Using GDAL to Translate US Topo GeoPDFs

5/3/2016
Andrew Burnes
aburnes@usgs.gov
US Topo customers are requesting other raster formats.
about **US TOPO**
US Topo (2009 - Present) ≠ USGS Historical Topographic Maps (1947-1992)
Inspired by Historic Topographic Maps

Automated 7.5-min Quadrangle Maps

Updated on 3 Year Production Cycle

Creates a Layered GeoPDF Map
People 😍 US Topo

→ ~10,000,000 downloads per year
→ ~55,000 maps across US & Territories
→ ~18,000 updated maps & data per year
→ Reboot of popular USGS topo maps
Some GeoPDF

- Large File Size ~22MB per file
- Proprietary Software TerraGo Plugin
- Poor Mobile Experience for phone/tablet
- Add’l Processing for GIS people to use
To GDAL

→ Translate to multiple raster formats

→ Customize maps to datasets

→ Clip data to neatline

→ Mosaic multiple topos maps
Setup

→ GDAL>=1.10.0 *(PDF read & write support)*

→ PDF Driver *(select one)*
  - Poppler
  - PDFium *(support starts at GDAL>2.1.0)*
  - PoDoFo
Setup

⇒ GDAL>=1.10.0 (PDF read & write support)

⇒ PDF Driver (select one)
  - Poppler
  - PDFium (support starts at GDAL>2.1.0)
  - PoDoFo
# Check that install has PDF support

$ gdalinfo --formats

[...]
  PDF -raster,vector- (rw+vs): Geospatial PDF
[...]

science for a changing world
# View PDF metadata and available data layers.

```bash
$ gdalinfo -mdd LAYERS ustopo.pdf
```

[...]

Metadata (LAYERS):

- LAYER_00_NAME=Map_Collar
- LAYER_01_NAME=Map_Collar.Map_Elements
- LAYER_02_NAME=Map_Frame
- LAYER_03_NAME=Map_Frame.Projections_and(Grids
- LAYER_04_NAME=Map_Frame.Geographic_Names
- LAYER_05_NAME=Map_Frame.Structure_Features_and_Names
- LAYER_06_NAME=Map_Frame.Boundary_Features_and_Names
- LAYER_07_NAME=Map_Frame.Transportation
- LAYER_08_NAME=Map_Frame.Transportation.Road_Names_and_Shields
- LAYER_09_NAME=Map_Frame.Transportation.Road_Features
[...]
```
customize US Topo
# Translating all US Topo layers to GeoTiff

$ gdal_translate -of "GTiff" ustopo.pdf ustopo.tif

# Custom US Topo to GeoTiff WITHOUT the raster datasets

$ gdal_translate -of "GTiff" \\n   --config GDAL_PDF_LAYERS_OFF "Images,Map_Frame.Terrain.Shaded_Relief" \\n   ustopo.pdf ustopo_vector.tif

# Custom US Topo to GeoTiff ONLY showing terrain datasets

$ gdal_translate -of "GTiff" \\n   --config GDAL_PDF_LAYERS "Map_Collar,Map_Frame.Terrain" \\n   ustopo.pdf ustopo_terrain.tif
# Get and assign neatline WKT geometry to variable

```
$ NLWKT=$(gdalinfo ustopo.pdf | grep 'NEATLINE' | sed 's/^.*POLYGON/POLYGON/')
```

# Using [https://github.com/mapbox/wellknown](https://github.com/mapbox/wellknown) to convert WKT to GeoJSON

```
$ echo $NLWKT | wellknown > nl_unprj.geojson
```

# Project neatline GeoJSON to US Topo coordinate system

```
$ ogr2ogr -f "GeoJSON" -a_srs [CoordSys] neatline.geojson nl_unprj.geojson
```

# Clip map collar from translated US Topo and set NoData to opaque

```
$ gdalwarp -dstalpha -cutline neatline.geojson ustopo.tif ustopo_clip.tif
```
# Recommended translation option flags for all US Topos

# Select or filter layers

```bash
--config [GDAL_PDF_LAYERS/GDAL_PDF_LAYERS_OFF]
```

# Minimize file size

```bash
-co COMPRESS=[LZW/DEFLATE/JPEG/...]
```

# Set output resolution

```bash
--config GDAL_PDF_DPI [integer]
```

# Set output bands for RGB or RGBA

```bash
--config GDAL_PDF_BANDS [3/4]
```
Caveats

Watch Out for Inconsistencies!

➔ **Pixel Resolution** (600dpi vs 300dpi)

➔ **Bands Count** (RGB vs RGBA)

➔ **Map Layers**

➔ **Naming** (Projection_and_Grids vs Projections_and_Grids)
going FORWARD
Possibilities

→ **On-Demand** translation & customization
→ **Storage** for US Topo is cloud bound
→ **Batch** process entire collection
→ **Webmap** built from rasterized US Topos
Benchmarks

Processing the Nation to GeoTIff

~60 secs/topo
~60 topos/hour (3600/60)
~917 total hours (55000/60)
~$97 total cost (917 * 0.105)

Tested on Amazon Web Services - Instance Type C4.Large - Instance Cost $0.105 per Hour
Resources


GDAL Translate - [http://www.gdal.org/gdal_translate.html](http://www.gdal.org/gdal_translate.html)

GDAL Warp - [http://www.gdal.org/gdalwarp.html](http://www.gdal.org/gdalwarp.html)

Ogr2Ogr - [http://www.gdal.org/ogr2ogr.html](http://www.gdal.org/ogr2ogr.html)

GDAL Cheat Sheet - [https://github.com/dwtkns/gdal-cheat-sheet](https://github.com/dwtkns/gdal-cheat-sheet)


GeoTiff Format - [http://www.gdal.org/frmt_gtiff.html](http://www.gdal.org/frmt_gtiff.html)

JPEG Format - [http://www.gdal.org/frmt_jpeg.html](http://www.gdal.org/frmt_jpeg.html)

Wellknown - [https://github.com/mapbox/wellknown](https://github.com/mapbox/wellknown)