# Rule-based or explicit storage of topology structure: a comparison case study

Marco Baars, Jantien Stoter, Peter van Oosterom,Edward Verbree

TU Delft

Delft, the Netherlands

m.t.l.baars@student.tudelft.nl, j.e.stoter@geo.tudelft.nl, oosterom@geo.tudelft.nl,
e.verbree@geo.tudelft.nl

## SUMMARY

*This paper presents the results of a comparison between the Geodatabase topology of ESRI and Radius Topology of Laser-Scan. The result of this comparison is a list of main differences, some conclusions and recommendations. The main conclusion is that in the Geodatabase concept, the user is able to decide which topological relationships have to be defined and what to do with objects that do not meet these rules defined by the user. In the concept of Radius Topology, the topological structure is fixed in a sense that it has to obey the implemented topological structures and the objects are persisted into a topological structure. The concept of Laser-Scan is interoperable.*

**KEYWORDS:** *geo-DBMS, topological structure, planar partition*

## INTRODUCTION

At TU Delft, a research has been carried out on the usability of Oracle Spatial within the Rijkswaterstaat organization (Zlatanova et al, 2003), responsible for public works in the Netherlands, which carries out the policy of the Ministry of Transport, Public Works and Water management. (Zlatanova et al, 2003) reports the results of the tests performed with three different datasets, organized in the currently available Oracle Spatial geometry types. These datasets are produced by Rijkswaterstaat and map all the rivers, lakes and other inland waterways in the Netherlands.

The conclusions of (Zlatanova et al, 2003) describe that topology is important to maintain a high level of data quality. The datasets contain errors. Many of the errors (invalid geometries, overlapping objects, non-complete coverage) can be easily solved, when using topology. Topologically structured objects have many advantages (Oosterom et al, 2002):

- It avoids redundant storage (more compact than storing full geometry).
- It is easier to maintain consistency of the data after editing.
- It is the natural data model for certain applications.
- It is efficient for certain visualizations or query operations (e.g. find neighbors).
- It can help detecting topological errors in the dataset.

Therefore, Zlatanova firmly recommends organizing the RWS data in a topological structure. There are some directions given in the report for further research about this topic. One of the directions is to investigate an appropriate topological structure for RWS, e.g. Laser-Scan Radius Topology vs. Oracle Spatial Topology Manager (release 10) vs. ESRI topology.

In this research, a comparison between the data model of ESRI and the data model of Laser-Scan, has been made. This comparison contains a functional analysis of how the management of the topology structure is handled. A functional analysis on both impementations has been carried out using one of the datasets of Rijkswaterstaat (DTB-nat), which was also used by Zlatanova. This dataset contains a lot of

topological errors, so a comparison can be made which data model can find these errors and what happens to the topological structure if you want to add polygons to or delete polygons from the dataset.

The main question of that this research tries to solve is:

*what are the main differences between a data model where topological consistency is implemented with a rule-based topological structure (ESRI) and a data model that applies an explicit topological structure (Laser-Scan)?*

## TOPOLOGY, THE THEORY

According to (Clementini et al, 1993), topological relationships are defined as follows: "*the subset of spatial relationships characterized by the property of being preserved under topological transformations, such as translation, rotation and scaling.»* Obviously, spatial queries can be easily processed if all the geometric relationships between the objects of interest are explicitly stored; however, such a choice is unsatisfactory since it requires a tremendous amount of disk space and, furthermore, it implies the execution of time-consuming maintenance procedures. It follows that instead of storing all spatial relationships among the objects of interest it is more convenient to store the relevant ones (e.g. neighbours) and compute the others. A case study was carried out on two implementations of topological structure. In the paper, the results of this research are described, which examined the differences between the way ESRI uses topological relationships and the way Laser-Scan uses it.

### ESRI Geodatabase topology, the concept

In the concept of the ESRI Geodatabase (not to be confused with the general concept of a Geo Database Management System, a GeoDBMS), the user defines topological conditions via a set of rules. This means that the user decides which relationships are important and in which way these rules have to be treated. "The topology is specified by zero or more topology rules. The topological primitives are not persisted as a special type of feature; instead, feature geometry is persisted and topological primitives (along with their geometry) are inferred. The process of topological integration (validation) results in vertex equality where features share underlying topological primitives. Given vertex equality, reconstruction of topological primitives is straightforward. Vertices on feature geometries in this scheme play the same role as assigned to embedded foreign keys in data structures that explicitly model topological primitives." (Hoel et al, 2003)

The data is stored as objects, with geometry, represented as a list of coordinates. The topology has to be created with a set of topology rules, for instance "Parcels must not overlap" or "buildings must not overlap with water". If you have created these topological rules, you can perform a topology check on a dataset. Topological errors will be put in an "error-table" and suggestions will be given to solve these errors.

As said before, topological integrity is defined a collection of topology rules. Topology rules are used to define constraints on the permissible topological relationships between features in one or more feature classes that participate in the topology. The collection of topology rules that are associated with the topology, are selected on the basis of which topological relationships are important for the user's model. There is a set of 25 topology rules from which the user may select zero or more rules for a specific case.

The validation process is a fundamental operation on a topology performed by a topology engine. The validation process is responsible for checking all specified topology rules and generating topology errors at locations where rules are violated. The validation process does not need to span all features within the topology dataset. A validation can be performed on a subset of the space spanned by the dataset.

A topology can have an associated dirty area – a dirty area corresponds to the regions within the topology extent where features participating in the topology have been modified (added, deleted or updated) but have yet to be validated. In order to ensure that the topology is correct, the topology in the dirty areas will need to be validated

**Laser-Scan Radius Topology, the concept**

The concept of Laser-Scan Radius Topology is different than the Geodatabase topology. Not only the feature geometry is persisted, also the topological structure is persisted. Radius Topology provides the user with the option to store both the geometries and the topological primitives or to only store the topological primitives and derive the geometries. The user defines the topological structure and after that, the DBMS calculates the topological primitives and the topological relationships. In Radius Topology, the topology of a map can be represented as a linear (one-dimensional) network topology or as a two-dimensional planar topology. A network topology uses node and edge primitives to describe interconnected linear features and points on a map. Planar topology uses the face primitive in addition to nodes and edges to describe two-dimensional areas on a map. An area consists of one or more faces.

Often applications want data to be separated into coverages, e.g. administrative areas, vegetation type or land use. These separate topologies are termed *manifolds*. Manifolds allow the application to build and manipulate independent sets of topology in the same DBMS. If data is held in different spatial reference systems, then a manifold is required for each, since all data within a manifold must use the same spatial reference system. If you want to use different topology models, you have to use more than one manifold. A *feature class* enables topological relationships to be defined between separate groups of features.

Topological primitives are the first part of the topological structure within the dataset. The other part is topology rules. Within a manifold, data is structured by applying topological rules that determine the relationships between feature classes. *Node Formation Constraints* determine how nodes are created in the topology network to represent the feature objects. *Topological priorities* determine which object will move when snapping occurs between objects.

## ADVANTAGES AND DISADVANTAGES OF BOTH SOLUTIONS

Working with ArcGIS 8.3 and the Radius Topology, configured with Oracle 9i spatial, the functionality of both can be described. The functionality is tested and this results in a list of advantages and disadvantages of both ways of treating topological relationships. These advantages and disadvantages are listed below.

**Advantages and disadvantages of ESRI Geodatabase topology**

The main advantages of the Geodatabase topology are:

- The topology structure is not explicitly stored. The user decides which topology rules are important for his dataset. The user has to specify the conditions of the dataset.
- The user decides what should happen with spatial objects if they do not meet all the conditions. Topological errors are not solved automatically, but the user has to solve errors manually.
- The validation process is relatively short. It did not take too much time to validate the entire dataset in comparison to Radius topology. It should be noted that with this validation, only the errors are marked, not solved.
- After editing some objects in the geodatabase, not the entire dataset has to be validated again. The edited area is marked as a dirty area and only the dirty areas have to be validated again.
- The Geodatabase topology is easy to use. You don't need to have special knowledge about topology to create a set of logical conditions for the data. And also the wizard to implement the topological rules is very useful and straightforward.

The main disadvantages of the Geodatabase topology are:

- There is a lot of redundancy; all edges shared between neighbours are stored twice.
- After validation of the topology, the dataset still does not satisfy the conditions. Only the errors are marked. So, at least one extra step is necessary.

- If there appears an object that does not meet the conditions specified, you have to be aware that this does not have to mean that there's a topological error. It is also possible that the defined condition is not defined well.
- The toolkit in ArcGIS to solve the topological errors is not very extensive. Some errors can be solved, but for many errors a more advanced toolkit is necessary.
- Not all the possible topological rules can be implemented. For instance the rule "POINT_OBJECT must not be in POLYGON" cannot be implemented.

**Advantages and disadvantages of Laser-Scan Radius Topology**

Some advantages of Laser-Scan Radius Topology are:

- The topology is explicitly stored. The user can only put data into a database, when it's topologically correct or when automatic correction is possible.
- After putting data into the topological structure, the user does not have to worry about topology anymore. The topology is correct. This can be useful for instance when a company sells a dataset to customers. The customer and the seller are sure that the topological structure is correct.
- With SQL statements, you can query the database at DBMS level.
- In spite of the fixed structure of Radius topology, the user can define some elements of the topological structure in the manifold. For instance, the user can define which topological model he uses, which priorities he gives to objects (which object snaps to which other object), he can define tolerances, etcetera.
- In this topological model, an edge is only stored once in the database.
- Radius Topology is standards-based, open and interoperable and can be a component of a diverse enterprise-wide solution and is therefore not tied up to a single vendor's solution.

Some disadvantages of Laser-Scan Radius Topology are:

- The validation process takes a lot of time. But you have to do it only once (create once, use many).
- The elements of the topological structure that can be defined by the user are limited. For instance, topological rules based on semantics are not possible, except for organizing the data in a specific way in a manifold.
- Due to the topological structure and the multiple reference tables, a lot of storage is required.


# CONCLUSIONS AND RECOMMENDATIONS

ESRI and Laser-Scan have a different way of treating topology. After the comparison between both methods, the following conclusions can be drawn.

- ESRI Geodatabase Topology is rule-based topology, because the user decides which topological rules are important and what to do with errors.
- In Laser-Scan Radius Topology the influence of the user is limited. Once you defined a topological structure, organized in a manifold, you can hardly influence the way Radius Topology is treating topological errors, except for defining snapping priorities and tolerances.
- A combination of Geodatabase topology and Radius topology could also be useful. For instance when you work with the dataset used in the case study. You edit the data with a set of defined rules in the Geodatabase topology and solve the errors. When you are going to sell a new version of the dataset, you put it into Radius Topology and you are certain that the topology is correct.
- Both the described methods have their advantages and disadvantages. Which method has got more future, cannot be concluded and depends on the requirements of the application, e.g. when performing topological queries on large datasets explicit topological relationships might be preferred.

It is important to be aware of the following issues:

- If you have defined a topological rule in the Geodatabase concept and there are features that do not meet these rules, then you are not certain if the data is incorrect (a real error) or if the rule is not well

defined (not a real error). The user has to be aware of this uncertainty while he is using the Geodatabase concept.
- Whether you use persisted topological relationships or not, queries still give incorrect results when you have incorrect data compared to the real world. Both the Geodatabase concept and the Radius concpet give tools to minimize the amount of incorrect data. The only thing is that in the Geodatabase concept, the data is validated and then put into a topological structure by the user. In a persisted data structure like Radius Topology, the user has less influence, what makes it easier for the DBMS to check the correctness.

Future research should focus on the following aspects

- In this research, a comparison is made between Radius Topology and Geodatabase Topology. When the new version of Oracle (10g with topology) is released, it is recommended to compare the Oracle Topology with both Radius and Geodatabase topology for a complete overview of topology methods.
- Another recommendation for research is to do a performance test on a large dataset. This research was only a functionality test.

## ACKNOWLEDGMENT

## LITERATURE

Clementini, E., Di Felice, P., & van Oosterom, P. (1993). *A small set of formal topological relationships suitable for end-user interaction*. In D. Abel, & B. C. Ooi (Eds.), Advances in spatial databases - Third International Symposium, SSD'93, Singapore (pp. 277-295). Berlin: Springer.

Hoel, E., Menon, S., Morehouse, S., *Building a Robust Relational Implementation of Topology*. 8th International Symposium on Spatial and Temporal Databases. Santorini, Greece, 2003.

Louwsma, J.H., *Topology versus non-topology storage structures*. Case Study. (www.geo.tudelft.nl/gist, publications) Delft, 2003.

Oosterom, P. van, Stoter, J.E., Quak, C.W., Zlatanova, S., *The Balance Between Geometry and Topology*. In: Advances in Spatial Data Handling, 10th International Symposium on Spatial Data Handling, Dianne Richardson and Peter van Oosterom (eds.), ISBN 3-540-43802-5, Springer-Verlag, Berlin, 2002, pages 209-224.

Zlatanova, S., Tijssen, T.P.M., Oosterom, P.J.M. van, Quak, C.W., *Research on usability of Oracle Spatial within the RWS organization*. Delft, 2003.