Paper Number: 4226

How Cool Can Be The Migration of an Old GIS Corporate Dataset to a New System. From Coverages to Geospatial Databases.

Asato, C.G.

Geological and Mining Survey of Argentina (SEGEMAR), Av. General Paz 5445 (INTI) Ed. 25, B1650WAB San Martín. Argentina. g_asato2000@yahoo.com

---

Since the adoption of geographical information systems in the late 80 by geoscience institutes (GI), hundred of geospatial dataset were stored in professional formats like Arc-Info coverages. This kind digital geospatial format was preferred because it support topology, has a good description of geospatial relationships, good control of coordinates and cartographic references [1]. Since that time, GIS technology evolved into more integrated systems where different kind of information can be managed and published in a unified and consistent environment (e.g. by facilitating the integration with RDBMS and Web), for that reason, the migration to new formats and systems is a must.

The migration from old systems to new ones not only means a change of format, it is a better way to model and store our subject of study (e.g. geology) with all the complex relationships. In this context, data migration is also a change of concept: From data structure (files) to a complex data model. This process will involve extraction, filtering, validation, normalization and reorganization of data without loss of information. However, corporate systems must deal with high volume of information and in consequence the migration task could be a painstaking effort because it will be repeated for each file. Without the help of process automation this kind of effort cannot be efficiently done.

In this work a data migration technique for extraction, transformation and loading of huge quantities of data is presented. More than 100 geological dataset composed by six layers (geo, faults, contacts, folds, structural measurements) were transformed from Arc-Info to Spatial-SQLite [2] using a program made with Python programming language [3] and the Geospatial Data Abstraction Library (GDAL/OGR [4]). Look up Tables were transformed to Style Layer Descriptor format (SLD-OGC) and stored in SQLite, other tables like data dictionaries were simple translated from csv to SQL.

Python is a popular programming language among scientific community, it support geospatial and others scientific libraries. GDAL/OGR is an open source and cross platform translator library, supports over 50 raster and 20 vector formats. SQLite is an open source software library that implements a self-contained; sever-less, zero-configuration, transactional SQL database with OGC Simple Features support.

In short, the transformation process involves the creation of a SQLite geospatial database with a data schema (layers, CRS, dictionaries, relationships, views, etc.). Then the program loads the GDAL/OGR corresponding Arc-Info driver (AVCBin), it looks the data in the file system, check the data structure, maps the old items to new ones, filters undesired data, make the connection to SQLite database and load the data.

A quick and consistent data migration method without loss of information is a deeply concerns of GI. New geospatial technologies allows the management of information in unified and integrated system

and can store information as a complex data model. For that reason, automatic migration tools that permits the passage from data structures to complex data models are essential.

References

[1] Library of Congress. http://www.digitalpreservation.gov/formats/fdd/fdd000284.shtml

[2] Spatialite. https://www.gaia-gis.it/fossil/libspatialite/index

[3] G. van Rossum (1995) CWI Report CS-R9525

[4] GDAL/OGR. http://www.gdal.org/