Hierarchical Data Format for Earth Observing System Data Product Developer’s Guide

HDF-EOS Workshop XXI / The 2018 ESIP Summer Meeting

Hyo-Kyung (Joe) Lee
Software Engineer / The HDF Group
hyoklee@hdfgroup.org

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C. This document does not contain technology or Technical Data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.
Motivation and Related Work

• The work presented in this talk is done in support of *Data Product Developers Guide Working Group*

• WG Mission Statement: *Help Data Product developers make data usable for End Users*

• WG chairs
  — Hampapuram Ramapriyan ([hampapuram.ramapriyan@ssaihq.com](mailto:hampapuram.ramapriyan@ssaihq.com))
  — Peter Leonard ([pleonard@sesda3.com](mailto:pleonard@sesda3.com))

• WG POCs
  — Chris Lynnes ([chris.lynnes@nasa.gov](mailto:chris.lynnes@nasa.gov))
  — Nathan James ([nate.james@nasa.gov](mailto:nate.james@nasa.gov))
  — John Moses ([john.f.moses@nasa.gov](mailto:john.f.moses@nasa.gov))

• The HDF Group members
  — Joe Lee ([hyoklee@hdfgroup.org](mailto:hyoklee@hdfgroup.org)) and Aleksandar Jelenak ([ajelenak@hdfgroup.org](mailto:ajelenak@hdfgroup.org))
Broader HDF-EOS Definition

- **Hierarchical Data Format for Earth Observing System**
- Any Earth data stored in HDF format
  - HDF4, HDF5, and netCDF-4
HDF-EOS Data Product

• Data is a consumer product like food, clothing, and house.
• Design and package it well.
• Users (=consumers) will appreciate it.
What Users Ask through Help Desk

- Geolocation retrieval
- Sampling over region & time
- Creating plots (e.g., Journal publication)
- GDAL* tools (e.g., ESRI ArcGIS)
- netCDF tools (e.g., Panoply)
- Programming in MATLAB

*Geospatial Data Abstraction Library
Better Products = Less Questions

- Improve Earth data user experience
- Self-describing = self-serviceable data
- How to create better data products?
Guide I: Geo-location

• Add latitude/longitude variables
  – Regardless of projection parameters in metadata
• For grids and points, use 1D dataset.
• For swath, use 2D dataset.
  – This will help visualization tools.
• No 3D dataset / No fill value
• Use **units** attribute (e.g., degrees_east and degrees_north)
Why Geo-location?

• Integrated Data Viewer throws “No Gridded data found” error message.
• NCAR Command Line Language cannot plot data if lat / lon has fill values.
Guide II: Named Dimensions

• Essential for netCDF interoperability
• Have named dimensions.
• 1-D coordinate variable, use the same name as dataset name (COARDS*)
• Use netCDF APIs but store as netCDF-4/HDF5 (easy).
• Use HDF5 dimension scale APIs if you don’t want to use netCDF APIs (difficult).
• Check with netCDF-Java tools.

* Cooperative Ocean/Atmosphere Research Data Service
Why named dimensions?

• Strange phony_dim_0 will appear for netCDF tools.
• Dimension names are heavily used by netCDF-Java tools to identify feature types.
• If 1D variable name matches dimension name, it becomes a coordinate variable automatically.
Guide III: The CF Conventions

• CF: Climate and Forecast Metadata
• `long_name` attribute
• `units` attribute
• `coordinates` attribute
• Use templates
Why `long_name` and `units`?

Some tools utilize them automatically!

OMI-Aura_L3-OMTO3e_2017m0105_v003-2017m0203t091906.he5

NCAR Command Line Language
Image from http://hdfeos.org/zoo
Guide IV: Test with tools.

- MATLAB, Python
- Geospatial Data Abstraction Library (GDAL) tools (e.g., gdal_translate)
- NCAR Command Line Language (NCAR)
- toolsUI and Panoply
- Integrated Data Viewer (IDV)
- Interactive Data Language (IDL)
- OPeNDAP (e.g., Hyrax*, THREDDS**)

*Hyrax is the data server from OPeNDAP.
**Thematic Real-time Environmental Distributed Data Services
Question: any tool for guidelines?

Answer: HDF Product Designer (HPD) can help data producers!
HDF Product Designer (HPD)

- Design is key.
- Design twice, produce data once.
- Testing and validation is a must.
  - CF checker from JPL
  - Testing with netCDF-C tool (e.g., ncdump)
  - Testing with THREDDS / Hyrax
Why HDF Product Designer?

• Design and test product quickly.
• Graphical User Interface (GUI)
• Design Templates
  – CF feature types
  – Existing NASA HDF4/HDF5 products
• Testing and validation is built-in.
  – CF convention checker
  – Hyrax/THREDDS
HPD GUI & Design Template
Case Study: JAXA*(Before)
Case Study: JAXA (After 90 min.)
HPD References

- http://hpd.readthedocs.io
- http://youtube.com/hdfeos

HPD Future Work?

- Common Metadata Repository (CMR) integration
- Web-based GUI
This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.