

# Evaluation of Data Conversion of Vectorial Geographic Features in Topographic Maps using Free Software Tools

Helton Nogueira Uchôa<sup>1</sup>, Maurício Carvalho Mathias de Paulo<sup>1</sup>, Luiz Carlos Teixeira Coelho Filho<sup>1,2</sup>, Paulo Roberto Ferreira<sup>3</sup>

<sup>1</sup>OpenGEO Consulting Services  
Av. Marechal Floriano 38/1203 – Centro – Rio de Janeiro, Brazil

<sup>2</sup>The Rio de Janeiro State University – Post Graduation Programme in Geomatics  
Rua São Francisco Xavier, 524, Faculdade de Engenharia, 5º Andar, Bl. D, Sala 5022 – Maracanã – Rio de Janeiro, Brazil

<sup>3</sup>Federal University of Rio de Janeiro – Post Graduation Programme in Transportation Engineering  
Centro de Tecnologia Bloco H, Sala 106 – Cidade Universitária – Rio de Janeiro, Brazil  
{uchoa,mauricio,luizcoelho}@opengeo.com.br, roberto@pet.coppe.ufrj.br

***Abstract.** This paper is related to a common problem found when trying to convert huge geodatabases that were extracted according to old or proprietary techniques still employed in Geographic Institutes and Services around the World. The libraries used for this task pertain to the Geospatial Data Abstraction Library (GDAL), a free library kit for performing geographic operations. An application was developed in order to import traditional maps delivered in a proprietary vectorial format, and convert them trying to preserve their original features. The results of this approach are commented throughout the text.*

## 1. Introduction

In the mid-eighties, most geographic services around the World were converting their huge sets of cartographic databases to a digital format. It was a heavy task, though. Maps that covered large extensions of the Earth were digitized and converted to vectorial files. At that time, there were not many solutions to do this, and all of them were proprietary. Because of that, until nowadays, many users struggle with obsolete or closed formats in which the databases are provided.

It is, still, a heavy task to convert those files to an open format (or, at least to a format that serves as a “lingua franca”, like the ESRI shapefile). Vectors might be lost, erasing relevant information regarding the features that exist in the mapped area. This is still a matter in an era when geodatabases are becoming free and widespread.

This article describes an experiment undertaken in order to evaluate the quality of a conversion between standard DGN (Design File) formats (used by MicroStation, one of the most used Computer Aided Design - CAD software for map making) and shapefiles. For this task the GDAL (an acronym for Geospatial Data Abstraction Library), a free library for geographic operations, is used.

## 2. Geospatial Data Abstraction Library (GDAL)

The Geospatial Data Abstraction Library is a translator library for raster geospatial data formats. There is a related library, which lies underneath the GDAL Project, called the OGR (according to [Mitchel, 2005], the acronym for OGR is irrelevant today), provides a similar capability for simple features vector data.

The latest version of GDAL is the 1.3.1 one [OGR API Tutorial], and is currently used in the development of popular applications, such as ILWIS, Google Earth, MapServer, GRASS, Quantum GIS, among others. It is licensed according to the GPL and can be freely distributed through its main website, which is described in the bibliography.

## 3. Methodology

For this work, two topographic maps were used: the 1:50000 maps named 2648-4 – Miracema and 2649-3 – São João do Paraíso. Both were made by the Brazilian Institute of Geography and Statistics (IBGE) and their geographic location is shown in Figure 1.



Figure 1. Study area, comprising the topographic maps 2648-4 and 2649-3.

Those maps were acquired in vectorial format. They were already digitized, structured and validated by IBGE ([Mapoteca Topográfica Digital]). So, they have been through all phases that would, technically, prepare them for being loaded into a GIS (Geographical Information System). All validation procedures performed at IBGE followed the rules established in the MTD (Mapoteca Topográfica Digital – Digital Topographic Map Library). This is a document that describes how features should be digitally acquired and exported.

The entire process of preparing those maps was probably made using Bentley's Microstation CAD Suite. This is, until today, the most used CAD application for geographic purposes, and several cartographic databases are still delivered in Microstation's DGN (Design File) proprietary format. So, as usual, those maps were available for download at IBGE's website only as DGN files.

At first, a script was developed in order to import all layers contained in those DGN files, exporting them to a single shapefile. Shapefiles are also a proprietary

format, developed by ESRI (Environmental Systems Research Institute), but the popularity amongst the community of the ArcView/GIS series has turned this file format into a “de facto” standard across the geoprocessing community. Also, it is widely accepted by most free and open source packages and is easily convertible to more open formats [ShapeLib API Documentation].

Shapefiles, however, pose several limitations related to geometry. One of them was the fact that different kinds of geometry cannot be stored in the same file. So, for instance, points and lines could not be found on the same shapefile. This obliged the development team to modify the script in order to export all geographic information to multiple files.

Later, the original DGN files and the shapefiles obtained from the conversion tasks were compared thoroughly in order to understand what features were lost (if there were any) in the conversion process, and if this methodology is useful for the conversion of massive databases to free formats that might be more shareable through cooperative environments.

#### **4. Results and Analysis**

At first, the conversion process cannot be considered an easy task. DGNs and shapefiles are, in nature, deeply different from each other. DGNs are suitable for CAD design, and will present some functionalities that shapefiles will not. For example, for spatial features, most CAD formats are able to store information regarding styles, colors, strength and layers. Shapefiles are not able to preserve natively those features. On the other hand, shapefiles are suitable for GISs, and will have special features that most CAD formats will not. For instance, they are able to attach attributes to geometric features. These attributes can be defined by the user who is editing this file and are stored in a rudimentary database (“dbf” files that share the same filenames of their related shapefiles), which can be used to store information the user wants to be queryable from each feature. So, as a consequence, in order to preserve all the information that can be found in the original files, graphic attributes had to be transformed into fields of the database that takes part of the file structure itself.

Following this principle, the OGR library has converted seven attributes during script execution: type, level, graphic group, color index, weight, text and style. These attributes were stored as previously described. Most geometries have not the attribute text, though.

The script was developed in order to turn explicit which layers and geometry types were successfully converted, or not. A comparison (between the original DGN file and the extracted shapefiles) based on the script output was done with the objective of verifying if the geometry types OGR's documentation mentioned that could be read were correctly understood by the software. All non-geometric features (such as cells and complex geometries), which served only to organize or group features, could not be read, due to limitations the library still has.

Then, a feature count was performed for each geometry type and for each layer, with the main objective of knowing if there was any kind of spatial information loss. This step has proven that all relevant data (i.e. geometry features and their attributes) were not lost. This was an expected result, because geometries are represented in a way that they are composed by no more than an organized list of variables.

Figure 2 shows a comparison of the original files and the files generated by the script.

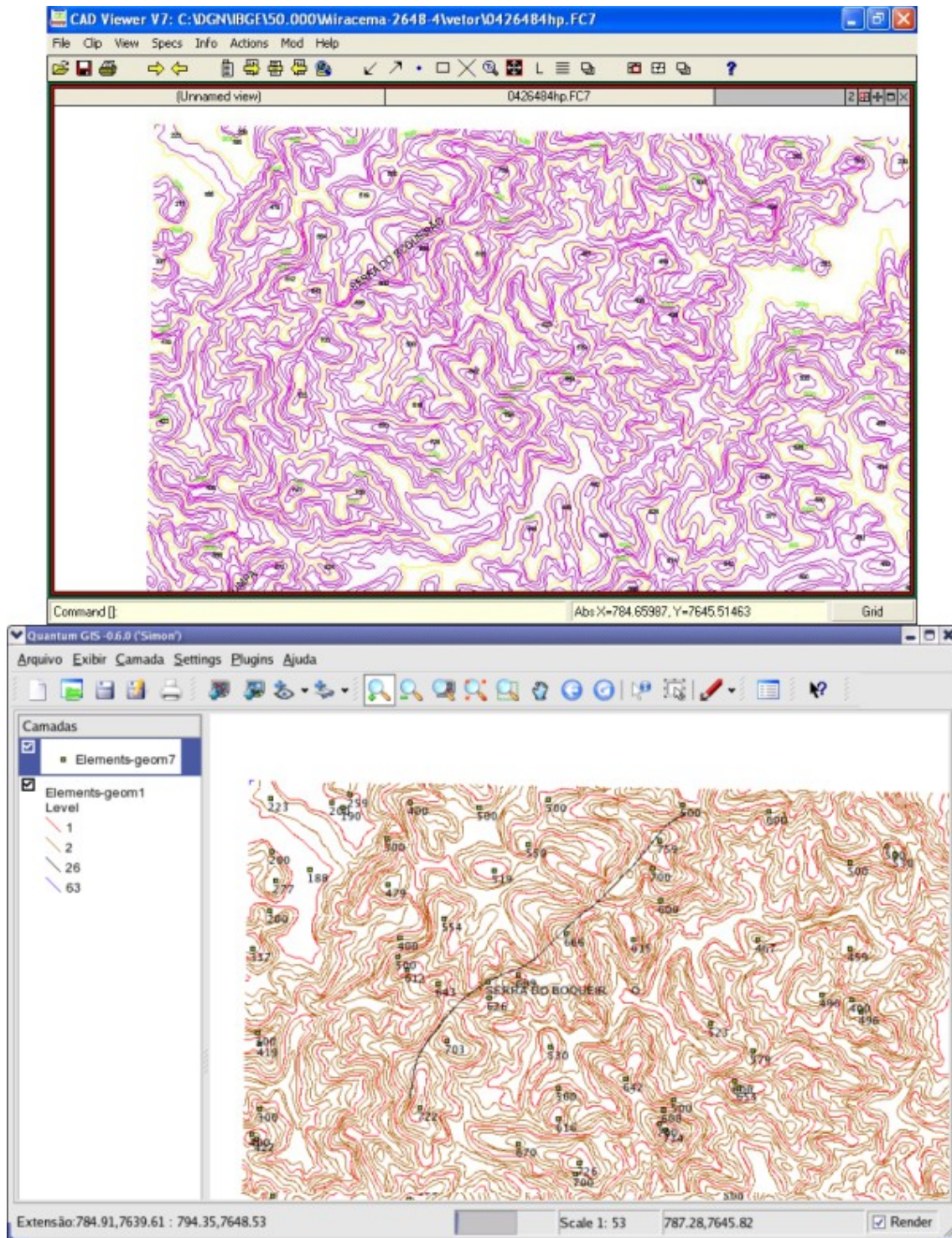


Figure 2. Screenshots taken from CadViewer (a freeware CAD viewer) and QGIS, showing the original files and the converted ones, respectively.

## 5. Conclusion

This work was a first attempt at creating a feasible way of converting cartographic databases stored in proprietary formats using exclusively free software. The main purpose of this task was to provide an inexpensive way of turning those databases accessible for the main public.

According to the methodology that was used, all steps were done with the aid of free software. GDAL's OGR libraries were used for reading and writing data, and other free applications, such as QuantumGIS, were used for viewing and analyzing the final result. Also, the scripting was all done in Python, which is a free scripting language "per se".

The results, albeit conclusive, still need further refinements. However, it can be concluded that this conversion is feasible, and that it could be automated and used both for internal and external purposes in Geographic Institutes and Services. Recent changes have turned geographic information much more widespread than it was some years ago, so, implementing an easier way of providing this migration process to institutions interested in it is still a challenge that has to be overcome.

Further projects on the same line of research might include the creation of a CGI script to turn this task worldwide accessible and the conversion of DGN files directly to a database management system.

## References

- CGI MapServer Reference Manual, <http://mapserver.gis.umn.edu/docs/reference/cgi/>, December/05.
- DGNLib Reference Documentation, [http://dgnlib.maptools.org/libhtml/dgnlib\\_8h.html](http://dgnlib.maptools.org/libhtml/dgnlib_8h.html), March/04.
- Mapoteca Topográfica Digital, IBGE, 4<sup>th</sup> edition, August/99.
- Mitchell, T. (2005), *Web Mapping*, O' Reilly, 1<sup>st</sup> edition.
- OGR API Tutorial, [http://ogr.maptools.org/ogr\\_apitut.html](http://ogr.maptools.org/ogr_apitut.html), February/06.
- Python Beginner's Guide, <http://wiki.python.org/moin/BeginnersGuide>, December/05.
- ShapeLib API Documentation, [http://shapelib.maptools.org/shp\\_api.html](http://shapelib.maptools.org/shp_api.html).
- Uchoa, H. N., Ferreira, P. R., Coutinho, R. J. C., Coelho Filho, L. C. T. and Brito, J. L. N. e S. (2005). *Arquitetura OpenGIS baseada em Software Livre para Solução de Geoprocessamento*. In *Anais do XXII Congresso Brasileiro de Cartografia*.