

#### 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 <u>https://wiki.earthdata.nasa.gov/display/ESDSWG/Data+Product+Developer</u> <u>s+Guide+Working+Group</u>
- WG Mission Statement: Help Data Product developers make data usable for End Users
- WG chairs
  - Hampapuram Ramapriyan (<u>hampapuram.ramapriyan@ssaihq.com</u>)
  - Peter Leonard (<u>pleonard@sesda3.com</u>)
- WG POCs
  - Chris Lynnes (<u>chris.lynnes@nasa.gov</u>)
  - Nathan James (<u>nate.james@nasa.gov</u>)
  - John Moses (john.f.moses@nasa.gov)
- The HDF Group members
  - Joe Lee (<u>hyoklee@hdfgroup.org</u>) and Aleksandar Jelenak (<u>ajelenak@hdfgroup.org</u>)



### **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.

Data Source Type:	Grid files (netCDF/GRIB/OPeNDAP/GEMPAK)
URI :	an internet and the second and a long and the second and the secon
0.121	ap.jpi.nasa.gov:ou/opendap/aiiData/oscar/preview/L4/oscar_triird_deg/osca
🥨 N	o Gridded Data
No g	ridded data found for:
htt	p://podaac-opendap.jpl.nasa.gov:80/opendap/allData/oscar/preview/L4/oscar
Do y	ou want to try to load this as another data type?
	OK Cancel



### 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





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

- 🗆 × HDF Product Designer Project Design Group Dimension Dataset Attribute Tools Help 🗒 Templates im NODC\_grid (Group: /) ė-**–** / String Conventions = CF-1.6, ACDD-1.3 string featureType = grid 🛷 string cdm\_data\_type = Grid 🛷 string standard name vocabulary = NetCDF Climate and Forecast (CF) Metadata Convention Standard Name Table 28 🛷 string title = 🛷 string summary = 🛷 string source = string platform = platform variable string instrument = instrument\_parameter\_variable 🛷 string uuid = 🛷 string id = String naming authority = Available designs X Ø string time\_coverage\_start = String time\_coverage\_end = String time\_coverage\_resolution = Select design: float32 geospatial lat min = 0.0 float32 geospatial\_lat\_max = 0.0 NODC arid String geospatial\_lat\_units = degrees\_north NODC point NODC profileIncomplete 🛷 string geospatial lat\_resolution = NODC profileOrthogonal Interpretation of the second secon NODC timeSeriesIncomplete float32 geospatial\_lon\_max = 0.0 NODC timeSeriesOrthogonal string geospatial\_lon\_units = degrees east NODC timeSeriesProfileIncomVIncomT 🛷 string geospatial\_lon\_resolution = NODC timeSeriesProfileIncomVOrthoT float32 geospatial vertical min = 0.0 NODC timeSeriesProfileOrthoVIncomT float32 geospatial\_vertical\_max = 0.0 NODC timeSeriesProfileOrthoVOrthoT T. string geospatial\_vertical units = string geospatial\_vertical\_resolution = string geospatial\_vertical\_positive = OK Cancel 🛷 string institution = string creator\_name = string creator\_url = hvoklee@hdfaroup.org 667de325-1d28-4f0a-8c53-c032efa6d4f2



#### Case Study: JAXA\* (Before)

	HDF Product Designer
	🖉 vlstring NumberOfMissingScans = ['0']
	Ø vlstring AntennaRotationVelocity = ['40.0']
	Ø vlstring ECSDataModel = ['B.0']
	Ø vlstring NumberOfPackets = ["]
	Ø vlstring NumberOfInputFiles = ['2']
	🖉 vlstring NumberMissingPackets = ['0']
	Ø vlstring NumberOfGoodPackets = ['31680']
	Ø vlstring OverlapScans = ['0']
	🖉 vlstring QALocationOfPacketDiscontinuity = ['Continuation']
	🖉 vlstring EphemerisQA = ['OK']
	Ø vlstring AutomaticQAFlag = ['Good']
	🖉 vlstring ScienceQualityFlag = [' 25 [%]']
	🖉 vlstring ScienceQualityFlagExplanation = ['Science Quality Flag [%] = (number of successfully retrieved pixels) / (number of targeted pixels) * 100.
	🖉 vlstring AutomaticQAFlagExplanation = ['1.MissingScanQA:Less than 21 is available->OK,2.MissingDataQA:Less than 321 is available->OK,3.Ante
	Ø vlstring QAPercentMissingData = ['0']
	🖉 vlstring QAPercentOutofBoundsData = ['0']
	Ø vlstring QAPercentParityErrorData = ['0']
	vlstring ProcessingQADescription = ['PROC_COMP']
	Ø vlstring ProcessingQAAttribute = ['']
	🖉 vlstring GlobalMeteorologicalDataType = ['None']
	🔗 vlstring AncillaryDataInformation = [' TiePoint ->TiePoint_20180125_3300300.tp']
$\overline{\mathbf{v}}$	int16 Geophysical Data[1980][243][1](fv=0)
	Ioat32 SCALE FACTOR = [0.10000000149011612]
	🖉 string UNIT = [%]
►	Iloat32 Latitude of Observation Point[1980][243](fv=0.0)
$\overline{\mathbf{v}}$	float32 Longitude of Observation Point[1980][243](fv=0.0)
►	iii uint8 Pixel Data Quality[1][1980][243](fv=0)
►	🕡 float64 Position in Orbit[1980](fv=0.0)
	in float64 Scan Time[1980](fv=0.0)



### 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.

### Raytheon

in partnership with



