EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

NASA Dataset Interoperability Recommendations for Earth Science

Aleksandar Jelenak[1], Peter J. T. Leonard[2], Charlie Zender[3]

[1]The HDF Group, Inc.

[2]ADNET Systems, Inc.

[3]University of California, Irvine

This work was supported by NASA/GSFC under
Raytheon Co. contract number NNG15HZ39C

TBD

NASA ESDSWG

- Earth Science Data System Working Groups (ESDSWG) established under the auspices of NASA Headquarters in 2004.
- Chartered to explore, deliberate, and develop recommendations relevant to NASA's heterogeneous and distributed Earth science data systems.
- Working Groups operate for a 12-month term and are expected to finalize their deliverables during that time.

Dataset Interoperability Working Group (DIWG)

- Formed in 2012 and re-formed every year since.
- Mission: Improve interoperability of NASA-stewarded datasets by:
  - recommending best practices for their structure and content;
  - specifying how to check compliance with various conventions and software tools;
  - modifying or extending relevant conventions.

DIWG Stakeholders

- Data producers (mission or science teams, etc.)
- Library developers (HDF[1], netCDF[2])
- Vetting conventions (CF[3], ACDD[4], HDF-EOS[5])
- Software tool developers
- Other Working Groups

1: Hierarchical Data Format
2: Network Common Data Form
3: Climate Forecast Convention
4: Attribute Convention for Data Discovery
5: Hierarchical Data Format – Earth Observing System

DIWG Recommendations

- Published by NASA Earth Science Data and Information System Standards Office as Suggested Practice documents.
- Two currently published:
  - ESDS-RFC-028 in 2016 with 12 recommendations
  - ESDS-RFC-036 in 2018 with 11 recommendations
- These recommendations are applicable to the Earth Science community at large.

SELECTED DIWG RECOMMENDATIONS

Maximize HDF5/netCDF4 interoperability via API accessibility

We recommend that Earth Science data product files in HDF5 be designed to maximize netCDF4 interoperability by making such HDF5 files accessible from the netCDF4 API to the extent that this is possible.

HDF5 files can be made nearly indistinguishable from netCDF4 files by adding dimension scales to HDF5 files in a way that mimics netCDF shared dimensions.

Include Basic CF Attributes

We recommend that these basic Climate and Forecast (CF) Convention attributes be included in future NASA Earth Science data products where applicable:

Conventions	standard_name	valid_max
units	_FillValue	valid_range
long_name	valid_min	scale_factor
add_offset	coordinates	

Include time coordinate even when having just a single value

We recommend that [HDF5] datasets in grid/swath structures include a time dimension, even if time is degenerate (i.e., includes only one value) for the cases when the entire grid/swath has one time range or time stamp.

Order dimensions of datasets to facilitate readability of grid/swath datasets

We recommend that the dimensions in grid/swath structure [HDF5] datasets be ordered in a manner that facilitates readability for the anticipated end users.

Consider “balanced” chunking for 3D datasets in grid structures

We recommend that "balanced" chunking be considered for three-dimensional datasets in grid structures.

“Balanced chunking” is a chunking (tiling) scheme which attempts to balance the chunk (tile) size in order to achieve similar performance for both temporal and spatial subsetting data access.

Use the units attribute only for variables with physical units

We recommend adhering to the CF convention's guidance on the use of the units attribute with the following clarifications:

- Unitless (dimensionless in the physical sense) property of the data in a variable is indicated by the lack of a units attribute, unless:
  - appropriate physical units do exist;
  - use of dimensionless units identifiers is common practice in the target user community.
- Values of the units attribute should be supported by the UDUNITS-2 library.
- A variable used in any context other than data storage must never have the units attribute.

Keep coordinate values in coordinate variables

We recommend that all coordinate values be stored in coordinate variables. No coordinate values, or any part thereof, should be stored in attributes, variable names or group names.

Avoid encoding time coordinate data in group hierarchies like 2017/01/30 (these are three groups named "2017", "01", and "30", respectively).

Include Georeference Information with Geospatial Coordinates

We recommend that Earth Science dataset granules be produced with complete georeferencing information for all their geospatial coordinates. This georeference information should be encoded in an interoperable way based on the CF convention and the following specific guidelines:

- Granules are required to contain the most applicable type of geospatial coordinates for their data. The decision whether to provide any additional type of geospatial coordinates is left to the data producer.
- The georeference information should be given as both grid mapping variable attributes and OGC Well-Known Text (WKT), whenever possible.
- In all other cases the georeference information should be given in either of the two formats that supports it.
- The preference when processing georeference information should be given to the WKT content if available.

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.

Raytheon

in partnership with

ati

Bellfore Consulting GROUP

Element 84

GENEX SYSTEMS

INNOVIM The Power of Innovation

MARYLAND HAWK CORPORATION

OPeNDAP

REAN CLOUD

Red Canyon Engineering & Software

STINGER GHAFFARIAN TECHNOLOGIES

The HDF Group

VALADOR

EOSDIS