ESDS-RFC-XXXv1.0
Aleksandar Jelenak, Peter J.T. Leonard, Charles S. Zender, et al.
Suggested Practice
February 2024

Dataset Interoperability Recommendations for Earth Science: Part 3

Status of this RFC

This RFC provides information to the NASA Earth Science community. This RFC does not specify an Earth Science Data Systems (ESDS) standard. Distribution of this memo is unlimited.

Change Explanation

Version 1.0 February 2024 Original document.

Copyright Notice

Abstract

This document contains thirteen new recommendations officially adopted by the NASA Earth Science Data System (ESDS) Dataset Interoperability Working Group (DIWG). They are the continuation of the DIWG recommendations published as ESDS-RFC-028 and ESDS-RFC-036 with the same goal of improving the interoperability of Earth Science dataset files. The DIWG recommendations here represent best practice instructions covering a diverse range of issues related to the content of dataset files and how some of the data's properties are conveyed in their file names. Some of the recommendations capture already prevailing community practices while others clarify or simplify among several possible options.

Table of Contents

1 Introduction

The Earth Science Data System Working Groups (ESDSWG) is a NASA organization established under the auspices of NASA Headquarters in 2004. The chartered role of the ESDSWG focuses on the exploration and development of recommendations derived from pertinent community insights of NASA's heterogeneous and distributed Earth Science data systems.

The purpose of the Dataset Interoperability Working Group (DIWG) is to formulate, deliberate and make recommendations to help ensure that Earth Science datasets smoothly interoperate with each other regardless of their origin. The DIWG recommendations herein embody best practices to reduce and bridge gaps between geoscience dataset formats widely used at NASA and elsewhere, and to improve dataset compliance, discoverability, and extensibility with relevant metadata conventions.

This document presents thirteen new recommendations from the DIWG regarding dataset interoperability.

All of the DIWG Recommendations, both old and new, can be found at https://wiki.earthdata.nasa.gov/display/ESDSWG/Dataset+Interoperability+Recommendations+for+Earth+Science

2 Dataset Interoperability Recommendations for Earth Science: Part 1

The DIWG Recommendations Part 1 document (ESDS-RFC-028) was initially released in 2016 (Version 1.0), and was most recently revised in 2018 (Version 1.3). The Part 1 document includes 12 Recommendations numbered 2.1 through 2.12:

2.1 Maximize HDF5/NetCDF4 Interoperability via API Accessibility

2.2 Include Basic CF Attributes
2.3 Use CF “bounds” Attribute
2.4 Verify CF Compliance
2.5 Distinguish Clearly Between HDF and NetCDF Packing Conventions
2.6 When to Employ Packing Attributes
2.7 Mapping Between ACDD and ISO
2.8 Make HDF5 Files NetCDF4-Compatible and CF-Compliant Within Groups
2.9 Include time Dimension in Grid Structured Data
2.10 Order Dimensions to Facilitate Readability of Grid Structure Datasets
2.11 Consider "Balanced" Chunking for 3-D Datasets in Grid Structures
2.12 Include Datum Attributes for Data in Grid Structures

3 Dataset Interoperability Recommendations for Earth Science: Part 2

The DIWG Recommendations Part 2 document (ESDS-RFC-036) was initially released in 2019 (Version 1.0), and was most recently revised in 2020 (Version 1.2). This document includes 11 Recommendations numbered 3.1 through 3.11:

3.1 Character Set for User-Defined Group, Dataset, and Attribute Names
3.2 Consistent units Attribute Value for Variables Across One Data Collection
3.3 Use the units Attribute Only for Variables with Physical Units
3.4 Include time Coordinate in Swath Structured Data
3.5 Keep Coordinate Values in Coordinate Variables
3.6 Include Georeference Information with Geospatial Coordinates
3.7 Not-a-Number (NaN) Value
3.8 Standardize File Extensions for HDF5/NetCDF Files
3.9 Ensure Granule's Filename Uniqueness Across Different Dataset Releases
3.10 Adopt Semantically Rich Dataset Release Identifiers
3.11 Date-Time Information in Granule Filenames

4 New Dataset Interoperability Recommendations for Earth Science

Presented here are 13 new DIWG Recommendations numbered 4.1 through 4.13:

4.1 Attach the CF flag_values or flag_masks Attributes Along With the CF flag_meanings Attribute to Each Flag Variable
4.2 Avoid Use of the missing_value Attribute
4.3 Define the Projection Ellipsoid to Match the Reference Datum
4.4 Document Missing Granules for Instruments That Acquire Data on a Regular Basis
4.5 Include Only One Variable per GeoTIFF File
4.6 Indicate in CRS Metadata the Order of Elements in Horizontal Coordinate Pairs
4.7 Make a Variable's Valid Data Range Consistent Within Each Product Release
4.8 Make a Variable's Valid Data Range Useful
4.9 Use a Number Outside of the Valid Data Range for a Variable's Fill Value

4.1 Attach the CF flag_values or flag_masks Attributes Along With the CF flag_meanings Attribute to Each Flag Variable

Recommendation:

Attach the CF flag_values or flag_masks attributes along with the CF flag_meanings attribute to each flag variable in an Earth Science data product. The choice of which to use depends on the use case.

Recommendation Details:

A common practice in Earth Science data products is to associate a flag variable with a science data variable to describe the quality of each successful retrieval and to explain each occurrence of the fill value - we recommend that the CF flag_values and flag_meanings attributes be attached to the flag variable in cases where the coded values are mutually exclusive. In the case of a range of conditions (more than two) for each retrieval is expected flag_masks should be used instead of flag_values. In practice, there are often two flag variables for a physical retrieval; one that contains the scalar flag_values (i.e., the overall quality of each pixel), and another with the flag_masks that contains more specific flags for algorithm performance and outlier detection.

In rare (and complex) flagging cases they can be used together.

For an example of such usage consult the CF documentation
(Example 3.5
https://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/build/ch03s05.html
, or here
https://cfconventions.org/Data/cf-conventions/cf-conventions-1.10/cf-conventions.html#flags
).

flag_values/flag_meanings example:

Use flag_values where a single status condition is appropriate to flag every retrieval.

For example, a science data variable named

total_column_ozone

with fill value

_FillValue = -999.9

could be accompanied by an associated flag variable (type byte) named

total_column_ozone_flags

to describe the quality of each successful retrieval as an enumerated list of status flags and to explain each occurrence of the fill value. The total_column_ozone_flags variable should have the CF flag_values and flag_meanings attributes attached to specify the flag values and meanings. For example,

flag_values = 0b, 1b, 2b, 3b, 4b, 5b;

and

flag_meanings = "good_sample glint_contamination high_sza non_convergence row_anomaly_error missing_input_data";

Note that there is a simple one-to-one mapping between the values of these two attributes. Also note that the values of flag_values are comma-separated, while the values of flag_meanings are space-separated, with an underscore being used as the word separator within each value of flag_meanings.

flag_masks/flag_meanings example:

Use flag_masks where a number of independent Boolean (binary) conditions using bit field notation are appropriate to describe a possible range of conditions for each retrieval.

For example, a science data variable named

sea_surface_temperature

could be accompanied by an associated flag variable (type byte) named

condition_flags

to describe the conditions of each retrieval as a set of boolean status flags . The condition_flags variable should have the CF flag_masks and flag_meanings attributes attached to specify the flag values and their meanings. For example,

flag_masks = 1b, 2b, 4b, 8b, 16b;

and

flag_meanings = "ocean land ice lake river";

As an example of implementation, a retrieval could have both ocean and detected ice in it, a condition that would be flagged with the value 1b + 4b = 5 (i.e., 2^0 + 2^2).

4.2 Avoid Use of the missing_value Attribute

Recommendation:

Avoid use of the missing_value attribute in new Earth Science data products.

Recommendation Details:

The missing_value attribute has been semi deprecated, and so this attribute should not be used in new Earth Science data products.

Historically, the missing_value attribute has been used both as a scalar (single-value) and as an array (multi-value) in Earth Science data products.

Use the CF _FillValue attribute instead of using missing_value as a scalar attribute.

Use a flag variable with the CF flag_meanings plus flag_values and/or flag_masks attributes attached (as explained in Recommendation 4.1) instead of using missing_value as an array attribute.

It is acceptable to continue to use the missing_value attribute in already published Earth Science data products, especially in cases where the downstream software specifically makes use of the missing_value attribute, though it would be a worthwhile improvement to include, in new releases of such products, the CF _FillValue attribute, and, where appropriate, flag variables with CF flag_meanings plus flag_values and/or flag_masks attributes attached.

4.3 Define the Projection Ellipsoid to Match the Reference Datum

Recommendation:

Define the projection ellipsoid to match the reference datum to minimize potential errors in geolocation and reprojection.

Recommendation Details:

When producing geolocated image data derived from satellite-based or airborne remote sensing instruments, we recommend defining the projection ellipsoid to be the same as the datum used

by the remote sensing system to define geodetic latitude and longitude. Specific details depend on the selected file format and metadata conventions. Examples provided in this recommendation use GeoTIFF terminology, but the recommendation also applies to other formats and metadata conventions. For example, when using GeoTIFF to represent content in a Projected Coordinate Reference System (PCRS), the projection ellipsoid should be the same as the Geodetic Reference Frame (datum) used by the remote sensing system. For many currently operating satellite instruments, the reported geolocation is referenced to the World Geodetic System (WGS) 1984 datum. Airborne instruments that are geolocated using GPS instruments are also referenced to WGS 84. When geolocated data from one of these instruments are used to create derived geophysical products, data producers may choose a PCRS that includes a map projection based on a reference ellipsoid. To ensure maximum interoperability when transforming such data products, we recommend choosing the PCRS map projection ellipsoid to match the underlying Geodetic Reference Frame (datum). This will minimize potential for geolocation error with overlays of related geolocated information such as coastlines or comparison data products.

In the following discussion, we use the ISO19111 terminology explained in Section 2.1.2 and Appendix A of Iliffe and Lott (2008):

Image transformation of projected data products may require a coordinate conversion or a coordinate transformation. A coordinate conversion is a change of coordinates from one Coordinate Reference System (CRS) to another. The CRSs can be based on the same datum, or, if the datums are different, no algorithm is applied to transform the coordinates of one datum to the other. Iliffe and Lott (2008) contend that, since coordinate conversions are considered to be exact, there is no loss of positional accuracy when a coordinate conversion is performed without transforming differing datum coordinates. A coordinate transformation is a change of coordinates from one CRS to another, in which the CRSs are based on different datums. In this case, a coordinate transformation algorithm is applied to convert the coordinates of one CRS to conform to the datum of the other CRS. (Further discussion and a case study example are included in Section 2. of Brodzik et al. 2012).

It is possible to properly encode both the PCRS map projection ellipsoid and the Geodetic Reference Frame datum in, for example, GeoTIFF metadata. However, some software packages may either incorrectly assume they are the same, or require that they be the same, in order to perform accurate coordinate conversions. Depending on spatial resolution of the image content, the effects of performing an incorrect conversion and/or transformation may not be visually apparent, or may only be apparent if the data include an obvious feature like a subtly shifted coastline when overlaid with independently-derived coastline vectors. A detailed example depicting NASA Operation Ice Bridge flightlines on incorrectly transformed NASA Blue Marble imagery is included in Brodzik et al. (2012).

Defining the PCRS map projection ellipsoid to match the underlying Geodetic Reference Frame ensures that software packages making this assumption will do the right thing, and will eliminate

the time that users might otherwise have to spend to direct the software to only perform the requisite coordinate conversion or coordinate transformation.

In the case of GeoTIFF, this recommendation may become obsolete, given the more specific user-defined details articulated in the OGC GeoTIFF Standard v1.1 (2019). We note that section B.2.3 of this document explicitly acknowledges historical examples that have used a spherical projection ellipsoid, but discourages the use of spheroids for modern applications. Given the recent date of this standard, it remains to be seen in practice how closely software packages adhere to encoded projection ellipsoids and reference datums when performing coordinate conversions and transformations.

4.4 Document Missing Granules for Instruments That Acquire Data on a Regular Basis

Recommendation:

A list of all missing granules should be provided on a permanent Web site (e.g., the DOI landing page) for each Earth Science data product.

Recommendation Details:

It is not uncommon for an Earth Science product to be missing some data granules. For example, data gaps can occur because of missing L0/1 data (e.g., the instrument was not operating, the instrument was in a mode that did not produce useful observations, or a permanent loss of telemetry occurred).

A list of all missing granules should be provided on a permanent Web site (e.g., the DOI landing page) for each Earth Science data product, because it can save users much time if the missing data are clearly identified and explained. This list should also explain the reason for each and every missing granule. A list of missing granules that is machine readable would be very useful.

This recommendation applies to instruments that are expected to acquire data on a regular basis (e.g., MODIS, MOPITT, CERES). This recommendation does not apply to instruments like ASTER, which have an irregular observation schedule.

4.5 Include Only One Variable per GeoTIFF File

Recommendation:

Include only one variable per GeoTIFF File.

Recommendation Details:

The GeoTIFF format was initially developed during the early 1990's with the objective being to leverage a mature platform independent file format (TIFF) by adding metadata required for describing and using geographic image data (OGC, 2019).

The term "Band" is associated with the GeoTIFF format. As defined in the OGC GeoTIFF Standard version 1.1 (OGC, 2019), Band represents a "range of wavelengths of electromagnetic radiation that produce a single response by a sensing device."

With the ease of use and strong support from both commercial and open source software (ESO, 2019), the GeoTIFF format has been widely used in the Earth science community (and more) to store data beyond what the GeoTIFF format was originally designed for. Some data producers tend to pack multiple different variables as different bands in a GeoTIFF file, even though these different variables do not necessarily represent "response" from measurements with "range of wavelengths". One such example is the "ABoVE: AirSWOT Ka-band Radar over Surface Waters of Alaska and Canada, 2017" dataset archived at the ORNL DAAC. Each of these GeoTIFF files contains 6 bands, representing 6 different variables (elevation, incidence_angle, magnitude, interferometric correlation, height_sensitivity, and error_bar).

We Discourage Packing Multiple Different Variables in a Multi-Band GeoTIFF File:

At this moment, there is no commonly-agreed-upon approach to tag GeoTIFF files at the band-level to associate metadata (e.g. variable name and units) to each individual band. Even though there are libraries (like GDAL) that support adding custom-defined metadata tags to individual bands of GeoTIFF files, many tools and applications still lack the capability to add or use such custom-defined metadata tags at the band-level. Such GeoTIFF files become less self-descriptive and thus decreasing their interoperability. Different data variables must be in the same data type (e.g. Int16 and Float32) once being physically stacked as multiple bands in a GeoTIFF file.

Some Advice Regarding Storing Complex Data in GeoTIFF Files:

Rather than storing complex data (e.g., multiple data variables and multi-dimensional data) in GeoTIFF files, please consider using formats like HDF5 and/or netCDF-4, which have established community standards/conventions to embed variable- and file-level metadata inside the files to make such files self-descriptive and interoperable, and the data from such files can be used to generate single-variable GeoTIFF files.

If the GeoTIFF format is really preferred, then please include only one variable per GeoTIFF file, and consider an approach, such as GDAL's Virtual Dataset (VRT) format, to "virtually" aggregate multiple single-variable GeoTIFF files together into a metadata-rich data file/asset without having to change their native data types.

4.6 Indicate in CRS Metadata the Order of Elements in Horizontal Coordinate Pairs

Recommendation:

Indicate in CRS metadata the order of latitude and longitude in coordinate pairs in Earth Science data products.

Recommendation Details:

There is no universal agreement regarding the order of horizontal coordinate pairs (i.e., (longitude, latitude) vs. (latitude, longitude)) in Earth Science data products. Axis ordering may be specified in the full description of the Coordinate Reference System (CRS) as given in a registry such as EPSG. If the order is not specified in a registered CRS, or the CRS is not in a registry, we recommend using the optional axis order keyword in the well-known text (WKT) representation of a CRS (ISO 19162:2019). The order keyword can be added after the mandatory direction keyword as shown in this example:

AXIS["longitude",east,ORDER[1]],

AXIS["latitude",north,ORDER[2]],

If a CRS is not specified using WKT, then we recommend that ISO 6709 be followed, which states that the following shall apply when no CRS is provided:

a) within a coordinate tuple, the latitude value shall precede the longitude value;

b) latitudes on or north of the equator shall be positive, latitudes south of the equator shall be negative;

c) longitudes on or east of the prime meridian shall be positive, longitudes west of the prime meridian shall be negative. The 180th meridian shall be negative. The prime meridian shall be Greenwich;

4.7 Make a Variable's Valid Data Range Consistent Within Each Product Release

Recommendation:

The valid data range for each variable in an Earth Science data product should be made consistent within each product release, and should not vary file-to-file within a given product release.

Recommendation Details:

There are cases of published Earth Science data products with valid data ranges for some variables (specified via the CF valid_min and valid_max attributes, or via the CF valid_range

attribute) that vary file-to-file, based on the actual data range for each particular variable within each particular product file - we do not agree with this approach.

The valid data range for each variable in an Earth Science data product should be made consistent within each product release, and should not vary file-to-file within a given product release.

The valid data range for any given variable should be based upon the relevant physics for that variable, and, possibly, constraints placed upon the variable based on the characteristics of the sensor(s) that collected the data. The relevant physics should not vary file-to-file. The characteristics of the sensor(s) can change with time, in which case we recommend that the widest valid data range consistent with the changing characteristics of the sensors(s) be used for each variable within a given product release.

We note that the choice of the valid data range for any given variable places a constraint on the choice of the fill value for that variable, because the fill value must be a number outside of the valid data range (as explained in Recommendation 4.9).

4.8 Make a Variable's Valid Data Range Useful

Recommendation:

The valid range for each variable in an Earth Science data product should put useful constraints on the data.

Recommendation Details:

Declaring the valid range of a variable's data according to the CF metadata conventions is part of an earlier DIWG recommendation (see Recommendation 2.1 of ESDS-RFC-028). The data value range can be specified either by two CF attributes, valid_min and valid_max , or via the valid_range CF attribute. Only one of these approaches should be used for a given variable.

The data ranges declared using these attributes are dependent on the type of data and their intended application, and should be chosen to place meaningful constraints on the possible data values. The CF metadata conventions require that any data value representing missing data or the variable's fill value must be excluded from the valid data range.

Generic range values are discouraged unless the actual data range is poorly understood. For example, we strongly discourage using the limits of a specific computer data type (e.g., floating-point single or double precision) as the valid range. If the valid range is poorly understood for a particular variable, then it would be best not to include the valid range attribute(s) for that variable.

A useful valid range allows scientists and other users to filter out values that violate physics or known characteristics of the sensor. It also allows visualization programs to either ignore such points or display them with a special style to warn users of the constraint violation.

4.9 Use a Number Outside of the Valid Data Range for a Variable's Fill Value

Recommendation:

The fill value of a variable should be a number outside its valid data range.

Recommendation Details:

The CF _FillValue attribute is used to indicate missing or invalid data for a variable. Also, the value of the CF _FillValue attribute should match the actual fill value used for the variable in the file.

The value of the CF _FillValue attribute should be a mathematically valid number that lies outside the valid range for a variable. Please note that NaN (Not-a-Number) is neither a number nor is it mathematically valid, and, thus, should not be used as the fill value (see Recommendation 3.7 of ESDS-RFC-036).

If possible, using zero as the fill value should be avoided, because zero looks too much like a physically realistic value, and this can be confusing to the product users.

There should only be one fill value per variable. We recommend using a quality flag variable along with the CF flag_values and flag_meanings attributes to explain the various reasons for using the fill value, instead of using several special values in the variable.

4.10 Use DOIs for Referencing Documentation

Recommendation:

A space-separated list of documentation DOIs should be used in the CF references attribute in Earth Science data products, both globally and for specific variables.

Recommendation Details:

The CF references attribute is useful for storing information regarding documentation in Earth Science data products. The CF references attribute can exist both globally and at the variable level. The most concise way to reference a document is via its DOI. We suggest that a space-separated list of documentation DOIs should be used in the CF references attribute in Earth Science data products. Use of the URL form of the DOI is strongly recommended. Also, URLs of relevant documents that do not have DOIs can be used in the CF references attribute.

4.11 Use Double Precision When Archiving Time in Seconds Since a Specific Epoch

Recommendation:

Use double precision when archiving time in seconds since a specific epoch.

Recommendation Details:

Earth Science data products must preserve time-related information with sufficient precision to resolve all timescales relevant to the data itself, to other data with which it may be intercompared, and to conventions for the numeric representation of time, such as Coordinated Universal Time (UTC). Geoscientific datasets commonly report time in intervals (such as seconds) measured from a particular epoch. Resolving one second on the 50+-year timescale from the UNIX/POSIX epoch (00:00:00 UTC on 1 January 1970) to the present day can require up to ten significant digits of temporal resolution, whereas the IEEE-754 single-precision (32 bit) floating point representations preserves at most seven significant digits. Resolving time to the nearest microsecond can require up to six more digits, for a total of sixteen digits, approximately the maximum precision of an IEEE-754 double-precision (64-bit) floating point number. Therefore, preserving sufficient temporal precision to label, store, and intercompare geoscientific data requires double-precision storage.

The most straightforward way of implementing this recommendation in an Earth Science data product is to make use of a double-precision time variable.

A somewhat less straightforward, but perfectly legitimate, way of implementing this recommendation in an Earth Science data product is to

1) include in each Earth Science data product file a double-precision reference time in seconds since a specific epoch

and

2) provide within the Earth Science data product file the time of each individual observation via a single-precision (perhaps scaled) integer w.r.t. the double-precision reference time.

Combining 1) and 2) results in time in seconds since a specific epoch in double precision. The limitation of this approach depends on how many digits to the right of the decimal place must be included. If time is to be reported to the nearest 0.01 second, then a single-precision integer would require a scale factor of 0.01 for conversion to seconds, in which case a single-precision integer could only hold one day's worth of information w.r.t. the reference time, and would limit this approach to sub-daily and daily product files.

4.12 Use Only Officially Supported Compression Filters on NetCDF-4 and NetCDF-4-Compatible HDF5 Data

Recommendation:

Only compression filters that are officially supported by a default installation of the current netCDF-4 software distribution should be used in Earth Science data products in netCDF-4 or netCDF-4-compatible HDF5 formats.

Recommendation Details:

NetCDF-4 has enabled access to non-default (i.e., non-DEFLATE) HDF5 compression filters starting from version 4.7.0. However, the filter identification and access are currently obscure (~five digit IDs) and non-portable (no guarantees client software will be able to decompress them). DEFLATE is currently the only compression filter that is guaranteed to work with default (non-customized) netCDF-4 installations, and so DEFLATE is the only compression filter that should be used in interoperable Earth Science data products in netCDF-4 or netCDF-4-compatible HDF5 formats. Use of the shuffle filter is not prohibited since it is not a compression filter and is supported by the netCDF-4 default installation. Combining the shuffle and the DEFLATE filters can noticeably improve the data compression ratio.

This recommendation may evolve as netCDF-4 extends its default compression filter support to other filters. Description of netCDF-4 filter support can be found at https://docs.unidata.ucar.edu/netcdf-c/current/filters.html .

4.13 Use the Null Character as the Fill Value for String Data and Metadata

Recommendation:

Use the null character as the fill value for string data and metadata in an Earth Science data product.

Recommendation Details:

Inaccurate values of string data or metadata, whether static or dynamic, should be avoided in an Earth Science data product, and the simplest solution is to replace such values with the appropriate fill value.

We recommend that the null character be used as the fill value for string data and metadata in an Earth Science data product.

The null character is represented by "achar(0)" in the Fortran 90 programming language.

The null character is represented by "\0" in the C programming language.

The null character is represented by "nf90_fill_char" in the netCDF-4 library Fortran 90 interface.

The null character is represented by "nc_fill_char" in the netCDF-4 library C interface.

5 Families of Dataset Interoperability Recommendations for Earth Science

With the advent of the new set of DIWG Recommendations presented in Section 4 it became apparent that there are two Families and one Super Family of Dataset Interoperability Recommendations for Earth Science.

5.1 "Valid Range" Family of Recommendations

There are three DIWG Recommendations related to the CF valid_* attributes:

1) (the valid_* part of) Recommendation 2.2 "Include Basic CF Attributes",
2) Recommendation 4.7 "Make a Variable's Valid Data Range Consistent Within Each product Release", and
3) Recommendation 4.8 "Make a Variable's Valid Data Range Useful".

These three recommendations should be considered together when developing an Earth Science data product.

5.2 "Fill Value" Family of Recommendations

There are two DIWG Recommendations directly related to the _FillValue attribute:

1) (the _FillValue part of) Recommendation 2.2 "Include Basic CF Attributes", and
2) Recommendation 4.9 "Use a Number Outside of the Valid Data Range for a Variable's Fill Value".

The DIWG has agreed that these two Recommendations alone are enough to represent a Family of Recommendations.

However, there are two additional Recommendations that mention the _FillValue attribute:

1) Recommendation 3.7 "Not-a-Number (NaN) Value", which suggests that NaN should not be used as the fill value, and
2) Recommendation 4.2 "Avoid Use of the missing_value Attribute", which suggests that _FillValue should be used instead of missing_value.

These four recommendations should be considered together when developing an Earth Science data product.

Also, in the future, the DIWG might develop a Recommendation that suggests specific values for the _FillValue attribute for each data type, which would bring the total number to five in the "Fill Value" Family of Recommendations.

5.3 “Valid Range and Fill Value” Super Family of Recommendations

The CF valid_* attributes are also related to the _FillValue attribute, because the fill value must be a number outside of the valid data range, and so the "Valid Range" Family of Recommendations and the "Fill Value" Family of Recommendations together represent a Super Family of DIWG Recommendations.

6 Endorsement of the Data Product Development Guide for Data Producers

The Data Product Development Guide for Data Producers (DPDG) is an aid aimed primarily at data producers who develop Earth Science data products that are to be archived at an EOSDIS DAAC, though product developers who do not archive their products at an EOSDIS DAAC may also find the DPDG to be useful.

The DIWG endorses the entire DPDG, especially Appendix D on global attributes and Appendix E on variable-level attributes.

Also, Section 3.3 of the DPDG on Cloud-Optimized Formats and Services, which is new in Version 2.0 of the DPDG, should be helpful for the development of data products for use in a cloud environment.

7 References

A. NASA ESDS Dataset Interoperability Working Group
"Dataset Interoperability Recommendations for Earth Science"
https://wiki.earthdata.nasa.gov/display/ESDSWG/Dataset+Interoperability+Recommendations+for+Earth+Science

B. Charles S. Zender, Peter J.T. Leonard, et al.
"Dataset Interoperability Recommendations for Earth Science" 2016
https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-028v1.3.pdf

C. Aleksandar Jelenak, Peter J.T. Leonard, et al.
"Dataset Interoperability Recommendations for Earth Science: Part 2" 2020
https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-036v1.2.pdf

D. Climate and Forecast Metadata Conventions
Section 3.5 Flags
https://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/build/ch03s05.html
and
https://cfconventions.org/Data/cf-conventions/cf-conventions-1.10/cf-conventions.html#flags

E. Iliffe, J. and Lott, R.

"Datums and Map Projections for Remote Sensing, GIS and Surveying, 2nd Edition" 2008 May 14
Whittles Publishing, Dunbeath, Scotland

F. Brodzik, M. J., B. Billingsley, T. Haran, B. Raup, and M. H. Savoie
"EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets" 2012
ISPRS International Journal of Geo-Information, Volume 1, Number 1, Pages 32-45
https://doi.org/10.3390/ijgi1010032

G. Open Geospatial Consortium
"OGC GeoTIFF Standard, Version 1.1" 2019 September 14
OGC Document Number 19-008r4
http://docs.opengeospatial.org/is/19-008r4/19-008r4.html

H. Fayne, J.V., L.C. Smith, L.H. Pitcher, and T.M. Pavelsky
"ABoVE: AirSWOT Ka-band Radar over Surface Waters of Alaska and Canada, 2017" 2019
ORNL DAAC, Oak Ridge, Tennessee, USA
https://doi.org/10.3334/ORNLDAAC/1646

I. ESDIS Standards Coordination Office
"ESCO Recommended Standard, OGC GeoTIFF Standard, Version 1.1" 2019
https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/geotiff
https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-040v1.1.pdf

J. International Organization for Standardization
ISO 19162:2019
Geographic information
Well-known text representation of coordinate reference systems
2019-07

K. International Organization for Standardization
ISO 6709:2022
Standard representation of geographic point location by coordinates
2022-09

L. Unidata Network Common Data Form (NetCDF)
"NetCDF-4 Filter Support"
https://docs.unidata.ucar.edu/netcdf-c/current/filters.html

M. Ramapriyan, H. K., and P. J. T. Leonard.
"Data Product Development Guide (DPDG) for Data Producers Version 1.1" 2021
NASA Earth Science Data and Information System Standards Coordination Office
https://doi.org/10.5067/DOC/ESO/RFC-041VERSION1

N. Ramapriyan, H. K., P. J. T. Leonard, E. M. Armstrong, et al.
"Data Product Development Guide (DPDG) for Data Producers Version 2.0" 2023
NASA Earth Science Data and Information System Standards Coordination Office
https://doi.org/10.5067/DOC/ESCO/RFC-041VERSION2

8 Authors

8.1 DIWG Leadership:

Aleksandar Jelenak
ajelenak@hdfgroup.org

Peter J.T. Leonard
peter.j.leonard@nasa.gov

Charles S. Zender
zender@uci.edu

8.2 DIWG Recommendation Suggesters, Drafters and Re-Drafters:

(Excludes Anyone Listed Previously)

Edward M. Armstrong
edward.m.armstrong@jpl.nasa.gov

Walter E. Baskin
(Retired)

Mary J. Brodzik
brodzik@colorado.edu

Siri Jodha S. Khalsa
khalsa@colorado.edu

James E. Johnson
james.e.johnson-1@nasa.gov

Wen-Hao Li
wen-hao.li@jpl.nasa.gov

Yaxing Wei
weiy@ornl.gov

8.3 DIWG Recommendation Major Commentators:

(Excludes Anyone Listed Previously)

Marty Brewer
brewer@remss.com

Allan Doyle
(Retired)

Christopher Lynnes
(Retired)

David F. Moroni
david.f.moroni@jpl.nasa.gov

Ewan O'Sullivan
eosullivan@cfa.harvard.edu

Byron V. Peters
byron.v.peters@nasa.gov

Hampapuram K. Ramapriyan
hampapuram.ramapriya@ssaihq.com

Vardis M. Tsontos
vardis.m.tsontos@jpl.nasa.gov

9 Abbreviations and Acronyms

ACDD - Attribute Convention for Data Discovery
API - Application Program Interface
ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer
CERES - Clouds and the Earth's Radiant Energy System
CF - Climate and Forecast Metadata Conventions
CRS - Coordinate Reference System
DAAC - Distributed Active Archive System
DEFLATE - A data compression algorithm based on Huffman coding and LZ77 compression.
DIWG - Dataset Interoperability Working Group
DOI - Digital Object Identifier
DPDG - Data Product Development Guide for Data Producers
EOSDIS - Earth Observing System Data and Information System
EPSG - European Petroleum Survey Group

ESCO - ESDIS Standards Coordination Office
ESDIS - Earth Science Data and Information System
ESDS - Earth Science Data System
ESDSWG - Earth Science Data System Working Groups
ESO - EDSIS Standards Office (renamed to ESCO)
GDAL - Geospatial Data Abstraction Library
GeoTIFF - Georeferenced Tagged Image File Format
HDF - Hierarchical Data Format
HDF5 - Hierarchical Data Format Version 5
IEEE - Institute of Electrical and Electronics Engineers
ISO - International Organization for Standardization
LZ77 - A lossless compression algorithm invented by Abraham Lempel and Jacob Ziv in 1977.
MODIS - Moderate Resolution Imaging Spectroradiometer
MOPITT - Measurements Of Pollution In The Troposphere
NaN - Not-a-Number
NASA - National Aeronautics and Space Administration
NetCDF - Network Common Data Form
NetCDF-4 - Network Common Data Form Version 4
OGC - Open Geospatial Consortium
ORNL - Oak Ridge National Laboratory
PCRS - Projected Coordinate Reference System
POSIX - Portable Operating System Interface
RFC - Request For Comments
TIFF - Tagged Image File Format
UNIX - A date and time representation widely used in computing.
URL - Universal Resource Locator
UTC - Coordinated Universal Time
VRT - Virtual Raster Format
WGS - World Geodetic System
WGS 84 - World Geodetic System 1984
WKT - Well-Known Text