Management of multiple representations in spatial DBMSs

S. Zlatanova, J.E. Stoter and W.Quak
Section GIS technology, Delft University of Technology
Delft, The Netherlands
Tel.: +31 15 2784268, Fax: +31 15 2782745
{s.zlatanova|j.e.stoter|w.quak}@otb.tudelft.nl

ABSTRACT

Different representations of the same real world objects are often maintained in separated DBMSs because datasets are created and maintained for different purposes. When the links between these representations are explicitly defined and managed in a DBMS environment, data consistency and integrity can be significantly improved. Furthermore the quality of the data content can be improved because of update propagation and error checks between different data sets. This paper discusses important issues that need to be solved before a management system for multiple representations can be established.

The paper addresses several important aspects related to maintenance of multiple representations, automatic interrelations between two and more representations of one real-world object and automatic generation of low-resolution representations from high-resolution. The results of a case study are described in detail. The paper concludes on topics for further research.

KEYWORDS: multi-representations, spatial DBMS, relationships, spatial operations

INTRODUCTION

An increasing number of DBMSs offer management of spatial objects and increasing number of applications make use of spatial objects maintained in a DBMS. Therefore the representation of real world objects in a DBMS is becoming a very important issue in modelling the real world.

In general, spatial data sets are created in order to provide information for specific applications. Usually, these sets are specific representations of real world objects, which are collected by distinct organisations. Each set of objects is a selection of real world objects (only the objects of interests are modelled) and therefore subjective (the objects are represented from the point of view of the specific application). Consequently, different representations of the same real world objects can be found in a variety of data sets. The intuitive question then is 'how distinct applications can make a maximal use of these representations'?

A lot of research on multiple presentations and generalisation related to (spatial) DBMS is already available in literature (Buttenfield and DeLotto, 1989; Devogele et al., 1996; Friis-Christensen et al., 2002; Grefen and Widom, 1997; Jones et al., 1996; Li and McLeod, 1992; Sheth and Larson, 1990; Spaccapietra et al., 1999), which is an indication for the significance of the issue.

In this paper we argue that data consistency and integrity would be significantly improved if the management of multiple representations is organised in a DBMS, at least at conceptual level. This means that different representations of the same real world object are stored and maintained in different DBMSs and interrelated together with the rules ensuring the integrity, consistency and transition from one representation to another. Both the users and the system should be aware of multiple representations of the same real world objects. Data consistency of different representations will be enforced through appropriate, automatic update propagation.

This paper discusses several important issues, which need to be solved before a management system for multiple representations can be established. The paper concentrates on both geometric representations and thematic meaning that are related to the same real world objects in multi-scale and multi-purpose environments (i.e. environments in which several date sets represent the same real world objects at different scales and for different purposes). The paper addresses three important aspects related to:

- Approaches for maintenance of multiple representations in a DBMS environment, i.e. one detailed representation from which all low-resolution representation can be derived; vs. maintenance of several different representations.
- Automatic generation of interrelationships between datasets by applying spatial operations available in DBMS.
- Automatic generation of low-resolution datasets from high-resolution representations using spatial operations offered by DBMS.

These aspects have been examined by carrying out a case study. The results of the case study will be described to illustrate how multiple representations can be managed in a DBMS. The case study has been based on an object-relational DBMS approach using Oracle Spatial 9i (Oracle, 2001).

MAINTENANCE OF MULTIPLE REPRESENTATIONS

Datasets are responsibilities of mapping agencies and governmental bodies (e.g. Municipalities, Water Boards, Provinces, Cadastres). At this moment many datasets already exist that differ not only in scale (geometry) but also in thematic aspects (since all of them are dependent on the specific goal of the dataset). One typical example is representation of houses in a Building Registration and in a Cadastral Registration. In reality, both registrations may refer to the same real world objects. However since the objective of both applications is different, the houses in the Building Registration will be divided into living units while the houses in the Cadastral Registration will be divided into property units. The datasets may therefore maintain spatial objects of the same real world objects (even at the same scale), which cannot (always) directly be derived from each other by general algorithms.

The optimal solution from a DBMS point of view may be having only one representation (at scale 1:1) of all real world objects in a centralised DBMS. Datasets at lower resolutions can be derived and represented in DBMS-views by applying generalisation algorithms. However, the problem is not as simple as that. Spatial objects are represented in the database with their geometry, relationships (topology) and have a specific theme. Thus the generalisation can be performed according to the geometry, theme, relationships or combinations of them. Furthermore, depending on the scale, the geometry may change drastically. For example, a road in scale 1:1000 will be represented as polygon (2D), while it will shrink in a line (1D) in scale 1:50 000.

Moreover, even looking only at simplification of geometry (without theme differences), no ultimate set of algorithms exists at this moment, which can generate a dataset at a required small scale based on a data set at a larger scale.

The existence of different datasets (created for different applications) as well as current technical generalisation problems, require research how to establish and manage a fundamental link between the spatial objects in different data sets and eventually provide mechanism to derive them from each other (if and when possible).

AUTOMATIC GENERATION OF INTERRELATIONSHIPS

Spatial operators can be used to find the spatial relationship between two different representations of the same real world object (see Figure 1).

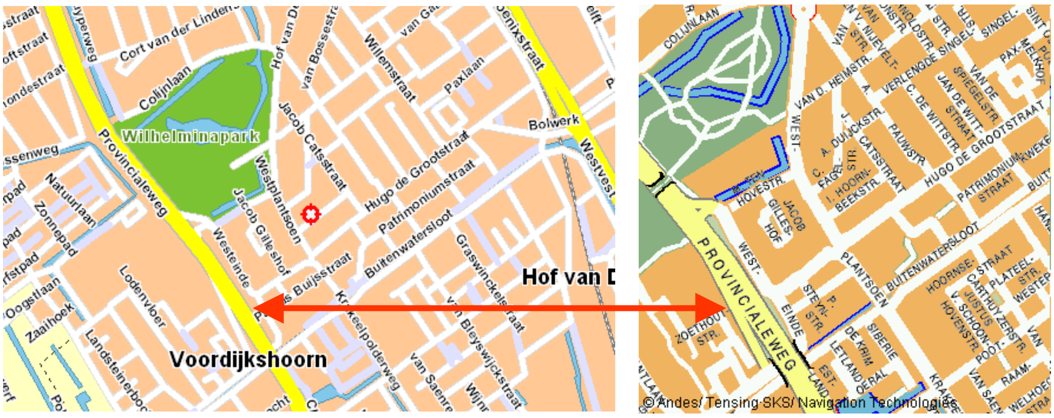

Figure 1: Relationship between two spatial objects, which refer to the same real world object (i.e. a road).

The OpenGIS Abstract Specifications (OGC, 1999) distinguish seven binary spatial relationships (contains, intersects, equals, union, intersection, difference, symmetricDifference) which are suggested for the geometry schema. Concerning binary topological relationships three frameworks are accepted by OGC as fundamental:

- Boolean set of operations (considering intersections between closure and exterior)
- Egenhofer operators (taking into account exterior, interior and boundary of objects)
- Clementini operators using the same topological primitives as Egenhofer but considering the type of the intersection

The currently available Implementation Specifications for SQL (OGC, 2001) implement topological relationships on the geometrical structure. The Implementations Specifications specify eight relationships based on the Egenhofer framework, i.e. equals, disjoint, intersects, touches, crosses, within, contains and overlaps.

Mainstream DBMSs have implemented the OGC Implementation Specifications which means that the spatial operators defined by the Implementations Specifications are available within DBMSs (Stoter and Zlatanova, 2002). These spatial operators can be used to find the relationship between two representations of the same real world object. This will be illustrated in the case study.

GENERATING LOW RESOLUTION SPATIAL OBJECTS WITHIN A DBMS

Although it was argued that research on multiple representations should focus on establishing links between existing datasets, research how to produce spatial datasets based on a spatial dataset at a higher resolution will always be necessary, for example within one mapping agency. Depending on the application a building can for example be represented by a point, an area or even a 3D representation (see Figure 2) (Stoter and Ploeger, 2003).

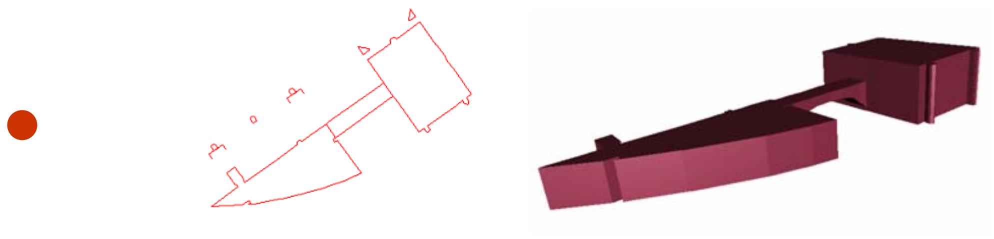

Figure 2: Different representations of the same building (point, polygon, polyhedron)

Traditionally, cartographers have to deal with producing maps at different scales. Till recently, datasets at different scales were maintained in separated databases. The lack of interrelationships between objects in different datasets referring to the same real world object led to the problem of not being able to keep consistent datasets at different scales. Moreover the datasets are produced separately, which requires the same (measuring and/or) digitising effort for different datasets while spatial objects referring to the same real world objects are produced.

Algorithms needed for geometric and thematic generalisation should take into account several aspects: change of dimension (e.g. going from a polygon to a line or a point), selection of the most important features (e.g. polygons need to have a certain area, otherwise they are deleted), simplification of boundaries of objects (e.g. polygons in scale 1:1000 should contain larger number of points compared to the same polygons in scale 1:50 000), aggregation of a collection of geometries, e.g. based on similar themes. Aggregate functions are not defined within OGC but implemented already in some mainstream DBMS. As an example the following aggregate functions are supported by Oracle Spatial 9i:

- SDO_AGGR_CENTROID: Returns a geometry object that is the centroid ("center of gravity") of the specified geometry objects
- SDO_AGGR_CONVEXHULL: Returns a geometry object that is the convexhull of the specified geometry objects
- SDO_AGGR_MBR: Returns the minimum bounding rectangle of the specified geometry objects
- SDO_AGGR_UNION: Returns a geometry object that is the topological union (OR operation) of the specified geometry objects.

CASE STUDY

The aim of the case study (Zlatanova et al 2003, Binkhorst and Zlatanova, 2003) was to show: a) how interrelationships between spatial objects in two existing datasets referring to the same real world objects can be established using a DBMS approach and b) how a small scale map can be derived automatically from a dataset at a larger scale using operators which are available in DBMSs.

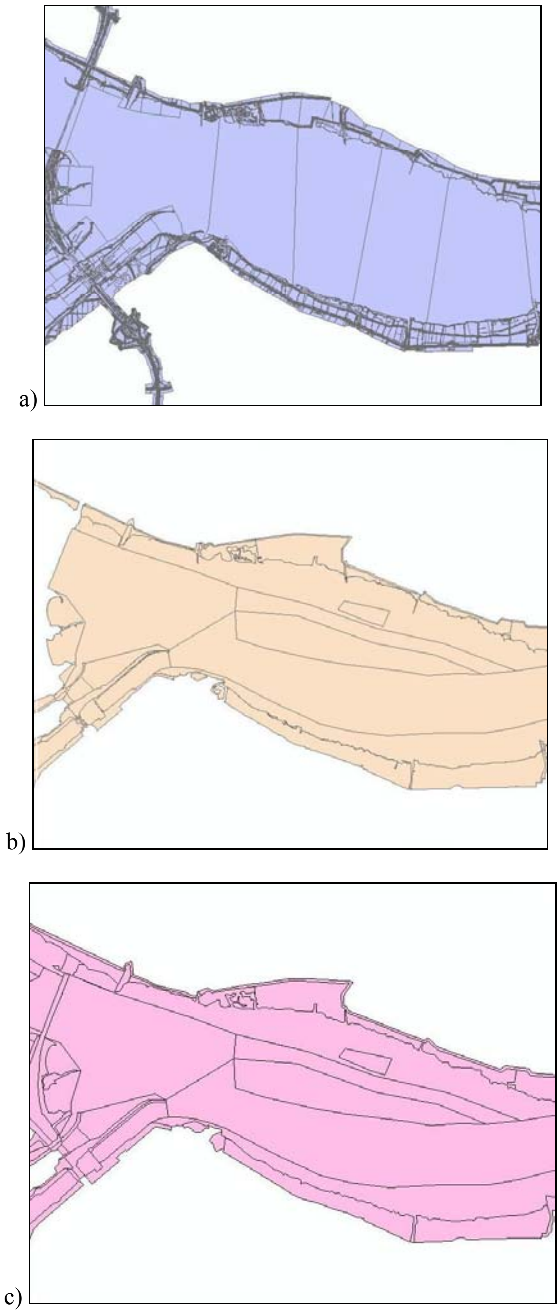

Figure 3: Data sets: a) DTB-Nat, b) Beheerkaart-Nat, c) Regiokaart-Nat

The three data sets used in the test are (see Figure 3):

a. DTB-Nat, large-scale topographic map (1:1000)

b. Beheerkaart-Nat large-scale maintenance map (1:1000)
c. Regiokaart-Nat small-scale maintenance map (1:50 000)

The three datasets are produced and maintained by three different organisations within the Ministry of Transport and Public Works in order to support work-processes concerning infrastructure. The differences between the different data sets can be specified as follows.

DTB-Nat and Beheerkaart-Nat have different objects and different geometries, but the scale is the same (thus the geometries may overlap somewhere). DTB-Nat has been created on the basis of topographic boundaries (grass, river, forest, etc.), while the Beheerkaart-Nat has been created on the basis of maintenance characteristics (road, facilities, etc.). In general, it is possible to have boundaries in Beheerkaart-Nat that do not follow topographic boundaries.

Beheerkaart-Nat and Regiokaart are both designed for maintenance purposes. Consequently the theme of the objects is the same (given with a unique code) however the scale is different (see Figure 4). This is to say that the polygons of the Regiokaart-Nat are defined with much less points compared to the polygons in Beheerkaart-Nat.

Case study 1: Establishing link between objects
The first important question was related to establishing a link between the objects in the three datasets. The test was performed in two steps:

- Establishing a link between the Beheerkaart-Nat and Regiokaart-Nat
- Establishing link between the Beheerkart-Nat and the DTB-Nat

1. Establishing a link between Beheerkaart-Nat and Regiokaart-Nat
Between the objects in Beheerkaart-Nat and Regiokaart-Nat a link was established using two approaches:

- comparison of geometries (by finding and computing percentages of overlapping areas)
- comparison of theme (by comparing codes)

Based on geometric properties The first method was performed following a procedure (written in PL/SQL) based on the Oracle Spatial function SDO_ANYINTERACT. All the objects from one of the representations were checked against interaction with the objects from the second map. Then the geometry of the overlapping parts is calculated and their area is computed. These areas are further compared with the areas of the objects and the percentage is computed. The results are stored in a separate Oracle table

From out experiments the conclusion can be drawn that objects with overlapping areas larger than 60% can be considered as one object. The statistic shows that using this procedure about 85% of the objects of the two representations can be automatically linked, i.e. one object from the Beheerkaart-Nat is matched with only one object from the Regiokaart-Nat. How many objects from the Regiokaart-Nat are not matched with any objects is difficult to check, since the area covered by the Regiokaart-Nat was much larger. There were several reasons for not matching the remaining 15%:

- one of the objects in one of the representations covers two objects in the other representation
- one of the objects does not exist as a separate object in the second representation
- the overlapping area was smaller that 60% although the objects were the same. Such a case may appear in long, narrow objects.

Based on thematic properties The second method (comparing codes) was less successful due to lack of codes in some of the maps or wrongly introduced codes. The two tests however clearly showed that objects from different representations can be automatically linked using the spatial operations offered by DBMS.

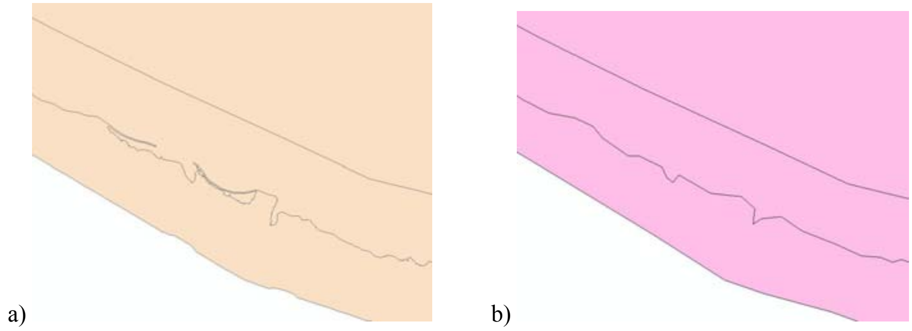

Figure 4: Difference between the resolution of a) Beheerkaart-Nat and b) Regiokaart-Nat

2.Establishing link between Beheerkaart-Nat and DTB-Nat

For establishing a link between the objects in Beheerkaart-Nat and DTB-Nat again two methods were used:

- comparison of geometries
- assembling geometries with similar theme. Since the existing themes in both maps are different, only a limited number of objects can be compared according to the theme (e.g. 'bottom of rivers' in the Beheerkaat corresponds to 'rivers' in the DTB-Nat).

Based on geometric properties The test based on comparing geometries was based on the same approach as in the case of linking Beheerkaart-Nat and Regiokaart, i.e. PL/SQL script that compares all the objects from the two representations for overlap. In this case however, the scale of the two maps was the same (i.e. 1:1000) but the thematic definition of objects was different. Therefore it was difficult to decide on the criteria for overlapping. For example an object "river side" from Beheerkaart-Nat may overlap with "grass" and "trees" from the DTB-Nat.

Clearly, when a DTB-Nat object has 100% overlap with Beheerkaart-Nat object then the object is completely inside the Beheerkaart-Nat object and can be linked to it. It has to be a multi-step process in which first objects that have a 100% match can be linked, then link "obvious" objects (those which are e.g. 60% or 70% or more inside a matching object). The remaining objects will require some additional "rules" (e.g. assign the object to the matching object with the largest overlap if this overlap is more than 50% of the input object). Probably some objects will remain which cannot be matched. This may require to handle the match between the objects in a different way. The last step of the algorithm was not tested. We have matched all the objects that have an overlap of 100% (58%), larger than 80% (70%) and larger than 60% (74%).

Based on thematic properties The second approach (aggregating DTB-Nat objects where possible) showed to be more promising, i.e. many (aggregate of) objects could be linked to objects of the Beheerkaart-Nat. Bearing in mind the possible operations in Oracle Spatial, it is possible to compute the geometry of an object that is a union of two objects. Using this function, all the objects with the same thematic meaning that have common boundaries can be recursively aggregated in one large object. However, in case of rivers this process can result in rather large objects because the river will not stop at a bridge, since the bridge in these data sets is in another layer. To avoid such objects we used the objects of Beheerkaart-Nat, which does separate rivers at the location of bridges, in the aggregation. Since the separations of rives in the Beheerkaart-Nat is for a particular river-region, the parts of the river of the DTB-Nat were also linked to these river-regions.

Case study 2: Deriving a small scale map from a larger scale map
The second part of the case study was related to the possibility to obtain the Beheerkaart-Nat automatically from the DTB-Nat and the Regiokaart-Nat. Since the Beheerkaart-Nat has themes identical to the Regiokaart-Nat and scale identical to DTB-Nat, the questions was whether the Beheerkaart-Nat can be derived from the two other datasets. For the purpose a set of scripts was developed that performed the following operations:

- select all the objects from the DTB-Nat that have geometries overlapping more than 60% with the objects of the Regiokaart-Nat and similar themes (e.g. ‘rivers’ and ‘bottom of river’)
- within this selection, select all DTB-Nat objects that ‘touch’
- create an ‘aggregation’ of touching objects and
- compute the outer polygon of the aggregation and assign to the new object the corresponding theme code from the Regiokaart-Nat

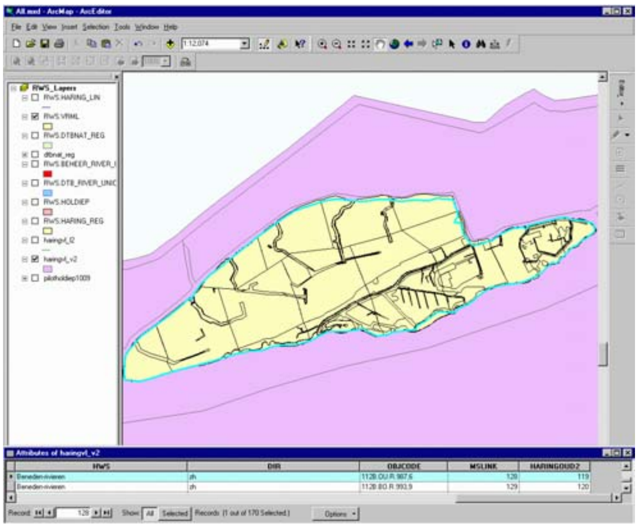

Figure 5: Generation of a new object: the two data sets overlap completely.

Figure 5 and 6 present some results of the algorithm described above. All the steps were completed with spatial functions available in Oracle Spatial.

Conclusion of case studies
Interesting conclusions can be drawn from the results of the tests. The tests based on geometries give more reliable results compared to the tests based on theme properties (codes and description). Since the theme is specified manually, the possibility of introducing error is much higher in comparison with the geometry. The case study has also shown that the spatial functions offered by Oracle Spatial (and which are available in other mainstream DBMSs) can facilitate establishment of links between

spatial objects from different data sets referring to the same real world objects. The results of the case study also show promising possibilities to obtain new spatial data sets on the basis of existing ones.

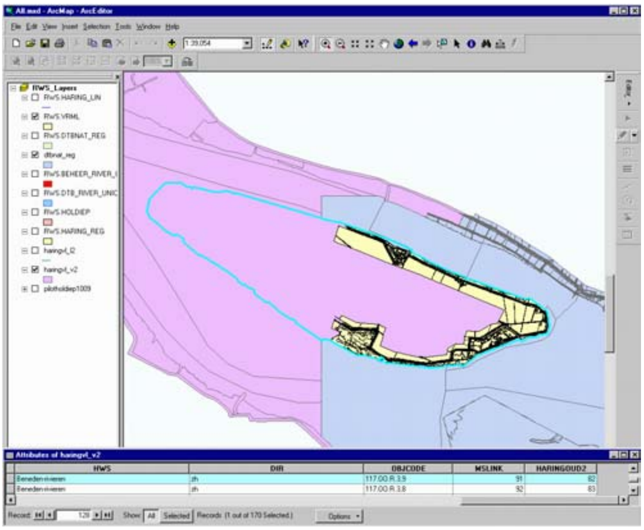

Figure 6: Generation of a new object: the two data sets do not overlap completely.

CONCLUSION

Multiple representations of real world data currently exist and will continue to exist in the future due to the diversity of applications or distinct requirements of one application (e.g. Level of Detail representations for fast visualisation of 3D city models). Many questions related to data consistency and update propagation using a DBMS approach need further explorations and appropriate solutions.

In this paper we argued that the optimal case would be to have one DBMS representation of a real world object whereupon other representations could be created on the fly. However, since complete generalisation algorithms do not (yet) exist, and since several organisations are responsible for (already existing) different data sets, support for multiple representations in DBMSs is indispensable. In this context two functionalities were further explored: 1. automatically establish links between different data sets and 2. automatic generation of low-resolution data sets from high-resolution data sets.

To test current DBMS technologies with respect to multiple representations in DBMSs we did a case study. Our tests showed that the functionality currently offered by DBMSs allows automation of most of the needed operations.

One of the findings of the case study was that the aggregation of several objects into a new object according to thematic characteristics is a relatively simple and straightforward process, which can be easily completed utilising available spatial functions in the DBMS. This can be used as a practical solution for comparing objects from different applications. The geometries of the objects can be aggregated according to a hierarchical classification up to a level at which the aggregated objects have the same meaning. In the example of the Building registration and the Cadastral registration, the difference in the partitions can be resolved by theme aggregation up to a building and establishing a link between the aggregates.

Further research will focus on the following aspects:

- Data structures for management of multiple representations. The correspondence between spatial objects in a multi-representation environment has to be further organised in appropriate data structures. The simplest way to register the link in a relational DBMS is creating a new relational table with the id's of the objects from the different data sets. On-going research seeks for approaches to incorporate procedures that will ensure consistency management of all representations (see also Friis-Christensen and Jensen, 2003).
- Functions for consistency checks and functions for updating propagation between different datasets to support management of multiple representations, all implemented within the DBMS
- Generalisation procedures for geometry – simplifications of boundaries, changes in dimensions of objects, aggregation of several small objects, selection of significant objects.
- Generalisation procedures with respect to the theme: aggregations of geometry based on theme properties and computation of a new object.
- As the case study showed, functions offered by DBMSs can be readily used for obtaining new geometries. However, further research is needed towards consistent organisation of new objects in the structures used for multi-resolution management, e.g. DBMS-views could be used for representing data sets using generalisation procedures for geometry and theme properties once these procedures have been implemented in the DBMS.

AKNOWLEDGEMENTS

This research is sponsored by the Rijkswaterstaat, Ministry of Transport and the research programme Sustainable Urban Areas at OTB, Delft University of Technology, The Netherlands.

BIBLIOGRAPHY

Binkhorst, V. and S. Zlatanova, De mogelijgeden van Oracle Spatial voor de natte sector, GeoNieuws 2003, No.2, pp. 6-9, 2003 (in Dutch)

OGC, 1999, OpenGIS Consortium. The OpenGIS Abstract Specification, Topic 1: Feature Geometry (ISO 19107 Spatial Schema), Version 5. Technical Report OpenGIS Project Document Number 01-101, Wayland, Mass., VS, 2001.

OGC, 2001, OpenGIS Consortium. Opengis simple features specification for sql. Technical Report Revision 1.1, OpenGIS Project Document 99-049, OpenGIS Consortium, Wayland, Mass., VS, 2001.

Oracle, 2001, Oracle spatial user's guide and reference release 9.0.1 part number a88805-01. Technical report, ORACLE, June 2001.

Friis-Christensen A., and C.S. Jensen, 2003, Object-Relational Management of Multiply Represented Geographic Entities.

Buttenfield, B.P. and J. S. DeLotto. Multiple Representations – Scientific Report for the SpecialistMeeting. Report 89-3, NCGIA, Department of Geography, SUNY Buffalo, USA, 1989.

Devogele, T., J. Trevisan, and L. Raynal. Building a Multiscale Database with Scale-transition Relationships. In International Symposium on Spatial Data Handling, pp. 337–351, 1996.

Friis-Christensen, A., D. Skogan, C. S. Jensen, G. Skagestein, and N. Tryfona. Management of Multiply Represented Geographic Entities. In IDEAS, pp. 150–159, 2002.

Grefen, P. W. P. J., and J. Widom. Protocols for Integrity Constraint Checking in Federated Databases. Distributed and Parallel Databases, 5(4):327–355, 1997.

Jones, C. and D. Kidner, L. Luo, G. Bundy, and J. Ware. Database Design for a Multi-scale Spatial Information System. International Journal of Geographic Information Systems, 10(8):901–920, 1996.

Li, Q. and D. McLeod. Managing Interdependencies among Objects in Federated Databases. In IFIP Database Semantics Conference on Interoperable Database Systems, pp. 331–347, 1992.

Sheth, A.P. and J. A. Larson. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. ACM Computing Surveys, 22(3):183–236, 1990.

Spaccapietra,S., C. Vangenot, C. Parent, and E. Zimanyi. MurMur: A Research Agenda on Multiple Representations. In International Symposium on Database Applications in Non-Traditional Environments, pp. 373–384, 1999.

Stoter, J.E. and H.D. Ploeger, 2003, Property in 3D-registration of multiple use of space: current practice in Holland and the need for a 3D cadastre. In: Computers, Environment and Urban Systems, Volume 27, November 2003, pp. 553-570.

Stoter, J.E. and S. Zlatanova, Workshop on 3D Cadastres and Large Scale 3D Urban modeling, Organised during UDMS 2002, October 2002, Prague, Tsjech Republic.

Zlatanova, S., T.P.M. Tijssen, P.J.M. van Oosterom and W. C. Quak, Research on usability of Oracle Spatial within RWS Organisation, GISt No.21, ISSN: 1569-0245, ISBN: 90-77029-07-9, AGI-GAG-2003-21, Delft, The Netherlands, 75 p. 2003.