HotdoGProposal

HotdoG Proposal

Abstract

The HotdoG project is an effort and suite of tools to convert HDF/HDF-EOS format into GeoTIFF format. This is the first of a potential series of incoming projects originating from The HDF Group – the non-profit organization is interested in evaluating the ASF as a potential home for many of its projects. The HDF Group is an independently funded organization that started many years ago with major investment from NASA as the Hierarchical Data Format (HDF), version 4 and now version 5, is the *de facto* remote sensing data format for NASA missions, and an increasing number of other disciplines including bio medicine, radio astronomy, climate science, and other domains.

HDF is both a data and metadata format, as well as a model for representing and access information. There are numerous downstream tools that can read and write HDF data, including a growing number of Geospatial data tools (ESRI-based, and also OpenGeo and other community led efforts). In addition, major interoperability efforts are also occurring between the remote sensing community and the climate modeling community (which has traditionally favored NetCDF as opposed to HDF) because of the efforts in HDF5 to leverage a common data format and model.

HotdoG is poised to be a first of its kind in the form of bringing one of the major 2 data formats for science to the ASF (the other being the NetCDF format).

Proposal

HotdoG is a software converter that converts Earth Science data in HDF/HDF-EOS format into GeoTIFF format. Doing so easily enables users of remote sensing data to interoperate with common GIS tools (like WebGIS, Web Processing, image analysis, and geo computational tools). We feel that the project is an incremental step, and an appropriate focus with tangible success possibilities by restraining our focus to HDF/HDF-EOS to GeoTIFF conversion.

There are numerous interesting paths that we can take the toolkit in – as conversion from remotely sensed data to GeoTIFF involves ensuring that the HDF-EOS metadata elements can be appropriately represented using GeoTIFF headers, and the associated format. Furthermore, capturing the HDF's appropriate geo datum in GeoTIFF will be another important challenge.

Background

GeoTIFF is a data and metadata standard for Earth science applications. It is based on binary Tagged Image File Format (TIFF). A GeoTIFF file has geographic (or cartographic) data embedded as tags within the TIFF file that are used to geo-locate the image. This is required for correct integration of the image in Geographic Information Systems (GIS) and other popular tools like Google Earth Pro.

In the recent years, GeoTIFF has gained popularity as a visualization format among NASA HDF Earth science user communities according to the NASA data user's survey. However, the conversion from HDF to GeoTIFF is not straightforward for end users because NASA HDF data products are diverse and organized in many different ways. For example, go to http://hdfeos.org/zoo and you'll see many scripting language examples because no single script can correctly visualize all NASA HDF data.

Rationale

The HEG tool is limited to some NASA HDF-EOS2 products (no support for HDF-EOS5 products) and it is not an open-source tool. The latest GDAL (version 1.9.2 and above) is an open-source tool but it cannot handle many non-HDF-EOS NASA HDF products such as TRMM (pure HDF4) and Aquarius (pure HDF5) correctly and automatically.

Initial Goals

We'll improve GDAL to support NASA HDF products better by handling geo-location information and physical meaning of data correctly and automatically.

We'll handle NASA products intelligently so novice users don't have to supply many options or figure out the details about the data products. For advanced users, we'll give a full control of accessing HDF products in many different ways so that the converted GeoTIFF file is scientifically valid and meaningful.

We aim to provide command line tools first and evolve them into a GUI tool.

Current Status

We're looking for people interested in HotdoG and to move the project into the Apache Community.

Meritocracy

We will discuss the milestone and the future plan in an open forum. We plan to encourage an environment that supports a meritocracy. The contributors will have different privileges according to their contributions.

Community

GIS and Earth Science

Core developers

- H. Joe Lee <hyoklee AT hdfgroup DOT org>
- Denis Nadeau <denis DOT nadeau AT gmail DOT com>
- Andrey Kiselev <dron AT ak4719 DOT spb DOT edu>
- Pedro Vicente <pvicente AT uci DOt edu>
- John Evans <john DOT g DOT evans DOT ne AT gmail DOT com>

Alignment

HotdoG employs H4CF Conversion Toolkit for reading data and GDAL for writing GeoTIFF. In addition, we plan to integrate HotdoG with other products from new missions such as SMAP and IceSAT-2.

If HEG becomes available as open source, HotdoG will employ it as well for HDF-EOS2 products.

Known Risks

HDF-EOS2 and HDF-EOS5 libraries are distributed under the NOSA license that may not be compatible with Apache License. This will be vetted with the Legal Affairs Committee and if found incompatible we will find a different way to read HDF-EOS products.

HDF is a flexible format which affords for unusual objects such as Point & Zonal Averages (HDF-EOS), VData (HDF4), and compound datatype (HDF5) that present challenges in mapping to GeoTIFF. To mitigate this challenge the initial goals will focus on those objects which are more common place and easily map to GeoTIFF.

Documentation

User's guide documentation will be provided via Doxygen. The guide will contain specific GeoTIFF conversion examples for each NASA HDF data product.

External Dependencies

- H4CF Conversion Toolkit
- GDAL
- HDF-EOS2 / HDF-EOS5
- HDF4 / HDF5

Required Resources

Mailing List

- hotdog-private (with moderated subscription)
- hotdog-dev
- hotdog-commits

Issue Tracking

JIRA HotdoG (HotdoG)

Other Resources

- Wiki
    - http://wiki.apache.org/HotdoG

Initial Committers

- H. Joe Lee (The HDF Group)
- Mike Folk (The HDF Group)
- Paul Ramirez (NASA JPL)
- Chris Mattmann (NASA JPL)
- Lewis John McGibbney (Stanford University)
- Denis Nadeau (NASA NCCS)
- Pedro Vicente (University of California at Irvine)
- Babak Behzad (University of Illinois at Urbana-Champaign)
- Nawajish Noman (ESRI)
- John Evans
- Andrey Kiselev (SRCES RAS)
- Adam Estrada

Affiliations

- H. Joe Lee (The HDF Group)
- Mike Folk (The HDF Group)
- Paul Ramirez (NASA JPL)
- Chris Mattmann (NASA JPL)
- Lewis John McGibbney (Stanford University)
- Denis Nadeau (NASA NCCS)
- Pedro Vicente (University of California at Irvine)
- Babak Behzad (University of Illinois at Urbana-Champaign)
- Nawajish Noman (ESRI)
- Andrey Kiselev (Scientific Research Center for Ecological Safety Russian Academy of Science)
- Adam Estrada (MDA Information Systems)

Champion

- Paul Ramirez <paul DOT m DOT ramirez AT jpl DOT nasa DOT gov>

Nominated Mentors

- Chris Mattmann
- Paul Ramirez
- Joe Brockmeier
- Greg Reddin

Sponsoring Entity

- Apache Incubator