Using country boundaries in ArcGIS with native Oracle Spatial format, a lesson on tolerance settings and seamless mapping

Alfred de Jager
Joint Research Centre of the European Commission
Via Fermi 2749, 21027 Ispra, Italy
alfred.de-jager@jrc.ec.europa.eu

Abstract

We test a new feature of ArcGIS 10.3; it is to store and display data in Oracle Spatial format. For the test we use the country borders of the world. We find that we can reach a goal of interoperability but that we also should pay additional care on explicitly setting geometry quality values. Finally we discover that we still cannot display the world as a sphere.
Keywords: Borders, Oracle Spatial, ArcGIS, Countries, Tolerance.

1   Introduction

ESRI's ArcGIS version 10.3 can read Oracle Spatial data using the native Oracle Spatial format. ArcGIS requires some additional descriptive tables for displaying the data using ESRI tools, the so-called Geodatabase. These tables are maintained by ArcGIS, but applications using only Oracle Spatial can read the spatial data from the tables without using ESRI middleware. For organisations like the European Commission, which collaborate with a large number of data providers and data editors, this opening up of the GIS market is an important step. We can now physically integrate our databases containing geographic information and move away from the need to have a copy for each individual software provider using a slightly different storage format for the same information. As a consequence, our most used and discussed datasets were converted into an Oracle Spatial structure maintained by ArcGIS, which at the same time can be queried by Mapserver and Oracle SQL applications. The dataset we use for the experiment described in this paper contains the global country boundaries, a dataset that is provided by the GISCO bureau centrally to all organisations and affiliates of the Commission. While the data could be converted readily, the storage and topographic rules of the two packages required some additional manual intervention.

2   The country boundaries

The country boundaries are a dataset that is used in almost all GIS applications of the Commission. Since the end of the cold war these boundaries are not anymore static lines. The break-up of Yugoslavia, Sudan, and the USSR followed by a number of border wars along the Russian border, the split of Cyprus and the creation of a wall in the Spanish Sahara are but a few examples. The new borderlines require specific rules on when and how these borders are displayed, since the Commission is not recognizing the official status of all proposed borders.

The country boundaries form a complex dataset. The lines and polygons are delivered with a set of scales. Each smaller scale is derived from the larger-scale dataset. In order to reduce redundancy, descriptive attributes such as names and statistics are provided in tables that can be linked to the country ISO 2 character codes.

2.1   Conversion preparation

As input a File Geodatabase, the native ArcGIS format, is used, containing a layer with the country boundaries as polygon features. In order to convert these polygons to an Oracle Database some initial settings in the Oracle database have to be defined. Roughly these settings concern the creation of an Oracle user called SDE. Subsequently ArcGIS creates a set of tables, which are referred to as the Oracle Geodatabase. Subsequently we can drag and drop the tables from the file geodatabase to the Oracle Geodatabase data container and the system will create the Oracle tables and spatial indexes and fill them.

2.2   Conversion issues

After successful creation of the tables also the spatial index is defined immediately. This working method provokes that geometries that are not valid for the Oracle Spatial index are rejected and not entered into the database during conversion. In our experiment invalid geometries were reported by ArcGIS messages but also some geometry that entered the table remained invalid according to the Oracle MDSYS.SDO_GEOM. VALIDATE_GEOMETRY function. After completing the conversion we could distinguish 2 types of issues that had to be solved using additional PL/SQL scripts.

2.3   Spheroid datamodel of Oracle Spatial.

Some features were invalid according to Oracle Spatial rules because the same feature was sharing a border. (Polygon

boundary crosses itself). This happened in Russia, Fiji and Antarctica. These features all cross the -180 180 date border and such features are stored differently in Oracle Spatial compared to ArcGIS. This problem can be solved by merging the polygon components into one feature, but subsequently ArcGIS software is unable to visualize the object.

Also note in Figure 1 that the right map ends at -180 decimal degrees and the left map at 180 degrees. We had to copy and paste them together in order to make an overview map of the archipelago.

Figure 1 a b: Pacific Ocean with Fiji Islands. The stripes are caused by limitations of the software to map seamless features across the date boundary
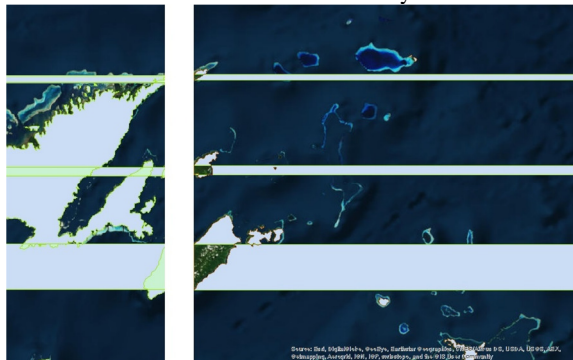
Source: ArcGIS software with Earthstar TerraColor backdrop.

2.4 Vertex Tolerances

The tolerance, which is the distance at which two points in a line or polygon can be apart and still considered the same, contains different default settings in ArcGIS and Oracle Spatial. Spatial stores decimal degree data up to 6 digits and therefore the tolerance will default to 0.05 or 5 centimeters [1]. Since ArcGIS does not convert the decimal degrees to the metric system it defaults the tolerance to 0.000005 which is subsequently overwritten by the Oracle default value of 0.05 meters. In order to be valid for Oracle Spatial, features are to be rectified. For Canada, for example, the rectifications lead to 2677 vertices out of 374868 being removed. In total we rectified 47 countries and territories out of 256 using this method for the initial ArcGIS tolerance of 0.000005. Subsequently we set the tolerance to 5 cm and removed an additional 2895 vertices in 60 countries. The remaining dataset contains 5908483 vertices in 256 countries or territories.

Related to this problem were errors of the previous type (Polygon Boundary crossing itself) which we encountered in the generalized versions of the dataset. The errors were detected when a generalized line touches a neighboring line within the more strict tolerance of Oracle Spatial. We encountered such cases for example along the coast of Bangladesh, containing several islands in a river mouth. The generated islands were lying too close to each other to be considered a new feature according to the tolerance settings of Oracle Spatial. The solution here was to remove such generalized islands if they were closer than a certain distance to the mainland (for example 100m.). This implies that a generalization algorithm needs to take into account also the minimum distance between two new resulting features otherwise the features ought to amalgamate intelligently.

3 Discussion

Conversion of geometric data is can be a cumbersome task due to specific storage rules of GIS software packages. With this test using ArcGIS to insert and display data in native Oracle Spatial format we can now be confident that we can serve a variety of GIS platforms using one physical database storage location only.

We learned that we should be more consistent on tolerance settings and avoid the storage and serving up of vertices that are not visible in practical use cases. This might also count for small geometric features such as islands around a coast when we want to represent a country.

Display of features lying across the arbitrary date border such as the Fiji islands and the Chukchi Peninsula should be made possible in new GIS platforms, since the storage of their geometry is seamlessly possible for more than 10 years.

The storage of derived small map scale features as separate layers causes the need for additional rules of generalization. The distance between resulting lines should further be set as an explicit parameter in order to avoid generating lines that are too close to each other to be visually separable and the minimum size of the resulting features should be specified. New algorithms should be developed to take into account such rules.

References

[1] C. Murray et.al. Oracle Spatial User's Guide and Reference, 10g Release 2(10.2), 592 pages, Oracle, 2005.

[2] A.L. de Jager. Preparing CORINE Land Cover Data for use. In EUR report 25163 EN, 17 pages. Luxembourg, 2011.