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.

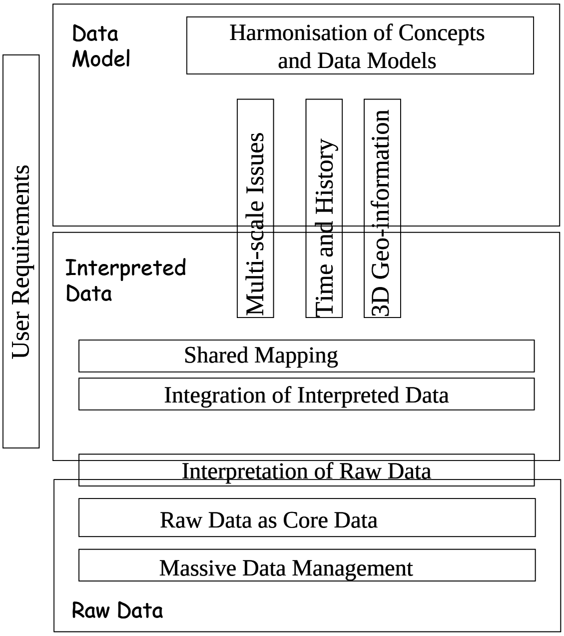
Currently, eight scientists and experts from universities, government agencies and companies cooperate in this Subcommission. The members are: Drs. R. van Essen (Tele Atlas), Ir. L. Heres (RWS-DID), Drs.Ir. A.J. Klijnjan (Dutch Land Registry Office (Kadaster)), Ir. R.G.A. Kroon (Ingenieursbureau Geodelta 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 meta-data 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.