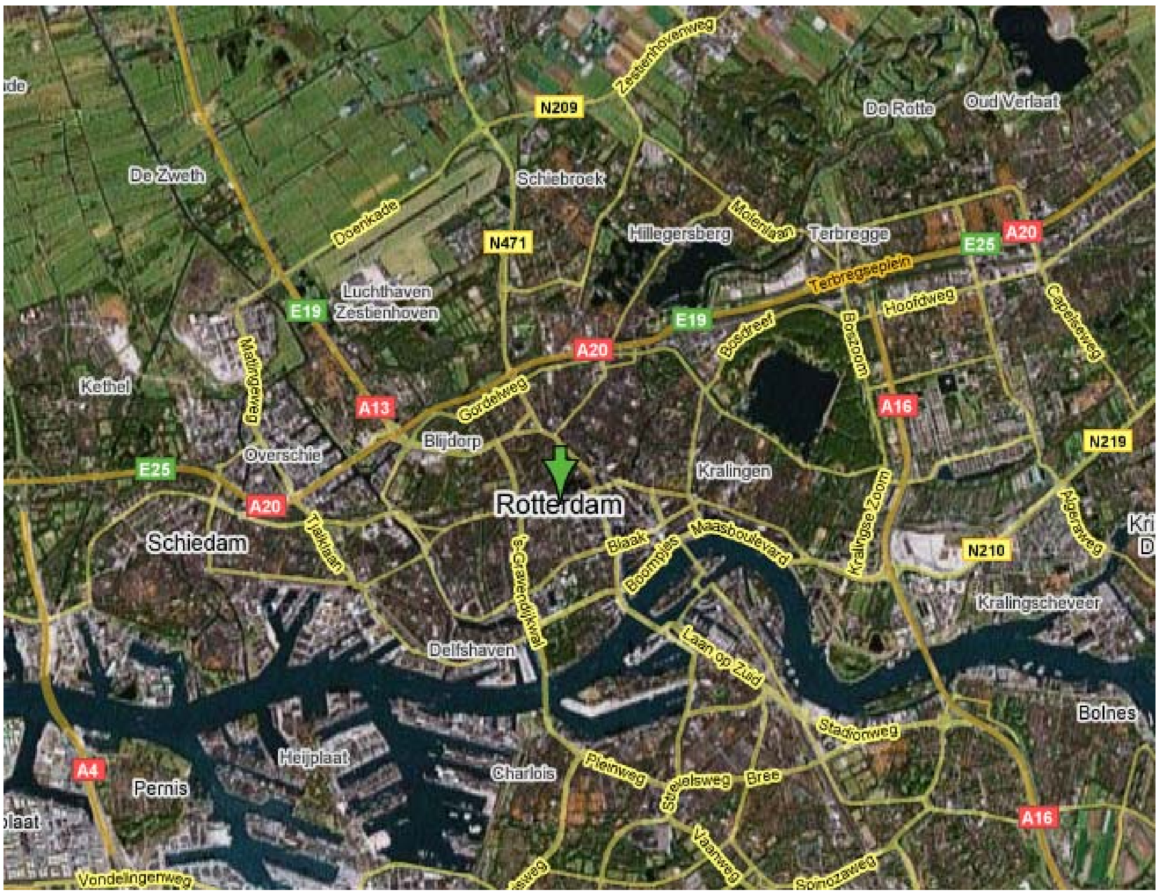
Application of generalisation in Rotterdam

A literature study



Arjen Hofman
1049992

Table of contents

Preface

The report you are about to read is titled ‘Application of generalisation in Rotterdam’. This report has been written as an individual assignment and will be used as part of the author’s master thesis on the same subject. This report is the literature study prior to the start of the actual thesis work. The author of this report is student at TU Delft, studying Geomatics. The thesis work will be done at the department of public works (Gemeentewerken) of the municipality of Rotterdam.

This report is about generalisation. The subject will be treated from literature as well as from the situation in which the author will be doing his master thesis; the theoretical and practical viewpoints on generalisation will meet here.
As a reader I would like to wish you a good time reading this report. It has been a pleasure to make it and I hope this pleasure will be contagious to the reader.

Thanks goes to Arta Dilo for her support during the writing process. Also thanks to Jan Haunert for his input and to my supervisors in Rotterdam: Nicole, Michiel and Evert. Thanks to my girlfriend Martine for her moral support.

Rotterdam, September 18th 2007
Arjen Hofman

Summary

This report is a literature study, which is made to find an answer to the question how the geographical data in the municipality of Rotterdam can be generalised. After this literature study the author's master thesis will go deeper into generalisation. The main question in this report is: *What is the best method to perform map generalisation for the municipality of Rotterdam?*

The municipality of Rotterdam wants to find a way to be able to only update the large scale base map (GBKN) and derive the medium scale base map (Top10NL) from this GBKN. To be able to do this the municipality wants to work with new standards, which are object oriented. The main standard for geo-information NEN3610 is translated into a model for large scale topography, IMGeo. This model will form the basis for the generalisation described in this report.

Three methods are investigated to see if they are possibly able to solve this matter. The generalisation methods are described below:

- The Multiple Representation Database (MRDB) focuses on data consistency. It stores objects at different levels of detail and links them to each other. The drawback of this method is the amount of storage capacity it takes.

- The on-the-fly generalisation methods focus on the reduction of storage space. The geo-information is only stored at the largest scale and, when data of another scale is requested, the algorithms are executed directly. The drawback of this method is that it performs slowly.

- The topological Generalized Area Partition (tGAP) structure stores the geometry at the largest scale and it stores the relations to new objects at their importance level. If a certain level is requested the database can select the right objects for this level of detail and send the requested objects to the client.

Conclusions

Through a criteria analysis the best method to perform generalisation in Rotterdam has been investigated. It turned out that the tGAP structure is the most suitable option to apply generalisation in Rotterdam. The tGAP is fast, scale independent and stores the geometry of objects only once. These conditions make the tGAP very suitable.

On the other hand there is a lot of innovation to be done to the tGAP to really implement it in an environment like Rotterdam. This means that enough new research content can be added by the master thesis research following this study.

The proposal for the thesis research is that a constrained tGAP structure will be built. The constraint tGAP doesn't build one tGAP tree, but many of them until the scale of the constraint is reached. The IMGeo map will serve as the basis for this with the Top10NL as initial constraint.

1. Introduction

In a map the producer has to ensure that the information he wants to communicate is clear enough. For this reason the information in a map often doesn't represent the real world, but the map as the producer wants it to be communicated. If this happens generalisation is applied. Generalisation can be defined as a selected and simplified representation of the real world appropriate to the scale and purpose of the map.

In the municipality of Rotterdam plans are made to derive the medium scale base map (1:10,000) through generalisation from the large scale base map (1:1,000). Because of the fact that the specifications of the data models and the information they want to represent differ, this is not a very easy task. In a Master Thesis following this literature study the problem of generalisation will be thoroughly investigated.

This report wants to investigate what the current possibilities are for map generalisation based on literature. With this background the report wants to give an answer to the following research question:

What is the best method to perform map generalisation for the municipality of Rotterdam?

To answer this question this report will first go into the challenge of map generalisation within the municipality of Rotterdam in chapter 2. In chapter 3 the mainstreams for generalisation are introduced and actual research will be described. Chapter 4 will give the application of the third chapter; based on the literature a preliminary choice with respect to the generalisation method will be derived using a criteria analysis. Finally, in chapter 5 the conclusions will be drawn.

2. Current situation in Rotterdam

The department of Gemeentewerken Rotterdam (Public Works) is the organization that has offered the Master Thesis project for which this literature study is made. This chapter will first describe the current situation of the organization in section 2.1 and describe current and future developments in section 2.2. In section 2.2 also the models IMGeo and Top10NL will be introduced.

2.1 Current situation at Gemeentewerken Rotterdam

Gemeentewerken Rotterdam has about 1800 employees. The department in which this master thesis project will take place is the department of Surveying (Landmeten). This department has about 90 employees. The section Topography produces two main products. The first is the Large Scale Base Map (GBKN) with a scale of 1:1,000 and the second is the Medium Scale Base Map (KBK Rotterdam) with a scale of 1:10,000. In the next two subsections the two data sets will be described.

2.1.1 Large Scale Topographic Base Map

The Dutch Large Scale Topographical Base Map (GBKN) is a map product with scales varying from 1:500 to 1:5,000 (productinformatie GBKN, 2006). The scale that is mostly used for this product is 1:1,000. All municipalities in The Netherlands are obliged to keep up this GBKN. Its use varies a lot; it is for example used as under layer for maintenance services within municipalities or as base map for utility companies. Also the collection rules vary per municipality.

Most municipalities use the GBKN as a basis for all soft and hard topography, the collection and maintenance is done in two different ways:

- Terrestrial (field) data collection
- (Stereographic) Aerial photographic data collection

Terrestrial field data collection is more accurate, but also more expensive. In lots of municipalities large parts of the GBKN is not measured terrestrially. Instead, they chose for the less accurate aerial photography to fill up the total municipal map. In Rotterdam this is not the case, all data for the GBKN is collected terrestrially. The Rotterdam version of the GBKN is called GBK-Rotterdam or GBK-R for short.

2.1.2 Medium scale Topographic Base Map

The medium scale topographic base map, as the municipality of Rotterdam maintains it, is not a regular map. Most municipalities only maintain their 1:1,000 map and leave the maintenance of a 1:10,000 map to the topographical service of the Dutch Cadastre (TD Kadaster). The TD Kadaster is the responsible party for this 1:10,000 map, called Top10NL or Top10Vector, its earlier version.

The 1:10,000 map of the municipality of Rotterdam (KBK-Rotterdam, or KBK-R for short) shows very much resemblance with the Top10NL. There is only a small amount of differences between the data models and the way information is presented. The 1:10,000 map is collected and drawn from aerial photographs and is therefore far less accurate than the 1:1,000 GBKN.

To get a better impression of how the models currently, take a look at figure 2.1.

Figure 2.1: Fragment of the GBKN (left) and KBK-R (right) of Rotterdam

Both these data models are currently in a process of change. In the next subsection we take a closer look at these changes.

2.2 Current developments within Gemeentewerken Rotterdam

This section will introduce the main developments in the organisation of Gemeentewerken Rotterdam. Besides the processes that will be described in this section also a large reorganisation is going on. This section will deal with the new models IMGeo in 2.2.1 and Top10NL in 2.2.2. Also attention will be given to how these developments fit in the idea of the Dutch Geo Information Infrastructure (NGII).

The current developments within the organization of Gemeentewerken are that both products are being upgraded. The GBK-R is currently a line based map. With the new standard IMGeo (Information Model Geography) they want to upgrade this to an object oriented map.

The setting of the 1:10,000 map will change as well. This 1:10,000 map for Rotterdam is currently produced by both the TD Kadaster and the municipality of Rotterdam. While this is very inefficient and the law on authentic registrations is coming, this will soon change. Rotterdam is applying for the deliverance of the 1:10,000 map to TD Kadaster. To do this deliverance the process of making the 1:10,000 map has to be changed.

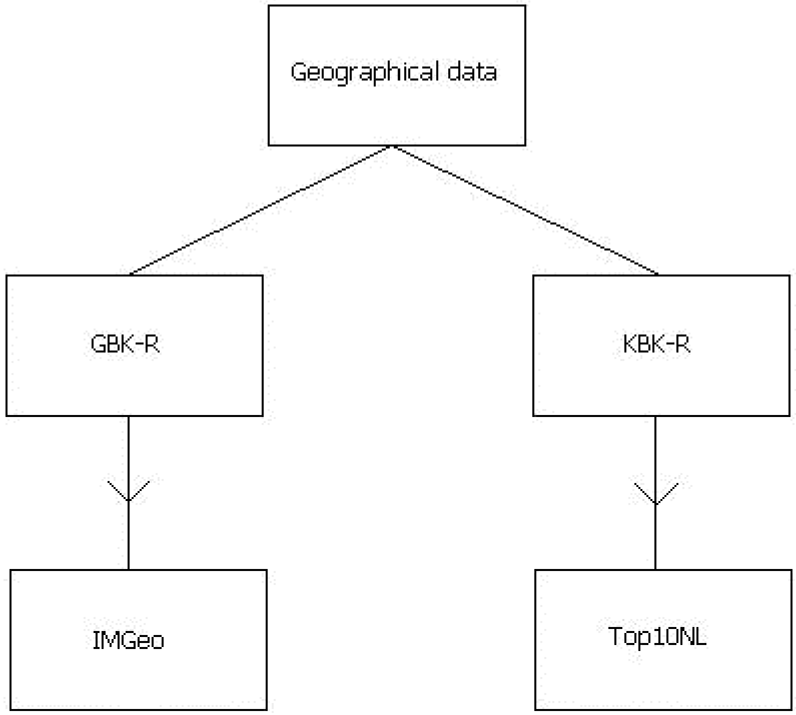


Figure 2.2 Current developments within the municipality of Rotterdam

In figure 2.2 the current developments are shown. The objective of the following Master Thesis is to apply generalisation on the object oriented model IMGeo to achieve a map as close as possible to the new Top10NL format.

Both these new models, what objects they have, will be extendedly described in 2.2.1 (IMGeo) and 2.2.2 (Top10NL).

2.2.1 IMGeo

The text in this subsection comes from the report on IMGeo, version 2.3. The author of this report translated and edited the text.

IMGeo is a standard which in the first place has been created because 4 large municipalities in the Netherlands felt the need for large scale object-oriented geo-information. These municipalities initiated to make an information model, which would define and standardize the exchange of objects.

IMGeo is based on the Dutch Basic Model for Geo-information; this is called NEN3610. This describes terms, definitions and rules for the exchange of geo-information. This model is considered to be the basic model for all geo-information sectors. Because this basic model is too general, a more detailed version of this information model for large scale topography had to be defined, which became IMGeo.

IMGeo is an object oriented model, in which meaningful polygons are made instead of lines. This object orientation enables the data collector also to attach attributes to the objects. Figure 2.3 shows what the differences are in IMGeo compared to the same line GBKN as in figure 2.1.

Figure 2.3: Line GBKN (left) and polygon GBKN (right)

Structure

IMGeo objects are used in the format of figure 2.4. The figure represents the form of an object class in IMGeo.

Klassenaam
+attribuutnaam : <attribuutdomein> [multipliciteit]

Figure 2.4 Representation of an object in IMGeo

Explanation of figure 2.4

- ‘Klassenaam’ = the name of the object class;
- ‘attribuutnaam’ = attributes defined for this object class.
- <attribuutdomein> = a reference to the acceptable values for this attribute, the domain.
- [multipliciteit] = the number of values the attribute can take.

For every object class there is a table inserted with definitions and other class information. The whole UML diagram of the IMGeo model, which defines how the objects are related to each other, is represented in Appendix A.

For every object lots of attributes are stored. For a full overview of these attributes the reader is referred to the complete description of IMGeo (IMGeo, 2007).

Objects

In IMGeo a lot of different objects are taken into account. Some of them are polygon objects; others are line or point objects. All available main objects will be listed. For every class the English translation is added between brackets.

IMG_ GeoObject (object)

This is the super class under which every object in IMGeo is situated. Its attributes are linked to every object in the model.

Weg (road), Wegdeel (road part)

This class defines all kinds of roads in the model; the object class 'weg' is the super class here.

Spoorbaan (railroads), spoorbaandeel (railroad part)

This class defines all railroads in the model; the object class 'spoorbaan' is the super class here.

Water (water), waterdeel (water part)

This class defines all water in the model; the object class 'water' is the super class here.

Terrein (terrain), terreindeel (terrain part)

This class defines all terrain parts in the model; the object class 'terrein' is the super class here.

Kunstwerk (civil works), kunstwerkdeel (civil works part)

This class defines all civil works in the model; the object class 'kunstwerk' is the super class here.

Pand, verblijfsobject (buildings)

This class defines all buildings in the model. All definitions in this class are according to the regulations of the registration for buildings and addresses (BAG). In IMGeo only the geometry of the building and the ID of the buildings and their associated residence objects appear.

Inrichtingselement (topographical elements)

This class defines all elements which fit up the area. All kinds of elements are meant here, for example traffic lights and lamp-posts. For the complete list of elements the reader is referred to the full report of IMGeo (IMGeo, 2007)

Registratief gebied (registration area)

The last class is the class of the registration area. Here is a subdivision in provinces, municipalities, places of residence, neighbourhoods, etc.

2.2.2 Top10NL

Within the organization of Gemeentewerken Rotterdam there is also a process going on to come to deliver the Medium Scale Base Map to the TD Kadaster. The KBK-R and Top10NL are both maps drawn from aerial photographs for the scale 1:10,000. Within the KBK-R polygons are drawn and given a specific colour, while in Top10NL attributes are connected to the objects.

Gemeentewerken Rotterdam is obliged to adopt Top10NL from the year 2010, because by that time Top10NL is the authentic registration on medium scale topography. This means that every governmental organisation is obliged to use Top10NL. Gemeentewerken Rotterdam tries to deliver Top10NL to the TD Kadaster and make an extra layer for own purposes and customers.

In the Top10NL attributes are connected to objects; this is not the case in the KBK-R. This means that a good amount of progress would be made if the KBK-R would be made with the rules of Top10NL. On the other hand there are some features mapped in the current KBK, which are not in Top10NL. Also the actuality of the KBK is better than Top10NL. How the municipality turns to Top10NL is still subject of discussion, because some customers of the KBK still want the specifications of this map, even if it is transformed to Top10NL.

Currently tests are done at Gemeentewerken Rotterdam to see how the Top10NL data structure can be fitted to the situation in Rotterdam. The way of working and the choice of objects in Rotterdam slightly differ from the Top10NL. The possibility of an extra layer is currently being investigated in a project within Gemeentewerken Rotterdam.

Resemblance

The question that rises is how the objects of the Top10NL resemble to the objects in IMGeo. Because Top10NL is based on the same standard NEN3610 the major object classes resemble very much to the major object classes in IMGeo. Figure 2.5 shows these classes.

The full UML diagram of Top10NL is shown in Appendix B. Even though IMGeo and Top10NL might be based on the same standard (NEN3610), the Top10NL model differs from IMGeo in some elementary points. It would be too extensive to describe these differences here. For this the reader is referred to the Master Thesis following this report.

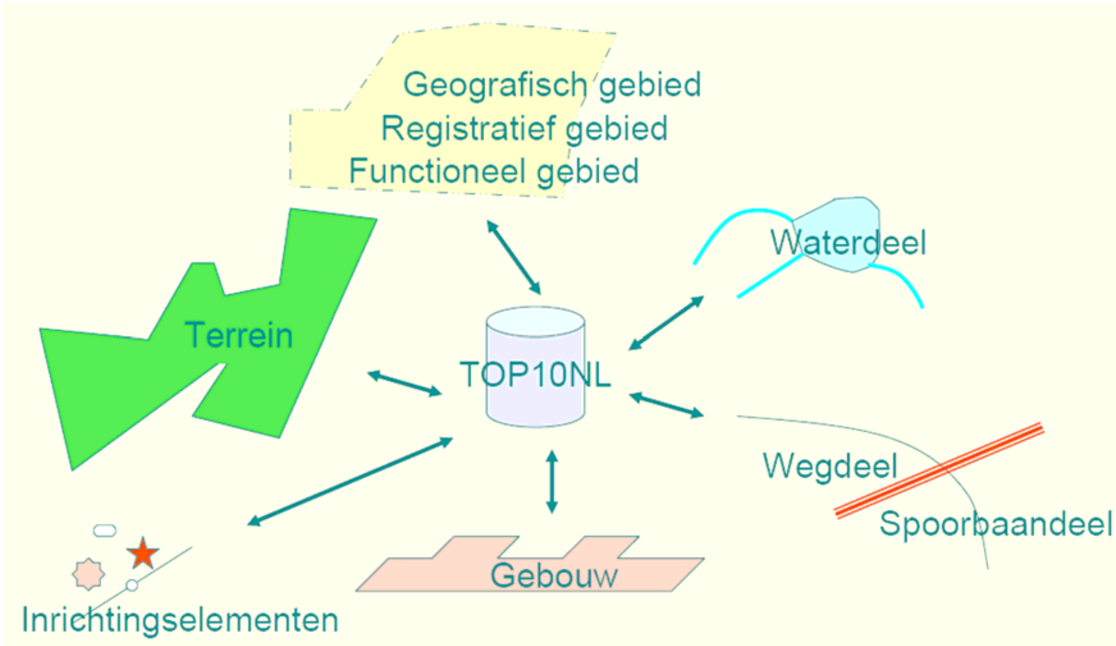


Figure 2.5: Objects in Top10NL

2.2.3 Nationale Geo Informatie Infrastructuur

The Nationale Geo Informatie Infrastructuur (NGII) is the Dutch national Geo Information Infrastructure (GII). Within the framework of a GII data should be collected once and used for multiple purposes.

Within the framework of the project 'Stroomlijning Basisgegevens' 6 authentic registrations were indicated. They had to be connected to each other. One of these authentic registrations is the registration of Topography. An authentic registration is said to be the only official governmental registration on that particular area. For the registration on topography the government chose for Top10NL as authentic registration.

In figure 2.6 it is shown how the 6 authentic registrations are related. It is shown that the building registration and the cadastral registration are both related to the topography registration. The problem with the 1:10,000 map is that buildings are not detailed enough in the Top10NL to connect them to the BGR and BRA (together these to form the authentic registration on buildings and addresses, called BAG). The connection with the GBKN could be made, but this is not an authentic registration, because it is partly financed by private parties.

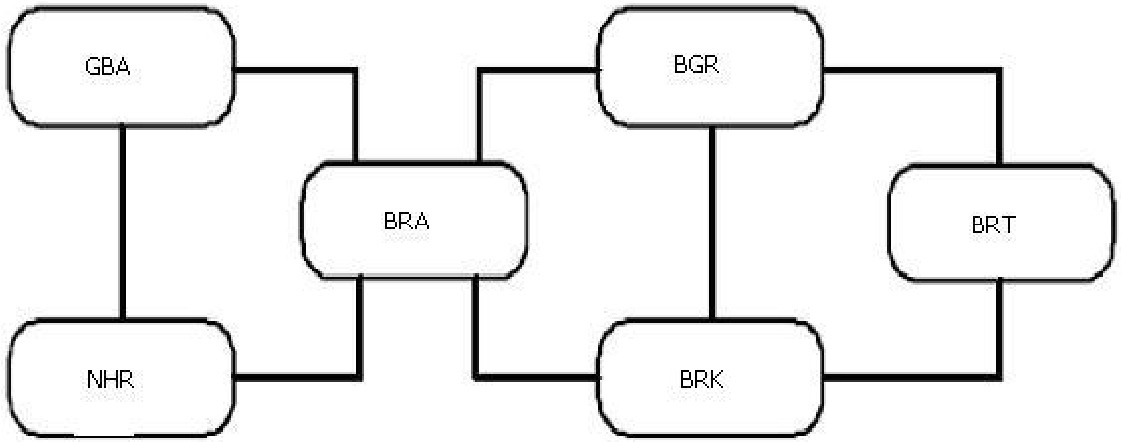


Figure 2.6 Authentic registrations in The Netherlands

The 6 authentic registrations in figure 2.6 are:

1. Municipal administration of citizens (GBA)
2. New company register (NHR)
3. Address registration (BRA)
4. Building registration (BGR)
5. Cadastral registration (BRK)
6. Topographical registration (BRT)

More and more registrations are applying for the title ‘authentic registration’. Nowadays there are about 10 registrations and candidate registrations. The GBKN is one of these candidate authentic registrations. Not to replace the Top10NL as an authentic registration, but to become a separate authentic registration on large scale topography beside the medium scale authentic registration (LSV GBKN, 2002).

In this case a strange situation would appear, because the vision of the authentic registrations is to collect the data once and to use it in all possible situations. When using the GBKN as well as the Top10NL the strange situation would appear that 2 separate authentic registrations show the same data at different scales with different specifications.

If the decision to make the GBKN an authentic registration is made, this would mean that the government has to take full control of the GBKN and has to formalise it. This would also mean that the government still uses two different datasets to make two different maps.

The municipality of Rotterdam wants to satisfy the requirements of the authentic registrations and wants to be as efficient as possible. To be able to connect the building registration to the topographic registration the GBKN is obviously needed, whereas the Top10NL has to be collected as well. The topographic registration could still be collected one time instead of multiple times when generalisation from the GBKN to the Top10NL is applied. In this way the map only has to be collected at the largest scale and from this other scales can be derived.

In chapter 3 we will see different methods to solve this generalisation.

3. Mainstreams of map generalisation

This chapter will give an introduction to map generalisation and with that introduce the best known scientific methods of map generalisation. These mainstreams of map generalisation are:

- Multiple representation databases
- On-the-fly generalisation
- tGAP structure

This chapter will give an overview of these types of generalisation and will take a look at current methods to perform generalisation. To be able to do so this chapter starts with an overview of the problems we face in non-generalised maps in section 3.1. Also the mainstream methods of map generalisation will be introduced in this section. Section 3.2 shows actual solutions to the problems indicated in 3.1 and open research issues. Finally section 3.3 gives an overview of the open issues and projects on generalisation in The Netherlands.

3.1 Introduction to map generalisation

This section gives an introduction to map generalisation and the main solution directions to the problem. Subsection 3.1.1 gives a general introduction to map generalisation, whereas in subsection 3.1.2 the three main directions of automatic map generalisation will be described.

3.1.1 Map generalisation

To be able to say something about which generalisation methods are needed in Rotterdam, it is necessary to first describe what generalisation is. The formal definition given by ICA is the following (ICA, 1973):

“Generalisation is the selection and simplified representation of detail appropriate to the scale and/or purpose of the map”.

The right level of detail in which information is presented at a certain scale makes the map readable and appropriate for its purpose. Generalisation is the way to obtain this.

Generalisation used to be performed by humans. This is very labour intensive work which we would very much like to pass on to computers. The problem with this is that the generalisation work done by human operators is very complex and subjective; therefore it is hard to formulate it in computer rules. The cartographer’s decisions should be really understood in order to automate the generalisation process (Mc Master & Shea, 1992).

Manual generalisation can be applied in many different forms. Therefore subjectivity is an important issue in map generalisation. Figure 3.1 shows 2 examples of different forms of generalisation in the delta of the river Mississippi and in the Isle of Skye. We see that different mapmakers make different decisions about the importance of certain aspects. This makes it hard to set very specific rules for automatic generalisation. Everyone applies generalisation according to his own preferences.

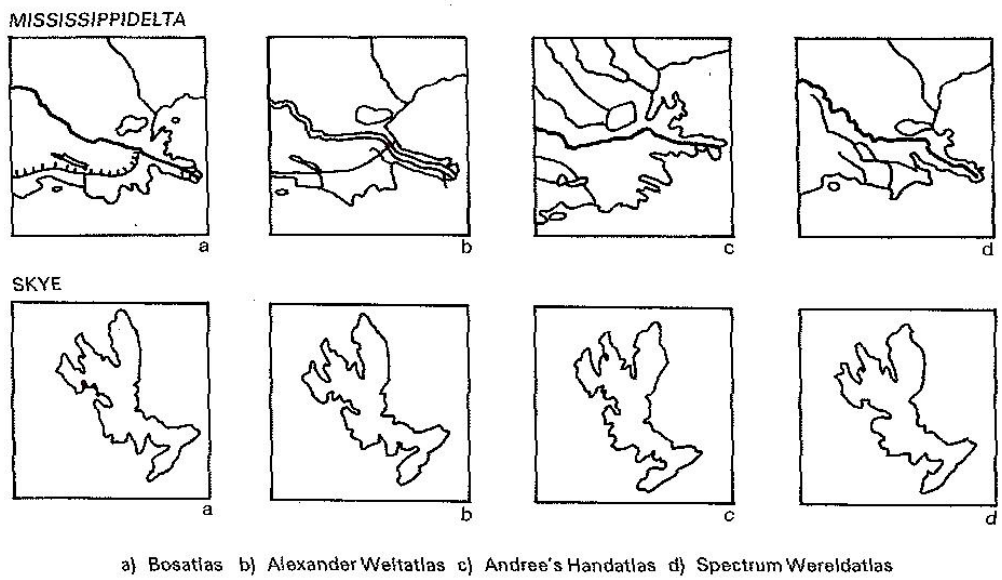


Figure 3.1: The same places according to different atlases. (Courtesy: Ormeling & Kraak, 1993)

We see that automatic map generalisation should be about a set of rules, which are not applied in the same way by every cartographer. But how to define these rules? For this we need operators. To be able to say something about how automatic generalisation should be applied, we need to know more about the operators.

The main difficulty in automatic digital generalisation is not the making of these rules to which a geographic data set has to apply, but the order in which they should be executed. According to Mc Master & Shea there are six theoretical elements to which generalisation rules have to apply:

- Reducing complexity
- Maintaining spatial accuracy
- Maintaining attribute accuracy
- Maintaining aesthetic quality
- Maintaining a logical hierarchy
- Consistently applying generalisation rules

With a lower scale level the *complexity* of the map increases. Too much and too complex graphical elements have a negative impact on the efficacy of the map; it then doesn't communicate effectively the main things.

The *spatial accuracy* of objects should be maintained as good as possible. In small scale maps this is impractical, because the map should communicate the most important objects in a clear way and make other objects subordinate to that, even in terms of spatial accuracy. This can mean that an important object could push another object aside because of its relevance. With this spatial accuracy can not always be maintained.

Maintaining attribute accuracy is about keeping the right attributes to the right objects. In case of a city all objects in the city can be generalized to one object with the city name, they then have to be categorized as that city name in order to let the generalisation perform well.

The art in maps should be retained. Therefore the *aesthetic quality* of the map should be maintained during generalisation. Cartographers deal with making guidelines for this, but so far in generalisation only general guidelines for the aesthetic quality are applied.

A map should maintain its *logical hierarchy*, an ordering of mapped features. With this hierarchy the reader should be able to know which objects are more relevant than others.

To *consistently apply the generalisation rules* not only the algorithms need to be good. It needs to be processed good as well. This is where most of the errors are made. To obtain a consistent generalisation the cartographer needs to determine three things:

- Which algorithms to use
- The order in which to apply these algorithms
- The input parameters needed to obtain a given result at a given scale

As we already saw in figure 3.1 the subjectivity of generalisation rules already shows that consistency is the main challenge in generalisation. Figure 3.2 explains different types of operators (Galanda, 2003; quoted in Meijers, 2006). In the table all different operators are mentioned and their action is briefly explained. However, having operators doesn't automatically mean we now how to automatically generalise a map. Having defined operators doesn't mean we directly have algorithms. For some of the operators a lot of algorithms are defined in literature. The way to produce algorithms and the order of execution of the algorithms is subject to research.

After this general introduction to map generalisation we go deeper in the material. In the upcoming subparagraph the main directions of current generalisation research are stated to get an impression on how to solve this problem.

Reclassification	Changes the category an object belongs to and possibly combines it with neighbouring objects of the same class
Aggregation	Combines an object with other objects of the same or a similar class to a new object
Displacement	Denotes the movement of an entire object; its shape remains unchanged
Enlargement	Denotes a global increase of an object
Exaggeration	Defines a local increase of an object, its shape is distorted
Collapse	Type of geometry is changed, possible changes are a polygon to a line or to a point and a line to a point
Elimination	Removes an object from the data set, the freed space is assigned to other categories
Simplification	Reduces the granularity of an outline
Smoothing	Improves the visual appearance of an object's outline
Typification	Reduces the complexity of a group of objects by removing, displacing, enlarging and aggregating single objects, maintaining the typical object arrangement

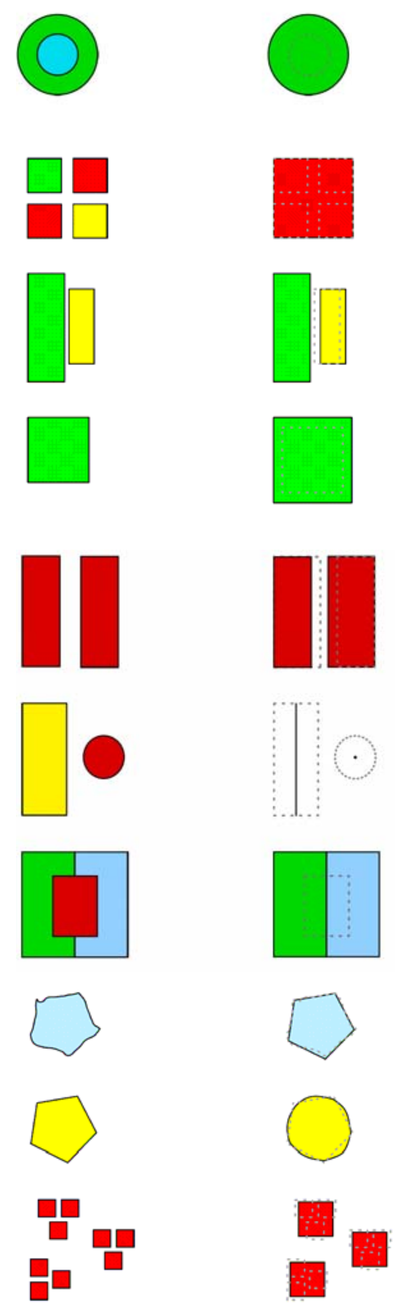


Figure 3.2: Different generalisation operators (Courtesy: Galanda, 2003 and Meijers, 2006)

3.1.2 Main map generalisation approaches

In this subsection we take a look at three general approaches to solve the automatic generalisation problem. These approaches are not yet ready-made solutions. The solutions, given in section 3.2, are derivations of the general approaches described in this subsection.

The general approaches that will be mentioned in this subsection are:

- Multi-Representation Data Base
- On-the-fly generalisation
- tGAP structure

Multiple Representation Data Bases

Multiple Representation Databases (MRDB) are databases that want to represent and manage geographical data at different scale levels. In an MRDB there are different representations of the same reality.

Some map producers, in Rotterdam as well, keep a separate database per scale level, because they are not able to link these levels to each other. This leads to problems like a lack of consistency and uncertain update propagation between the scales (MurMur, 1999). In Rotterdam the scale levels are totally not related, which means that there is no propagation of updates between the scale levels and therefore there is a difference in actuality. The solution is to have a database which has the same information stored in different scale levels with links in-between.

Within a Multiple Representation Database you can store data collected at the largest source and represent it multiple times in a database at several other scales or linked representations. The optimal representation of the dataset at several scales is stored and when requested the database returns the representation at that level of detail closest to a scale saved in the database.

With an adequate number of representations stored this should enable fast and accurate representation of geo-information for every purpose. Typically, the appearance changes so much when the scale is increased with a factor two to four, that intermediate layers are necessary (Hampe et al., 2003).

If we take this rule into account there should be more intermediate layers between our IMGeo and Top10NL than currently available. This would mean that we need an intermediate level at around 1:5,000. This is not necessary according to the specifications of the GBKN, which say that the GBKN is collected for the scale level 1:1,000, but can be used within the range of 1:500 to 1:5,000. According to Hampe it is also necessary to have intermediate scales within this range.

Another main principle of the MRDB is the relation. Objects are stored in several layers, but are related to each other. Of course the objects that are stored differently in the database have a natural correlation, because they represent the same real world phenomena.

The necessity of links between layers is described in (Hampe et al., 2003). For GIS-analysis via different scales links between different scales offer more functionality for the user. For the user that only wants to visualise the spatial data, separate scales may be sufficient.

What is very important is that links between objects can provide the possibility to update maps only in the largest scale and let this change propagate through all different layers. In the WIPKA project an approach is done to enable updates through an MRDB (Anders & Bobrich, 2004).

On-the-fly generalisation

The primary element of on-the-fly generalisation is that generalisation algorithms are computed if a map at a certain scale is requested by the user. Like in all other cases the geographical data can be collected at the largest scale, but it is not stored at multiple scales as in the MRDB.

This makes that for on-the-fly generalisation not too much storage space is used. In the case of an MRDB geometry is stored many times to speed up the process; this is the disadvantage of on-the-fly generalisation. Because all the algorithms have to be processed 'on-the-fly' the data transfer can be very slow.

The research on on-the-fly generalisation can't be seen without the research on MRDB. Where MRDB focuses on fast presentation at some essential scales, on-the-fly generalisation focuses on the fast real-time execution of generalisation algorithms. The algorithms used do not have to differ from other generalisation algorithms; the difference is that have to be executed every time a certain scale is requested.

The problem of update propagation using this kind of generalisation doesn't exist. The data is inserted in the largest scale and from there the algorithms are only executed when requested. One of the main problems in on-the-fly generalisation is that the algorithms have to be executed in a specific order. They can't be executed at the same time. This gives problems in prioritising different algorithms because a lot of algorithm combinations are possible, which is an optimisation problem (Ware et al., 2003).

tGAP generalisation

The topological Generalised Area Partition (tGAP) structure is a generalisation method which is being developed at the TU Delft. This method is an implementation of batch generalisation, but it combines several aspects of the previous methods. There are links between the different scale levels as in the MRDB, but the map is not stored at all these different levels. Because the links are stored, the computations do not have to be done all the time again.

The tGAP structure supports a progressive data transfer. The tGAP structure is based on the GAP-tree (van Oosterom, 1993), which assigns all objects in a database a weight or importance factor, with which they compete for occupation with other objects. The moment an object is absorbed into another object is stored in the tree structure. The function to define which neighbour absorbs the erased object is called the class compatibility function. With all objects stored in a tree, the map is in the end generalised to only one object. The situation from a given importance ratio can be recalled by looking at what was the object with an importance lower than the one required. By looking at this aspect the situation with that importance can be viewed.

The tGAP structure is composed of a face tree, an edge forest, and Binary Line Generalisation (BLG) trees - one for each edge. Geometry is stored only for edges, not faces, in order to remove redundancy. References to left and right faces are stored for each edge, and are being used to build face geometry (for visualisation). When objects are modified a new object is created to replace the old objects with a new importance factor, computed from the old importance of the objects. With this change also the values for adjacent objects changes. To keep the objects in the tree structure unique this is necessary.

The face tree can be sorted in order of importance, the edge forest as well.
When entering a certain level of importance from the client side, the server only returns those objects to the client that are of the required importance.

3.2 Current solutions to the automatic map generalisation problem

This section provides an overview of current solutions to the automatic generalisation problem. The subsection starts with methods described in literature in 3.2.1 and ends with open problems the authors mention in their articles in 3.2.2.

3.2.1 Solutions from literature

The solutions that will be described in this subsection are:

- Least Squares Adjustment
- CHANGE
- AGENT
- GiMoDig
- WIPKA
- MurMur
- Web server for tGAP

Least Squares Adjustment

In geodesy the Least-squares Adjustment is a well known principle. The observations are written as a function of the unknowns; this results in a matrix, which connects the observables as good as possible to the unknowns. The least squares method wants to enforce that the residuals, which are the result of the modelling, are as small as possible. It is possible to assign a weight matrix to the function as well, for which also the covariance matrix can be used.

With the functions derived from this method lots of algorithms can be formed (Sester, 2005). When rules for generalisation are given these are inserted in the function as constraints in a least-squares algorithm, they are also called control parameters.

For example building simplification, exaggeration and displacement can be well executed with this method.

CHANGE

The program CHANGE is a generalisation program from the University of Hannover which has been studied by the municipality of Rotterdam as a possible solution to their generalisation problem (Sester, 2003).

The program CHANGE (**C**artography – University of **HAN**nover – **GE**neralisation Software) generalises buildings and road networks in different modules. The generalisation algorithms that are used are simplification, smoothing and filtering (Bobrich, 2001).

For the buildings first a plausibility check is done which checks for self crossing objects. After the execution of the algorithms problem areas are detected automatically which can be generalised manually.
For the road network the program generates a node and edge structure, which is the basis for smoothing. Only the middle axes of the roads are used for this procedure, this middle axis is automatically extracted from the road objects.

The methods have proven to be successful in the federal state of Niedersachsen, where the 1:25,000 map is currently produced according to this method from available 1:5,000 and 1:10,000 maps.

AGENT

The project AGENT (**A**utomated **GE**neralisation **N**ew **T**echnology) handles all objects on its own and let them think for their selves. The objects are active agents with a set of constraints. After applying several generalisation algorithms every agent can measure to what extend the constraints are met or violated. Two types of agents can be distinguished: micro-agents, which monitor individual objects and meso-agents, which monitor objects in larger contexts.

Actually, the project AGENT ended in a method which is not ready to be implemented, but is suitable as a basis for applications. One example of an AGENT-based application is Clarity, which makes use of an implementation in Java (Neuffer et al., 2004).

The AGENT project is especially interesting, because it takes into account all aspects of generalisation, including cartographic generalisation operators like enlargement, exaggeration and displacement.

GiMoDig

One of the applications of the MRDB is the EU-project GiMoDig (**G**eospatial **i**nfo-**Mo**bility Service by Real-time **D**ata **i**ntegration and **g**eneralisation).

The vision of this project is to let a mobile user in Europe receive on-line information about the environment. This requires access to national databases and representation of these data on different scales in order to give the user flexibility in orienting himself on the environment.

The cartographic data in GiMoDig is stored in an object-relational PostgreSQL database. The largest level of detail is the 1:10.000 scale, the derived scales are 1:25k, 1:50k and 1:100k.

Some applications of GiMoDig are described in literature: the variable scale map, which in this case means that objects are projected in one screen with different scales (the scale is larger in the centre of the map than it is on the edges), the emphasizing of objects through symbolization and the assigning of attributes to linked objects are examples of these applications. GiMoDig mainly focuses on the use of small display applications (Hampe et al., 2004).

WIPKA
The project WIPKA is a funded German project from the German Federal Agency of Cartography and Geodesy (BKG). The project is executed by the universities from Munich and Hannover. The project has already implemented an MRDB in ESRI ArcGIS after developing a tool in VBA (Visual Basic Applications for Microsoft) and is researching possibilities to do update propagation in the MRDB (Anders & Bobrich, 2004).

MurMur
MurMur is a generalisation project executed by a large consortium of universities. The focus in this research is on developing methods for multiple representation and multiple resolution.

The overall objective of the MurMur proposal is to demonstrate that current functionality, provided by commercial data management software (DBMS or GIS), can be extended to support more flexible representation schemes. In this case the project makes the distinction between semantic and cartographic flexibility. Semantic flexibility is about the multiple representations of the same real-world data at the same scale; here some attributes can be for example exaggerated to communicate that part of the information very clear to the user. Cartographic flexibility focuses on flexibility in scale.

Web server for tGAP
Recently tests with the tGAP data structure are done by Meijers during his master thesis research (Meijers, 2006). The objective of this research was 'to verify the GAP-face tree and GAP-edge forest in theory and test the functionality and performance, in terms of time and storage requirements, of the tGAP structure for on-the-fly database map generalization.'

The implementation is done with test datasets from the Dutch Cadastre and the municipality of Amsterdam. Furthermore a visualisation of the data in Google Earth is made. For this purpose code has been written which can be applied to all kinds of different datasets. Meijers is currently assigned to the TU Delft as a PhD researcher to extend this structure.

3.2.2 Open problems

The open problems that still exist in automatic map generalisation will be mentioned in this subsection. It takes as starting point the open issues mentioned by Galanda (Galanda, 2003), but will also mention other non-solved problems.

In his thesis Galanda (2003) mentions all kinds of generalisation operators, which are listed in section 3.1. In the previous subsection we already saw that some of these operators are well covered by current solutions. Simplification (Least squares, CHANGE) and displacement (AGENT) are operators which are well defined in literature.

Galanda points in his thesis that there is a need to set up shape simplification algorithms and not only algorithms that simplify the individual lines surrounding the shapes. Galanda also mentions that a typification algorithm would benefit to automatic generalisation.

The order of execution of the generalisation algorithms still stays a point of interest according to Mc Master and Shea (1992). The subjectivity of the cartographers also doesn't make it easy to create rules that are fixed (Ormeling and Kraak, 1993).

When talking about on-the-fly generalisation mostly new algorithms are needed to progress in research. The order of execution can be considered to be a part of the algorithm-making. Open problems in the MRDB are mainly on update propagation (Anders and Bobrich, 2004) and also on creating new algorithms.

The tGAP structure is now developed for a small range of operators. It is still an object of research within the group GIS technology at TU Delft. Not only more algorithms will be added to this structure to make it more complete, also the way to define importance values to objects and compatible neighbour classes is still part of this research.

As shown in this subsection there is enough research possible within the whole spectrum of automatic map generalisation. Not only fundamental research is done in the field of map generalisation, also applied research is done in The Netherlands with the datasets presented in chapter 2. The next section will take a look at this research.

3.3 Related projects in The Netherlands

This section takes a look at two projects in the field of map generalisation in The Netherlands. These projects are both executed by consortia, which will be introduced in this section. Subsection 3.3.1 is about the project TopNL, subsection 3.3.2 is about a part of the Ruimte voor Geo-informatie (RGI)-project DURP ondergronden.

3.3.1 TopNL

The project TopNL is a joint project by the Dutch Cadastre (Kadaster), ITC Enschede (International Institute for Geo-Information Science and Earth Observation) and TU Delft. Its aim is to combine the small scale datasets of the Kadaster through generalisation. All object classes should form at certain scale levels a logical and consistent set of topographical elements. The model therefore needs to know what classes need to be displayed at what scale and what level of detail is requested for an object class at a certain scale.

The scales that have to be modelled within TopNL are at least 1:10k, 1:25k, 1:50k, 1:100k, 1:250k, 1:500k and 1:1000k. These scales are necessary for the law on the authentic registrations, the scales 1:250k and 1:1000k are also needed for the European products respectively EuroRegionalMap and EuroGlobalmap.

The project has defined some requirements to which the model has to satisfy. A requirement of the Kadaster is that the model of Top10NL should be unchanged. Other requirements are requirements with respect to the generalisation procedures. One other requirement is that TopNL should not only be suitable for the scales mentioned, but should have a possible vario-scale output for future products (Stoter e.a., 2007).

The master thesis project of the author of this report is about generalisation of even larger scales than mentioned in this project. Because generalisation from a level 1:1,000 is even more interesting than when starting from 1:10,000 the master thesis project is seen as an interesting addition to this generalisation project.

3.3.2 DURP Ondergronden

Another interesting development is to be found in the project DURP Ondergronden. This project is mainly about generalising topographical planning maps. It is executed by a large number of parties, which are:

- Bentley Systems Netherlands
- ESRI Nederland
- ITC Enschede
- Kadaster
- Landelijk Samenwerkingsverband GBKN (LSV-GBKN)
- NedGraphics
- Sense Organisatie & Coaching
- Technische Universiteit Delft

The objective of this research project is “to generate and use base maps for integrated querying of digital physical plans”. Because the research was mainly focussing on maps of the Kadaster a subproject is defined in which the LSV-GBKN can participate. This subproject is about generating Top10NL from IMGeo, which is exactly the theme of the master thesis as well.

4. Analysis of the Rotterdam situation

This chapter will discuss the findings of the previous chapters. The information collected in this chapter is brought together here. For the generalisation problem of the municipality of Rotterdam three generalisation methods have been described in chapter 3, these are:

- Multiple Representation Database (MRDB)
- On-the-fly generalisation
- tGAP

In this chapter a choice between the methods will be made and explained. In section 4.1 criteria will be chosen; they will be analysed in section 4.2. In this section also the choice for one of the methods is made. Section 4.3 introduces the proposed way to implement this method during the master thesis.

4.1 Criteria choice

The criteria on which the decision to continue with a certain generalisation method are derived from the descriptions in chapter 2 of this report, in which the current situation of the geo-information provisions in Rotterdam were described. Also criteria are added which make the research suitable for graduation at TU Delft

Because not all criteria will be equally important, every criterion gets assigned a weight in the decision making. This weight and the associated reasons will be explained below. The criteria on which the best generalisation method for Rotterdam will be based are:

- Speed
- Scale independence
- Object linking
- Storage space
- Update propagation
- Contribution to research
- Flexibility
- Direct contact for research assistance
- Connection to TopNL and DURP

The speed of the application is very important. The application has to handle huge amounts of data, if the data should perform very slowly when requested at a certain scale, this would be worse to work with than it is now.

Scale independence is important, because generalisation wouldn't be necessary if maps weren't dependent on scales. A map is called scale independent if the scale and the representation are always complementary. This is not the case within MRDB, which has fixed scales.

The object linking is only done in the MRDB and tGAP and not in the on-the-fly generalisation. Within on-the-fly generalisation the linking of objects can be seen as optimal, because the link is made directly when requested.

The amount of storage space can be optimalised through generalisation. Currently the maps are made at 4 different scales. This might be reduced through this master thesis.

The update propagation is important, because we want to update only the largest scale data set in the end. These updates should propagate in other scales and representations.

There should be enough new research content in the master thesis. As shown in chapter 3 there are enough open issues to get a good research content with all methods.

The flexibility of a method defines whether it is possible to change things to the structure. As shown there are always research possibilities, but can new ideas be implemented in the structure?

If a method is chosen it would be good to have direct access to assistance to implement it. In this respect it would be very good to opt for the tGAP, as this method is developed at the TU Delft. This criterion is valued very high.

It would be very good if the chosen method would be somehow related to the projects mentioned in section 3.3, TopNL and DURP Ondergronden.

In section 4.2 these criteria will be analysed and a final judgement will be given.

4.2 Criteria analysis and method choice

This section describes how a choice for one of the three methods is made based on the prior mentioned criteria. The judgements mentioned are the authors judgements based on the knowledge of earlier chapters. The criteria are mentioned in the order of importance, which means that the most important factor is to be mentioned first.

One of the most important factors in the decision making is the *speed* of the application. Because Rotterdam is dealing with an enormous amount of data, the organisation can't afford to have a slow operating application. On this criterion the on-the-fly generalisation methods score worse than the others. The MRDB is fast, because of the predefined representations; the tGAP structure is fast, because of its progressive data structure.

The *contribution to research* and the direct access to *research assistance* are taken into account together and are valued very high. Because research at TU Delft focuses on the tGAP structure and assistance in implementation are therefore very nearby, the tGAP structure is very suitable for this master thesis. Because generalisation is mostly performed in projects which are done for third parties, algorithms are not always available for further research.

Scale independence is of importance, because a vario-scale solution gives far more freedom within the representations. On-the-fly generalisation and the tGAP are in this respect better than the MRDB.

The *update propagation* and the *object linking* are somehow related to each other. In the on-the-fly generalisation methods this is not necessary, because in these applications the data is only stored at the largest scale. The target scale is defined on-the-fly and the update propagation is no problem, because algorithms are only executed with the data of the latest update at the largest scale.
In MRDB objects are specifically linked to each other, this makes that this criterion is fulfilled for MRDB as well. On update propagation in MRDB's some research has been done in the project WIPKA, but this is still a research topic. If the MRDB is to be implemented in Rotterdam, the links between the different object scales first have to be established, for there are none.
The tGAP structure also links objects to each other in the BLG-tree structure. On update propagation nothing has been done yet.

On *storage space* the MRDB scores worst, because for every representation all objects are stored. Currently this is also the case in Rotterdam, but it doesn't lead to large problems if we leave it this way. On-the-fly generalisation and tGAP are better at this point; in both cases the data is only stored at the largest scale. In the tGAP structure also the links are stored to be able to do the progressive data transfer.

The *flexibility* of the structure is not very good developed within the tGAP structure. This is a method which is under development, but which can't handle very much generalisation operations at this moment. Although some operators are being developed at this moment, like a collapse operator (Dilo, 2007), it is currently unknown whether operators like displacement will perform well within the tGAP structure.

The *projects TopNL and DURP* are not part of the master thesis, but if it is possible to cooperate this is a good thing to do. At this moment it is not known to which of the methods the projects most resemble, because they have just started.

Final judgement
The above judgement is based upon the theoretical background of chapter 3. The judgements are weighed qualitatively and based on the prioritising of criteria described in this section.

The result of the analysis is that the tGAP structure turns out to be the most promising method to perform research on in Rotterdam.

The tGAP structure is a promising method with a lot of research possibilities. In section 4.3 the ideas for new research to the tGAP are introduced.

4.3 The constrained tGAP structure

The constrained tGAP structure is an idea developed by Jan Haunert from the University of Hannover and the TU Delft. The idea of the constrained tGAP tree is that the tGAP is not built from the largest scale alone, but that it is built between two independent map scales. Between these map scales the tGAP structure can than be built to show which objects are aggregated.

Statistics can be made from the overlay of the map fragments; they show which object combinations are aggregated more than others. With this information on aggregations the importance factors can be better determined which in the current tGAP have to be estimated.

From this constrained tGAP structure general rules can be defined for the generalisation parameters. The related goal in this is to get to a situation in which the tGAP tree can be better built and maintained without constraints.

For the master thesis the 1:1,000 IMGeo map and the 1:10,000 Top10NL are used to build this constrained tGAP. IMGeo is in this case the basis and Top10NL forms the initial constraint. The representation resulting from the constrained tGAP at 1:10,000 should than be acceptable with respect to the current 1:10,000 map.

5. Conclusions

The aim of this report was to answer the following question:

What is the best method to perform map generalisation for the municipality of Rotterdam?

The municipality of Rotterdam wants to find a way to be able to only update the large scale base map (GBKN), based on IMGeo and from that derive the small scale base map (Top10NL) through generalisation.

Three methods have been investigated to see if they are possibly able to solve this matter. These generalisation methods are:

- Multiple Representation Database
- On-the-fly generalisation
- tGAP structure

Through a criteria analysis the best method to perform generalisation in Rotterdam has found out to be the tGAP structure.

The most important advantages of the tGAP structure are:
- Fast (allowing progressive data transfer)
- Scale independent
- Geometry is stored only once

The technology is developed at the TU Delft, which gives good opportunities for assistance during research. As a new contribution to research the following master thesis will investigate the possibilities of a constrained tGAP structure, in which IMGeo is the basis and Top10NL the initial constraint.

References

Anders, K.-H., Bobrich, J., *MRDB approach for automatic incremental update.* ICA Workshop on Generalisation and Multiple Representation, Leicester 2004, 2004.

Bobrich, J., *Cartographic Map Generalisation in Urban Districts.* In Proceedings of GIS Research UK, 9th Annual Conference, pp. 513-515, 2001.

CHANGE Kurzdokumentation. Hannover: Institut für Kartographie und Geoinformatik, (on-line). Universität Hannover, 2003.

Dilo, A., *Server side technology for tGAP.* Technical report for MobiMap project, 2007.

Galanda, M., *Automated Polygon Generalization in a Multi Agent System.* PhD thesis, University of Zürich, 2003.

Productinformation GBKN (in Dutch), In: www.gbkn.nl, LSV GBKN (on-line), 2006.

Hampe, M., Anders, K.-H., Sester, M., *MRDB applications for data revision and real-time generalisation.* In: Proceedings of the 21st International Cartographic Conference. Durban, 10.–16. August 2003, pp.192–201, 2003.

Hampe, M., Sester, M., Harrie, L., *Multiple representation databases to support visualisation on mobile devices.* In: Proceedings of the 20th ISPRS Congress, July12-23, 2004, Istanbul, Turkey: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 135-140, 2004.

ICA, *Multilingual Dictionary of Technical Terms in Cartography.* Franz Steiner Verlag, Wiesbaden, 1973.

IMGeo version 2.3 (in Dutch), Concept; not published yet, 2007.

Landelijk Samenwerkingsverband GBKN, *GBKN towards an authentic registration? (in Dutch),* Documentnr. 02.05/053, 2002.

Mc Master, R.B., Shea, K.S., *Generalization in digital cartography,* Association of American Geographers, 1992.

Meijers, M., *Implementation and testing of variable scale topological data structures,* Master Thesis TU Delft, 2006.

Müller, J., Lagrange, J., and Weibel, R., *GIS and Generalization: Methodology and Practice.* Number 1 in GISDATA. Taylor & Francis, 1995.

Neuffer, D., Hopewell, T., Woodsford, P., *Integration of Agent-based Generalisation with Mainstream Technologies and other System Components,* ICA Workshop on Generalisation and Multiple Representation, Leicester 2004, 2004.

Ormeling, F.J., Kraak, M.J., *Cartography: Visualisation of Spatial Data (in Dutch),* Delft: Delftse Universitaire Pers, 1993.

Oosterom, P. van, *The GAP-tree, an approach to "on-the-fly" map generalisation of an area partitioning,* In: volume GIS and Generalization: Methodology and Practice of GISDATA, chapter 9, pages 120–132. Taylor & Francis, London, 1993.

Oosterom, P. van, *Variable-scale topological data structures suitable for progressive data transfer: the GAP-face tree and GAP-edge forest*, In: Cartography and geographic information science, 32, pp. 331-346, 2005.

Oosterom, P. van, Vries M. de, Meijers, M., *Vario-scale data server in a web service context,* In: Proceedings of the ICA Commission Workshop on Map Generalisation and Multiple Representation, June 25th 2006, 14 p., 2006.

Rietdijk, M., Verhoef, J.J., *Address unknown (in Dutch),* Uitgever Programma Stroomlijning Basisgegevens, Den Haag, 2002.

Sester, M., *Optimization approaches for generalization and data abstraction.* International Journal of Geographical Information Science 19(8-9), pp. 871-897, 2005.

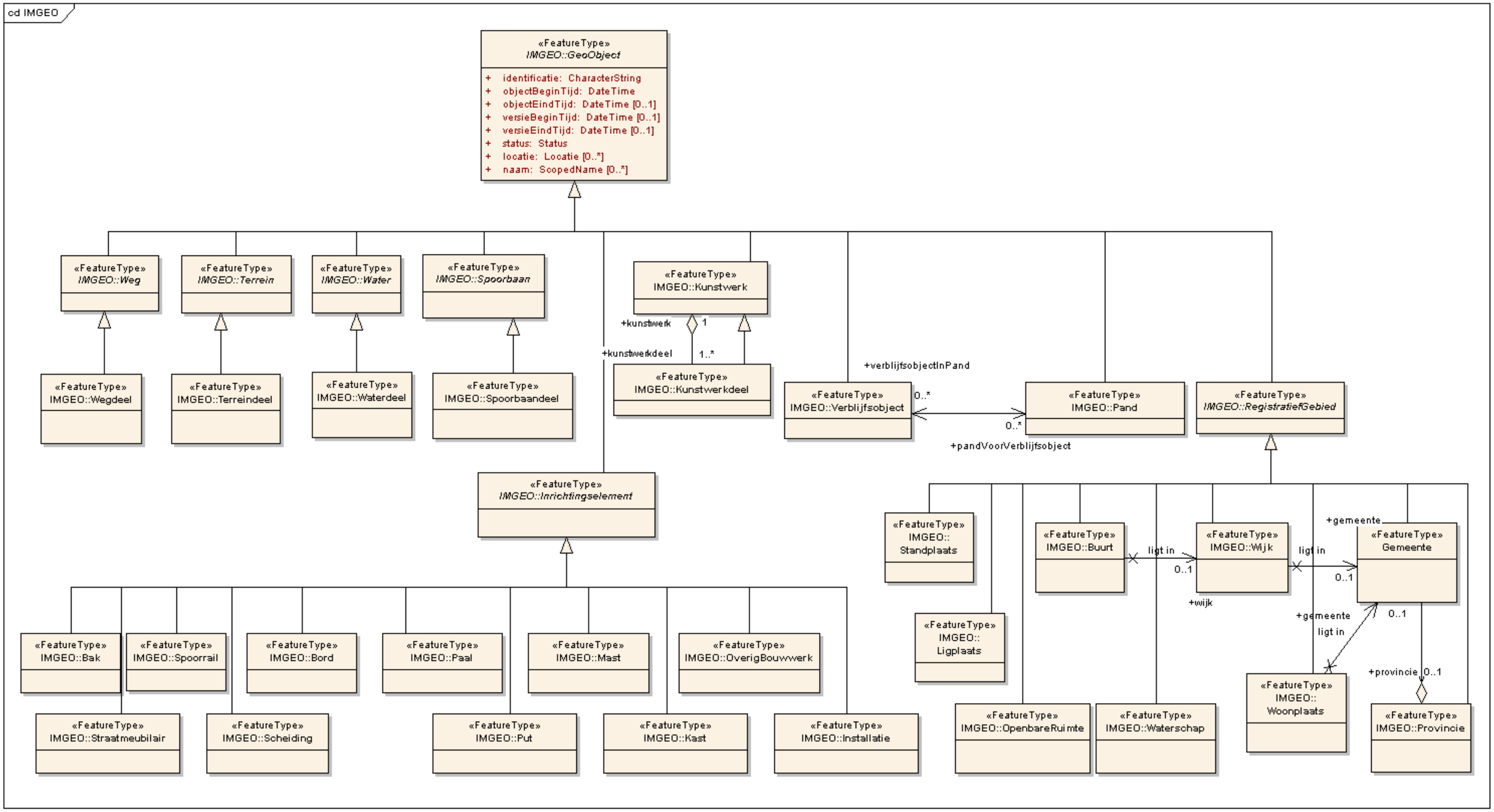
Spaccapietra, S., Vangenot, C., Parent, C., Zimányi, E*., MurMur: A Research Agenda on Multiple Representations.* In: International Symposium on Database Applications in Non-Traditional Environments, DANTE 1999. 28-30 November 1999, Kyoto, Japan. IEEE Computer Society 1999, ISBN 0-7695-0496-5, pp. 373-384, 1999.

Stoter, J., Quak, W., Van Oosterom, P.J.M., Meijers, M., Lemmens, R., Uitermark, H., *Considerations for the design of a semantic data model for a multi-representation topographical database.* In: Proceedings, ISGI 2007, International Symposium on Generalisation of Information, 2007.

Stoter, J., Lemmens, R., Köbben, B., *Preliminary investigation TopNL (in Dutch),* International Institute for Geo-information Science and Earth Observation (ITC), 2007.

Ware, M.J., Jones, C.B., Thomas, N., *Automated map generalization with multiple operators: a simulated annealing approach.* International Journal of Geographical Information Science, vol 17(no. 8): pages 743-769. December 2003, 2003.

Appendix A: UML diagram of IMGeo



Appendix B: UML diagram of Top10NL