

GDAL and Complex Datasets

Frank Warmerdam
OSGeo / Google

Overview

- Introduction to GDAL/OGR
- GDAL Data Model
- Focus on Subdatasets
- HDF4
- HDF5
- NetCDF
- Discussion

Who am I?

- Worked at PCI Geomatics (1991-1998)
 - GeoGateway, EASI+, ImageWorks
- Founder of GDAL/OGR project
- Consultant (1998-2011, GDAL, MapServer)
- maintainer: libtiff, libgeotiff, PROJ.4
- Founding director OSGeo
- Now working in Geo at Google

GDAL/OGR Introduction

- Geospatial Data Abstraction Library
- Raster (GDAL) and Vector (OGR)
- read/write access to many geospatial formats
- Widely used (FOSS+proprietary): GRASS, MapServer, QGIS, FME, ArcGIS, G. Earth
- 13 years old
- 40+ committers (perhaps 10 active)
- A project of OSGeo
- MIT/X Open Source license (non-reciprocal)

GDAL Features

- Coordinate systems around OGC WKT.
- Utilities for translation, warping, subsetting,...
- Efficient support for large images - tiling, overviews
- Written in C++ with C-linkage wrapper
- Language bindings: python, perl, c#, java,...

GDAL Formats

- Plain Raster: jpeg, png, gif
- Geospatial Files: GeoTIFF, .img, NITF
- Wavelet: jpeg2000, ecw, mrsid
- RDBMS: Oracle Raster, PostGIS Raster
- Web Service: WMS, WCS, OPeNDAP
- Radar: CEOS, Envisat
- Elevation: DTED, USGS DEM
- Containers: HDF4, HDF5, NetCDF
- Special: In-memory, VRT

Over 100 Formats

GDAL Data Model - Overview

Dataset:

- XSize/YSize (in pixels)
- Coordinate System
- Georeferencing (...)
- Metadata (...)
- Driver
- FileList
- Bands (0+)

GDAL Data Model - Bands

- XSize/YSize in pixels (same as Dataset)
- Pixel Type: Byte, UInt16, Int16, UInt32, Int32, Float32, Float64, and the complex types CInt16, CInt32, CFloat32, and CFloat64.
- Block Size (for efficient access)
- Metadata (...)
- Description
- Nodata pixel value
- Nodata mask band
- Category Names
- minimum and maximum value.
- offset and scale
- Raster Units (ie. "meters", "pascals").
- Color Interpretation (ie. red, grey, paletted)
- Color Table

GDAL Data Model - Georeferencing

- GeoTransform: Affine (regular, rotation, shear)
- Ground Control Points
 - pixel/line
 - geo x, geo y, elevation
 - name
 - use to create polynomial or thin plate spline transform
- RPCs (ratio of polynomials)
- Geolocation Grids (auxiliary grids of x/y)

GDAL Data Model - Georeferencing

Geolocation:

```
LINE_OFFSET=0
```

```
LINE_STEP=210
```

```
PIXEL_OFFSET=0
```

```
PIXEL_STEP=249
```

```
SRS=PROJCS["UTM Zone 49, Northern Hemisphere",GEOGCS[...],UNIT["Meter",1]]
```

```
X_BAND=1
```

```
X_DATASET=HDF4_EOS:EOS_SWATH_GEOL:"AST_...hdf1":SurfaceRadianceSWIR:Longitude
```

```
Y_BAND=1
```

```
Y_DATASET=HDF4_EOS:EOS_SWATH_GEOL:"AST...hdf1":SurfaceRadianceSWIR:GeodeticLatitude
```

GDAL Data Model - Metadata

- string name/value pairs
- on dataset, and band objects
- segregated into subdomains
 - generic: default, IMAGE_STRUCTURE, RPC, GEOLOCATION, SUBDATASETS
 - driver specific: NITF_DES_METADATA, CGM
 - xml: xml:ESRI, xml:XMP

Subdatasets

- Default assumption: one file has at most one dataset
- Subdatasets: more than one dataset per file
- Areas within file given a pseudo-filename.
- Exposed as metadata when opening the file.

Subdataset Examples

Subdatasets:

```
SUBDATASET_1_NAME=HDF4_EOS:EOS_GRID:"MOD09Q1G_EVI.A2006233.h07v03.005.2008338
190308.hdf":MODIS_NACP_EVI:MODIS_EVI
SUBDATASET_1_DESC=[4800x4800] MODIS_EVI MODIS_NACP_EVI (16-bit integer)
SUBDATASET_2_NAME=HDF4_EOS:EOS_GRID:"MOD09Q1G_EVI.A2006233.h07v03.005.2008338
190308.hdf":MODIS_NACP_EVI:Smoothed_EVI
SUBDATASET_2_DESC=[4800x4800] Smoothed_EVI MODIS_NACP_EVI (16-bit integer)
...
% gdalinfo HDF4_EOS:EOS_GRID:"MOD09Q1G_EVI.A2006233.h07v03.005.2008338190
308.hdf":MODIS_NACP_EVI:MODIS_EVI
...
Driver: HDF4Image/HDF4 Dataset
Files: MOD09Q1G_EVI.A2006233.h07v03.005.2008338190308.hdf
Size is 4800, 4800
...
Band 1 Block=4800x4800 Type=Int16, ColorInterp=Gray
  Description = MODIS EVI Computed from MOD09A1
  NoData Value=32767
  Offset: 0,   Scale:0.0001
```

HDF4

- Primary Types:
 - HDF4_EOS, HDF4_GR, HDF4_SDS, HDF4_GR
- Subtypes:
 - EOS_GRID, EOS_SWATH, SEAWIFS_L3, HYPERION_L1
- Rules For: ASTER L1A, L1B, L2, L3, MODIS L1B, L2, Coastwatch, Naval Research Lab
- Uses NCSA HDF4 library
- Uses "GDAL adapted" portion of HDFEOS lib
- Preliminary work by Andrey Kiselev

HDF4 - Swath

- GEOLOCATION arrays *and* GCPs
- GCPs can be NONE, PARTIAL or FULL (GEOL_AS_GCPS config option)
- Coordinate system handling very product specific
- "Lattice" disables geolocation
- Dimension 2 or 3

HDF4 - Grid

- GDprojinfo to get GCTP projection def
- GDgridinfo for affine transform
- dimension 2, 3 or 4 (unrolled as bands)

HDF4 - GDAL

- GDAL created files are SDS (Scientific Data Set) not HDF-EOS
- SRS in "Projection" attribute
- geotransform in "TransformationMatix" attr
- Other metadata written with SDsetattr()
- No support for gcps, geolocation on write

HDF5

- Just using HDF5 library (no HDFEOS lib)
- Subdataset name are path-within-file
- All rank 2/3 H5G_DATASETs treated as subdatasets.
- "Latitude" and "Longitude" used to create GCPs (ought to be geolocation grids)
- WGS84 SRS assumed
- No affine geotransform support!
- No product specific logic
- Preliminary work by Denis Nadeau (NASA!)

HDF5 - Bag

- HDF5 profile for marine data such as bathymetry
- Custom driver using hdf5 library
- Supports ISO 19115 grid description (georeferencing, coordinate system)
- No SUBDATASETs - a fairly predictable product
- Preliminary work by Frank Warmerdam

NetCDF

http://trac.osgeo.org/gdal/wiki/NetCDF_Improvements

- Variables of rank 2+ treated as datasets
- Dimensions 3+ unrolled as bands (time/z/...)
- Uses netcdf library (hdf4/netcdf conflict!)
- UNIDATA CF Conventions (ie. COARDS)
- Good coordinate systems
- Affine / Geolocation / 1D Geolocation?
- NetCDF4 is HDF5 - netcdf preferred driver
- Now maintained by Etienne Tourigny
- Separate GMT-in-netcdf driver.

Areas for Work

- Geolocation support (HDF5 at least)
- Need more samples, improve SRS handling
- Additional maintainer/commmitter?

Summary

- GDAL data model supports georeferencing
- Not all tools utilize GCPs, Geolocation
- Metadata model is flat
- Unrolling 3/4 dim data as bands not great
- Drivers are often deficient
- HDF4 products vary widely, hard to support

Questions?

See Also:

<http://www.gdal.org/>

http://www.gdal.org/frmt_hdf4.html

http://www.gdal.org/frmt_hdf5.html

http://www.gdal.org/frmt_netcdf.html