommission Core Spatial Data (Vosselman, Schröder, Van Essen, Heres, Klijnjan, Kroon, Van Oosterom, and Van Rossem) present their first research agenda consisting of 10 different themes. Next Frank van den Heuvel (CycloMedia) will address an another important type of core spatial data: aerial and ground-based imagery. The special role of Cadastral information among core spatial data is discussed by Jaap Zevenbergen (TU Delft & ITC), Harry Uitermark and Chrit Lemmen (Kadaster & ITC). Finally, Robert Kroon (Geodelta) will make clear that all these data do not come for free and without pain in the last contribution "Laissez-faire in the air, a real nightmare".

Enjoy reading the different contributions in this publication!

References

Michel Grothe and Jan Kooijmans, Editors (2008), Sensor Web Enablement, NCG publication 45, Delft, 2008. 87 pages.

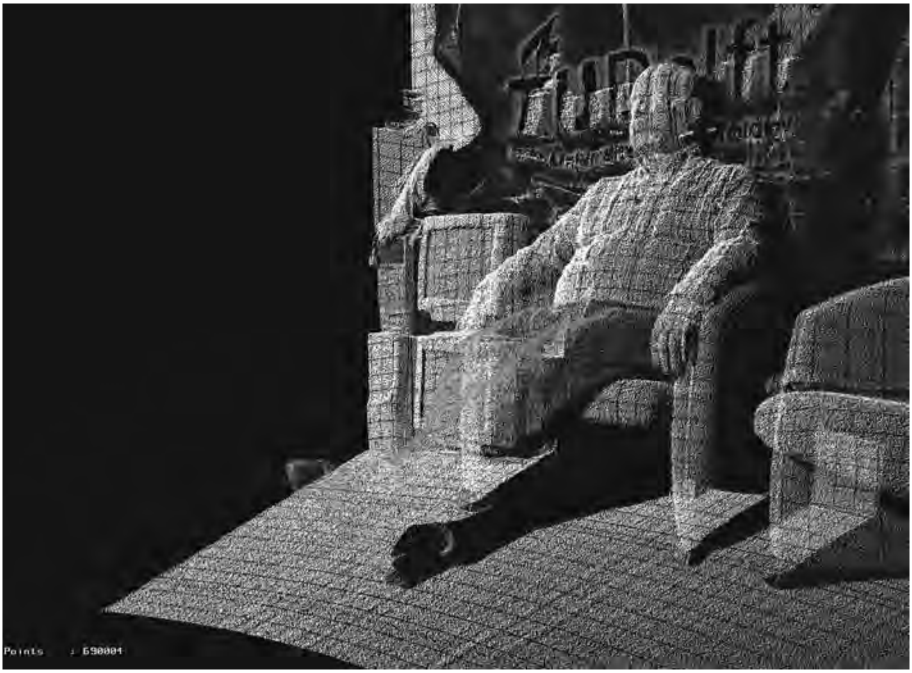
L. Heres, redactie (2004), GeoMetaMatica. Inleidingen van de studiedag over metagegevens voor geografische informatie. Subcommissie Geo-Informatie Modellen, Utrecht, 18 mei 2004. NCG publication 40, Delft, 2004. 62 pages.

M. Molenaar, M.G. Vosselman, E. Dolle, E. Kolk, M.-J. Kraak, R.J.G.A. Kroon, P.M. Laarakker, and P. J.M. van Oosterom (2006). Ruimtelijke basisgegevens 2010, Rapport van de Taakgroep Ruimtelijke Basisgegevens 2010 van de NCG, Version March 2006.

Peter J.M. van Oosterom, Editor (2005), Geo-information Standards in Action, NCG/GIN Farewell Seminar Henri J.G.L. Aalders, Delft, 17 November 2004, NCG publication 42, Delft, 2005. 98 pages.

Peter J.M. van Oosterom and Marc. J. van Kreveld, Editors (2006), Geo-information and computational geometry, NCG publication 44, Delft, 2006. 62 pages.

Figure 2. 3D laser scan point cloud of Tjeu Lemmens (thanks to Ben Gorte, TU Delft, Faculty of Aerospace Engineering for creating this point cloud).



Towards one domain model and one key register topography

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

Abstract

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

1. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

2. Background

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

2.1 NEN3610: Base Model Geo-information

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

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

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

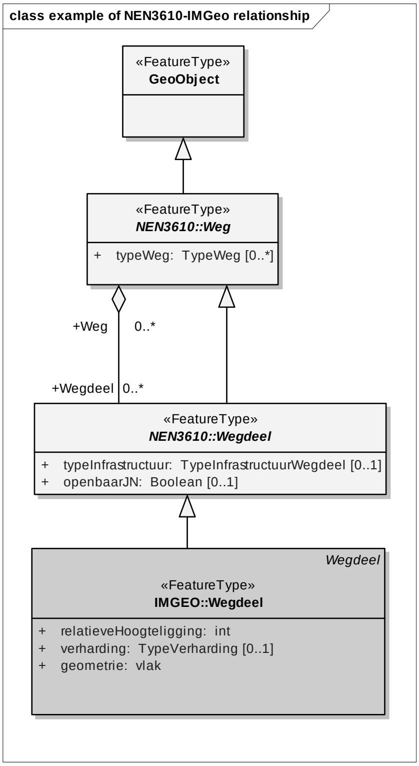


Figure 1. Relationship between abstract class 'PartOfRoad' (Wegdeel) in NEN3610 and the same class in the domain model IMGeo.

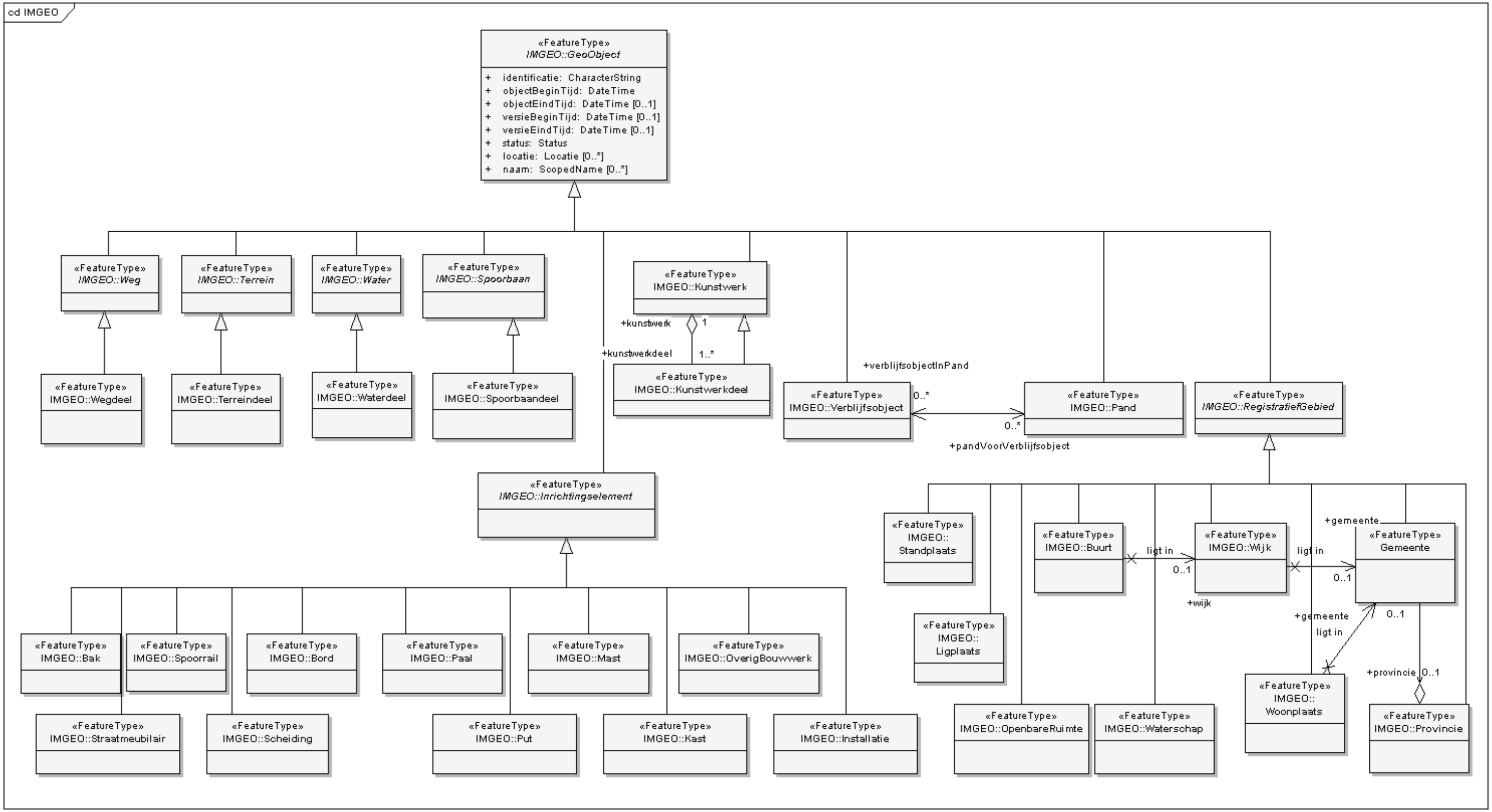


Figure 2. UML class diagram of IMGeo (IMGeo, 2007).

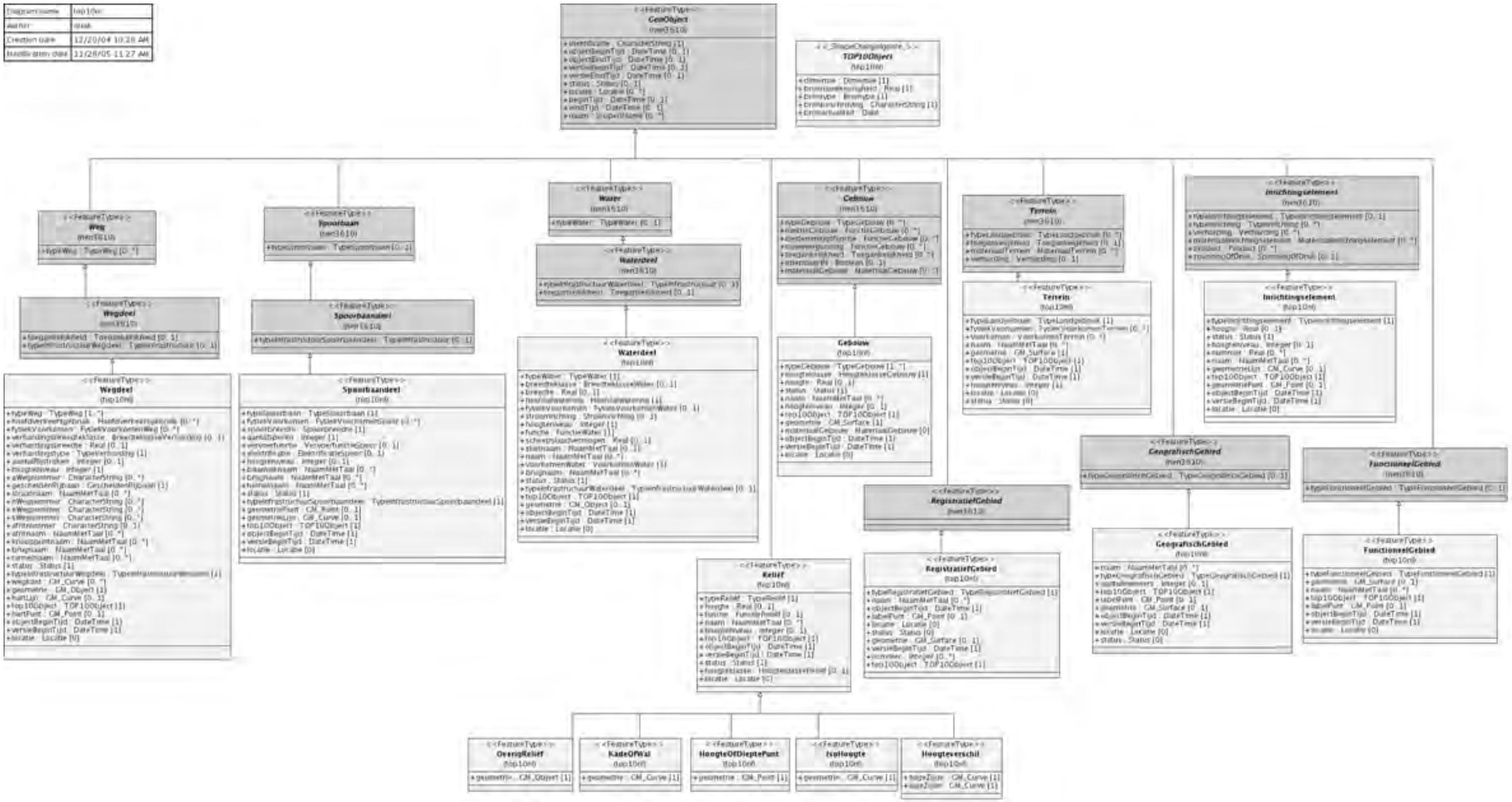


Figure 3. UML class diagram of TOP10NL (TOP10NL, 2005).

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

2.3 TOP10NL, IMTOP and key register topography (BRT)

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

3. IMGeo and TOP10NL: differences and simularities

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

3.1 General comparison

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

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

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

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

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

Geographical Area, Functional Area and Relief are only modelled in TOP10NL (Relief not available in NEN3610). Geographical Area is used to link toponyms in TOP10NL to geographical objects. Functional Area is used to group several objects into one object, for example a sport-area consisting of roads, building complexes and grass. Relief is used for topographical objects such as quays, peaks, isotopes and height differences.
IMGeo distinguishes Engineering Structure for infrastructural engineering structures such as bridges, viaducts, locks and dams, represented with polygon geometry (also available in NEN3610). In TOP10NL these classes are modelled as a specific type of infrastructural objects (PartOfWater, PartOfRailway or PartOfRoad) or as a Layout Element.
TOP10NL models much more attributes for its classes. The reason is firstly because these attributes are needed to visually distinguish different objects within one class. IMGeo is mainly an exchange model and therefore does not need these kinds of attributes. Secondly, IMGeo does not define more attributes than available in the underlying GBKN data.

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

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

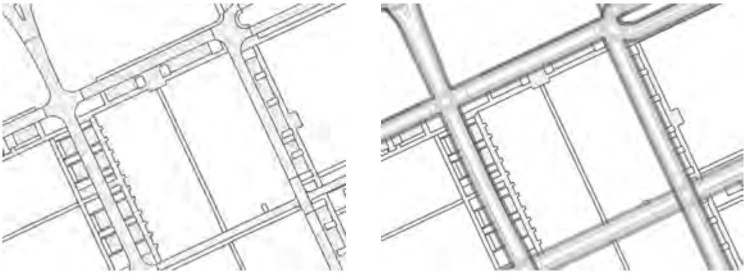
The object catalogue of IMGeo (IMGeo, 2007) indicates that all objects with polygon geometry and relativeHeight '0' divide the terrain into objects that do not overlap. IMGeo does not specify which classes do (or which classes do not) contribute to the partition as TOP10NL does. For some classes with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

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

3.2 Comparison in detail

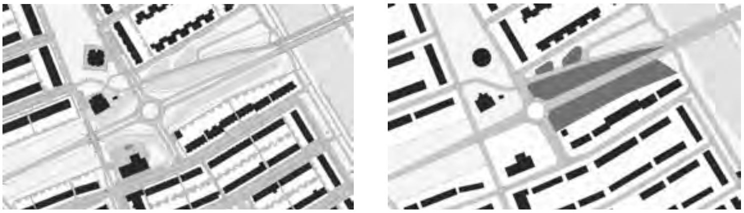
TOP10NL was established before IMGeo. However TOP10NL was not used as starting point for IMGeo. Consequently there are no relationships formulated between the two models to show which classes, attributes and attribute values model the same concepts. This section studies for a selection of concepts how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

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



IMGeo roads.

TOP10NL roads transparently projected on IMGeo roads.



IMGeo data.

TOP10NL data.

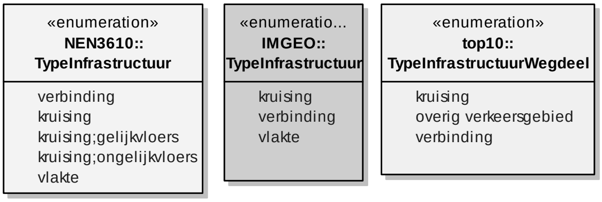
Figure 4. Visualisation of IMGeo data (courtesy of Municipality of Almere) and TOP10NL data.

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

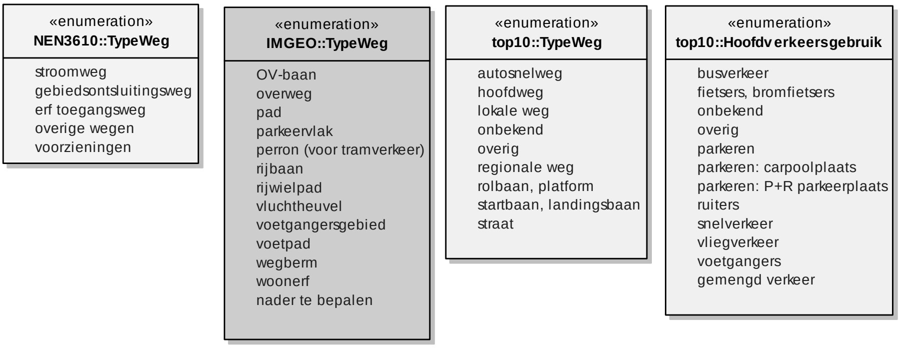
3.2.1 Road

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

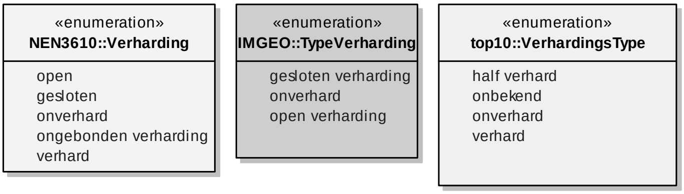
Another significant difference is that verge is considered PartofRoad in IMGeo but in TOP10NL verge is identified as Terrain, often with land use 'gras'. It would be possible to define this difference in a derivation rule, i.e. IMGeo PartOfRoad-verge is converted into TOP10NL Terrain-'gras'. However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



a. Attribute values for typeInfrastructuur (typeOfInfrastructure).



b. Attribute values for attribute typeWeg (typeOfRoad).



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

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

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

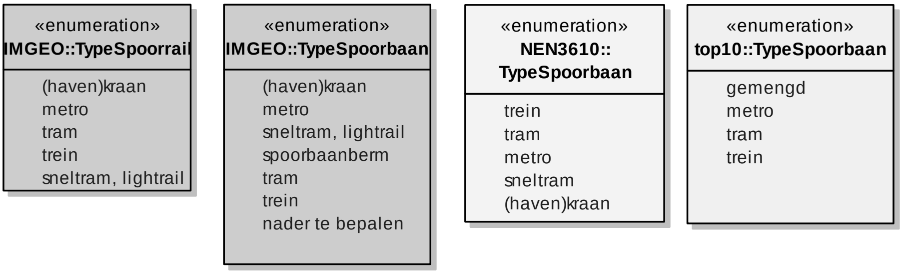
Another difference in road definition is the attribute typeOfRoad (*typeWeg*) used in a different way in both models, see Figure 5b. IMGeo uses the attribute to indicate different objects contributing to a road for example parking area, public transport-lane, footpath, verge, roadway, cycle path, pedestrian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

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

3.2.2 Railway

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

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



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

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

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

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

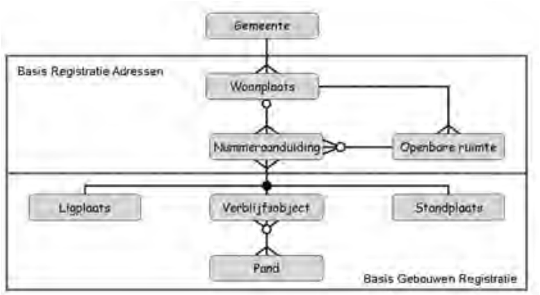
As for the road concept, we can conclude that the railway concept is differently defined in IMGeo and TOP10NL. In addition the defined types of railway differ (even between the two classes Rail and Railway both defined in IMGeo). As in the case of pavement type and type of infrastructure for roads, it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

3.2.3 Building

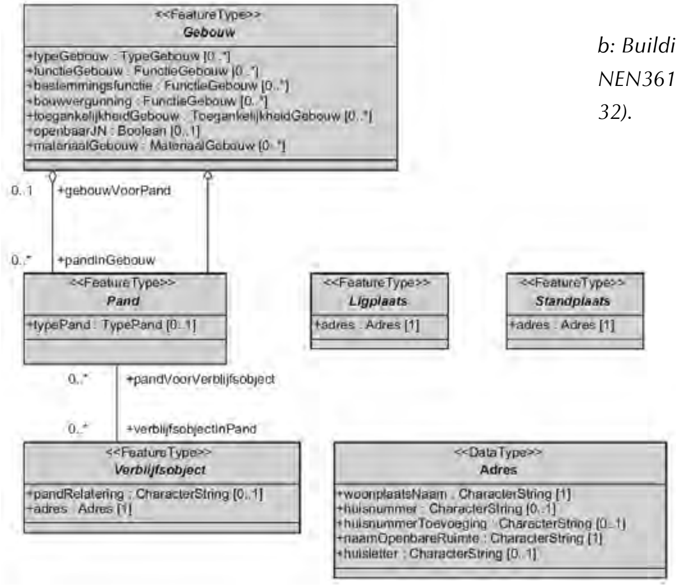
There are three building related classes defined in the models: Building Complex (*Gebouw*), Building (*Pand*) and Living Unit (*Verblijfsobject*). The last two are prescribed by BAG. BAG does not contain a class for Building Complex (*Gebouw*; see Figure 7a), since the designers of BAG could not find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

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

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



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



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

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

As can be seen in Figure 4, IMGeo buildings have a higher precision than TOP10NL buildings. Since TOP10NL will use GBKN buildings in the future, the differences will largely disappear (Kadaster, 2005).

We can conclude from above that the classes Building and Building Complex have a different meaning in NEN3610, IMGeo/BAG and TOP10NL.

3.2.4 Terrain

For the class Terrain (IMGeo also models PartOfTerrain) the attributes typePartOfTerrain (IMGeo) and typeOfLandUse (TOP10NL, and also NEN3610) model the same concept. Table 3 compares the possible terrain types in both models; also all NEN3610 types are shown. An important observation is that none of the types mentioned in IMGeo has exactly the same name as a TOP10NL type. A few types are presumably the same (gras and gras-land; 'nature and landscape' and heather).

Another observation is that IMGeo models one type of forest where TOP10NL models four types of forest. Also for IMGeo green object we can observe four possible values in TOP10NL. Apparently higher level of detail is required here for the smaller scale dataset.

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

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

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

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

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

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

3.2.5 Layout Element

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

4. Towards one domain model and one key register topography

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

4.1 Towards one domain model topography

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

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

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

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

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

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

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

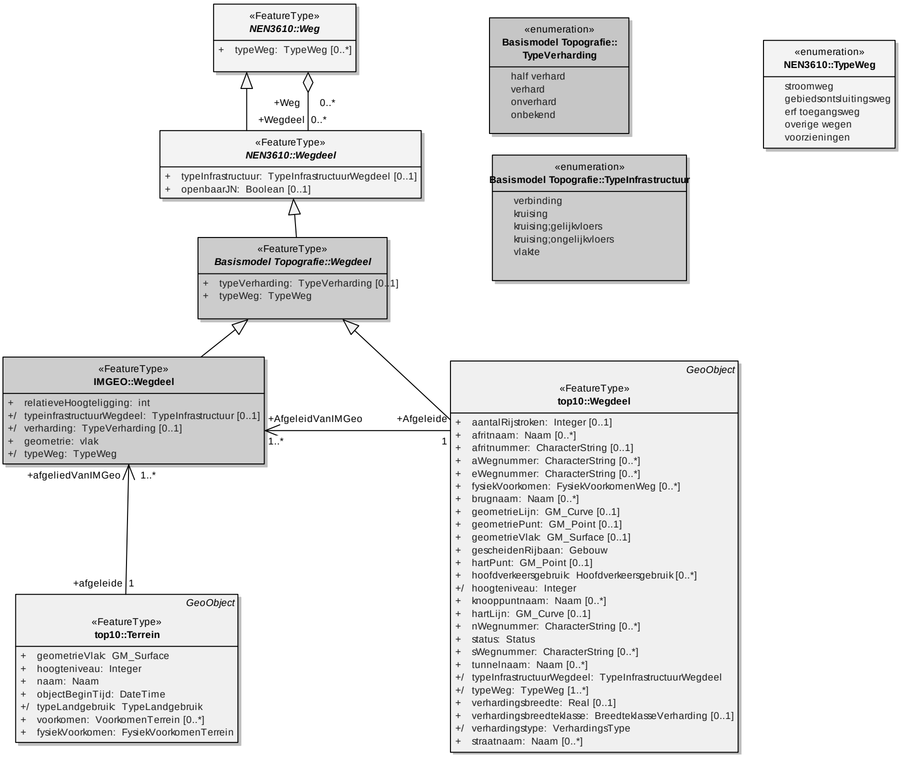


Figure 8. Concept PartOfRoad (Wegdeel), modelled in Base Model Topography.

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

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

4.2 Towards one key register topography

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

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

5. Conclusion

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

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

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

Acknowledgements

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

References

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

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

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

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

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

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

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

LSV GBKN (2007). GBKN Handboek, versie 2.1, Stichting LSV GBKN, Documentnr. 07.05/052 2007. http://www.gbkn.nl/nieuwesite/downloads/07.05.065%20GBKN%20handboek%20VIPU2.1.pdf.

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

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

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

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

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

AHN in perspective:
Dutch Digital Elevation Data (AHN) Core Spatial Data?

Ing. S.J. Flos MSc MMI *
SJF projects & support, the Netherlands

Summary

The completion of the first countrywide digital elevation dataset of the Netherlands (AHN-1) in 2003 marks an important turning point in the collection of elevation data. Quality, density, temporal consistency and covered area, all are dramatically improved using remote sensing laser scanning techniques. With more and more users relying on the availability of digital elevation data and with a second AHN in the making the significance of this development is important. Therefore the potential of AHN as a core spatial dataset needs to be explored: what makes the AHN Core Spatial Data? To be able to review this potential, it is important to view this digital elevation dataset in perspective of technical (supply), usage (demand) and organisational (institutional) dimensions and offset this to core spatial data requirements. First of all we should ask ourselves what Core Spatial Data is and what its requirements are? Secondly we should realize that a country wide dataset can not exist without an organisation responsible for aligning supply and demand and safeguarding its sustained existence.

1. Introduction

Various government agencies in the Netherlands hold a staggering amount of data. This data is mostly administrative data, but a specific set is geographic or spatial data. Estimates range in the order of 15.000 – 30.000 individual spatial datasets [BDO, 1998]. Several trends force the various government agencies to cooperate and consolidate the most important datasets: consolidation of lower government organizations such as waterboards and municipalities, the legal formation of authentic (spatial) records [www.VROM.nl] and the European requirements for (environmental) data exchange (p.e. INSPIRE: Infrastructure for Spatial Information in the European Community [http://inspire.jrc.ec.europa.eu]). These trends seem to drive a process of consolidation of individual spatial datasets into larger, preferably national datasets.

In between the national authentic spatial datasets anchored in law and maintained through formal government organisations on one side and the various unorganized datasets on the other side of the spectrum, consensus is growing around a special type of geo-information: important frequently used spatial datasets at the core of government and professional use: Core Spatial Data (CSD).

---

* The contribution is made on a personal basis, based on 10 years involvement with the collection, use and organization of digital elevation data in the Netherlands. S. J. Flos works a private consultant in the field of water (safety) management and ICT and was the secretary of the Dutch AHN steering committee from 2003 - 2008.

The Subcommission Core Spatial Data (Ruimtelijke Basisgegevens) of the Netherlands Geodetic Commission is focussing on these important spatial datasets, perceived as playing a core role in developing a spatial-data infrastructure. [http://www.ncg.knaw.nl/ subcieruimtelijkebasisgegevens.html]. However, a clear definition and matching criteria to distinguish CSD from 'normal' Spatial Datasets is not available.

One example put forward by the NCG as important spatial data is (digital) elevation data. Recent technological developments have lead to large scale collection of high density elevation data in the Netherlands using Lidar scanning techniques. In 2003 a first country-wide digital elevation dataset in the Netherlands was created: the Actueel Hoogtebestand Nederland (AHN-1). Beside this dataset, various government agencies hold and actively collect digital elevation data in addition to the AHN dataset.

As part of the 2008 annual NCG conference organized by the CSD Subcommission, a presentation was prepared exploring the possibilities and impossibilities of the AHN as a core spatial dataset. This paper is based on that presentation presented at the annual 2008 NCG study day in December 2008 and various discussions within and outside the NCG. Clearly CSD require much more than only a focus on technical supply potential and user possibilities. Beautiful pictures of elevation data, striking examples of its potential for use, per example in the field of archaeology or stunning 3D animations are not enough to establish and maintain a CSD set. An organization has to align supply and demand effectively.

The aim of this contribution in the annual NCG publication is to provide an argumentation as input in discussions within (and outside) the NCG committee related to CSD, its requirements in general and digital elevation data in particular. To do so, first the requirements of a Core Spatial Dataset will be explored. Secondly the AHN dataset will be reviewed in terms of technical, organisational and user dimensions and subsequently the AHN will be put in a CSD perspective. Finally some conclusions and recommendations are presented.

This is not a technical review or a scientific paper. The background of this paper is based on more than ten years of working with Lidar data and direct involvement with the AHN organisation, first as client (representing a waterboard) and later representing the AHN organization holding its secretariat for five years. This involvement has resulted in the observation that technical developments on the supply side of Lidar scanning data out run user potential. A country wide elevation data set has even more potential, because knowledge for using the data can be applied throughout the country. An effective organization representing core (government) users is needed to align supply and demand and this seems to be the bottleneck. This contribution is based on this observation.

Preamble

The annual 2008 NCG meting on Core Spatial Data was a special occasion celebrating Dr.Ir. Tjeu Lemmens 25th year connection with the TU Delft, spanning a period of rapid technological development in the field spatial data collection. Dr. Lemmens was involved at the onset of laser scanning in general and the development of the AHN dataset in particular.

To illustrate the developments in the geo-information field a comparison is made with the rapid developments in the music business in the late 50's of last century. The end of the 1950's saw the conversion from mono to stereo long-play recordings and the onset of popular music.

Prior to the 50's, musical events were mostly a live, one on one experience. With the advent of recorded music a technological revolution was taking place in the background: stereo single and long play recordings transformed the music experience in a 'record once and listen many times' experience. This enabled a popular, demand driven music industry. Mario Lanza was one of the last icons of the popular classic music genre of the past at the RCA label. He was followed up (in 1959) by a new young singer at RCA: Elvis Presley, marking the onset of a new era: pop music.

A newly released SACD recording of this singer, made exactly 50 years ago in December 1959 was used to illustrate the importance of innovative quality recordings. The unique content of these first stereo recordings (made using three microphones for enhanced stereo quality) can now be heard in its original form. The master tapes are invaluable and irreplaceable documents in time.

The rapid conversion within spatial information technology from paper tot digital format and from point sources to remote scanning techniques can be compared with the transition of the music industry in the late 50's and the start of popular demand driven (geo-information) industry. Thus focus will shift from recording technique to popular user demand. It will require a new type of organisation capable of aligning supply and demand effectively without compromising quality. More importantly: information needs to be delivered (consumed) instantly almost real time. Time to market is crucial. Building a historical dataset is just spin-off.

2. Core Spatial Data: requirements

One of the basic questions to be answered first in order to be able to review the AHN as CSD is the question what Core Spatial Data is? How does CSD stand out from other spatial data sets and what are it's distinct characteristics and criteria.

The NCG paper 'Core Spatial Data 2010' ('Ruimtelijke Basisgegevens 2010') [NCG 2006] provides an overview of current rapid technical developments of spatial data collection and use over the past 10 years. This NCG publication on CSD however, does not provide a clear definition what CSD is and how to recognize it, but it contains many implicit clues about CSD requirements.

At a general (conceptual) level it is obvious that certain spatial data types belong to the core of spatial data. Height data, or z-data, is inextricably a part of a position, making x,y,z measurements the core of all geo-information. That z-data is not always used and most representations are 2D is not important in this respect. Elevation data is core data in general.

At a specific level CSD will be have to be related to a specific spatial data set or product. CSD seem to be positioned in between 'normal' unregulated spatial data sets and authentic (established) core spatial data sets. Authentic spatial data is regulated by law and maintained by a structural organisation to provide specifically defined end products. The most prominent authentic spatial datasets are cadastral maps as maintained by the cadastre and national scale topographic maps such as the top10 (1:10,000) topographic map and the 1:1,000 GBKN topographic map.

CSD could be seen important basic geo-information comparable with authentic spatial data but without the legal requirement. Thus its characteristics are comparable, being data that is country wide and sustainably collected regardless of administrative boundaries, has a uniform (known) described quality, is reliable and complete. The data is linked to metadata providing trace back information to the data source (pe; time and method of collection, quality and precision, ownership etc). The data is collected by or for a cross range of government organisations with a focus on professional government use. There is well a defined historical use and supply base.

For the established authentic spatial products these requirements are obvious, because the established product is the result of a requirement by law to construct, manage and maintain the dataset. The result is a unique dataset for official government use, authentic and thus core spatial dataset, regardless of other (or small scale) datasets. Interestingly authentic datasets start with putting an organization in place, for without a formalized responsibility for maintaining the CSD product nothing would materialize. Requirements state that control over the dataset is laid down with a specific organization and responsibility for realization and maintenance of the dataset resides under a cabinet minister.

CSD therefore are comparable with authentic records and share the same requirements. As defined by the ministry of housing, spatial planning and the environment (VROM) there are 12 requirements for an authentic record (www.vrom.nl). Per example the dataset should be compulsory use by government organisations. Agreements and procedures should be in place between holder and users and accessibility should be well organized. Procedures are put in place for interaction and end-user involvement in decision making regarding the dataset. The quality of the (end) product is well maintained and the position of the dataset within the framework of datasets is clearly described.

One other requirement of authentic data is also important to note. It states that a dataset should be well defined in terms of content and reach of the dataset [www.vrom.nl]. In other words, it should be clear what the dataset is and what purpose it is made for. The dataset should have a specific purpose, related to specific government tasks, procedures and responsibilities. There is no room for competition with related datasets regarding the authenticity or unique qualities of the dataset. Therefore it can be concluded that CSD as with authentic spatial data should be purpose driven: its use and function, mainly within government use will drive its existence and its position within the general (geo-information) framework should be clear.

Linking core spatial data with core spatial data use from a government perspective is vital for developing sustainable geo-information datasets. Not the technical possibilities (how it's made) are likely to be a driver but the core data requirement itself: it's purpose to the professional user. The users are simply not interested or specialized to acquire the data themselves since their primary focus is with using the data for their core processes. The data is 'must have' data for a core range of professional users. In order to provide value to the professional (government) end users per example, standard interpreted datasets should be defined which are not linked to a certain technology or scanner brand. This requires standard end products that can be maintained regardless of developments in technology and require little to no extra input to prepare the data for their core processes.

It can be concluded that the definition and requirements of a core spatial dataset are comparable with the definition and requirements of an authentic dataset, minus the legal basis. In other words: a CSD set is organized as if it were an authentic dataset. This makes the establishment of a CSD set a function of organization, more than that of technical supply or user demand.

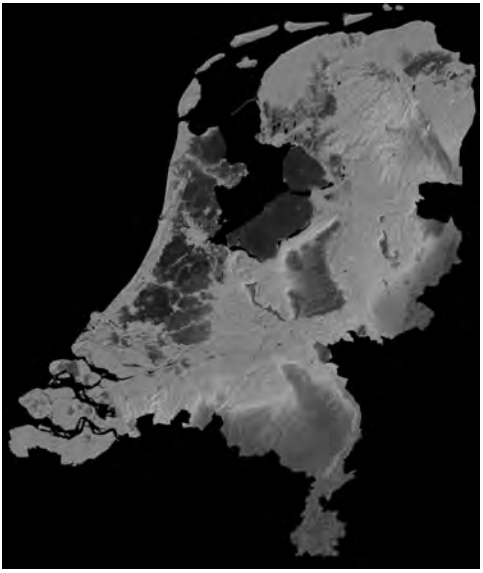
3. AHN: data collection, use and organisation

Elevation data play an important role in many aspects of managing land, water and infrastructure. Especially for a flat, low lying country such as the Netherlands, elevation data is essential data related to water and dyke management. Traditionally the waterboards, the water management organisations in the Netherlands are wholesale users of elevation data in both flat polder area's and related to dyke management. The data is used to establish the average depth to water table in a polder or calculate the minimum height and safety level and structural form of man-made dykes structures.

Over the centuries a variety of techniques and standards have been used to measure elevation. These techniques are deployed in the field and require trained personnel to hand-pick the data in situ. The process is time consuming and often require elaborate preparations in order to get permission to access private land in the project area.

With the onset of developing airborne laser scanning techniques in the 1990's the remote sensing aspect of this technique was seen as the driver for commissioning the first projects. Using helicopters and airplanes had the advantage that no prior access permission was needed and that more data could be collected in a shorter period. In 1995 the first projects were commissioned by waterboards with involvement of the ministry of public works (RWS).

In 1997, now more than ten years ago, the first steps were undertaken to construct a national digital elevation map of the Netherlands, named Actueel Hoogtebestand Nederland (updated digital elevation map of the Netherlands). The idea behind AHN-1 was to collect the elevation data in a concerted method involving the individual government

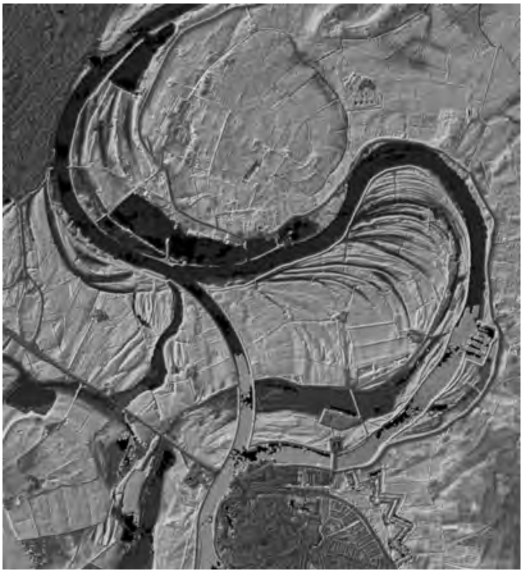


organisations and thus save costs. This project turned out to be the first of its kind in the world using laser scanning techniques to complete a country wide dataset.

The project was completed in 2003 and although not all expectations were met regarding the quality of the data, the completed AHN marks an important step in the development from single in-situ point measurements towards remote sensing scanning of elevation data. The resulting AHN-1 dramatically improved quality, density, temporal consistency and covered area. In addition, all data was available at a central point.
Because the AHN data is available to users outside the organizations of the owners, a wide array of users has been able to experience the value and unique qualities of large scale digital elevation data. The potential for data use is very promising and the AHN data is in high demand. At this moment RWS and the water boards are in a process to update the existing data: AHN version 2 with higher point density, higher point accuracy and higher consistency of the complete set.

Much of the success of the AHN is related to the scale of the dataset itself and to one very important new aspect: visualisation of relative height differences, providing an accurate insight in the micro relief of the natural landscape. Relative height differences prove to be more resilient to change as compared to the exact height data. An area might slightly subside in exact terms, its relative height distribution remains more or less the same. This enables users to view AHN visualisations as a window in time: the historic traces in the landscape are remarkably apparent.

The exponent of the success of relative height differences is demonstrated by the use in archaeological desktop studies. Current practices require that prior to archaeological field research the AHN visualisation of micro relief should be studied first. The visualisation of relief can be further enhanced by using shadowing technology. Differences in height



Shadow relief produced from AHN-1 data of the river IJssel near Doesburg.

in the range of centimetres can be visualized because of the geographical context: small differences stand out in contrast with its surroundings.

Parallel to developments of AHN in 1997, water boards also started using laser scanning technology to collect elevation data of dikes and embankments form 1998 onwards. In this case the dikes were scanned in corridors or lines covering only a limited elongated area, following the path of the dike structure and its immediate surroundings. These projects were executed by the individual water boards and point density and accuracy were higher. Typically point densities of 10 to 20 points/m^2 were obtained and an accuracy of about 5 cm per point were accepted. Moreover, the underlying datasets are more consistent in temporal quality as every section is recorded in one and the same flight.
In the current AHN-2 projects both demands, for water management and dike management for elevation data are combined in a standard of approximately 10 points per m^2 and 5 cm accuracy. More importantly, the standard is set from a traditional water and dike perspective. The new standard will provide more potential for a wide use of the data, outside the traditional water corner.

With the establishment of AHN-1 the data is also available to other users. One very important new user group, outside the traditional water corner of users, are the archaeologists. By looking at the subtle differences in relative elevation that landscape of hundreds to thousands of years ago can be indicated. Therefore it is possible to efficiently indicate possible archaeological sites of interest, since these are located on higher grounds.

The AHN organisation itself is merely a project organisation. At first aimed at establishing AHN-1 and continued to establish AHN-2. The organisation is typically confined to the water corner and there is no legal or institutional basis or framework for the organisation. The organisation is best described as a project organisation for jointly purchasing (whole sale) elevation data for waterboards and the ministry of public works. Efforts were made to organise user participation outside the users of the participating organisations, but these efforts were abandoned. Currently there is no formal organisation and structural involvement of users. AHN-2 data is currently not available for users outside the participating organisations.

It can be concluded that the AHN-1 has been and still is a big success. This is mostly the result of the country wide scale of the dataset and its central availability at the ministry of public works. With a new AHN in the making the potential is huge. The AHN organisation seems to be the weakest link. It's confined to the water corner of public sector organisations and functions merely as a project organisation for joint purchase of elevation data.

Conclusion and Recommendations

This short review is presented to question the possibilities of digital elevation data (AHN) as a Core Spatial Dataset. It is obvious that elevation data form a general, conceptual point of view is and will be core spatial data. This is a straight forward as x, y and z.

For a spatial dataset to be a Core Spatial Dataset more is needed than a conceptual basis. In this contribution it is presumed that a CSD set shares the requirements of an authentic dataset, minus the legal requirement. It seems therefore, that a CSD set is merely the result of institutional aspects: of establishing an organization capable of aligning supply with core data demand from a neutral position, representing all core users. For the AHN as a product to move in the direction of a CSD set it will have to come out of the 'water corner' and move to a neutral position.

The NCG Subcommission on CSD can assist it in this transition. It is recommended that the NCG Subcommission focus more on the institutional aspects and specific organisational requirements of CSD. In other words, not only what and why CSD but also who. The establishment of a country wide, sustainably maintained dataset requires a specific organisation, capable of sustained alignment of supply and demand, realistically, effectively and (cost) efficiently. Not an easy task.

The NCG Subcommission could per example engage with the current AHN organisation, get involved in the establishment of a user group and make AHN as a CSD set a target in it's working plan. Moreover the NCG is in a position to demonstrate the strategic importance of country wide high quality and high density for government use within and outside our country and help pull AHN out of the water corner.

Like the historic quality recordings of classical music, made in the 40's and 50's the historic LIDAR data collected in the past 10 years will remain of value for a long time to come. But for the current multi-user demand, new, reliable and quality datasets should become available and time to market should be short. Otherwise the dataset will be out of sinc with demand, driving the formation of small-scale datasets, of which there are so many.

References

NCG 2006: Ruimtelijke basisgegevens 2010, Rapport van de Taakgroep Ruimtelijke Basisgegevens 2010 van de NCG, Delft versie maart 2006.

BDO consultants, Elektronische bestanden van het bestuur, 1998.

AHN 2008: Bestek aanbesteding AHN-2, RWS Delft 2008.

Flos 2003: FLOS, SJ. P. Franken, F. Hoefsloot, R. Verhage, AHN is niet zomaar geschikt voor peilbesluiten, VI Matrix Eindhoven maart 2003.

VIW 2008-06: Laseraltimetrie voor waterkeringbeheer: ontwikkelingen, gewenste specificaties, procesbeschrijving en evaluatie AHN-2 proef, STOWA Utrecht 2008.

www.AHN.nl: Website van de stuurgroep AHN: www.ahn.nl.

www.VROM.nl.

The research plan of the NCG Subcommission Spatial Core Data

Prof.Dr.Ir. M.G. Vosselman, F.H. Schröder, Drs. R. van Essen, Ir. L. Heres, Ir.Drs. A.J. Klijnjan, Ir. R.J.G.A. Kroon, Dr. J.E. Stoter, Prof.Dr.Ir. P.J.M. van Oosterom, Ir. R.P.E. van Rossem
NGC Subcommission Spatial Core Data, the Netherlands

Abstract

In 2007 the Netherlands Geodetic Commission (NCG) installed the Subcommission Spatial Core Data to discuss, co-ordinate and initiate research in the field of acquisition, representation and usage of the spatial core data. This document describes the areas in which the Subcommission wants to be active and identifies the open research questions.

Introduction

The Subcommission Spatial Core Data of the Netherlands Geodetic Commission was installed in 2007 to discuss, co-ordinate and initiate research in the field of acquisition, representation and usage of the spatial core data.

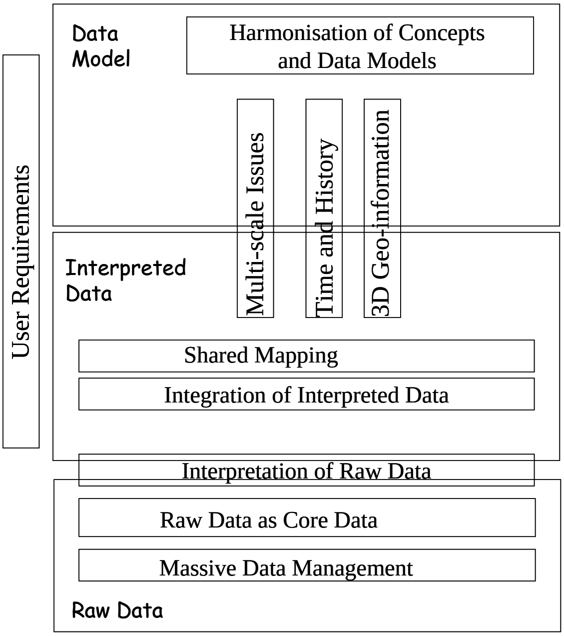
Currently, eight scientists and experts from universities, government agencies and companies cooperate in this Subcommission. The members are: Drs. R. van Essen (Tele Atlas), Ir. L. Heres (RWS-DID), Drs.Ir. A.J. Klijnjan (Dutch Land Registry Office (Kadaster)), Ir. R.G.A. Kroon (Ingenieursbureau Geodelta B.V.), Prof.Dr.Ir. P.J.M. van Oosterom (TU Delft), Ir. R.P.E. van Rossem (Ministry of Housing, Spatial Planning and the Environment), Dr. J.E. Stoter (TU Delft), and Prof.Dr.Ir. M.G. Vosselman (chairman, ITC).

The field of research of the Subcommission is sketched in this document and elaborated in ten themes. The themes can roughly be divided into two categories: research on interpreted core data (most often vector data) and research on raw or uninterpreted data as acquired by various kinds of sensors.

Theme 1. User Requirements

Research on raw core data
Theme 2. Raw Data as Core Data
Theme 3. Massive Data Management
Theme 4. Interpretation of Raw Data

Research on interpreted core data
Theme 5. Harmonisation of Concepts and Data Models
Theme 6. Integration of Interpreted Data
Theme 7. Multi-scale Issues
Theme 8. Time and History
Theme 9. 3D Geo-information
Theme 10. Shared Mapping



Relationships between the research themes on spatial core data.

These themes are mutually related. Their relationships are illustrated in the diagram above, around the notions of *Data Model, Interpreted Data and Raw Data*.

Theme 1. User Requirements

User Categories

A distinction can be made between professionals, light professionals and consumers. Consumers traditionally use geographic data in the form of paper maps. It is only since the rise of the Internet and navigation systems, that consumers have started using digital maps. They are typically end users and normally do not process, edit, or adapt the data.

Professionals use geographic data in their working environment. They also have a long tradition of using paper maps, but already started using digital maps in the late sixties. They have different roles regarding these digital maps. Many of them, here called the light professionals, are end users and have more or less the same requirements as consumers. Some of the, here called the heavy professionals, are producers of data and use source data as a semi-manufactured article, integrate it with other source data in order to produce a new dataset. These different user groups will have different requirements.

Topic Groups

Other important dimensions are national versus international and private versus public. This leads together to the following topic groups:

- European and national developments;
- Professional market;
- Consumer market;
- Private versus public.

European and national developments

In a European context, INSPIRE is currently the most important driver. It serves as a basis for harmonisation of the content of basic data sets. International standards provided by ISO and the Open Geospatial Consortium (OGC) are used for the dissemination and transfer.

At the national level, the programme 'Stroomlijning Basisgegevens' (Streamlining Core Data) is the most important one. It aims to realise six so-called Core Registrations: Persons, Enterprises, Buildings, Addresses, Topography and Cadastre.

Professional market

For the professional market up-to-date-ness, quality and accessibility are important issues. The possibility to integrate attribute data with topographic data will become more and more a conditio sine qua non.

Keywords are furthermore: standardisation, object orientation, leaving data at the source, data integration, 3-D and simulation (serious gaming).

Another important topic, in particular for private enterprises, is the issue of copyrights.

Consumer market

Consumers will use more and more geographic information in digital form. This is driven by technical developments such as car navigation systems, location based services, video games, internet applications such as Google Maps. Consumer requirements will therefore play an increasing role. The game industry (flight and drive simulators) evokes a demand for realistic and detailed landscapes models. For applications as virtual town walks and city tours these models have to be completed with terrestrial images.

Private versus public

In a situation where public authorities are active in the same field as private enterprises, there is a risk that they may disturb the competition relations. Therefore the Ministry of Interior is developing a new policy regarding to this subject. Public authorities will get more and more a co-ordinating and stimulating role. To this co-ordinating role belongs the provision of reference data. Enrichment of these reference data and the development of applications will be the role of private companies.

Research questions

Within the scene sketched above the NCG Subcommission will focus on the following research questions concerning the user requirements to core data.

- What are the emerging application fields?
- What are the user requirements related to these fields?
- How to translate user requirements into product specifications?

What are the emerging application fields?

In order to focus the research efforts on the right themes, it is important to have an overall insight in how the new technologies will be used in the well-known application areas as well as in new application areas. A survey of these potential application areas is therefore a useful meta-activity

What are the user requirements related to these fields?
Once the application fields have been identified, the user requirement should be investigated. Examining these user requirements is therefore a sensible investment.

How to translate user requirements into product specifications?
The third and last step in this research triad is the investigation of the product specification as function of the user requirements identified in the second step.

Theme 2. Using raw data as core data

Several parties collect geo-data with similar or at least related content. The reason for doing so was (and still is) that domain specific geo-information was (and still is) connected to domain specific user requirements for domain organized public and private organisations. Examples are the acquisition of large scale stereo aerial images for the update of the topographic contents of large scale topographic databases, the acquisition of small scale aerial stereo images for the update of small scale topographical contents and the acquisition of 'in between' scale aerial images for specific purposes like agricultural monitoring.

The question arises if the INSPIRE key issue with respect to efficiency and consistency ('collect once, use many times') can also be an advantage during the acquisition of geo-information.

Technological developments in the last years have resulted in a rapid change in the way data can be collected. Airborne digital photogrammetric cameras, SAR-equipment and airborne laser scanners produce very detailed and hence very large raster datasets.
Traditionally these datasets serve as intermediate products in the production of specific topographic information. But today more and more new uses are found for these intermediate products. An example of this is the widespread use of satellite and aerial imagery in Google Earth and Microsoft's Live Earth. The use of imagery in these applications shows that the intermediate products have become products unto themselves.
These datasets can be regarded as 'raw core data', or 'uninterpreted data'. Raw core data is data that has been acquired to serve multiple purposes. It is covering a large area and preferably has a nationwide availability. Depending on specific user requirements raw core data will be further processed into tailor made products.

This leads to the following topics for research:

- Which geo-data can be considered as raw core data?
- Which metadata form part of a raw core dataset?
- Which technical specifications should a raw core dataset comply with?
- Which organization model is a preferred one for the periodic acquisition of raw core data, the quality control and the distribution of the data?

Which geo-data can be considered as raw core data?

Data can be categorized as raw core data if the data is of the agreed high quality, if it is up-to-date and if it can serve as a skeleton for geo-information applications for a large variety of users. Examples of possible raw core data are a nationwide geodetic reference frame, a nationwide laser altimetry height dataset and a nationwide set of high resolution digital aerial images. It may be important to not only look at the demands of the heavy professionals but also at the demands of light professionals and even consumers. As an example a high resolution geo-positioned set of stereo images of our cities might not only be a valuable source for the mapping industry but also for developers of scene realistic computer games. More investigation is needed in order to make a proper decision which datasets can be regarded as raw core data and which not.

Which metadata form part of a raw core dataset?

Raw core data is more than acquiring spatial sensor data. Also metadata form an important part of the raw core dataset. A nationwide coverage of orthoimages might be a possible raw core dataset. However the quality of orthoimages depends on the quality of the attitude and positioning parameters of aerial images and the quality of a digital elevation model. It could be a better solution to give original acquired aerial images a raw core data status provided additional quality controlled metadata like the aforementioned position and attitude parameters, acquisition time, camera type, camera calibration parameters, etc. are all part of the raw core data set. Raw core data together with a proper set of metadata parameters opens the way to process the data to all kind of customer driven special products. More investigation is needed to define which meta-datasets should be collected and with which accuracy.

Which technical specifications should a raw core dataset comply with?

The required quality and the level of detail of raw core data including the technical way to provide the information need to be specified. Here it is important to not only look at what's needed at this moment, but to also anticipating on future new technologies that improve resolution or quality of raw core data or enable acquisition of core data through sensor webs. What do the users expect from raw core data sets? Which raw core dataset quality is feasible with present and near future technology? Does this match? Investigation is needed to get a clear view on both user demands and technological possibilities so that the technical specifications of each raw core dataset can be specified.

Which organisation model is a preferred one for the periodic acquisition of raw core data, the quality control and the distribution of the data?

Both the public sector and the private sector need periodically acquired data of known quality. The public sector uses these data for all kinds of planning and monitoring purposes. The private sector is the preferred party to acquire the data. In addition the private sector can develop applications and deliver services to add value to the raw core datasets. Customers will be the public sector and the private sector. Traditionally the geo-branche is a sector with many public organisations and relatively few private organisations involved. In recent years the role of the private sector has increased considerably. Investigation is needed which organisational model is needed to guarantee a regular acquisition of raw core data which satisfies prescribed quality criteria.

Theme 3. Massive data management

When acquiring raw core data with nation wide coverage the data volume easily amounts to many terabytes. This poses various questions on how to handle such massive data volumes. Currently, the Subcommission does not have the expertise to work on this issue, but the need to address the management of massive data volumes has been identified. The Subcommission plans to work out the research issues in a later stage. We briefly distinguish three research problems.

- How to browse through such large amounts of data for interactive visualisation?
- How to compress the data without loosing the original data?
- How to reduce the amount of data such that the relevant information (e.g. terrain shape in a airborne laser scanning point cloud) is preserved?

Theme 4. Extraction of geo-information

Sensor developments in the past years led to a large increase in the amount of data that can be acquired. In the air, digital cameras can now operate with high percentage of forward overlap. Airborne laser scanners can collect over 250.000 points per second. On the ground, camera's and laser scanners on a tripod have been complemented with (panoramic) cameras and scanners on mobile platforms, allowing efficient acquisition of cities from the street level. Image matching algorithms improved considerably in the last years and now take advantage of the high amount of overlap between photographs, leading to more robust estimates of corresponding points.

These developments now enable an efficient acquisition of high resolution datasets. While visual inspection of these datasets is already providing much information on the recorded area, many applications require the extraction of object oriented data. Considering the large amount of data, automation to extract information is indispensable.

This leads to the following topics for research:

- Object recognition;
- Change detection;
- Semi-automated mapping;
- Quality analysis of raw data and extracted information.

Object recognition

The automation of interpretation of aerial imagery has proven to be an extremely difficult task. Although humans often easily identify buildings, roads and terrain in imagery, it is complex to model the knowledge we use for this purpose. With the advent of airborne laser scanners as well as the progress in dense surface matching in aerial images with high overlap percentages height information becomes available to assist in the task of data (image) classification. Using height, the classification into the categories of ground, vegetation and buildings becomes much more reliable. Clearly vegetation and buildings can be considered as objects above the ground surface. Vegetation and buildings can, however, also be separated by considering the local height variations. In addition they can be supported by the analysis of multiple echoes or full waveforms (in the case of laser scanners) or by colour infrared information (in the case of optical imagery). Further research is required to improve and analyse the quality of data classifications making use of these new features.

Change detection

As most mapping activities nowadays update existing maps (and do not start from scratch), the detection of changes becomes an important aspect of topographic mapping. This is in particular true for production processes with a short map revision cycle. Here the time spent on detecting changes may even exceed the time required to update the changed features in the database. Like for the classification, height may play an important role in the automation of change detection. While it is obvious that construction or demolition of buildings leads to a significant height change in the surface model, construction of roads also involves earth works that will be visible when comparing surface models from before and after the construction. As the recognition of buildings in point clouds and imagery also becomes more reliable, results of building detection in a single data set may also be used for comparison with objects in a database to be updated.

Semi-automated mapping

The new data sources at high spatial resolution are also expected to enable a larger automation in mapping, i.e. the actual outlining, of features like buildings, roads, rail roads, and trees. This extraction of boundary descriptions is traditionally only done in two dimensions (the X–Y plane). Advancements in geo-information technology nowadays enable communication with three-dimensional (urban) environments. Research is required to further develop interactive methods for the efficient production of such 3D environments from sensor data.

Quality analysis of raw data and extracted information

As the sensor resolutions are improving and enabling new types of information to be extracted a careful analysis is required of the quality of both the raw sensor data (point clouds, high resolution imagery) as well as the information extracted from this data. This will also lead to new quality control procedures as well as criteria for the acceptance of data offered by data providers.

Theme 5. Harmonisation of Concepts and Data Models

Background

The study on semantics focuses on the meaning of concepts. Semantics concerns the mutual relationships between concepts (tree – chestnut) and the representation of these terms by lexical symbols (boom – baum – arbre – dendron – tree). Geographers and cartographers have traditionally spent a lot of attention to this aspect. Geographers are often involved in creating taxonomies; e.g. a soil classification. Cartographers are involved in the mapping of concepts to graphic symbols. The map legend is traditionally the place where the graphic symbols (signs) and meaning meet each other.

In the early days of GIS the functional and technical aspects of information processing got most of the attention. When there was a need to use data from other sources, exchanges formats were defined. These were mainly limited to specifying the syntax (structure) of the files. The true meaning of the content is outside the scope of these exchange formats (with exception of some fundamental concepts such as coordinates, reference datum). Now the use of each others data is getting more and more common and the basic technology is no limitation anymore (also influenced by Internet developments (XML related standards)), a new problem arises: how to find the right information and combine these in a useful manner. The activities in this area will have to include development of agreed domain information models (based on ontologies), exchanging of the repositories and investigating the related methods and techniques.

Agreeing on concepts of spatial data and the development of systems handling these is the first step towards spatial information infrastructures (SII). OGC and ISO/TC211 have developed a rich set of standards in this area (independent of specific themes or domains). Parallel to this development has been the growth of the Internet and all its protocols that have created the foundation of the SII. This does not mean that we understand each other's information, as for this we also have to agree on the domain (or thematic) models. In the context of these models the data get more meaning, and it is fair to state that data become information. Today these models are often expressed as UML class diagrams, often limited to just the data side (not including operations).

Topics of interest in this theme include:

- Definition of basic spatio-temporal concepts;
- Creating and using ontologies;
- Creation and harmonization of domain models;
- Methods and languages.

Definition of basic spatio-temporal concepts

Point, line and area may seam to be concepts where there is no more need to define anything further. However, when these concepts are implemented in a system, then also relationships between the concepts and more precise rules need to be defined; e.g. what is a valid representation of an area (polygon). Here still significant differences occur in reality. This is even more true for complex geometry types (B-splines, NURBS, polyhedrons, etc.) and temporal concepts. Though quite some work has been done in this area (ISO, OGC and others), still work remains to be done in order to get a consistent set of definitions for basic spatio-temporal concepts that ensures the absence of conflicting implementations in Geo-DBMS, GIS and CAD systems. So, more research is needed here.

The Subcommission Spatial Core Data will stimulate R&D activities in this field and keep in touch with relevant research groups.

Creating and using ontologies

In a large number of domains (sectors, application areas, themes, ...) there is a need to standardize the set of used concepts, often indicated with the term information model or the related term ontology, which also includes the classifications/taxonomies (is-a) and partonomies (part-of).

The Subcommission Spatial Core Data will contribute to standardisation efforts in this field and participate in the NEN3610 system consultation groups, chaired by Geonovum.

Creation and harmonization of domain models

A domain model is an information model for a specific domain, such as: topography, soil, geology, cadastre, pipelines and cables, cultural history, water, spatial planning, etc. Not only the hierarchical classification of the concepts is of concern, also the mutual association between the concepts and their cardinality is important. Further, the definition of the attributes (names and types) and the constraints associated with the model are important aspects of a domain model. Most of the time is not required to develop of a new model, but to making an implicit existing model explicit and perhaps even more often it is the harmonization of two independently developed models within the same domain; e.g. obtaining an agreement between the similar models in different countries.

Two important advantages of agreeing on domain models are (1) it becomes easier to understand the information of others within the domain and (2) system developments may be shared as many partners base their systems on the same model. The benefit of domain models (and ontologies) for facilitating information discovery and building knowledge-based systems is clear. However, independent domain models for different geo-information themes are still difficult to be harmonized between themes (perhaps confusing overlap and also double work). Anyhow, it will not stimulate interoperability between these themes as needed for a wide spatial semantic web. The development of thematic (semantically meaningful) models is the future of geo-information standardization. Recently there are a number of large initiatives that have started to develop harmonized (interoperable) model specifications covering many themes. For example, within INSPIRE, 34 different themes are covered; see http://inspire.jrc.it/. It will be an incredible challenge for the 27 countries of the European Union to realize this: first agree on the harmonized models and next deliver information according to these models. Cleary, supportive research is needed here.

Examples of a domain model of is NEN3610, Sub-models (sector-models) of NEN3610 are: IMWA, IMRO, IMGEO, IMTOP, IMKAD, IMBAG, etc.

Methods and languages

There are a number of different methods to perform information analysis and to design data models. There are even larger numbers of options available to describe and document the designed models. Some of these approaches have their origin in the (relational) database design corner, others have their roots in the Artificial Intelligence (AI) research and yet others are originating from the discipline of object oriented (OO) design and

programming. The most recent developments stem from the Internet: the Semantic Web and W3C. With several languages developed in this context, conceptual schemes can be described and exchanged. In database terminology the location where the conceptual schema (and the derived logical and physical schema's) are maintained is called the 'repository', actually containing (model/content related) metadata. Examples of these languages are: Object Role Modeling (ORM), Unified Modeling Language (UML), Object Constraint Language (OCL), Resource Description Framework (RDF), Web Ontology Language (OWL) and Formal Concept Analysis (FCA).
The research on methods and languages should result in showing the possibilities and limitations of these techniques and languages to define, harmonise and use information models and to integrate information which may originate form different sources.

The role of the Subcommission is to investigate which of these methods and languages are relevant for our geo-information discipline and how they could be put to best use.

Organisations like ISO, INSPIRE, Geonovum already play a co-ordinating and stimulating role with respect to the establishment and the harmonisation of these models. The Subcommission Spatial Core data will support these organisations in their task by focusing on the scientific aspects of these standards, e.g. by looking at questions as.

- What is the best methodology to document all these models (including storage and dissemination)? See also the section on methods and languages.
- Which other models are required?
- How to harmonise all information models concerning topography?
- How to organise the 'Stelsel van basisregistraties' for topography at different scales?

Furthermore, the Subcommission will advise organisation in the transition from one model to another one (or incorporating elements of another model) how this can be done in a cost-effective way.

Theme 6. Integration of interpreted data

Background

Two different interpreted data sets may be called 'integrated' when they behave as one single information base. An alternative term for 'integrated' is 'fused'.

It is not always the final goal to completely fuse two data sets into one single data set, as it may be needed to keep the original two data sets separately and explicitly store the correspondences (matches) between the two sets of instances. Therefore a distinction is made between the following two cases: 1. a complete fusion and 2. a 'LAT'-relationship ('Living Apart Together'). Both a complete fusion and a LAT-relationship require that the data models of both datasets are harmonised. This condition is the subject of theme 5. In case of a complete fusion also the following two conditions need to be fulfilled:

- The object populations of corresponding object types have to be equalised.
- The object identifications of corresponding object instances have to be aligned.

These both conditions are further explained in the sections underneath. In case of a LAT-relationship between the datasets, these conditions are not required. Corresponding object instances in both datasets are related by means of relationship table. This may result in 1:1 relationships but in some situations also in 1-to-many or even many-to-many correspondences.

The required research in this field is the search for optimal methods to find corresponding object instances.

Equalising object populations

Two different (interpreted) datasets may differ in content and accuracy, even when they are based on the same domain model and use the same surveying specifications. This has to do with differences in interpretation, levels-of-detail (scale), scope (relevant attributes) and up-to-date-ness. Considering two geographical interpreted datasets (collection of features), this may be seen as a difference in *population*. Equalising is the activity of adapting both datasets in such a way that these population differences disappear (or are hidden). This equalisation process normally is a time consuming task. Reason that organisations exhibit hesitating behaviour and are often postponing this process. An example of datasets, which populations are worth to be equalised, are the GBKN and the various local datasets (GBR/DTB, Pro Rail GBKN) that other organisations maintain and that partly use the same domain model as GKBN does. Another example is Top10NL and NWB. These Information Bases are partly based on the same model, but they show nevertheless differences in those object populations (e.g. Junctions) that theoretically could be the same.

The required research in this field is searching for methods that can make this task easier.

Aligning object identifications

Two (interpreted) datasets can be considered as being integrated when they behave as one single (interpreted) dataset (though physically it may be distributed). Equalisation of the object populations is a necessary condition, but not sufficient. Identical objects need to be identified in a unique an unambiguous way so that they can be referenced in an unambiguous and consistent manner. Sometimes a combination of attributes can play this role, but in most case use has to be made of an 'artificial' identifier (number or name) which is especially created and assigned for that purpose. Due to the fact that the assignment is independently done in two different information bases, corresponding objects in these information bases will have different identifiers, even in the case that the populations have been equalised. The main task of integrating two datasets consists therefore of caring that the identifiers of the objects in one dataset linked to those of the others, or, which amounts to the same, that a look-up table between those identifiers is created. Relating object identifiers is a time consuming task. This task can be relieved by techniques using the above-mentioned 'identification by means of corresponding attributes and relations'. In particular the shape and position of objects can serve as a powerful identification tool. This technique never leads to a 100% reliable identification (which is the reason that it can not replace artificial identifiers) but it can be used to do the main of the work in relating those artificial ids.

Research is needed to contribute to solving this 'identifier matching' question.

Theme 7. Multi-scale issues

Several parties collect geo-data with similar or at least related content, at both similar and different scale levels. Acquiring and maintaining consistent geodatabases is a heavy task. For efficiency and consistency reasons the key issue of INSPIRE, as well of core spatial data sets is: collect once, use many times. Conceptually this approach seems very logic for disseminating geo-information. This concept starts with storing a very detailed version of the spatial data. In a next step any required data set at a less detailed level is automatically derived from it, at the moment needed. This process is called 'generalisation'. Automated generalisation has received a lot of attention in research from the time geo-data became digitally available. However full automated generalisation is still not possible and some argue that it will never be (at least not for topographical data), since some human interpretation as applied in generalisation can never be automated. Often it is required to store data at several scales in a multi-scale database. In order to avoid inconsistencies and in order to use the large scale data to update small scale data, multi-scale knowledge should be available in both data models, as well as database and generalisation applications.

Projects such as Magnet have shown potentials of an object oriented approach for automated generalisation. These solutions should be applied and extended for Dutch cases (e.g. generalisation of TOP50NL from TOP10NL), to get insight into feasibility of automated generalisation for INSPIRE and key registrations as well as into future research issues. Future research in generalisation and multi-representation should focus on the following aspects.

- Generalisation of specific data sets.
- Multi-scale database.
- Scaleless data sets.

Generalisation of specific data sets

Available knowledge in the area of generalisation needs to be consolidated and applied to IMGEO- and TOP10NL-compliant data sets in order to derive topographical databases and maps such as TOP25NL, TOP100NL etc. For significant progress in automated derivation of these products more insight is required to answer the question whether there should be a separation between database and map (in current production line there is no separation) and how this separation should look like. That is, a different representation of the instances of the Digital Landscape model (DLM) and Digital Cartographic Model (DCM). Formalising requirements for generalisation (covering both maps and databases) is extremely important for machine-based solutions.

Also efficient algorithms for generalisation still need further development, specifically ones that take the context of objects into account. Examples of retained problems in automated generalisation are building generalisation in urban zones, solving overlapping conflicts in locally dense networks, pruning of artificial networks, and ensuring consistency between themes in particular areas such as coastal zones. Other type of algorithms that needs more attention, are algorithms generalising data with a temporal component where both the spatial as well as the temporal component need to be generalised.

Evaluation methodologies need to be developed to assess the outputs of automated generalisation processes.

Multi-scale database

In situations where automated generalisation is not feasible a solution should be studied for a multi-scale data-base, where derivations of several generalisation steps are maintained and supported by applications. Knowledge on scale transitions should be formalised and modelled within the multi-scale database (e.g. which object at a small scale correspond to an object at a large scale; how do object classes and specific instances behave at scale transitions). Such knowledge can be a result of an analysis of human decisions in interactive generalisation processes completed with context dependent information.

Multi-scale spatial analyses need to be developed for multi-scale databases, e.g. typical GIS analyses in which several data sets are combined to generate new information. The multi-scale database should also be supported by functionality enabling querying the multi-scale database.

Scaleless data sets

Scaleless (or vario-scale) data sets are another research area for generalisation. Scaleless data structures enable objects to be stored once and to be displayed at any arbitrary scale via the use of supporting data structures. These data structures then contain a lot of the generalization 'decisions' (computed at pre processing time), avoiding starting from scratch when deriving a smaller scale representation from a large scale source. Vario-scale data structures do avoid multiple representations (as much as possible) and are therefore less sensitive for inconsistencies between multiple representations. A step-wise process should show the feasibility of such an approach for practical applications in the context of INSPIRE and key registrations, that is, in some situations a 'second' representation is created; e.g. in case of complex situations (many objects participating in an aggregation or other generalization operation; costly geometric computations; etc.) Vario-scale data structures could be used to realize smooth zooming, making sure that the user is 'not lost' in the step from one scale to the next scale. Vario-scale data structures could further support progressive transfer in a network setting when transferring data from the server to the client (streaming mode): show rough representation first, which is then gradually refined when more detailed data is being received.

Theme 8. Time and history

Adding time to a spatial information base makes data handling a lot more complicated. Nevertheless there are important reasons to add one – or even more – time dimensions to a geographical database; examples of such reasons are: monitoring a spatial phenomenon (climate change), monitoring the changes in a set of related features (merging or splitting parcels in planar partition), monitoring the changes in the characteristics of a particular feature (number of passing cars per hours at road junction), and transfer the changes from one database to another (move from the newer to the older database). Note that in theme 6, differences in population of two datasets may be due to difference in actuality (time) and equalisation my require synchronisation of the two dataset (which should then occur at a regular basis). The following issues arise frequently in a temporal database:

- Temporal primitives;
- Continuous versus discrete representation;
- Time and space separate or integrated attribute(s).

Temporal primitives

Similar to the spatial representation, there are a number of temporal primitives used in modelling spatio-temporal information. Relevant aspects of the temporal representation include: valid time versus system time (real world versus database time), moments versus periods (time intervals), measuring time (units) and notational aspects, and temporal granularity. A bi-temporal spatial database supports both valid and system time. With respect to granularity of the pieces of data to which temporal information is attached, this can range from course to fine pieces of data: map or universe/whole data set (e.g. every 6-years revised), object class (e.g. all roads every 2-years), object instance (e.g individual parcel on a cadastral map), or attribute level (e.g. ground water level at fixed station/point location). In general the more course the granularity the higher the redundancy (because also unchanged data is replicated). However, the more fine the granularity the more complicated the temporal models becomes.

Continuous versus discrete

Time is, like space, a continuous concept, not only from a mathematical but also from a physical perspective: between the life time of an elementary particle and the universe are 25 powers of 10 (10^{25}). The storage of time in an information base however, is necessarily finite, which means that one has to choose for a smallest time unit. This however may cause problems when one starts to calculate with time (e.g. a route planner that plans a route and wants to predict the traffic intensity in a given location P at the moment of passing that location). Representing continuous changing phenomena (e.g. salinity in the ocean) require other temporal representation techniques (based on sampling) then discrete changes (e.g. splitting a parcel and selling one of the parts). This differences is often aligned with the difference between natural and human-conducted processes.

Time and (up to 3D) space separate or integrated attribute(s)

Deep integrated treatment of (up to 3D) space and time in one internal 4D data type representation might have some benefits for the future realization of a (3D) spatio-temporal information systems. Deep integration implies that an object does not have separate attributes for its spatial characteristics and its temporal characteristics, but only an integrated spatial-temporal description. Some of the potential benefits are: optimal efficient 4D searching (specifying both space and time in same query), true 4D data types provide parent-child relationships between parcels (the lineage) as neighbour queries in a topological structure (neighbours for which at least the time attribute did change), 4D analysis: (e.g. do two moving groups of fish have spatio-temporal overlap/touch?), but most important, several applications might require a conceptual full (4D) partition (of 3D space + time, no overlaps, no gaps) as our foundation for the system; e.g. 4D Cadastre, having true 4D geometry and topology (space and time integrated) is the most solid foundation. However, there are also a number of arguments, which can be made in favour of separate treatment of space and time: current (new, but state of the art) technology can be used to implement the separated approach while for support for true 4D geometry and topology further R&D activities will be required.

Theme 9. 3D geo-information

There is an increasing need for 3D geo-information in general and 3D topography in specific. This is caused on the one hand because 3D technologies for collecting 3D information and for building 3D models and using these in 3D applications are maturing and therefore these become available to be applied in spatial applications. On the other hand the intensive use of our environment as well as the growing awareness for the environment require more precise registrations of spatial situations as well as more accurate predictions of the impact of pollutions and disasters on the environment. This search for improved accuracy and precision triggers the increasing need for 3D information and applications. Several research topics in the area of 3D geo-information can be defined:

- Gap between 3D research and non-ad hoc applications/real user requirements;
- 2D and 3D functionality in one seamless environment;
- 3D raw data acquisition models and 3D interpreted models;
- The complete 3D chain, including interaction and visualization;
- Integrating 3D with time and scale dimension.

Gap between 3D research and non-ad hoc applications/real user requirements

Firstly a gap can be identified between achievements in research and the hesitations of organisations from practice for introducing 3D applications. There seems to be a mismatch between current research efforts and user expectations and needs. Consequently prototypical applications need to be analysed with regard to user requirements for 3D geo-information. Insight into those requirements should lead to the definition of a 3D topographical model that can serve applications, together with new methods and techniques for data collection, storage and analysis.

2D and 3D functionality in one seamless environment

Related to research on user requirements for 3D geo-information is a second research topic on how 3D analysis and 3D simulation techniques can extend the possibilities of 2D spatial applications. Also it may be desired to have representations which are capable of merging 2D and 3D data in one environment and also to do processing in this environment. Often there are already a large 2D data sets available (and at many locations this can be sufficient) and it is sufficient to have only a limited number of areas represented in 3D. however, such an integrated 2D/3D environment may from the conceptual point of view more complicated that a pure 3D environment (but requires less from new data acquisition).

3D raw data acquisition models and 3D interpreted models

A third research issue for 3D geo-information relates to the rapid developments in sensor techniques. Because of these developments more and more 3D data becomes available. Effective algorithms for (semi) automatic object reconstruction are required. Integration of existing 2D objects with height data is a non-trivial process and needs further research. The resulting 3D models can be maintained in several types of 3D models: TEN (Tetrahedral Network), Constructive Solid Geometry (CSG) models, Regular Polytopes, TIN Boundary representation and 3D volume quad edge structure, layered/topology models, voxel based models, 3D models used in urban planning/polyhedrons, and n-dimensional models including time. Research is needed to see what applications can be served best by what kind of model.

The complete 3D chain, including interaction and visualization

A multidisciplinary approach for research on 3D geo-information is required. 3D geo-information covers a wide range of research areas such as requirements analysis, data collection and modeling (advanced approaches for 3D data collection, reconstruction and methods for representation, linking CAD and GIS), data management (topological, geometrical and network models for maintenance of 3D geo-information), data analysis (frameworks for representing 3D spatial relationships, 3D spatial analysis and algorithms for navigation, interpolation, 3Dfuntionalities etc) and visualisation (Advanced Virtual Reality and Augmented Reality visualisations). Considerable progresses in 3D applications can only be assured if the interdisciplinary aspect of 3D geo-information is acknowledged in scientific research.

Integrating 3D with time and scale dimension

As for 2D data, also scale (level-of-detail) aspects and temporal (including versioning, history) aspects are relevant; see themes 4 and 5. There may be good motivations to integrate the 3D spatial dimensions in a representation also supporting the temporal and level-of-detail aspects.

Theme 10. Shared mapping

Background

Currentness of maps has been in the attention of mapping companies since the beginning as a main explaining factor for map errors. Traditionally this has led to updating concepts whereby updated versions of a map were released on regular intervals, typically in the order of magnitude of several years. Clearly, this was considered less a problem for largely static map contents than for content with a high rate of change. As such topographic maps of areas of high economic growth and a corresponding high degree of topographic changes were updated with intervals of 1 to 2 years while more remote regions were updated with intervals which could exceed periods of 5 years. Traditionally (paper) commercial road maps typically have an updating interval of one year or longer.

With the onset of digital mapping the principles of map updating did not change. Also here a map was updated by releasing an updated map. Digital maps are typically updated with a release schedule of twice or four times a year.

Digital maps, unlike traditional paper maps typically are not used stand-alone but in a system which delivers a service to the user on basis of the map information. The consequence of this is that the map user is less forgiving with regards to map deficiencies. Rather than considering it as an inevitable aspect of maps, it is considered as a system malfunction. In in-car systems, the safety aspect of map deficiencies is recognised more and more. Generally these systems are considered to contribute positively to traffic safety. Map deficiencies leading to incorrect advice to the car driver however decrease this effect. The safety aspect becomes more prominent with the onset of the use of maps in in-car safety systems (aka ADAS) which assist the driver to drive safely. Clearly, the consequences of map errors resulting in system malfunction are less acceptable than with previous ways of map use. These factors have are putting a bigger emphasis on map updating and are calling for more advanced ways which in the end will enable daily or weekly release schedules of updates.

Looking to improving traditional ways of map updating, i.e. extending field survey frequencies, flying the area more often, processing external source data more frequently etc., to solve this problems is both from an economic point of view and from an organizational/ logistical point of view often problematic. A promising alternative is to leave map making not only to the professional map makers but to involve also other stake holders of the mapping process in this process. These stake holders are generally referred to as the map community.

The role of the map community is not restricted to providing updates. It can also play a role in the generation of new information, i.e. they can add their own information to existing maps. In this way, owners of business can make sure their business is (correctly) added to an existing map. And the community can even go one step further. It can create its own map in a joint exercise, from scratch or on basis of an available open source type of map. Generally this process is referred to as Open Source Mapping. The OpenStreetMap initiative is the most obvious example of open source mapping.

The map community

Two sub-groups in the map community are of particular importance for the map updating process. First there are the users of the maps. These are confronted with map anomalies while they are using the maps. Enabling them to report about map anomalies has a big potential for map updating. This group also will contain the people who are interested in adding new information to the map. The second group are the people or organisations who are responsible for the change process of the reality represented in the maps, i.e. the process which causes the map to be out-of-date. Road authorities are in charge of a large portion of the change in the road network which makes them of particular importance for road maps. Municipalities are another category of organisation in charge of the change the reality reflected in road maps. They submit building permits for new construction and related to that are in charge of the design and implementation of the related road networks.

The community contributing to Open Source Mapping is in principle unrestricted. In practise however it the contributors will be part of the group who use the open source map.

Map users as source for map updating Open Source Mapping

There is clear evidence, as for instance the on-line encyclopaedia Wikipedia clearly shows, that communities of users are able to generate a high quality results through a community process. In order to deliver high quality map updates additional measures are likely necessary though. More attention needs to be given to the user interface through which the user can report his updates. In fact, such a user interface should be a reflection of the map model. In such a way, the user will be forced to report updates in a way which is meaningful in the map context and gross misconceptions will be avoided. Also statistical processing of the updates is important. This will avoid that updates which, wilfully or by accident contain false information. The relevance of statistical processing also is an indication that the community should be of a certain minimum size which points again to the relevance of the user interface which should appeal to the user. Another relevant aspect is the way in which the community is stimulated to provide its input. Clearly members of the community have an interest in providing their input and care should be taken to fulfil this interest.

Apart from active reporting of updates also passive reporting of so called Floating Car Data (FCD) can be used to provide update information. The use of anonymous position data will prove to be a powerful source for map updates as well as for the generation of dynamic traffic information. The potential of this needs further research.

Open Source Mapping is a largely autonomous process. Quality levels are more the result of a process than a requirement. This is also likely to be true for the data model which defines the structure of the map. Research is to focus on describing the process and its outcome rather than on how certain requirements can be fulfilled. Ownership aspects and rights-of-use is another topic for research.

Road Authorities and Municipalities as source for map updating

Road Authorities and Municipal Authorities are in charge of a large portion of the information contained in digital road maps and as a consequence also of the changes therein. Information from road authorities or authorities in general is already since long an important component of the updating process of digital road maps. However, this process is characterised by its informal nature and non-standardised information flows. For a truly efficient map it is necessary that the following measures are addressed:

1. Formalization of the process.
2. Development of standard information protocols and interfaces.
3. Integration of the information protocols and interfaces in the planning and execution processes of the authorities.

The relevance of the role of Road Authorities in the updating process of digital road maps has been recognised by the EC. Working Group 11 of the eSafety forum has issued a final report (www.esafetysupport.org) recommending a European infrastructure to enable active involvement of Road Authorities in the supply of update information to producers of digital road maps. The EC-FP7 project ROSATTE (FP_-ICT-2007-1-213467) is currently designing and prototyping such an infrastructure.

CycloMedia's aerial and ground-based image databases

Dr.Ir. F.A. van den Heuvel
CycloMedia Technology bv, the Netherlands

1. Introduction

Driven by the motto "an image says more than a thousand words" CycloMedia is building large-scale image databases with both aerial and ground-based imagery. CycloMedia is a spin-off company from the Delft University of Technology that has a long term experience in systematic and large-scale visualisation of the environment through 360° panoramic imagery taken from all public roads. This limits the view on the environment to those parts visible from the road. This disadvantage has been relieved by the recent extension of CycloMedia's portfolio with aerial imagery. As a result, the environment not visible from the road can be inspected from above.

The aerial images are taken with a state of the art photogrammetric camera while ground-based images – so-called Cycloramas – are taken with a panoramic camera. Both types of cameras are calibrated and thus facilitate photogrammetric measurement and both types of images have a number of common characteristics that make them suitable for a wide range of applications. Examples of applications are supporting property taxation, or the implementation of the BAG ('Basisregistratie Adressen en Gebouwen' or 'Address and Building Base Register'). The common characteristic that characterizes these images as core spatial data is the fact that the location of each image is known in the national coordinate system.

This paper discusses the national aerial and ground-based image databases of CycloMedia. It focuses on the image acquisition techniques and workflow with emphasis on the new mobile mapping system DCR7 and its calibration. The aerial image database is discussed in section 2 with details on the workflow (section 2.1) and the approach to quality control (section 2.2). Section 3 addresses the ground-based image database with emphasis on

Figure 1. Airborne and ground-based platforms at airport Lelystad.

Figure 2. Sample aerial image (left) and Vexcel camera (right).

the mobile mapping system (section 3.1) and its calibration (section 3.2), the Cyclorama production workflow (section 3.3), possible applications of Cycloramas (section 3.4), and a promising new research field of automated image content analysis (section 3.5).

2. Aerial imagery

Aerial imaging is contracted out by CycloMedia to a photogrammetric company that applies state-of-the-art technology. In 2008 Blom Aerofilms, a UK-based photogrammetric company was selected for a unique aerial imaging project for nation-wide stereo- and orthophoto production of the Netherlands. The stereo imagery is acquired with 10 cm ground sampling distance (GSD) and covers an area of over 40,000 km^2. A yearly update rate is aimed at.

New for the Dutch aerial imaging market is that through this project aerial imagery can be bought off-the-shelf, while traditionally purchasing this type of data involved a tendering process that required a high level of domain knowledge, or the involvement of a party with an advisory role.

The photogrammetric aerial image acquisition and processing is a mature technology with well defined requirements. The workflow is therefore standardized to a high degree.

2.1 Workflow: from photoflight to orthophoto

Up to five aircrafts equipped with digital photogrammetric cameras have been deployed in order to maximise the chances of completing the photoflight of the whole country within the flying season of 2008. This is a challenging objective that never has been realised before. Especially the Schiphol area is a risk in this respect because of limited access due to the heavy air traffic. At the end of 2008 92% of the Netherlands had been photographed

Two cameras from the same product family are being deployed: the Vexcel UltramCam D (90 Mpixel) and the Ultracam X (140Mpixel). The stereo imagery is flown with 60% overlap in flying direction and 30% side overlap. For the complete Netherlands at a GSD

Figure 3. A full 360x180 degree spherical panorama or Cyclorama.

of 10 cm the required storage space is in the order of a hundred terabyte. The file format for the stereo imagery is untiled tiff-5.

Before the photoflight takes place, approximately 500 control points have been marked in the terrain and their locations were accurately measured using GPS-RTK. After the photoflight the raw imagery is converted to aerial photographs that are input to a semi-automatic aerotriangulation, followed by a bundle block adjustment. A detailed study on block adjustment with the UltracamD camera can be found in (Baz et al., 2007). The stereo imagery, camera calibration information, and exterior orientation data allow the customer to directly load the data in a photogrammetric workstation, and perform 3D feature extraction, city modelling, and mapping. The next step in the workflow is the production of an orthophoto, also with a GSD of 10 cm and delivered in blocks of 500 x 1000 meter in the geo-tiff format. The orthophoto is a seamless image with homogeneous scale, composed of semi-automatically selected parts of the stereo imagery. In this step use is made of a height model. Furthermore, cut lines are being set in such a way that the visual appearance of the orthophoto is optimized.

2.2 Quality control

The overall quality control is assured by an independent photogrammetric company: Ingenieursbureau Geodelta. Detailed specifications have been formulated for both the stereo imagery and the orthophoto, and all data and products produced are being verified by Geodelta. The results of the quality control are published on the Internet together with the progress of image acquisition, processing, and availability (http://www.cyclomedia.nl/page.php?id=324).

3. Ground-based imagery

For the acquisition of ground-based imagery CycloMedia has developed several generations of panoramic camera systems. With the latest system, referred to as mobile mapping system DCR7, 360-degree imagery free of parallax is taken from a driving car. With the standard interval of 5 meter the maximum speed of the car is 80 km/hr. This is equivalent to about 5 frames per second. Currently, a large number of systems is being built. These mobile mapping systems facilitate a regular update of the photography of the Netherlands, as well as CycloMedia's expansion to other countries in Europe and the Middle East.

Figure 4. The CycloMedia ground-based image acquisition platform DCR7.

In section 3.1 the mobile mapping system DCR7 is presented. The calibration procedure is presented in section 3.2, the Cyclorama production workflow in section 3.3, and possible applications of Cycloramas in section 3.4. Currently, a considerable research effort is devoted to the automated analysis of Cyclorama image content and is discussed in section 3.5.

3.1 The mobile mapping system

CycloMedia's mobile mapping system DCR7 contains the following sensors:

- 2 high-resolution digital cameras with electronic shutters;
- positioning system based on GPS and an Inertial Measurement Unit (IMU);
- 2 high-resolution Distance Measuring Instruments (DMI).

For positioning the SPAN (Synchronised Position, Attitude and Navigation) system of the Canadian company NovAtel was chosen. The positioning solution is based on an integral use and adjustment of GPS-, IMU-, and DMI-data. GPS-data consists of both data received by the GPS-antenna in the camera system and data registered at fixed GPS-stations in a reference network. In this way it is possible to determine the position continuously, even when there is temporarily bad or no GPS signal reception as between high buildings or in tunnels. Where in the past the position could be accurately determined to within a few metres, now an accuracy of a decimetre and 0.1° is possible. An important advantage of this methodology is that no control points are needed for the positioning and orientation.

The two cameras are equipped with fish eye lenses, and are triggered in such a way that the entire 360-degree image is recorded at one single position. Together with a proper camera calibration this allows the construction of a seamless panorama free of parallax.

3.2 Calibration of the panoramic camera system

A thorough calibration of the systems is essential for metric use of the recorded image data. In (Van den Heuvel et al., 2006) the calibration procedure of the previous version of

DCR is discussed. Here we elaborate on the latest, largely automated procedure designed for DCR7. This calibration procedure is repeated at least yearly to monitor the stability of each camera system.

The calibration of a DCR7 has a radiometric and a geometric component. The radiometric component of colour calibration consists of establishing the colour settings for each camera. Both of the system's cameras have a grey card visible in the image that give a reference for correcting the colour as much as possible when processing the recording. Below, we will concentrate chiefly on the geometric component of the calibration. This calibration consists of three steps:

1. Camera-lens calibration;
2. Boresight calibration;
3. Calibration verification.

1. Camera-lens calibration
In the first step, the internal geometry of every camera is determined. In photogrammetry, this is called the internal orientation. With the internal geometry, all image distortion resulting from properties and placement of the lens is known and can be eliminated. The camera model used for this purpose contains the focal length, the principal point and lens distortion parameters.
The camera-lens calibration is performed in CycloMedia's calibration room in which there is a field of more than 500 partly coded targets. This calibration field is photographed with every camera system in four different positions, and four different orientations. Then the images are measured and processed entirely automatically and the above-named parameters are calculated. The most important part of this processing is the least-squares adjustment for determining the parameters of the camera model.
A number of criteria are tested during the quality inspection, including the estimated standard deviations of the observations such as those determined by the adjustment. These are required to be under 0.2 pixel. The reliability of the camera lens calibration is high, owing to the considerable redundancy in the adjustment.

2. Boresight calibration
The second step in calibrating the system concerns the so-called boresight. This means that the relative orientation of both cameras in the system is determined in relation to the IMU. First a calibration run is carried out, and the images are automatically processed. Afterwards, the orientations of the individual images are calculated with a least-squares adjustment, also called a photogrammetric bundle adjustment. Comparison with the orientations measured with the IMU provides the boresight.
Here too, quality checks have a place in every step of the process. The results of the bundle adjustment are particularly analysed because the quality achieved in this step is indicative of the measuring precision that is possible with the images. The estimated measuring precision is required to be better than 1 pixel.

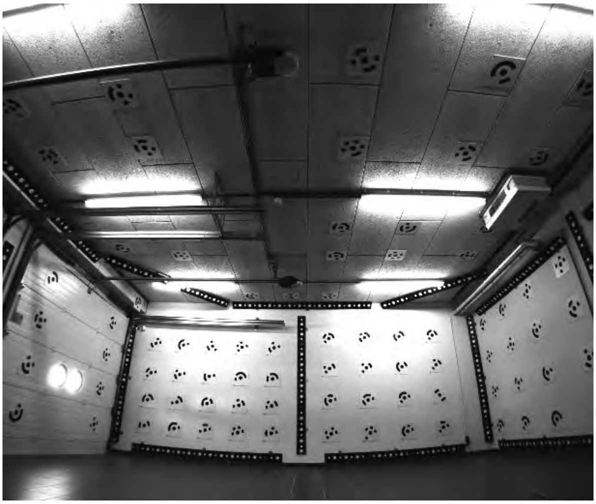


Figure 5. The calibration room with coded targets.

3. Checking the calibrations
When all parameters described above have been determined, they are saved in the memory of the camera system itself. This enables straightforward circulation of systems. The images made during a test drive are processed and analysed on a large number of characteristics such as: sharpness of the whole picture, exposure, colour, levelling, and parallax. The joining of both image segments should be nigh on invisible. In addition to this, another automatic bundle adjustment is performed with complete 360° panoramas in order to verify the accuracy of measurements.

3.3 Cyclorama production

Special real-time hardware regulates the triggering of the shots under the control of software that also controls communication with the operator. As an option, it is possible to let the recording system function fully autonomously. As a result of this, the operator is only responsible for driving the vehicle, whilst the system itself ensures the production of a single set of geographic images with set recording interval, adjustable for individual areas. The combination of the autonomous recording system and the camera triggering working independently of the speed makes a strong contribution to increased traffic safety.

Standard available hardware is utilized for saving the image data generated by the cameras (up to 80 MB/s). Due to the system having a modular structure, the recording vehicle does not have to be taken to a post-processing location: a courier simply exchanges the system's data disks at the production location.

Processing of the recorded image and positioning data takes place in CycloMedia's data centre. The result after post-processing, using correction data from a GPS reference network, are panorama images with the following properties:

- 360° field of view with a resolution of 0.075° per pixel, equivalent to 13 mm at a distance of 10 m from the camera;

Figure 6. *Sample ground-based image.*

- seamless, parallax-free, corrected for tilt (levelled);
- known image geometry (subpixel level);
- georeferenced (position 0.1 m, orientation 0.1°).

Because old images are also kept and remain available online, the collection now contains tens of millions of Cycloramas. These large quantities of data require storage space in the order of hundreds of terabytes.

3.4 Applications of ground-based imagery

Cycloramas are primarily used for visual inventories. However, the properties mentioned above allow determining the position and dimensions of an object visible in more than one panorama. Because of the large-scale and systematic recording methodology, Cycloramas form core data for making inventories. Furthermore, Cycloramas facilitate creation and texture mapping of 3D city models. For these applications, CycloMedia and third parties have developed viewing, retrieval, measuring and processing software. Much of the functionality can be used integrated into all leading GIS packages. It is also possible to access the data on mobile phones, and PDAs.

3.5 Image Content Analysis

Recording, processing, and hosting of image data is the core business of CycloMedia. However, CycloMedia's R&D department is putting more and more effort in the investi-

Figure 7. Integrated use of aerial image and Cycloramas for road asset management.

gation and development of tools for supporting a variety of applications in which object recognition plays a major role. Because of the vast amounts of image data these tools are required to show a high level of automation. Currently this information extraction is still a labour-intensive process in which user-friendly tools play a major role. An example for the road sign inventory application is shown in Figure 7. The location of Cycloramas is displayed on an aerial image, possibly in combination with a base map. Road signs are identified, classified, and located (measured) using multiple Cycloramas. The road signs are then added as a layer to the map data, and their symbols can be back projected into the imagery, just as any other layer of the map data. Optionally, an image of the traffic sign can be cut from a Cyclorama and added to the database.

Road sign inventory and road mark extraction are asset management applications with a great potential for automation (Herbschleb & de With, 2009). The integrated use of aerial and terrestrial images is expected to improve the efficiency, especially when the registration of the two types of imagery is of high fidelity (Tournaire et al., 2006). This application area is to be extended in the future towards the automatic detection and recognition of street furniture, house facades, and other objects. In fact, CycloMedia aims at a high level of automation of the mapping process. The automatic detection and matching of objects in multiple georeferenced images implies that the location of these objects on the map results. Furthermore, it is a first step towards automated 3D reconstruction and modelling of the environment for an application as city modelling. Using Cycloramas limits the use of resulting 3D models to a street level perspective. With the help of aerial stereo images this limitation can be eliminated. Therefore research aims at combining both image databases.

Our environment is not static and thus there is a need for a regular updating of the image databases. The current databases with Cycloramas dating back to 1990 can be regarded as a cultural heritage archive. Besides, answering the question *"What has changed?"* is a challenging one if these changes are to be detected in millions of images. Therefore automated change detection is a hot research topic.

4. Conclusions

Many millions of high-resolution images taken from the air as well as from the ground have been acquired by CycloMedia and hundred thousands are added every week. These represent a large amount of core spatial data and inherent challenges in image storage and retrieval that require advanced tools for querying this visual database. A seamless integration with available GIS-systems is essential for applications such as management of open space and urban planning. In these traditional viewing applications image content is of primary importance, geometry is secondary. However, the high quality of the georeferencing and the geometry of Cycloramas pave the road for new applications, especially when Cycloramas are combined with aerial images. In the first place updating of large-scale base maps like the GBKN is a promising application. Cycloramas offer the perspective of the surveyor which has a positive effect on the quality of the (photogrammetric) mapping and will reduce field work. Furthermore, research is conducted for the analysis of image content for an application as *road asset management*, and Cycloramas are suitable for the production and photorealistic texturing of 3D city models. Because of the large amounts of data, research aims at fully automatic processing in all applications. Many challenges are awaiting us, but the required core spatial image databases are available!

References

F.A. van den Heuvel, R. Verwaal, B. Beers, 2006. Calibration of fisheye camera systems and the Reduction of chromatic aberration, Proceedings ISPRS Commission V Symposium, IntArch-PhRS Vol. XXXVI, Part 5, Dresden, Germany. www.isprs.org/commission5/proceedings06/paper/1267_Dresden06.pdf.

E.D. Herbschleb, P.H.N. de With, 2009. Real-time Detection and Recognition of Traffic Signs, SPIE Electronic Imaging Conference, VCIP, Vol. 7257, San Jose, 18-22 January 2009 (accepted for publication, see: http://spiedl.aip.org/).

O. Tournaire, B. Soheilian, N. Paparoditis, 2006. Towards a sub-decimetric georeferencing of ground-based mobile mapping systems in urban areas: matching ground-based and aerial based imagery using roadmarks. in IntArchPhRS, Proc. ISPRS Commission I Symposium "From Sensors to Images", Paris, France.

I. Baz, G. Buyuksalih, K. Jacobsen, 2007. Bundle block adjustment with high resolution UltraCamD images, IntArchPhRS Vol. XXXVI. Band 1/W51. Hannover, Germany.

Cadastral information: more than base data

Prof.Dr. J.A. Zevenbergen [1]***, Dr.Ir. H.T. Uitermark*** [2]***, Ir. C.H.J. Lemmen*** [2]
[1] TU Delft, ITC; the Netherlands
[2] Cadastre, ITC; the Netherlands

Abstract

Cadastral information combines geometric and administrative data related to land, including data on ownership, use and value of the land. The cadastral map is a geometric description of parcels, determined by the area to which the same property right applies. In most cases use and user of the land are also connected to these parcels, and thus this dataset is an excellent base to link, analyze and visualize all kinds of socio-economic phenomena. The cadastre plays important functions in the land market, the physical planning and the taxation of a country, once it runs in an appropriate way.

1. Introduction

If something is labelled 'cadastral information' most of us will have an immediate idea about what information we are talking about. Nevertheless, the term 'cadastre' (as well as the somewhat more general term 'land administration') has many non-identical definitions, and it may include data on a variety of aspects, that are useable for a multiplicity of applications. We will give an overview in section 2 "The many faces of cadastral data".
The actual implementations of cadastral information systems in different countries also show much variety. Work on designing an (international) data reference model, that captures the core of cadastral information, has been ongoing since the International Federation of Surveyors (FIG) Congress in 2002 (Van Oosterom and Lemmen 2002). First under the name Core Cadastral Domain Model (CCDM), and more recently as Land Administration Domain Model (LADM), which has been submitted in 2008 by FIG to ISO for standardization. In section 3 "A reference model for cadastral data" we will elaborate on this development.

The Social Tenure Domain Model (STDM) is a pro-poor land administration tool. It is based as such on LADM. It covers also land administration in a broad sense including administrative and the spatial components. Traditional (or conventional) land administration systems relate names or addresses of persons to land parcels via rights. In the STDM, an alternative option for this is to relate personal identifiers such as fingerprints to a coordinate point inside the land in use by that person, via a social tenure relationship. This is further explained in section 4 "The social tenure domain model".

Cadastral information clearly plays a role in the geo-information infrastructure or spatial data infrastructure (SDI). For SDI a list of core layers is mentioned: topographic (elevation), cadastral data, geodetic control, and government/administrative boundaries (Onsrud 1998). In the Dutch context the cadastral information has been legally mandated as one of the key registers (with base data; sometimes called 'basic registers') in a new law, which became effective at the start of 2008. This means that the cadastral information, as far as a data element has been declared authentic, should be used within the public sector. Earlier this was determined for certain topics in specific laws (e.g. who has to pay the real estate tax?, who is mentioned in an expropriation order?). The consequences of a full implementation of the 'authentic data' notion for cadastral data will be far reaching when fully applied. In the realm of private law, however, the legal status of the cadastral information in relation to real estate transactions is rather limited, and is not supposed to change. In section 5 "Cadastre data as a key register" this situation will be discussed, with reference to the earlier situation and the (continuing) situation in the private law sphere. We end the paper in section 6 with a number of "Final Remarks".

2. The many faces of cadastral data

The concepts and definitions of cadastre, as well as the wider 'land administration' are introduced, as well as the notion of the cadastral parcel.

2.1 Cadastre

To give the definition of the term 'cadastre' is not possible. Even its linguistic roots are uncertain with both Greek and Latin being mentioned. It can be defined "as an official record of information about land parcels, including details of their bounds, tenure, use, and value" (McLaughlin and Nichols 1989: p. 82). It usually refers to a predominantly technical registration, which contains information on where a property is located, what its boundaries are and how large it is. The use of the term cadastre has been mainly found in continental Europe, where it has shifting meanings. In much of the Anglo-Saxon world the term was virtually unused, although the term cadastral surveys has been in use for the surveying of property boundaries. The term has been promoted at the international level by the FIG in 'The FIG Statement on the Cadastre', which contains the following description:

"A Cadastre is normally a parcel based, and up-to-date land information system containing a record of interests in land (e.g. rights, restrictions and responsibilities). It usually includes a geometric description of land parcels [*for example see Figure 1*] linked to other records describing the nature of the interests, the ownership or control of those interests, and often the value of the parcel and its improvements [*for example see Figure 2*]. It may be established for fiscal purposes (e.g. valuation and equitable taxation), legal purposes (conveyancing), to assist in the management of land and land use (e.g. for planning and other administrative purposes), and enables sustainable development and environmental protection." (FIG 1995: p. 1).

The different applications of the cadastre given in the last sentence, are also referred to as the fiscal, juridical (or legal) and multi-purpose cadastre (e.g. Dale and McLaughlin 1988: p. 13, McLaughlin and Nichols 1989: p. 82).

A cadastre usually consists of two parts; a geographic part ('map' or 'plan') and a descriptive part ('register' or 'indexes'). The relation between the two is of the utmost importance, and usually arranged through a so-called 'parcel identifier'. This can also be seen in the following definition:

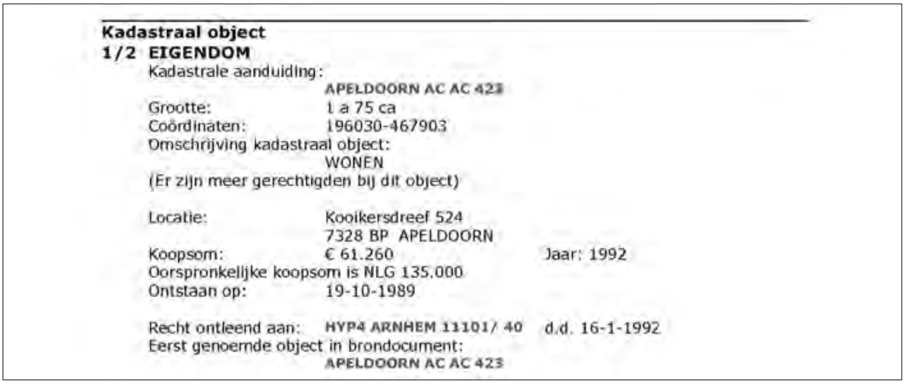
Figure 1. Excerpt from a modern parcel based cadastral map (boundaries fat line; buildings thin line).

"[a] cadastre is a methodically arranged public inventory of data concerning properties within a certain country or district, based on a survey of their boundaries. Such properties are systematically identified by means of some separate designation. The outlines or boundaries of the property and the parcel identifier are normally shown on large scale maps which, together with registers, may show for each separate property the nature, size, value and legal rights associated with the parcel. It gives an answer to the questions 'where' and 'how much'." (Henssen and Williamson 1990: 20).

It is often mentioned that the roots of cadastres have to be found with the taxation of real properties (e.g. Larsson 1991: p. 21; Simpson 1976: p. 111). Without wanting to dismiss the numerous (small scale) activities that had already taken place before, the major development in introducing cadastres (with maps) took place in the early 19th century. In 1807 Napoleon I, Emperor of France, instituted the cadastre in France and all the areas which at that time were under his rule (most of the South and West of continental Europe). In 1817 Francis I, Emperor of Austria, introduced a much improved cadastre for the whole Austrian-Hungarian Empire, which at that time covered most of Central Europe.

It is again often mentioned that initially the introduction of the (fiscal) cadastre did not influence the existing (juridical) land registration systems much, but that it became increasingly desirable to use the cadastral maps, which were compiled through systematic land survey, for identification of real properties in the land registration process (e.g. Simpson 1976: p. 122; Larsson 1991: p. 24). This later development in which the cadastre fulfils both a fiscal and juridical role had always been the intention of Napoleon I (see Zevenbergen 2002: p. 28). But in most countries the taxation side got all the attention during implementation, and the supporting role to the civil code was lost or remained underdeveloped.

Nowadays practically all countries which have both a cadastre and a land registry identify the property in the latter by its description in the cadastre, unless of course the cadastre was not complete (as was the case in Spain, Portugal and Latin America, where the land registry often missed a unique identification and is practically independent from the



```
Kadastraal object
1/2  EIGENDOM
     Kadastrale aanduiding:
                          APELDOORN AC AC 423
     Grootte:             1 a 75 ca
     Coördinaten:         196030-467903
     Omschrijving kadastraal object:
                          WONEN
     (Er zijn meer gerechtigden bij dit object)

     Locatie:             Kooikersdreef 524
                          7328 BP  APELDOORN
     Koopsom:             € 61.260                Jaar: 1992
     Oorspronkelijke koopsom is NLG 135.000
     Ontstaan op:         19-10-1989

     Recht ontleend aan:  HYP4 ARNHEM 11101/ 40   d.d. 16-1-1992
     Eerst genoemde object in brondocument:
                          APELDOORN AC AC 423
```

Figure 2. Excerpt from a modern record of interests (note the linking with the cadastral map in Figure 1 by the parcel number, highlighted in grey).

cadastre). This use of cadastral identification in land registration has been both used to enhance deeds registration and to facilitate the change from a deeds to a title registration system (Larsson 1991: pp. 25-26). On the other hand the cadastre can be kept much more up-to-date when the information on land transactions through land registration is made readily available (Zevenbergen 2002: p. 29). In for instance England, a cadastre in the above described sense was never set up, and the (pre)existing topographic large scale maps are used to identify and depict the land that is registered in the in the land register.

Therefore it is essential to consider land registration and cadastre together. They should at least cooperate and work closely together, something which is unfortunately not the case in many countries. Experts expressed that "there is a strong need to integrate and rationalize land title registry and cadastral systems" (UN 1996: p. 28), but very often historically grown situations and the vested power structures based on those prevent the merger of the two organizations involved, although several Western European countries made this transition recently (e.g. Belgium, Norway, Sweden, and Finland).

2.2 Land administration

Regardless whether land registration and cadastre are arranged in two organisations or not, the term land administration is used to indicate their close relation. It could be defined as follows:

"Land administration is the operational component of land tenure; land administration provides the mechanisms for allocating and enforcing rights and restrictions concerning land. Land administrative functions include regulating land development and use, gathering revenue from the land (through sale, leasing, and taxation), controlling land transactions, and providing information about the land. These functions are accomplished, in part, through the development of specific systems responsible for boundary delimitation and spatial organization of settlements, land registration, land valuation, and information management activities." (McLaughlin and Nichols 1989: p. 79).

Land administration can also be described as "the process whereby land and information about land may be efficiently managed" (UN ECE 1996). It includes the provision of information identifying those people who have interests in real estate; information about those interests e.g. nature and duration of rights, restrictions and responsibilities; information about the parcel, e.g. location, size, improvements, and value.

The goals attributed to land administration by Van der Molen are: a) improving land tenure security, b) regulating the land market, c) urban and rural land-use planning, and d) the taxation of land (Van der Molen 2001: pp. 4-5). They include both the legal security for the owner and purchaser, and are also focused on government and society at large. The information needed for the legal security of owner and purchaser forms an important subset of all the land information that can be found in land information systems (LIS). Both issues are important and interrelated, but Pryer (1993: p. 64) stresses that there is a wide gulf between those who see land registration as primarily for the benefit of the landowner and those who see it as an instrument of state control. The differences relate to how the processes (especially updating) are arranged, which data elements are included, and the balance between the different professionals involved (lawyers, surveyors, planners, economists, ...). Thus cadastral data looks different in different countries, but different professionals also *read* the cadastral data differently depending on their needs and expertise.

2.3 Cadastral parcels

Land as the object of property rights is different from most other types of property. In many cases there is not a 'logical' object. The object has to be defined, has to be legally constructed, and can change relatively easily (see Zevenbergen 2002: pp. 38-41; pp. 67-70). The objects are separated by boundaries which define where one landowner's territory ends and the next begins. Even though there often is a reasonably high congruity between topographic boundary features and legal extent, there is no necessary identity between the topography of a parcel and the legal extent of that parcel. The extent and boundaries of land parcels are a matter of legal definition. (Burdon 1998: p. 152) Parcels and boundaries are abstract concepts. This makes a very large difference with most other types of geo-information, which depict 'real life' geographical phenomena (compare Van der Molen 2001: p. 15). Title plans and parcel maps are legal documents in a graphical form and not just another dataset for a geographical information system (GIS) (Burdon 1998: p. 154). Similarly the rights in land are abstract concepts. These abstract concepts are very important to society, and instead of seeing them contrary to reality, they can be described as 'institutional reality'. This includes institutional facts which exist only by human agreement and are observer relative, as opposed to brute physical facts which exist in external reality independent of human observers and human intentions. More on these concepts, derived from J. Searle's 1995 book 'The Construction of Social Reality', and their application to cadastres can be found with (Smith et al 2008) and (Bittner et al 2000).

The traditional depiction of a parcel is on a two dimensional map, but of course the related rights and usage are three-dimensional. A conceptual description of the object is a prismatic volume from the centre of the earth into the sky, although in most countries specific legislation limits the power of the right holder up and downwards. The 2D representation of the object is the line where this volume intersects the surface of the earth. Such a 2D representation becomes too limited as soon as rights relate to objects that are above or below one another. The most common case of this is an apartment unit in a large complex. The efficient use of inner-city space has in recent times created more and more complex constructions with diverse usage and possessors. To posses a part of such a construction under the strongest rights, 3D property descriptions are needed, and where they are in use, the land administration system should be able to depict these in an appropriate way. Several approaches for a 3D cadastre to deal with this can be found in e.g. (Stoter 2004) and (Stoter and Van Oosterom 2006). Comparable issues relate to time-sharing of the same object. This is best known for holiday estates, but can also be thought of for seasonal agriculture with more than one crop per year. Work on a 4D cadastre to fully integrate this has started (Van Oosterom et al 2006; Doner et al 2008).

In the earlier given descriptions of land rights and parcels as abstract concepts, one could already see references to reality on the ground. The abstract concepts of land rights and boundaries find their most important day-to-day application in regulating use patterns. And use is a very real, and often very visible phenomenon. Where one person stops to use land, and another person starts to use it, people tend to erect physical features. Those can be purely practical (keeping domestic animals and/or children in, wild animals and strangers out, or blocking the view) or intended to mark 'the boundary'. In both cases these are 'real life' translations of the abstract 'boundary'. Physical features are not infinitesimally thin and semi-permanent at best. Walls, ditches, and hedges are rather thick. Fences fall down and are accidentally or deliberately erected in a (slightly) different position and hedges sometimes grow more in one direction than in the other. Deliberately placed boundary markers are usually rooted rather deep and of durable materials (long iron poles, concrete monuments). But they still can be displaced or removed accidentally or deliberately.

In all cases when a boundary 'alert' is visible in the terrain, it is often taken at face value, even if it is no longer in the original position. This even applies in most instances where the original position has been 'registered' by means of surveying and/or mapping techniques and is part of the cadastre; these days usually in the form of a database.

3. A reference model for cadastral data

The actual implementations of cadastral information systems in different countries are highly varied. Work on designing an (international) data reference model, that captures the core of cadastral information, has been ongoing since the FIG Congress in 2002 (Van Oosterom and Lemmen 2002). First under the name Core Cadastral Domain Model (CCDM), and more recently as Land Administration Domain Model (LADM), which has been submitted in February 2008 by FIG to the International Organization for Standardization (ISO) (ISO/TC211 2008).

3.1 Motivation for a reference model

Land administration is a large field; the focus of this standardization is on that part of land administration that is mainly connected to land (or water) and property ownership, and the geometrical (spatial) components thereof. The LADM provides a *reference model* which will serve at least two important goals: (1) to avoid reinventing and re-implementing the same functionality over and over again, but rather to provide an extensible basis for the development and refinement of efficient and effective land administration systems, based on a Model Driven Architecture (MDA), and (2) to enable involved parties, both within one country and between different countries, to communicate, based on the shared vocabulary (that is, an *ontology*) implied by the model. The second goal is important for creating standardized information services in an international context, where land administration domain *semantics* have to be shared between regions, or countries, in order to enable necessary translations. Important considerations during the design of the model were: it should cover the common aspects of land administration all over the world; it should be based on the conceptual framework of Cadastre 2014 (Kaufmann and Steudler 1998); it should follow ISO TC 211 standards; and, at the same time, the model should be as simple as possible, in order to be useful in practice.

3.2 A reference model and its implied functionality

Until now, most countries (or states, or provinces) have developed their own land administration system. One country operates deeds registration, another title registration. Some systems are centralized, and others decentralized. Some systems are based on a general boundaries approach, others on fixed boundaries. Some systems have a fiscal background, others a legal one. However, the separate implementation and maintenance of land administration systems is not cheap, especially if one considers the ever-changing requirements. Also, the different implementations (foundations) of the various land administration systems do not make meaningful communication across borders easy. Looking from a distance, one can observe that the different systems are in principle largely the same: they are all based on the relationships between people and land and property, linked by (property) rights, and are in most countries influenced by developments in Information and Communication Technology (ICT). Furthermore, the two main functions of every land administration and land registry are: (1) keeping the contents of these relationships up-to-date (based on legal and related transactions); and (2) providing information from the register.

3.3 A reference model including land tenure

The UN Land Administration Guidelines (UN ECE 2006) describe land administration as the 'process of determining, recording and disseminating information on ownership, value and use of land when implementing land management policies'. If *ownership* is understood as the mechanism through which rights to land are held, we can also speak about *land tenure*. A main characteristic of land tenure is that it reflects a social relationship regarding rights to land, which means that in a certain jurisdiction the relationship between people and land is recognised as a legally valid one (either formal or non-formal). These recognised rights are in principle eligible for registration, with the purpose being to assign a certain legal meaning to the registered right (e.g. a title). Therefore, land administration systems are not just 'handling geographic information', as they represent a lawfully meaningful relationship amongst people, and between people and land.

3.4 The strategic importance of ICT

As land administration activity on the one hand deals with huge amounts of data, which moreover are of a very dynamic nature, and on the other hand requires a continuous maintenance process, then the role of ICT is of strategic importance. Without the availability of information systems it will be difficult to guarantee good performance with respect to meeting changing customer demands. Organizations are now increasingly confronted with rapid developments in technology, a *technology push* (internet, spatial data bases, modelling standards, open systems, GIS), as well with a growing demand for new services, a *market pull* (e-governance, sustainable development, electronic conveyance, integration of public data and systems). Modelling is a basic tool facilitating appropriate system development and reengineering and, in addition, it forms the basis for meaningful communication between different (parts of the) systems.

Standardization has become a well-known process in the work of land administrations and land registries. In both paper-based systems and computerized systems, standards are required to identify objects, transactions, relationships between objects (e.g. parcels, more generally spatial units) and persons (e.g. subjects, more generally parties), classification of land use, land value, map representations of objects, and so on. Computerized systems require further standardization, when topology and the identification of single boundaries are introduced (Van Oosterom and Lemmen 2001). In existing land administrations and land registries, standardization is generally limited to the region, or jurisdiction, where the land administration or land registry is in operation. Open markets, globalization, and effective and efficient development and maintenance of flexible (generic) systems, require further standardization.

3.5 The scope of the reference model

The reference model defines a LADM covering all basic information-related components of Land Administration (including those over water as well as land, and elements above and below the surface). The LADM provides:

- an abstract, conceptual schema with five basic packages related to (1) parties (people and organizations); (2) spatial units (parcels); (3) rights, responsibilities, and restrictions (property rights); (4) spatial sources (surveying); and (5) spatial representations (geometry and topology);
- a terminology for land administration, based on various national and international systems, that is as simple as possible in order to be useful in practice;
- the terminology allows a shared description of different formal or informal practices and procedures in various jurisdictions, and
- a basis for national and regional profiles.

The LADM enables the combining of land administration information from different sources in a coherent manner.

The interference with (national) land administration laws, that might have any legal implications, is outside the scope of the LADM, as well as the construction of external databases with person data, address data, valuation data, land use data, land cover data, and taxation data. However, the LADM provides blueprint stereotype classes, which indicate what data LADM expects from these external sources, when available.

3.6 The core classes of LADM

Figure 3 shows the core LADM as a UML 2.1 class diagram. The core LADM is based on *four* classes:

1. Class LA_Party. An instance of LA_Party is a *party*: a person, or group of persons, that compose an identifiable single entity. It is associated to zero or more (0..*) instances of a subclass of LA_RRR.
2. Class LA_RRR. An instance of a subclass of LA_RRR is a *right*, *restriction*, or *responsibility*. It is associated to zero or one (0..1) instances of LA_Party, and to exactly one (1) instance of LA_LAUnit.
3. Class LA_LAUnit. An instance of LA_LAUnit, *launit*, concerns the administrative information of spatial units (see ad. 4). It is associated to one or more (1..*) instances of a subclass of LA_RRR, and to zero or more (0..*) instances of LA_SpatialUnit.
4. Class LA_SpatialUnit. An instance of LA_SpatialUnit is a *spatial unit*: a single area of land or, more specifically, a volume of space, under a homogeneous and unique right. It is associated to zero or more (0..*) instances of LA_LAUnit.

LA_Party is associated with LA_LAUnit, which means that a party might be a launit.

3.7 The packages of LADM

LADM contains five packages. This facilitates the maintenance of different datasets by different organizations. The complete model may be therefore implemented through a distributed set of (geo-)information systems, each supporting data maintenance activities and the provision of elements of the model. The model may also be implemented by one or more maintenance organizations operating at national, regional or local level. This underlines the relevance of the model: different organizations have their own responsibilities in data maintenance and supply, but may communicate on the basis of standardized administrative and technical update processes. For an overview of the packages see Figure 4.

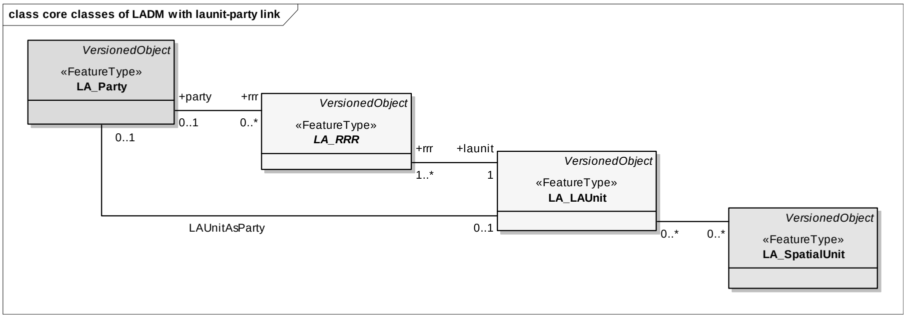


Figure 3. The core classes of LADM.

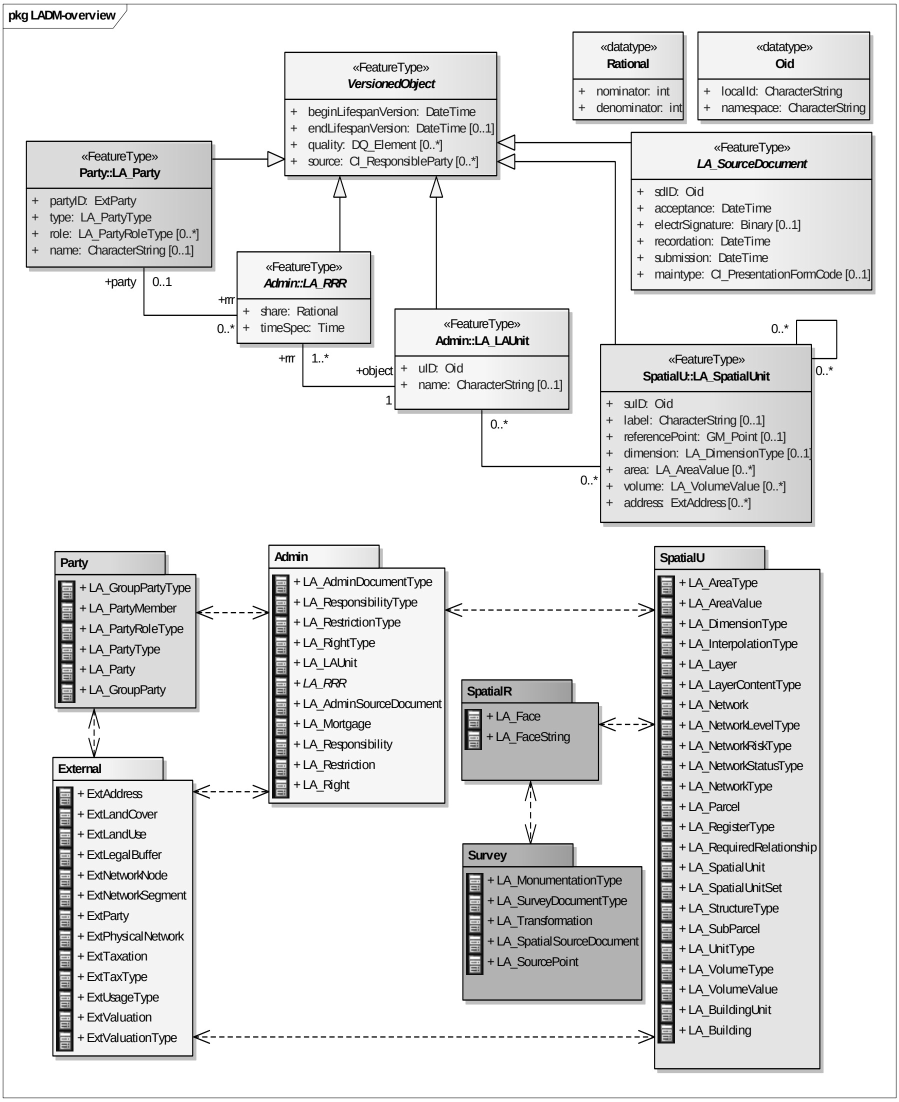


Figure 4. The five packages of LADM.

4. The social tenure domain model

The Social Tenure Domain Model (STDM) is a pro-poor land administration tool. It covers land administration in a broad sense including administrative and spatial components. Traditional (or conventional) land administration systems relate names (or addresses) of persons to land parcels via rights. In the STDM, an alternative option for this is to relate personal identifiers such as fingerprints to a coordinate point inside the land in use by that person, via a social tenure relationship. Depending on the local conditions, there can be a variety of social tenure relationship types and other rights. The STDM thus provides an extensible basis for efficient and effective system of land rights recording. The STDM is to be seen as equivalent to the Land Administration Domain Model, with partly its own terminology.

4.1 People – land relationships

The STDM describes the relationship between people and land whereby it strives to record all forms of land rights, social tenure relationships and overlapping claims or rights over land. STDM is designed to support land rights recording in areas where regular or formal registration of land rights is not the rule. That is, STDM makes it possible to record rights, which are not necessary registered rights, nor registerable, as well as claims, that need to be adjudicated both in terms of the 'who', the 'where' and the 'what' type. The focus is on recorded rights (or social tenure relationships) and not only registered rights. This means recording personal land use rights and not only real rights – this implies that real rights are included. STDM handles the impreciseness and possible ambiguity of the description of the rights, both in terms of 'who', 'what' and 'where'. STDM, therefore, records not only registered, but also the range of rights in the continuum simultaneously; e.g. there can be, apart from formal rights: non-formal and informal rights, customary types, indigenous rights, tenancy, possession. Financially, STDM records options such as group loan and micro credit.

Similarly, STDM records the types of person ('who', e.g.: a group with non-defined membership, a group of groups, natural persons, companies, municipalities, co-operatives, married couples, ministries, etc.). STDM also records a range of spatial units ('where', e.g. a piece of land which can be represented as a single point – inside a polygon, one point – street axes, a set of lines, as a polygon with low or high accuracy coordinates, as a 3D volume, etc.).

4.2 Data acquisition

The type of approach in data acquisition can vary from one area to another – both for collection of spatial or administrative data. For example in slum areas it may be sufficient as a start to relate informal people-land relationships to a single point. Then attributes such as photographs and fingerprints can be attached to the records. In a business centre a traditional cadastral map or register may be required – where in residential areas a map based on satellite images combined with formal descriptions may be suitable. There could be an overlap with areas with customary traditions. Satellite images are a very promising approach for data collection. A large-scale plot of such images can be used to identify the land use types by the people themselves. The World Bank funded a pilot in Ethiopia as a proof of concept. The results are encouraging. There are similar experiences from other countries, e.g. Rwanda (Sagashya and English 2009).

4.3 Prototype

The next logical step is the software development, starting with a prototype and experimentation with such software in a pilot project in a country which has slums, customary communities and overlapping land tenures and non polygon rights and claims. With the support of the Global Land Tool Network (GLTN), a prototype is under development at the International Institute for Geo-Information Science and Earth Observation (ITC, the Netherlands), in close co-operation with UN-HABITAT and FIG.

5. Cadastre data as a key register

In the Dutch context the cadastral information has been legally mandated as one of the key registers in a new law, which became effective at the start of 2008. The consequences of a full implementation of this notion of key register for cadastral data will be far reaching when fully applied. In this section this situation will be discussed, with reference to the earlier situation and the (continuing) situation in the private law sphere.

5.1 Key registers

Comparable to the list of core layers of an SDI as mentioned in the introduction, the Dutch administration has identified a number of base data sets needed to operate the public administration, which include personal data, data concerning immovable or movable property and similar data, which are essential for the public sector to function properly. The list is still occasionally expanded, and includes several geo-datasets, among those the cadastral data.

The idea is that the base data will be collected just once from individuals or businesses, and then be mandatory (re-)used throughout the public administration. Base data will be recorded in key registers, of which a (large) part of the data elements has been indicated as 'authentic' data. Key registers must meet certain criteria:

- Registration is regulated by law.
- The clients have an obligation to report mistakes and all tiers of government have an obligation to the data from the key register.
- There must be clear lines of accountability. The costs of realisation and operation must be within reason and unambiguously allocated.
- There must be transparency about the scope and content of the registers and firm agreements and procedures between the registrar and the clients.
- The procedures for accessing the key registers must be unequivocal and there must be a strict regime of quality control.
- Fixed procedures must be defined for the obligatory involvement of clients in the decision-making.

The position of a key register within the overall registration system and the connections with other key registers must be clearly defined. Authority over the key register must lie with a government agency and one minister will be responsible for realisation and operation. (Besemer et al 2006).

5.2 Act key registers cadastre and topography

The 'Act key registers cadastre and topography' passed the parliament early 2007, and came (largely) into effect on 1 January 2008. It is not a self standing act, but it only contains changes to other acts, most importantly to the Cadastre Act. That act already regulated the Dutch cadastral system (including the land registry) since 1992. In many respects

the Dutch cadastral system could already be seen as fulfilling the role and meeting the requirements of a key register. In this paper we will not discuss the second key register on topography, which is also legally introduced through the same Act.
The most noticeable change brought by the 2007 Act is the change in terminology. The new term 'key register cadastre' replaces the term 'cadastral registration', and includes both the administrative registration, as well as the national cadastral map, both in the form of a digital database.

A second change is not so much the notion of mandatory use, but its expansion to all public activities. Before 2008 a number of specific provisions in public land management acts already assumed the data from the cadastral registration as a priori correct (and decisions would be based on the data, including the name of the right holder to be addressed, although counterclaims could be made that did not hamper the procedures already started). It includes the Expropriation Act, the legislation relating to the real estate tax and the former procedure for voting among farmers on execution of a land consolidation project. (Van Rossem 2006). Interestingly the spatial planning legislation has never prescribed the use of the cadastral map in preparing a binding spatial plan, and still does not do so (digital access will soon become mandatory).

To underpin this mandatory use and assumed correctness, the data should be as reliable as possible. This means that if any of the mandatory users is confronted with clear indications that some data element might not be correct, this user will have to report this back to the dataset holder (i.c. the cadastral agency). The dataset holder has three options to react to this directly after the report has reached him. He can a) correct the data element within one day, b) reject the reported doubt within one day, or c) connect a message to the data element saying 'under review'. In the last case of course that review should be undertaken, and in due time lead to correction or rejection as well. During the time a data element is 'under review' it is not mandatory to use it, which is also allowed in case the use of the data would mean that the user would not be able to adequately fulfil his public duty. (art 7k, par. 2c Cadastre Law).

It is clearly stated in the Act that the legal effect of the key register in relation to the recorded deeds in the land registry will remain the same. The 'new' effect of becoming a key register is limited to the public administration, and also only applies to the data elements that have been declared authentic (and are marked as such in the key register). The data elements relating to mortgages for instance have not been declared authentic, since they are of little importance in the public administration. A detailed overview can be found in the product catalogue of the key register cadastre.

5.3 Position of the cadastral registration in land conveyancing

Within the Dutch system of land administration the cadastre and the juridical land registration have operated closely together since the mid 19th century. Nevertheless from a legal-dogmatic point of view, land registration operates as a 'registration of deeds'. Deeds are documents describing the details of a transfer or the establishment of a property right, and since the mid 20th century these documents need to be in the form of notarial deeds (safe a few exceptions). The legal moment of transfer of the property right takes place at the moment of offering the deed for registration. The registrar is legally obliged to accept the deed when it meets a few formal requirements, and can only give a warning if the deed is not concurrent with the pre-registered situation. Such a case is very rare, since the administering notary is under the obligation to check (and where necessary explain or repair) this. Furthermore the purchaser on the next transfer, or in reality the notary administering the next transfer, will check the pervious transfer and any other registered information since. In addition to all of this, the registrar will 'summarize' the deed by updating the cadastral registration. The automated cadastral registration (AKR), as its earlier paper predecessors, functions (legally) as an index to the underlying recorded notarial deeds. Nevertheless the 'summary' of the contents of the deed given in AKR is of high importance, certainly in practice.

Legally the title of a sold property changes hands the moment the notarial deed is offered for recordation (rather unusual in other countries operating registration of deeds). If such a deed transfers a property which differs from an existing parcel, the Cadastre will – after recordation – arrange for its surveyors to inquire about the boundary, to survey the boundary, to update the cadastral map and to replace temporary, administrative (sub) numbers with new (full) cadastral numbers.

The causal doctrine is applied in the Netherlands, which means that problems in the 'title' are not repaired by recordation, although bona fide third parties can to a large extent rely on what is and is not present in the public registers.

The notaries not only legalize the deeds, they also check to see if all prerequisites for the intended transfer are in order. They will look at the underlying sales contract with a legal eye, they will check the cadastral register and the previous deed, and often several other registers as well (e.g. the public registers, the company register and the marriage register) and all the money will go through them. They have an 'active care duty' in all of this, and are liable for mistakes (they have mandatory indemnity insurance). In 1995 the number of notaries was still fixed by the government at approximately 1250, although many of them have several highly qualified staff (including candidate notaries who after a year can replace the notary in his or her absence). The cooperation between Cadastre and notaries is very good, and usually an occasional mishap by one is (formally or informally) reported by the other, and quickly solved.

The daily practice seems to surpass what one might expect from the 'law in books'. Several of the (theoretical) 'negative' aspects of the system, are rarely – if ever – experienced in practice. (Zevenbergen 2003) Occasionally the description in the deed of a subdivision is rather vague, causing problems during the inquiry and survey in the field. This can be aggravated in cases where this field visit takes place long after the deed was signed (although the severe backlogs in several offices have been solved). But even in such cases notary and Cadastre usually manage to find a workable solution. (Zevenbergen 2002: pp. 135-136).

5.4 Consequences of having the key register cadastre

For the public administration the cadastral data, as far as authentic, should be used, and if it turns out to be incorrect or incomplete, this has to be reported back to the Cadastral Agency. One can wonder if the quality of especially the cadastral map in non-renovated old urban areas is up to this challenge. This is especially interesting since the Cadastral Agency has always claimed that the cadastral maps is merely an index for orientation, and should not be used to measure in or reconstruct boundaries from (this should be based on the underlying field notes).

There are some cases known where the parcellation on the map is nearly half a house off in housing rows as derived from recent areal imagery (see Figure 5). Following the act to the letter would mean that the governmental office noticing this on the one hand would need to refer to two parcel numbers for a decision effecting one house, and on the other hand would need to inform the Cadastral Agency of this situation. This does not appear to be very practical in the short run (see also Van der Meer 2006). The exception on mandatory use when it would harm the public duty will at least partly solve this problem.

It is clearly stated that the legal effect of the key register cadastre in relation to the recorded deeds in the land registry will remain the same. The 'new' effect of becoming a key register is limited to the public administration, and also only applies to the data elements that have been declared authentic (and are marked as such in the key register). The data elements relating to mortgages for instance have not been declared authentic, since they are of little importance in the public administration.

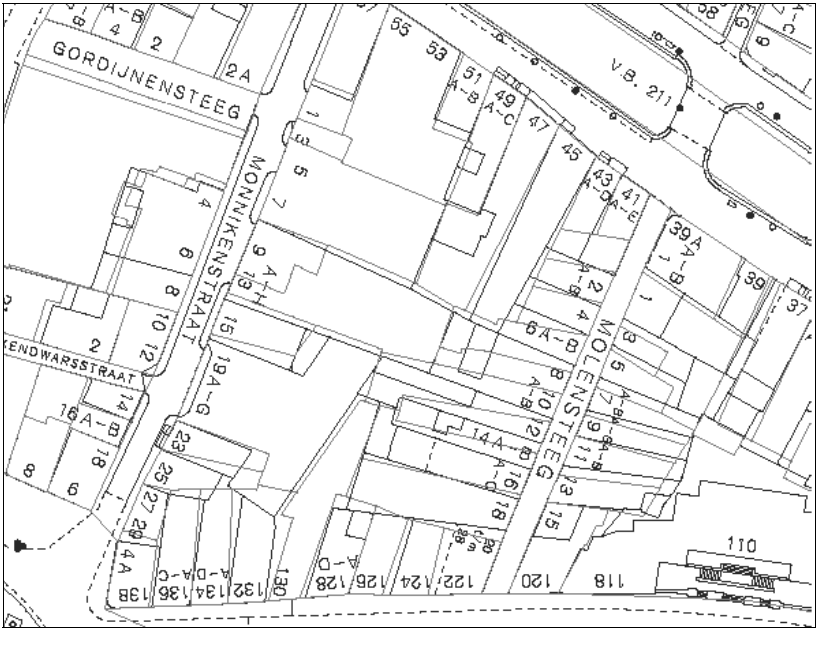


Figure 5. The difference between the cadastral map (black) and recent aerial imagery (grey).

Regardless of this official notion, in the future there will also be an effect in the practice within the private law domain. The 1992 legislation and the nearly parallel computerization of the cadastral registration have also influenced the practice within the private law domain, well beyond the legal provisions. Possible ways to formalize these developments (by changes in the legislation or the official statements made by the Cadastral Agency with regard to the information), have been the topic of both independent research for the Cadastral Agency (Huijgen et al 2006) and subsequent policy considerations (see Klaasse and Louwman 2007). But for the time being the (full) effects in practice will slowly show themselves.

6. Final remarks

It is clear that cadastral information is useable for many applications, on its own, or as part of a combination with other (geo-)data. However, the fact that it is data on socio-economic realities, limits it usability to be the base for many physical phenomena, and also has consequences for the way the data has to be collected. On the other hand cadastral data can be easily combined with data describing physical phenomena in all kind of (spatial) analysis. Together all these datasets, when organized in a comprehensive way, make up a geo information infrastructure or SDI:

- Increased geo-technologies are enhancing the possibilities for such integrated analysis. The more standardized the underlying domain models are, the more generic analytical tools and models will become available.
- But also for the (historical) primary function(s) of cadastre, quality improvements are possible through geo-technologies, data modelling and international comparison, assuming the institutional context (organizational structure, legal framework, business model) allow both for a strong foundation as well as for flexibility to profit from these developments.
- As the title of this paper already said, cadastral information is *more* than base data. On the one hand, it is the core data within a land administration system, serving both the land market and the implementation of land policy. On the other hand, it is one of the base datasets, focussing on 'human-centric' (socio-economic) phenomena, which can be combined with other geo-data for an enhanced spatial understanding.

References

Besemer, J., P. Laarakker, L. Murre, R. ten Kroode and N. Zuiderveen Borgesius (2006), Basic Registers for Geo-Information, FIG Conference, Munich 2006.

Bittner, Steffen; Annette von Wolff, and Andrew U. Frank. (2000). The Strucutre of Reality in a Cadaster, In: Proceedings of 23rd International Wittgenstein Symposium, Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society, p. 88-96.

Burdon, Ian (1998). Automated Registration of Title to Land, Registers of Scotland.

Dale, Peter F. and McLaughlin, John D (1988). Dale, Peter F. and McLaughlin, John D. (1988). Land Information Management, An introduction with special reference to cadastral problems in Third World countries, Oxford: Clarendon Press.

Doner, F, R. Thompson, J. Stoter, C. Lemmen, H. Ploeger, P. van Oosterom (2008). 4D Land Administration Solutions in the Context of the Spatial Information Infrastructure, FIG Working Week 2008 'Integrating Generations', Stockholm, Sweden 14-19 June 2008.

FIG (1995). International Federation of Surveyors (FIG), The Statement on the Cadastre, FIG Publication No. 11, Canberra.

Henssen, J.L.G. and Williamson, I.P. (1990). Land registration, cadastre and its interaction; a world perspective, Proceedings XIX FIG Congress, Commission 7, Paper 701.1, Helsinki 1990, p. 14-43.

Huijgen, W.G., J. de Jong, H.D. Ploeger, J.A. Zevenbergen (2006). Eindrapport 'Vergroting juridische betrouwbaarheid Kadaster', TU Delft / OTB, 11 December 2006 (64 pp).

ISO/TC 211 (2008). New work item proposal, Geographic information - Land administration domain model (LADM). Lysaker, Norway, ISO: 27 pp.

Kaufmann, J. and D. Steudler (1998). Cadastre 2014. A vision for a future cadastral system. FIG-Commission 7, Brighton, U.K.

Klaasse, Z. and W. Louwman (2007). Vergroting juridische betrouwbaarheid Kadaster, WPNR 6728, p. 885-889.

Larsson, Gerard (1991). Land Registration and Cadastral Systems: tools for land information and management, Harlow (Essex): Longman Scientific and Technical.

McLaughlin, J.D. and Nichols, S.E. (1989). Resource Management: The Land Administration and Cadastral Systems Component, In: Surveying and Mapping, No. 2, p. 77-86.

Onsrud, H.J. (1998). Survey of national and regional spatial data infrastructures activities around the globe. Paper presented at the Third Global Spatial Data Infrastructure Conference, Canberra, Australia.

Pryer, E. John (1993). Computer Application of Land Registration, In: Computer application to the cadastre and land registration in the Near East, Rome: FAO, p. 59-101.

Sagashya, D. and C. English (2009). Establishing a land administration system: Technical and Economic Analysis for Rwanda. In: Proceedings FIG - World Bank Conference: Land governance in support of the Millennium development goals, responding to new challenges, 9¬10 March, 2009, Washington, D.C., USA. http://www.fig.net/pub/fig_wb_2009/papers/sys/sys_2_english_sagashya.pdf.

Simpson, S. Rowton (1976). Land Law and Registration (book 1), London: Surveyors Publications.

Smith, Barry, David M. Mark, and Isaac Ehrlich (Eds.) (2008). The Mystery of Capital and the Construction of Social Reality, Open Court.

Stoter, J. (2004). 3D cadastre , PhD thesis Delft: NCG 2004.

Stoter, J., P. van Oosterom (2006). 3D Cadastre in an International Context : legal, organizational, and technological aspects, Boca Raton: CRC Press, 2006.

UN (1996). United Nations, Report of the United Nations Interregional Meeting of Experts on the Cadastre, Bogor.

UN ECE (2006), United Nations. Economic Commission for Europe. Land Administration Guidelines, Geneva, Switzerland, http://www.unece.org/hlm/wpla/publications/laguidelines.html (accessed 11 December 2008)

Van der Meer, A.J. (2006). Kanttekeningen bij het wetsvoorstel basisregistraties kadaster en topografie, Geo-Info 2006-6/7, p. 324-325.

Van der Molen, Paul (2001). Cadastres Revisited: The promised land of land administration, Inaugural Address ITC, Enschede.

Van Oosterom, P.J.M. and C.H.J. Lemmen, 2001, 'Spatial Data Management on a very large cadastral database', in: 'Computers, Environment and Urban Systems', Theme Issue 'Cadastral Systems', p. 509-528, Volume 25, number 4-5, 2001, Elsevier Science, New York.

Van Oosterom, P.J.M. and C.H.J. Lemmen (2002). Impact analysis of recent Geo-ICT developments on cadastral systems. FIG XXII International Congress, Washington, D.C., FIG.

Van Oosterom, P.J.M., Ploeger, H.D., Stoter, J.E., Thompson, R. and Lemmen, C.H.J. (2006) Aspects of a 4D cadastre: a first exploration. In: FIG 2006 : Proceedings of the conference : Shaping the change, XXXIII FIG congress, Munich, Germany, 8-13 October 2006. Frederiksberg: International Federation of Surveyors (FIG), 2006. ISBN 87-90907-52-3.23 pp.

Van Rossem, R.P.E. (2006). Basisregistraties Kadaster en Topografie, Geo-Info 2006-5, p. 200-205.

Zevenbergen, J.A. (2002). Systems of Land Registration – Aspects and Effects, PhD thesis Delft: NCG 2002.

Zevenbergen, Jaap (2003). The Dutch system of land registration, title registration with a 'privatized' registrar, EULIS Seminar on Conveyancing Practices - 15-16 May 2003, Lund (S) (15 pp).

Aspects concerning the use of spatial core data

Ir. R.J.G.A. Kroon
Ingenieursbureau Geodelta B.V., the Netherlands

Not so long ago the geo-information world was a static orderly world. Geographic data were collected for one specific application purpose, e.g. the production of a cadastral map. The final product was this map. The collected source data and any possible half-finished products were considered to be means of production without any further value.
Nowadays the geo-information field is a dynamic world in which many suppliers construct products for and offer services to all kind of users, ranging from professionals to consumers. Being source data or a half-finished product for one users group can be a final product for another group. This contribution presents ideas and suggestions how the present policy objective 'single collection, multiple use' can also be achieved for geo data.

1. Raw source data, raw core data and interpreted core data (information)

The concept 'raw source data' has been defined as collected geographical data that have undergone none of almost none processing. So it concerns data that has been prepared as product suitable for one certain application. To put it differently: the data has not been interpreted yet. Examples that can be considered as raw source data are aerial photographs and laser altimetry data. An example of what should not be considered raw source data is an orthophoto image. Such an image is produced from an aerial photograph by adding additional information; in this case a height model and the absolute stand and position parameters of the aerial photograph, expressed in relation to one or another co-ordinate system. Herewith the recording is processed into a recording which fits with a certain accuracy in the geometry of the topographic map. How well this conversion should be done depends to a great extent of the application for which the orthophoto images will be used.

The concept of 'raw core data' is defined as being raw source data of which additional information is known explicitly. It concerns the following information:

- Quality characteristics: Beforehand the raw core data should have a defined minimal technical quality;
- Actuality characteristics: Beforehand the raw core data should have a defined temporal actuality;
- Meta datasets: Beforehand the raw core data should have a defined meta-dataset.

Having available explicitly this additional information offers quite a number of advantages. For example: in this way it will be possible to make a well-founded consideration if, and if so how raw core data can be used for the own application.

In short:

Raw source data = Non-interpreted data;
Raw core data = Non-interpreted data + additional information.

Characteristic for raw core data is that we are dealing with data which are at the source of a large number of products and services. For this reason the device is 'single collection – multiple use' holds especially for collecting raw core data.

The information present in the raw core data can be interpreted. This results in topographic information (interpreted core data). An example of topographic information is vector information. Vector information exists in many shapes and sizes. Characteristic is that they are obtained via interpretation of the topography in the raw core data or via direct interpretation of topographic details in the terrain. The way of interpreting depends on the context for which the vector information is being made up.

A special category of vector information is the core information (interpreted core data). This interpreted set contains such a collection of information that it has for many users the status of an information frame to which they can link their relevant information. Even if it is a matter of interpreted data (that is, information), this core information plays more or less the role as raw core data, namely being a staring point for extracting or adding information for the purpose of an application of one's own. For core information (interpreted data) the same requirements regarding quality, actuality and metadata apply.

Figure 1. The decline of the glacier.

Figure 2. The results of core data not meeting the required quality criteria.

2, Incorrect core data

If core data are not collected at a regular basis, this influences the quality and actuality of the topographic infrastructure. It will gradually degenerate. It is a latent process. It can take years before it becomes clear that the described 'reality' in the topographic data does not match adequately with the actual reality any more. The process is comparable with the life cycle of a glacier. The glacier needs to be fed continuously with snow on the top. If this does not happen, first of all the glacier will slowly change as far as shape and size are concerned. A permanent lack of feeding will result in an increasingly faster decline of the condition of the glacier. Due to the present global warming of the earth this becomes more and more visible.

The consequences of an inadequate topographic infrastructure are big. In case core data not meet the required quality criteria or are not collected at regular intervals this can lead easily to wrong decisions. Examples are unreliable spatial analyses on the basis of non-complete geo-information, wrong shortest routes to accident situations, wrongly marked borders and errors in the positioning of civil technical engineering works (under construction).

Just like with the glacier the errors in core data will become visible after many years. For this reason it is vital that a continuous investment in collecting and making available good core data takes place.

3. When is core data core data?

Core data can only remain core data if:

- there exists a frequent supply of new data so that the core data preserve a well-defined quality and actuality;
- the quality of the data is guaranteed;
- core data are so-called frame data, also called 'greatest common factor' data.

The last point refers to the fact that core data are not produced as a matter-of-course for one organisation and for one application. Core data are data that can be deployed by a broad as possible group of users as starting point for a more efficient realization of their own business processes for which the core data have to be more are less upgraded.

In the Dutch situation de following source data can easily get the status of basis data:

- the system of RD coordinates;
- the system of NAP coordinates;
- the system of AHN2 height points;
- the absolutely oriented high resolution Stereo-10 digital aerial recordings fitted in the (RD, NAP) system[1].

Meanwhile the Basic Registration Large Scale Topography ('Basisregistratie Grootschalige Topografie', BGT) has obtained a legal status. This is considered core information.

When defining which data should get the predicate 'core data' it is tempting to give a large number of datasets this predicate. After all each person concerned would like to see that his or her very important data will be included in a core dataset. However it is important that basis data are only realized for those data of which it is has been proved beyond the shadow of a doubt that a financial, a technical and a quality advantage can be achieved.

4. An organizational model for the collection, the management, the control and the supply of core data

Core data are a part of a topographic infrastructure for which the authorities are responsible. In fact it is a basic need that sees to it that users of geo-information apply the same infrastructure. Therefore it is advisable that the authorities take the lead in obtaining core data. Next private market parties can collect and process the core data.
For this the authorities have to act as a professional director, that:

- knows very well what users expect from core data:
- is very well informed about the present methods and techniques used for the creation of core data;
- is capable of contracting out projects to private market parties in a professional way;
- has a good view of the field of action and is able to anticipate on technological innovations that justify changes in the specifications of the core data.

The world of geo-information changes rapidly. New technologies, new products and new services are changing the field of action at high speed. A part of the products offered van be used for free, e.g. Google-Earth, Google Streetview and the oblique recordings in Microsoft Virtual Earth. The benefit and the need of availability of core data have to be weighed over and over again against what market parties offer. As soon as the need is no

---

[1] Under the condition that these data are collected periodically.

longer obvious the production, processing and distribution of that core data should be stopped.

Besides the authorities with their directing role also other parties are necessary to realize core data. These are:

- a strong private sector, that collects and processes core data;
- an organisation that evaluates if the prepared core data meet all required technical specifications;
- an organisation that distributes the core data among the users;
- a strong private and public sector that has the capability to upgrade the core data to sector specific products and services if needed.

The figure below shows an organizational model of the various actors in the field of action of core data.

The private sector collects the core data and processes them into the products as outlined in the contract specifications. This part of the private sector can count on a continuous amount of work for keeping the core data up-to-date. The products are delivered to the control agency. This agency evaluates if the delivered core data meet all technical specifications. If so, the data are delivered to the organization responsible for the distribution of the core data. Of course core data are and remain core data that have to be upgraded for certain applications like the earlier mentioned sector specific data. This upgrade is done by specialists from the field of action concerned. This can be private as well as public organizations. It may happen that during this upgrade data are created that could get the predicate 'core information'. An example is the measurement of a topographic situation for the construction of a civil technical engineering work. The after these measurements actualized topographic situation could be delivered to the control agency as potential 'BGT' data. The control agency evaluates if this data can get the 'BGT' status and therefore can be distributed.

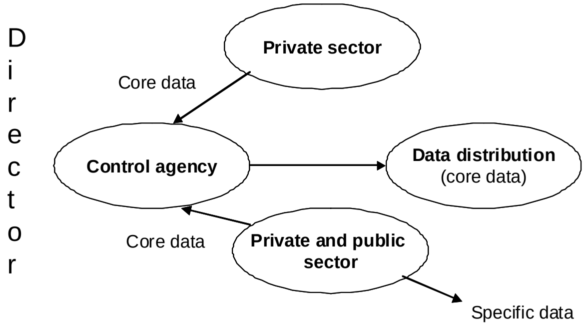


Figure 3. Organizational model showing the actors in the core data field.

5. Where to go from here?

During the past years the authorities have made an important move towards the creation of a system of legal key registers. At the centre of many of these administrations is, what should be available in the registration and not how the data should be obtained. The next step will be to make the collection and processing of geo-data more efficient. If it is possible to realize that data is collected only once whereas it can be used for multiple applications purchasing of geo-data by the authorities becomes cheaper. After all this will prevent duplication like for instance ordering several times the production of aerial photographs of the same area and in the same season by various public bodies. Moreover a more central steering will give better guarantees that the quality will meet the set requirements. By analogy with the philosophy behind key registers it is therefore desirable that the authorities for their need for actual geo-information not only look at their need for core information (interpreted data) but also at their need for raw core data.

For this reason it is advisable that further research will be carried out along two tracks:

- A technical track taking into account the following points:
    - Which source data are qualified for the predicate 'core data'?
    - Which quality requirements should the selected core data meet?
    - At which interval should the core data be collected?
- An organizational track in which attention is paid to the roles that the public and private sector can play with respect to collection, quality guarantee and supply of core data.