Geometrical and Syntactical Integration of Geospatial Data
Kristian Lunde
Ostfold University College

Geometrical integration is the process of integrating geometrical data into an already existing data set. Geometrical integration is similar to a map conflation method called boundary alignment. Map conflation is a difficult and unwieldy problem area, and boundary is the handiest of the conflation methods, other map conflation methods are out of scope in this article. The geometrical integration integrate a section of data to a already existing data set, this section of data are new and lie adjacent to the already existing data. This integration of new data into an already existing data set does not go painlessly. The problems that occur in such a process origins from differences in the structural build up of the data sets to errors in one or both data set. These errors result in that the data sets are not aligned. Line segments that cross data sets can and most likely will not be coincided. To solve this problem the data sets has to be aligned along their common borders. This means that the line segments that cross the border between the data set becomes coherent, and result in one data set, with coherent line structure. The alignment of line segments are a cumbersome process and require human assistance to secure a correct and valid process. Information on the geometrical integration process is as important as the integration process. This information is called metadata, and is stored together with the geometrical features. Metadata contain information that are relevant to the data set, for instance the date of the integration process, the coverage of the feature, the resolution of the feature, the format of the original data, description and creator. Metadata is important because of future integration processes, due to consistency. It is important to know the resolution of features, last updated and its coverage area.

The geometrical integration do the integration of two geometrical data sets, however it do not say anything about how these data should be integrated. Syntactical integration deals with the process of integrating data set on GML format. GML has standard components that define the geometrical features, however data sets can be built on different and several XML schemas. This complicates the process of the integration. To approach this problem a method called lazy integration has been used. The main objective with lazy integration is to preserve the structure of the integrated data set. Using this method the GML file will be able to handle storage of geodata that have different structure and information. The core of the lazy integration is to use several XML schemes and namespaces to uniquely distinguish between geometrical features from different data sources.

The purpose with this paper has been to study the inserting and replacement of geospatial data in a repository using geometrical and syntactical integration. Line segments or Linestrings of coastlines have been used as foundation data in this analysis. With the use of cases problems of conflation and integration has been highlighted, solutions to these problems have also been suggested. Both the problems and solutions suggestions are discussed in detail.

This analysis has based its work on the OneMap repository.
OneMap has a large data repository containing geospatial data of the world, the project has a lot of sub-projects going on, some of these projects result in retrieving new geospatial data that should be integrated or replaced in the repository. These retrieved geospatial datas do not fit into the existing data without some conflation processes.

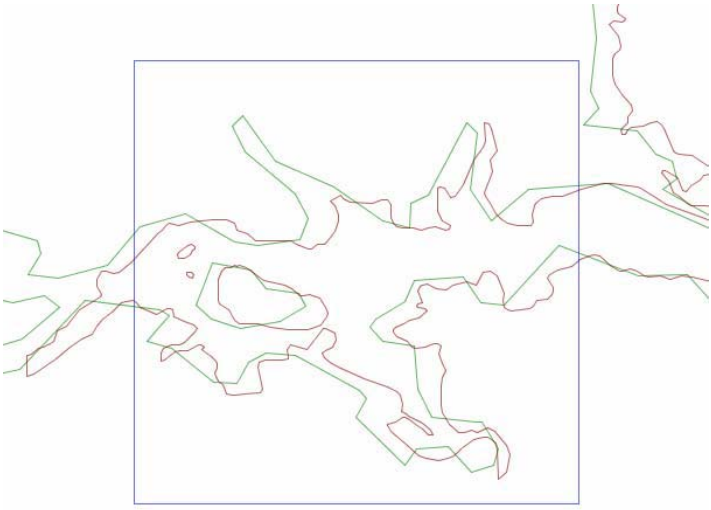

Figure (above): The green line is data retrieved from the map repository and are originally Vmap0 data, the red line is new data that should be integrated in the repository. As seen these dataset are not aligned.

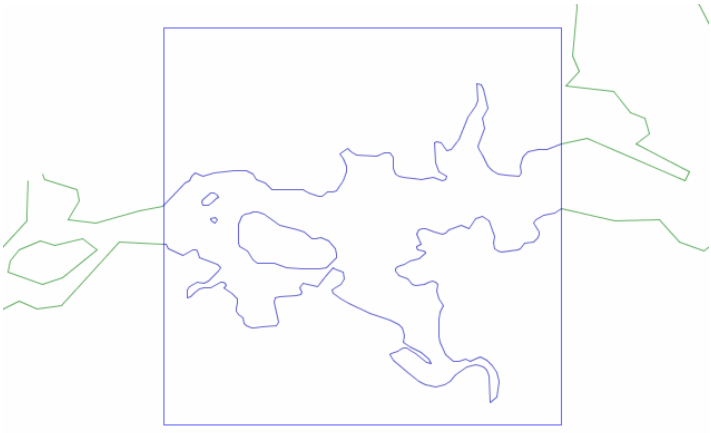

Figure (above): After a boundary alignment process are applied on the new dataset and the adjacent data from the repository the line work are coherent.