Real-time integration and generalization of spatial data for mobile applications

Mark HAMPE, Monika SESTER

Institute for cartography and geoinformatics, University of Hannover, Germany,
mark.hampe@ikg.uni-hannover.de, monika.sester@ikg.uni-hannover.de

Abstract:

The EU-project GiMoDig developes methods for delivering geospatial data to a mobile user. One part of this project will be the real-time data-integration and generalisation. This paper describes the ideas and approaches of these components. It depicts the intention for implementing an MRDB and the rules for developing the global schema for the datasets.

1 Introduction

The EU-project GiMoDig, an acronym for „Geospatial Info-mobility Service by Realtime Data-Integration and Generalisation", aims at developing the spatial data delivery from national primary geo-databases for mobile use. The project started in November 2001. The following partners are involved in this project:

- Finnish Geodetic Institute (FGI) as coordinator,
- University of Hannover, institute for cartography and geoinformatics (ikg),
- Federal Agency for Cartography and Geodesy (BKG),
- National Survey and Cadastre Denmark (KMS),
- National Land Survey of Sweden (LMV) and
- National Land Survey of Finland (NLS).

The main vision of GiMoDig is a mobile user, travelling within an European country and receiving on-line information of his or her environment on the mobile device. Even when crossing a border, the type of information presented does not change, thus having the situation, that seamless topographic information is available.

The project deals with issues like user requirements analysis and small-display cartography constraints. On the one hand the data from the national databases have to provided, on the other hand the user must be able to state his requirements. Because of the limited display size and resolution it is important to transmit only the required information to the user. This will include the selection of the desired objects and features as well as the desired resolution of the presentation.

Due to these requirements, one sub-objective of the project is the development of methods for generalising the graphic representation of geospatial data in real-time, to be suited for display of the data at varying scales on small, mobile devices. The presentation on the mobile display will be dependent on the special user requirements i.e. data resolution and content.
Because of the complexity of the processes involved going from a large scale to a small scale (say 1:10k to 1:Mio) it is obvious, that (at least today) the generation and visualisation of ad hoc personalised products of spatial data in arbitrary scales on a mobile platform cannot be solved without any pre-generalised datasets. Only the real-time generalisation can be restricted to operations of minor complexity, that can be solved completely automatically.

The second major issue in the project is the harmonisation of data sets of the partner National Mapping Agencies (NMA's) in order to allow for a seamless data provision (workpackage "Global Schema"). This presumes that (topographic) data sets of the different countries are analysed with respect to their contents and that a semantic harmonisation is performed in order to guarantee the same "world view" when traversing the borders. Also here, the necessity of providing different data sets with different views of the same physical entities, as well as means to link these data sets is given.

Thus, in order to solve the problems of data generalisation and harmonisation, the concept of a MRDB (Multi Resolution / Representation Database) is used. On the one hand, the MRDB (Multi Resolution Database) serves as a pre-generalised and pre-harmonised data structure with spatial data in given scales. To minimise the effort of computation work during the real-time generalisation process, the GiMoDig-service selects a scale close to the desired scale requested by the mobile user. Based on this neighbouring scale, only small scale transitions are necessary, that can be handled in real-time. In this way the need for complex algorithms, for example simplification or displacement, can be minimised or even excluded.

In the course of the project, the merits of pre-computed datasets with various levels of detail (LoD's) in a multiple representation database in conjunction with methods for real-time generalisation will be investigated.
In a first step, the structure of the MRDB has to be developed, in compliance with the future generalisation and harmonisation work.

The paper is structured as follows: after an overview on MRDB and related work in general, the aims and the approach of the GiMoDig project are described and first issues for the solution of the problems are presented. A summary and an overview on future work concludes the paper.

2 Multi-Representation Database

2.1 Related Work on MRDB

A MRDB can be described as a spatial database, which can be used to store the same real-world-phenomena at different levels of precision, accuracy and resolution (Devogele et al. (1996), Weibel (1999)). Furthermore these phenomena can be stored in different ways of presentation or symbolisation.

It can be understood as a *multiple representation database* as well as a *multiple resolution database*. Regarding to the project in the first sense, the data integration is treated, whereas in the second meaning, the provision of different levels of detail, i.e. generalisation will be worked on.

In a MRDB, different views on the same physical objects or phenomena can be stored and linked. This variety of sights can stem from different views of the world, different applications, as well as different resolutions. These lead to differences in the objects as such, i.e. in the semantics and in the geometry. Also the graphic representation can be taken into account, leading to geometric, semantic and graphic multiplicities (Bernier et al. (2002)).

There are two main features that characterise a MRDB (see Fig. 1):

a) different levels of detail (LoD's) are stored in one database and
b) the objects in the different levels are linked.

The first feature can be compared to the analogue map series of the NMA's: these maps of different scales exist separately, only linked by the common geometry. In the second case, however, individual objects are linked with each other and thus allow for a detailed inspection of individual objects via several scales: Badard et al. (2002) use the notion of "drilling" objects through the scale space.

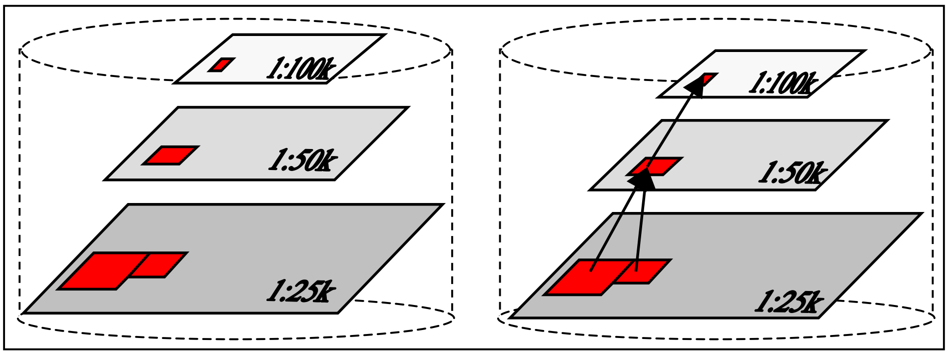


Fig. 1. a) MRDB with seperate scales b) objects linked between the LoD's

There are several existing concepts of MRDB's who depend on the specific needs and requirements.
Research projects aiming at combining real-time generalisation with multiple representation databases can be found in Badard (2002) and Cecconi (2002).
Vangenot et al. (2002) describe modelling concepts which support not only the multiresolution view but also the different views on the object features like object types, attributes and their values.
Kreiter (2002) describes the concept of a MRDB from the NMA's point of view.

Important questions concerning the implementation of a MRDB are:

(a) The number of layers to be provided:
It depends on the application area, typically at factor two to four the appearance of spatial phenomena changes so dramatically, that intermediate layers have to be introduced.
(b) The necessity of links between layers:
The question concerning whether explicit links should be established depends on the application; also the way these links are designed is of importance (uni-directional or bi-directional). If there is only the requirement to visualise the spatial data, seperate scales may be sufficient (see Fig. 1a). In order to offer more functionality (i.e. GIS-functionality) it *is* necessary to compute the link (Fig. 1b).

There are several reasons for introducing a MRDB: On the one hand it allows for a multi-scale analysis of the data: Information in one resolution can be analysed with respect to information given on another resolution. E.g. Gabay & Sester (2001) present an example where a topographic data of lower resolution containing only settlement areas is queried concerning the buildings in that area – an information that can be derived from a more detailed data set, whose objects are directly linked. A major reason for National Mapping Agencies to investigate and implement MRDB is the possibility of propagating updates between the scales: the actual information is only updated in the most detailed data set, this new information can then be propagated through the links in MRDB to all the other scales (Kilpeläinen (1997), Harrie (1999) et al.).
Cecconi (2002) discusses the possibility of creating intermediate scales based on a MRDB using a set of rules.

Devogele et al. (1997) classify three different stages when designing a MRDB:
(a) *Correspondence between abstractions*: Database schemata translate phenomena of the real world into abstracted instances of databases by focusing only on relevant parts of these phenomena; integration of abstractions thus requires methods for schema integration on the semantic level.
(b) *Correspondence between objects* of different representations: Data models are required to describe the links between corresponding individual objects of the different representations.
(c) *Defining the matching process* between objects: in order to identify corresponding (homologous) objects, two sets of geographical data must be searched for objects that represent the same real-world objects; methods for this purpose are subsumed under the term 'data matching' (Badard (1999), Sester et al. (1998)). In the case of multiple

resolutions, the correspondences between different objects can also be established by generalisation operations.

2.2 The GiMoDig Approach for Integration and Generalisation in a MRDB

Concerning the three stages of designing and populating a MRDB, the first problem (a) will be treated in GiMoDig in the workpackage "Global Schema" and will be described in section three. For every selected scale an own semantic schema has to be established. Besides the re-classification procedures are required for harmonisation of the conceptual models. The operations include also tools for adaptation of geometry (e.g. conversion of geometry type) and adaptation of resolution (e.g. aggregation of features). Such tools are elements of model generalisation, thus linking the work in workpackage "Global Schema" and workpackage "MRDB / real-time generalisation".
In section four ideas for establishing links between objects in different scales will be presented. The *ikg* participates mainly in these two workpackages dealing with semantic harmonisation ("Global Schema") and generalisation based on MRDB ("MRDB / real-time generalisation"). The research on real-time generalisation will be mainly accomplished at the Finnish Geodetic Institute (FGI) and the National Land Survey of Sweden (LMV).

To derive the objects of the less detailed LoDs two cases are conceivable:
- The first option is the case to combine existing datasets of different scales to an MRDB by a matching process, as pointed out in the third problem (c) above.
- The other is to derive the less detailed LoDs from a base-level of higher LoDs, in our case from the core datasets of the NMAs (i.e. ATKIS Basis-DLM in Germany). This will be the approach pursued in the project using generalisation operations to link different resolutions, which leads to the following problems.

By implementing the second solution, new problems occur (spotted so far):
- There are no existing generalisation-tools who allow an automatic generalisation of a given dataset,
- besides model generalisation also cartographic generalisation has to be taken into account, as the objects have to be clearly visualised and symbolised on the computer screen.

These are open research questions that also will be tackled within GiMoDig.

3 The global schema

The ideas presented in this section were mainly developed at the Federal Agency for Cartography and Geodesy (BKG) as the main partner in the workpackage "Global Schema".

The raw data for the MRDB in GiMoDig comes from the National Mapping Agencies (NMAs) of Germany, Sweden, Denmark and Finland, as the aim of GiMoDig is to provide access to the national (primary) topographic geo-databases.
Because of the dissimilitude of these datasets, the first step will be to transfer them into an integrative data-schema. This schema has not only to be developed for the highest LoD but also for every layer in the database.
All of these countries can provide national data at resolutions 1:5k to 1:10k, 1:250k and 1:1.000k. Data at resolution 1:25k to 1:50k is available in Finland and Sweden only, but is planned also in Germany and Denmark for the near future.
The specifications for the EuroRegionalMap and EuroGlobalMap, two projects developed by Eurogeographics (2002), the organisation of European National Mapping Agencies, do already pave the way for pan-European topographic databases at medium and low resolution (1: 250k and 1: 1.000k). Therefore, GiMoDig will concentrate on the higher resolution datasets up to a scale of 1: 100k.

The workpackage “global schema” deals with the data integration. Inspecting the databases of the NMAs reveals different views onto the same real world phenomena. The different feature-catalogues and representations of the same real world objects mirror these differences. Looking, for example, at the object “grassland”: In Germany “grassland” is an object of the objectclass “vegetation” and is presented as an area feature (to be found in the ATKIS-OK25). In Denmark land-use is not registered in the topographic data base and in the Swedish database “grassland” can be interpreted as all open land that is not defined as “forest” or “arable land”, so “grassland” is not mapped explicitly.

Because of their dissimilitude these data have to be harmonised before they can be used in the info-mobility service. To harmonise means in this case to design a common schema and to transform the various models and data-sets into this schema.
The preliminary global schema, devised by Illert & Afflerbach (2002) for the project, is based on the selection of object feature types from the national core databases.

The distributed feature groups will be as follows. They are ordered according to the EuroRegionalMap specification (Eurogeographics (2002)):

- Boundaries,
- Hydrography,
- Miscellaneous (including feature types like parks, buildings, cropland etc.),
- Settlements,
- Transport,
- Vegetation and soils.

For the harmonisation of the national core databases into a global schema several rules have to be defined in order to find a schema that minimises the harmonisation efforts and at the same time is consistent with the EuroRegionalMap. A proposal for the rules according to the principle of the least common denominator is the following set (defined in Illert & Afflerbach (2002)):

- Feature type: The semantic model is a subset of the semantic model of the EuroRegionalMap.
- Geometry type: That geometry type (point, line, area) with the majority among the national datasets is chosen.
- Collection criteria:
  - Case A: same selection expression but different thresholds.
    For example feature type *Park*:
    DE All green areas with a size of more than 10.000 m^2 are recorded
    DK Recreational areas larger than 2.500 m^2 are included
    FI The minimum size of a park is 5.000 m^2
    In this case the least common denominator has to be chosen, therefore all parks with more than 1 ha are recorded.
  - Case B: different selection expressions:
    Use only those collection criteria that are common to a majority of national databases and connect the expressions with Boolean OR.
- Attributes: Attributes of the EuroGlobalMap are only included to the GiMoDig global schema, if they are supported by the majority of the countries or are of great importance.

In part, this can imply, that data has to be captured by the respective partners, in order to comply with the common global schema. This (preliminary) data-schema with its objects and features serves as the data-schema for the base-level of the database.
Additional models have to be developed for the following levels of the database by model-generalisation.

4 The Multi-Resolution Database in GiMoDig

The workpackage "MRDB / real-time generalisation" relies on the schemas developed in the workpackage "Global Schema". Therefore, the semantic multiplicity will be eliminated before the data are stored in the project-database. As the data models will be harmonised the database will contain only a geometric and graphic multiplicity and can thus be defined as a Multi Resolution Database.

The MRDB in the project is used to facilitate the real-time generalisation process.
In the project, a seamless generalisation is aimed at, i.e. any scale can be generated depending on the application (See Figure 2). This will be achieved by integrating pre-generalised data and on-the-fly generalisation operations. The latter is based on using XML and XSLT (Lehto et al. (2000)).

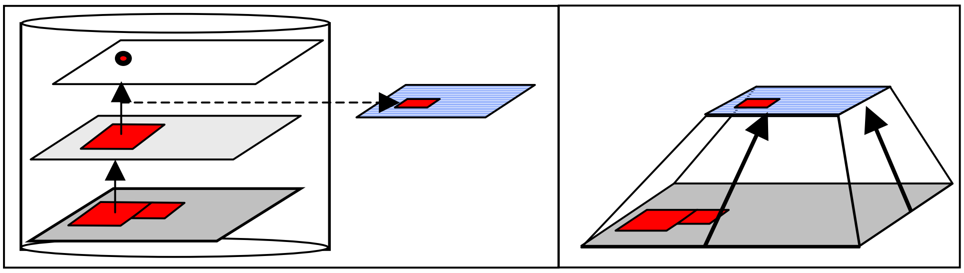


Fig. 2. realtime-generalisation by a) pre-generalised data or b) in one step

4.1 The database schema

The database-software used in the project is an open-source, unix-based and object-relational DBMS called *PostgreSQL* (postgresql (2002)). It supports almost all SQL constructs, including subselects, transactions and user-defined types and functions. It will be used together with an extension called *PostGIS* (postgis (2002)), which adds support for geographic objects to the PostgreSQL object-relational database. PostGIS follows the OpenGIS "Simple Features Specification for SQL" (opengis (2002)).

The work has started with the two feature types roads and building. This work and concept is inspired by work of N. Kreiter (2001).

For our first experiments we are using a cadastral data set of approximate scale 1:1k and the ATKIS Basis-DLM (1:25k) that will be integrated in the database (Figure 3).

Fig. 3. The stored ATKIS©-dataset. Presented as a SVG-graphic

The existing datasets, which are available in the ESRI-conform *shape*-format, can be imported to the database by the simple tool *shp2sql*, distributed by PostGIS. The *shp2sql* data loader converts ESRI Shape-files into SQL suitable for insertion into a

PostGIS/PostgreSQL database. This tool facilitates a conversion and upload to the database.
Figure 3 shows an example of the ATKIS data set. The corresponding representation in a PostGIS table is shown in Figure 4.

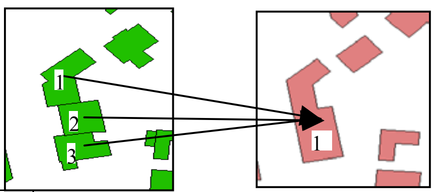
gid	aobjid	text	the_geom
53	N01ABAQ	N01ABAQ014N01AE	SRID=-1;MULTILINESTRING((3547632.93 580
54	N01ABAQ	N01ABAQ014N01AE	SRID=-1;MULTILINESTRING((3547619.32 580
55	N20GR7A	N01FUF7	SRID=-1;MULTILINESTRING((3547570.77 580
56	N01FUHH	KN0320100000400	SRID=-1;MULTILINESTRING((3548130.87 580
57	N01FUHG	GNOffensteinstraße	SRID=-1;MULTILINESTRING((3548065.06 580
58	N20C8N2	KN0320100001810	SRID=-1;MULTILINESTRING((3547811.35 580
59	N01FUH4	GNFröbelstraße	SRID=-1;MULTILINESTRING((3547927.09 580
60	N01FUHD	KN0320100002857	SRID=-1;MULTILINESTRING((3547982.2 5804
61	N01FUH6	KN0320100000435	SRID=-1;MULTILINESTRING((3547906.18 580
62	N01FUH7	KN0320100003259	SRID=-1;MULTILINESTRING((3547982.2 5804
63	N01FUHD	KN0320100002857	SRID=-1;MULTILINESTRING((3548018.92 580

To link the two datasets, first of all, the possible relations between objects at different scales have to be identified. The following kinds of links will be possible between the objects:

a) *1:1* relation: one object in scale *A* refers to one object in scale *B* (the object may however be simplified in shape)
b) *n:1* relation: more than one object in scale *A* will be aggregated to one object in scale *B*
c) *1:0* or *n:0* relation: one or more objects will be omitted in the lower LoD
d) *0:1* or *0:n* relation: objects will appear only in the lower LoD.
e) *1:n* relation: more than one object in a lower LoD arised from one object in the higher LoD: i.e. displacement

The links are represented in the tables, that have to be enhanced with additional attributes. The first idea is be to extend every table, representing the objects of one scale-level, by two more columns. An ID-Number of the object in the actual level and the ID-Number of the object in the following level has to be added.

ID1k	ID25k	the_geom	…
1	10	Linestring (…)	
2	10	Linestring (…)	
3	10	Linestring (…)	



ID50k	ID100k	the_geom	Dx	Dy	…
10	50	Linestring (…)	12,5	7,8	

11	50	Linestring (…)	3,7	5,5	
12		Linestring (…)	7,8	5,3	

Fig. 5. First possibility to compute the linking structure

In this case two problems can occur. There are empty fields for those objects that do not appear in the following LoD. The other problem is, that one or more objects in the level 1k can only be linked with one object in the level 25k, although there are cases where more than one link is required. For that case an additional line has to be inserted and that means that the whole object had to be duplicated.

Another approach is to generate an additional linking table for every link between two scales. Each row of this table indicates a relation between two objects.

ID1k	**ID25k**
1	10
2	10
3	10
4	20
4	30

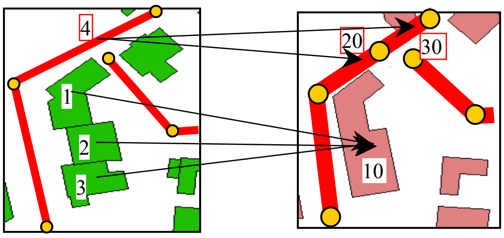


Fig. 6. Alternative possibility to link corresponding objects by using an extra table

This solution covers the case of a 1:n relation, shown in figure 6, as well. These cases occur for instance when two datasets of different sources with different data models are combined (i.e. ALK and ATKIS in Germany) as shown in Gabay & Sester (2002).
As shown in this table all kinds of relations can be modelled. In the case of a 1:0-relation, i.e. an object disappears, no relation has to be stored.

5 Conclusion and further work

The idea of the GiMoDig-project is to bring the topographic data of the national mapping agencies to a mobile user. This user will have the possibility to freely specify his / her needs for the delivered map, i.e. scale, accuracy and features.
It is shown in this paper, that for setting up such an info-mobility service there is a necessity to integrate the inhomogeneous data of the NMA's. In a first step these data have to be translated into a global schema.

To generate the required data as fast as possible the data should be generalised and harmonised in real-time. In order to this real-time-service the service-provider must have the possibility to resort to prepared data. Only a selection of features and minimal changings have to be computed, that can be accomplished in real-time. This paper showed the advantages of storing these data in a MRDB and an approach to model this database.

The paper could only show preliminary ideas and concepts. In the near future, these concepts will be implemented and tested. Firstly the data model will be established and populated with data, i.e. two linked datasets will be generated by manual linking.
After intensive testing of this structures, some selected datasets with streets and buildings have to be generalised automatic or semi-automatic. During this process the corresponding objects of two scales have to be linked automatically. The relationships are generated based on the generalisation operations performed on the objects.

6 References

Badard, T. (1999), On the automatic retrieval of updates in geographic databases based on geographic data matching tools, in: `Proceedings of the 19th International Cartographic Conference of the ICA', Ottawa, Canada.

Bédard, Y. & Bernier, E., (2002). Supporting multiple representation with spatial databases views management and the concept of "VUEL", ISPRS / ICA Joint Workshop on Multi-Scale Representations of Spatial Data, Ottawa, Canada, 2002.

Bédard, Y., Proulx, M.-J., Larrivée, S. & Bernier, E., (2002). Modelling multiple representations into spatial data warehouses: A UML-based approach, In: Proceedings of the ISPRS Joint International Symposium on Geospatial theory, processing and applications, Canada, 361.pdf, 2002.

Cecconi, A. , Weibel, R. & Barrault, M., (2002). Improving automated generalization for on-demand wepmapping by multiscale databases, 10th International Symposium on Spatial Data Handling, Ottawa, Canada, 2002

Eurogeographics (2002). www.eurogeographics.org/Projects/ProjectsIndexLevel1.htm

Gabay, Y. & Sester, M., (2001): Forming and utilizing communication between two spatial representations at different scales - a demonstration. Submitted to Geoinformatica.

Harrie, L. & Hellström, A.-K. (1999). A Case Study of Propagating Updates between Cartographic Data Sets, in: `Proceedings of the 19th International Cartographic Conference of the ICA', Ottawa, Canada.

Illert, A., Afflerbach, S., BKG, (2002). Differences between Databases. GiMoDig-project deliverable D5.1.2. unpublished

Illert, A., Afflerbach, S., BKG, (2002). Selection of Feature Types. GiMoDig-project deliverable D5.1.1. unpublished

Kilpeläinen, T., (1997). Multiple Representation and Generalization of Geo-Databases for Topographic Maps. Kirkkonummi 1997

Kreiter, N., (2002). Multirepräsentationsdatenbank für die Kartographie. Vertiefungsblockarbeit. Eidgenössische Technische Hochschule, Zürich. www.karto.ethz.ch/teaching/VTB-Kreiter1.pdf, 2002

Kreiter, N., (2002). Multirepräsentationsdatenbank als Basis von topografischen Landeskarten. Diploma thesis, unpublished, Institute of Cartography, Swiss Federal Institute of Technology Zurich, 2002.

Lehto, L. & Kilpeläinen, T.(2000). Real-Time Generalization of Geodata in the Web. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B4, Amsterdam, pp. 559-566.

Opengis (2002), OpenGIS® Simple Features Specification for SQL at http://www.opengis.org/techno/implementation.htm

Postgresql (2002), http://www.us.postgresql.org.

Postgis (2002), http://postgis.refractions.net

Sester, M., Anders, K.-H. & Walter, V. (1998). Linking Objects of Different Spatial Data Sets by Integration and Aggregation, GeoInformatica 2(4), 335-358.

Vangenot, C., Parent, C. & Spaccapietra, S., (2002). Modelling and Manipulating Multiple Representations of spatial data, 10th International Symposium on Spatial Data Handling, Ottawa, Canada, 2002

Weibel, R. & Dutton, G.,(1999). Generalising spatial data and dealing with multiple representations. In: P.A. Longley, M.F. Goodchild, D.J., Maguire & D.W. Rhind, editors, Geographic Information Systems – Principles and Technical Issues, volume 1,. John Wiley & Sons, 2 edition, 1999, pp 125–155.

Biographies

Dipl. Ing. Mark Hampe, born in 1973, studied Geodesy at the University of Hannover and obtained the Master's degree (Dipl.-Ing.) in 1999. His master thesis was about „animated maps and their advantages for the user". Between 1999 and 2001 he attended a trainee on public service. Since 2002 he is a scientific assistant at the institute for cartography and geoinformatics, university of Hannover. His primary exercise lies in the part of the University of Hannover in the EU-project Gimodig.

Dipl.-Ing. Mark Hampe
Institute for cartography and geoinformatic
University of Hannover
Appelstraße 9a
30167 Hannover

e-mail: mark.hampe@ikg.uni-hannover.de
fax: +49511/762-2780
page: www.ikg.uni-hannover.de

Professor, Dr. Monika Sester, born in 1961, studied Geodesy at the Technical University of Karlsruhe and obtained the Master's degree (Dipl.-Ing.) in 1986. Until October 2000 she was a staff member of the Institute for Photogrammetry of the University of Stuttgart, Germany, where she obtained her PhD in 1995. In 1994 she became the head of the GIS research group of the Institute, consisting of six researchers. Since November 2000 she is Professor and head of the Institute for Cartography at the University of Hanover. Her primary research interests lie in multi-scale approaches in GIS and image analysis, data generalisation and aggregation, data interpretation, and integration of data of heterogeneous origin and type. She is chair of the Working Group on "Data Generalisation and Data Mining" of Commission IV of the International Society for Photogrammetry and Remote Sensing (ISPRS). She is also active in the Commission on "Generalisation" of the ICA. Furthermore, she is head of the GIS-Group of the German Society for Photogrammetry and Remote Sensing (DGPF).

Prof. Ing. Monika Sester
Institute for cartography and geoinformatic
University of Hannover
Appelstraße 9a
30167 Hannover

e-mail: monika.sester@ikg.uni-hannover.de
fax: +49511/762-2780
page: www.ikg.uni-hannover.de