

Dataset Interoperability Recommendations for Earth Science: Part 2

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 1.2	June 2020	Add page numbers. Adjust formatting/indentation of 2 nd example in section 3.6 to improve clarity. Check all URLs, replace broken URL in [3] in section 1.3.1.
Version 1.1	October 2019	Add footnote to section 4 indicating change of status of CF Conventions. Fix minor grammatical and typographical errors.
Version 1.0	April 2019	Original document.

Copyright Notice

Copyright © 2019 United States Government as represented by the Administrator of the National Aeronautics and Space Administration. No copyright is claimed in the United States under Title 17, U.S. Code. All Other Rights Reserved.

Abstract

This document contains eleven recommendations officially adopted by the NASA Earth Science Data System (ESDS) Dataset Interoperability Working Group (DIWG) during its 2015/16, 2016/17 and 2017/18 terms. They are the continuation of the DIWG recommendations published as ESDS-RFC-028 [1] 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.

Keywords: ESDIS, ESO, ESDSWG, DIWG, HDF5, netCDF, interoperability, recommendations, data, dataset, metadata

Table of Contents

STATUS OF THIS RFC	1
CHANGE EXPLANATION	1
COPYRIGHT NOTICE	1
ABSTRACT	1
TABLE OF CONTENTS	1
1 PREFACE	2
1.1 Purpose.....	2
1.2 Scope.....	2

1.3	Related Documentation.....	2
1.3.1	Reference Documents (Informative).....	2
2	INTRODUCTION	4
3	RECOMMENDATIONS.....	5
3.1	Character set for user-defined group, dataset, and attribute names	5
3.2	Consistent units attribute value for variables across one data collection.....	6
3.3	Use the units attribute only for variables with physical units	6
3.4	Include time coordinate in swath structured data	6
3.5	Keep coordinate values in coordinate variables.....	7
3.6	Include Georeference Information with Geospatial Coordinates	7
3.7	Not-a-Number (NaN) Value	10
3.8	Standardize file extensions for HDF5/netCDF files	10
3.9	Ensure Granule's Filename Uniqueness Across Different Dataset Releases	11
3.10	Adopt Semantically Rich Dataset Release Identifiers	11
3.11	Date-Time Information in Granule Filenames.....	12
4	SUPPORT FOR PROPOSED CF CONVENTIONS EXTENSIONS.....	13
5	AUTHORS' ADDRESSES	14
	APPENDIX A - GLOSSARY.....	15

1 PREFACE

1.1 Purpose

This document contains eleven recommendations adopted by the NASA Earth Science Data System Dataset Interoperability Working Group (DIWG) during its 2015/16, 2016/17 and 2017/18 terms. These recommendations are published as a Suggested Practice by the ESDIS Standards Office (ESO).

1.2 Scope

This document provides information to the NASA Earth Science Division (ESD). This document contains interoperability recommendations for producers of NASA Earth science data products. Distribution of this document is unlimited.

1.3 Related Documentation

1.3.1 Reference Documents (Informative)

The following documents are not binding on the content but referenced herein and, amplify or clarify the information presented in this document.

- [1] C. S. Zender and P. J. Leonard, "Dataset Interoperability Recommendations for Earth Science," July 2016. [Online]. Available: <https://cdn.earthdata.nasa.gov/conduit/upload/9696/ESDS-RFC-028v1.3.pdf>. [Accessed 18 June 2020].
- [2] The HDF Group, "HDF5 User's Guide, Section 1.4.2, HDF5 Path Names and Navigation," [Online]. Available: <https://portal.hdfgroup.org/display/HDF5/HDF5+User's+Guide>. [Accessed 18 June 2020].
- [3] Unidata, "NetCDF User Guide: Characters in NetCDF Names," 15 March 2018. [Online]. Available: https://www.unidata.ucar.edu/software/netcdf/documentation/NUG/netcdf_data_set_components.html - Permitted. [Accessed 18 June 2020].
- [4] Climate and Forecast Metadata Convention, "Naming Conventions," 15 August 2017. [Online]. Available: http://cfconventions.org/cf-conventions/cf-conventions.html#_naming_conventions. [Accessed 18 June 2020].
- [5] Unidata, "UDUNITS 2.2.25 Manual," [Online]. Available: <https://www.unidata.ucar.edu/software/udunits/udunits-2.2.25/doc/udunits/udunits2.html>. [Accessed 18 June 2020].
- [6] Climate and Forecast Metadata Convention, "Units," 15 August 2017. [Online]. Available: <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#units>. [Accessed 18 June 2020].
- [7] Climate and Forecast Metadata Convention, "Horizontal Coordinate Reference Systems, Grid Mappings, and Projections," 15 August 2017. [Online]. Available: <http://cfconventions.org/cf-conventions/cf-conventions.html#grid-mappings-and-projections>. [Accessed 18 June 2020].
- [8] Climate and Forecast Metadata Convention, "Appendix F: Grid Mappings," 15 August 2017. [Online]. Available: <http://cfconventions.org/cf-conventions/cf-conventions.html#appendix-grid-mappings>. [Accessed 18 June 2020].
- [9] Open Geospatial Consortium, "Geographic information — Well-known text representation of coordinate reference systems," 2012. [Online]. Available: <http://docs.openeospatial.org/is/12-063r5/12-063r5.html>. [Accessed 18 June 2020].
- [10] IEEE, "754-2008 - IEEE Standard for Floating-Point Arithmetic," 2008. [Online]. Available: [doi://10.1109/IEEESTD.2008.4610935](https://doi.org/10.1109/IEEESTD.2008.4610935). [Accessed via <https://doi.org/10.1109/IEEESTD.2008.4610935> 18 June 2020].

- [11] International Organization for Standardization, "ISO 8601:2004. Data elements and interchange formats -- Information interchange -- Representation of dates and times," December 2004. [Online]. Available: <https://www.iso.org/standard/40874.html>. [Accessed 18 June 2020].
- [12] C. Zender, "CF2-Group: Draft Extension for Files with Groups," 16 April 2018. [Online]. Available: <https://github.com/diwg/cf2/blob/master/group/cf2-group.adoc>. [Accessed 18 June 2020].
- [13] A. Jelenak, D. Santek, K. Yang and R. Carp, "Encoding of Swath Data in the Climate and Forecast Convention," 6 March 2018. [Online]. Available: <https://github.com/Unidata/EC-netCDF-CF/blob/master/swath/swath.adoc>. [Accessed 18 June 2020].

2 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.

The recommendations from this document are developed to complement the recommendations published as ESDS-RFC-028 [1] listed below:

1. Maximize HDF5/netCDF4 interoperability via API accessibility
2. Include Basic CF Attributes
3. Use CF "bounds" attributes
4. Verify CF compliance
5. Distinguish clearly between HDF and netCDF packing conventions
6. When to employ packing attributes
7. Mapping between ACDD and ISO
8. Make HDF5 files netCDF4-Compatible and CF-compliant within Groups
9. Include time dimension in grid structured data
10. Order dimensions to facilitate readability of grid structure datasets
11. Consider "balanced" chunking for 3-D datasets in grid structures
12. Include datum attributes for data in grid structures

3 RECOMMENDATIONS

3.1 Character set for user-defined group, dataset, and attribute names

We recommend that user-defined group¹, variable², and attribute³ names follow the Climate and Forecast (CF) Conventions' specification. The names shall comply with this regular expression: "[A-Za-z][A-Za-z0-9_]".* Exempt are system-defined names for any of these objects that are required by various APIs or conventions.

Recommendation Details:

Character sets for naming objects in HDF5 or netCDF files differ. The most permissive is the HDF5 character set while most restrictive is the CF character set, hence, we recommend the CF rules to achieve the largest possible naming interoperability. Below are excerpts from the HDF5 and netCDF users guides, as well as the CF Conventions.

HDF5 User's Guide [2]:

A path name is a string of components separated by "/". Each component is the name of a link or the special character "." for the current group. Link names (components) can be any string of Unicode characters not containing "/" (except the string "." which is reserved). However, users are advised to avoid the use of punctuation and non-printing characters because they may create problems for other software.

NetCDF User Guide (NUG) [3]:

The names of dimensions, variables and attributes (and, in netCDF-4 files, groups, user-defined types, compound member names, and enumeration symbols) consist of arbitrary sequences of alphanumeric characters, underscore "_", period ".", plus "+", hyphen "-", or at sign "@", but beginning with an alphanumeric character or underscore. However, names commencing with underscore are reserved for system use.

Beginning with versions 3.6.3 and 4.0, names may also include UTF-8 encoded Unicode characters as well as other special characters, except for the character "/", which may not appear in a name. Names that have trailing space characters are also not permitted.

CF Metadata Conventions [4]:

Variable, dimension and attribute names should begin with a letter and be composed of letters, digits, and underscores. Note that this is in conformance with the COARDS conventions but is more restrictive than the netCDF interface which allows use of the hyphen character. The netCDF interface also allows leading underscores in names, but the NUG states that this is reserved for system use.

¹ HDF5 group or netCDF group

² HDF5 dataset or netCDF variable

³ HDF5 attribute or netCDF attribute

3.2 Consistent units attribute value for variables across one data collection

We recommend the units attribute's value for a particular variable if given should be the same across all files in a data collection.

Recommendation Details:

Knowing the physical units of data in a variable is vital for proper use. While the presence of the units attribute satisfies that requirement, its value may vary from one file to another for the same variable. For example, while these two values of the units attribute represent length: mm and μm , the software processing that variable's data from different files may not take the attribute's value change into account. Using the same units attribute value for one variable throughout a data collection decreases the chance of errors.

3.3 Use the units attribute only for variables with physical units

We recommend adhering to the CF Conventions' 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 [5].*
- *A variable used in any context other than data storage should never have the units attribute.*

Recommendation Details:

The units attribute is part of the CF Conventions [6] and is one of the recommended variable attributes in the NASA ESDIS Standards Office document, ESDS-RFC-028 [1]. The application of this attribute was ambiguous in the past, particularly in the case of unitless (dimensionless) data or in the choice of appropriate attribute values.

This recommendation standardizes on both of these issues. Values of the attribute should be supported by the UDUNITS-2 [5] library. For example: "km²" or "km2" are valid; "square km" is invalid. For dimensionless data, the use of the units attribute is allowed when there are physical units in common usage. For example: "milliradians", "g/kg", or "%". While all of these units represent a dimensionless physical quantity there are user communities that use them. The same applies to cases with fractional numbers as dimensionless units, such as: "1e-3" or "1e-6".

The use of units="1", units="" (empty string), or any similar construct as generic units for unitless data are strongly discouraged.

3.4 Include time coordinate in swath structured data

We recommend that swath dataset files always include a time coordinate even when there is only one value.

Recommendation Details:

A time coordinate is required for a swath file when it contains data from many time instances. Sometimes swath files with data from a single time instance could be without a time coordinate because each file records that specific time value in either the file name or file-level attributes, or both. A time coordinate should be defined and used in all data variables that vary in time, regardless of the number of time instances. This will allow downstream users to more easily and efficiently aggregate data across separate files.

3.5 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.

Recommendation Details:

Coordinate values are essential for putting all the other data in the proper physical domain context. Storing coordinate values in coordinate variables improves the consistency of data access, especially by software, which is directly related to data interoperability. Storing any part of coordinate values in attributes, variable names or group names is strongly discouraged. For example, avoid encoding time coordinate data in group hierarchies like 2017/01/30 (these are three groups named "2017", "01", and "30", respectively).

3.6 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 Conventions and the following specific guidelines:

- *Granules are required to contain the most applicable type of geospatial coordinates for the expected application of 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 CF 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.*

Following the above directions will support the widest range of software tools and at the same time control the file size by avoiding storage of redundant geospatial coordinate data.

Recommendation Details:

Note: This recommendation expands on the Recommendation 2.12 published in ESDS-RFC-028 [1].

With the increase in the variety of software tools analyzing or visualizing Earth Science datasets the need to accurately and unambiguously put their data into the proper geospatial context has become an essential aspect of data interoperability. Correctly encoding geospatial coordinates in granules is only the prerequisite as those coordinates must also be associated with appropriate coordinate reference systems (CRS). We assume the former and provide here instructions how to achieve the latter with the CF Conventions.

Georeference information is supplied via *grid mapping variables* [7]. Such variables do not hold any data and serve only to group attributes that describe a specific CRS. An attribute `grid_mapping_name` must always be present containing the string identifier of a supported CRS or map projection listed in Appendix F [8] of the CF Conventions. The particular CRS or map projection determines all other attributes that need to be attached to the grid mapping variable.

The example below illustrates the concept for the latitude and longitude CRS on the WGS-84 geoid.

```
dimensions:  
  lat = 180 ;  
  lon = 360 ;  
  
variables:  
  double lat(lat) ;  
    lat:standard_name = "latitude" ;  
    lat:units = "degrees_north" ;  
  
  double lon(lon) ;  
    lon:standard_name = "longitude" ;  
    lon:units = "degrees_east" ;  
  
  float eo_data(lat, lon) ;  
    eo_data:long_name = "Earth observation data" ;  
    eo_data:grid_mapping = "crs" ;  
  
  int crs ;  
    crs:grid_mapping_name = "latitude_longitude" ;  
    crs:longitude_of_prime_meridian = 0.0 ;  
    crs:semi_major_axis = 6378137.0 ;  
    crs:inverse_flattening = 298.257223563 ;
```

The `crs` variable is the grid mapping variable. The name and datatype of this variable are arbitrary and, for simplicity, it should always be a scalar (rank is zero). The `grid_mapping_name` attribute of the `crs` variable has the value of `latitude_longitude` which is the CF identifier for the generic latitude-longitude CRS on a geoid datum. The other attributes — `longitude_of_prime_meridian`, `semi_major_axis`, and `inverse_flattening` — serve to define the actual geoid parameters.

Starting from Version 1.7, the CF Conventions also permit georeferencing information as a string in the OGC WKT format [9]. The grid mapping variable `crs` from the last example is now:

```
int crs ;
  crs:grid_mapping_name = "latitude_longitude" ;
  crs:longitude_of_prime_meridian = 0.0 ;
  crs:semi_major_axis = 6378137.0 ;
  crs:inverse_flattening = 298.257223563 ;
  crs:crs_wkt = "GEODCRS[\"WGS 84\",
    DATUM[
      \"World Geodetic System 1984\",
      ELLIPSOID[
        \"WGS 84\",6378137,298.257223563,
        LENGTHUNIT[\"metre\",1.0]
      ]
    ],
    CS[ellipsoidal,2],
    AXIS[\"latitude\",north,ORDER[1]],
    AXIS[\"longitude\",east,ORDER[2]],
    ANGLEUNIT[\"degree\",0.01745329252],
    ID[\"EPSG\",4326]]" ;
```

The WKT description of the WGS-84 datum is stored in the new attribute: `crs_wkt`. Despite allowing the WKT format, the CF Conventions still requires that georeference information be primarily described via a number of attributes. Although the WKT format provides a more extensive georeference definition, an issue with including the WKT description is that there is no general translation between WKT parameters and CF grid mapping variable attributes.

The example below demonstrates georeferencing information for a map projection:

```
dimensions:
  y = 500 ;
  x = 500 ;
  time = 50 ;

variables:
  int crs ;
    crs:grid_mapping_name = "lambert_conformal_conic" ;
    crs:longitude_of_prime_meridian = 0.0 ;
    crs:semi_major_axis = 6378137.0 ;
    crs:inverse_flattening = 298.257223563 ;
    crs:standard_parallel = 25.0 ;
    crs:longitude_of_central_meridian = -100.0 ;
    crs:latitude_of_projection_origin = 25.0 ;
    crs:false_northing = 1500000.0 ;
    crs:false_easting = 5000000.0 ;

  double y(y);
    y:units = "km" ;
    y:standard_name = "projection_y_coordinate" ;

  double x(x);
    x:units = "km" ;
    x:standard_name = "projection_x_coordinate" ;

  int time(time) ;
```

```

time:units = "hours since 2001-06-23T22:00:00Z" ;

float eos_data(time, y, x) ;
eos_data:grid_mapping = "crs" ;
eos_data:long_name = "Earth observation data" ;

```

The grid mapping variable is `crs` and the map projection's definition is provided by its attributes: `standard_parallel`, `longitude_of_central_meridian`, and `latitude_of_projection_origin`. The `x` and `y` are the map projection coordinates, designated by their `standard_name` attribute values: `projection_x_coordinate` and `projection_y_coordinate`, respectively.

The above example is not strictly compliant with the CF Conventions because CF requires providing the latitude and longitude coordinates in addition to the projected coordinates `x` and `y`. Latitude and longitude coordinates are omitted from the example because they are not required to geospatially describe the data on the map projection and including them would make the file much larger.

3.7 Not-a-Number (NaN) Value

*We recommend Earth Science data products **avoid** using Not-a-Number (NaN) in any field values or as an indicator of missing or invalid data.*

Recommendation Details:

The Institute of Electrical and Electronics Engineers (IEEE) floating-point standard [10] defines the NaN (Not-a-Number) bit-patterns to represent results of illegal or undefined operations. Unless carefully written, any arithmetic operation involving NaN values can halt a program. Furthermore, any relational operator with at least one NaN value operand must evaluate to *False*. These properties make NaN values difficult to handle in numerical software and reduce the interoperability of datasets that contain NaN.

Some dataset producers adopt NaN as a missing or invalid datum (a.k.a. `_FillValue`) without appreciating the difficulty this poses to processing tools. It is more interoperable to use out-of-range values (e.g., `-9999`) to represent missing or invalid data (a.k.a. `_FillValue`) and to **avoid** NaN entirely.

3.8 Standardize file extensions for HDF5/netCDF files

We recommend using standardized file name extensions for HDF5 and netCDF files, as follows:

- *.h5 for files created with the HDF5 API;*
- *.nc for files created with the netCDF API; and*
- *.he5 for files created with the HDF-EOS5 API.*

Recommendation Details:

File name extensions are a traditional piece of metadata used to infer properties of the file's content or its planned use. HDF5 files can be created with several APIs (libraries) which provide different interface levels to their native data structures. Therefore, it is beneficial to know beforehand with which HDF5 API to open files. On the other hand, the netCDF API took a different approach and

offers the same interface for its supported storage file formats. Regardless of the recommended file name extensions, software developers are cautioned to not blindly trust the file extension and implement a heuristic process for determining whether the file meets software's expectations. Software tools should fail gracefully with a meaningful error message when the file type does not match the filename extension.

3.9 Ensure Granule's Filename Uniqueness Across Different Dataset Releases

We recommend that each granule belonging to an Earth Science dataset in a public archive have a unique file name across different dataset releases (versions, collections) to improve interoperability and avoid confusion. The minimum content to ensure unique granule file name consists of:

1. *a unique dataset identifier,*
2. *a unique identifier for each release (version, collection) of the dataset, and*
3. *the date-time, or any part thereof as applicable, of the first data observation in the file.*

These minimal elements should be easy for humans as well as machines to understand while maintaining the primary goal of file name uniqueness.

Recommendation Details:

There are non-DAAC archives that rely solely on a directory structure to uniquely identify different versions of Earth Science dataset granules – that is, the same granule but in a new dataset release (version) has the exact same name as its earlier version. The problem is that confusion can result once a granule file is removed from its directory structure. Inevitably, the day comes when some end user discovers that there are two identically named granule files that contain quite different data, and the inferior version of the granule has been used to carry out some sort of a scientific analysis. Ensuring that same granules from different dataset releases can be distinguished based on their file names would help eliminate this problem.

This recommendation does not:

1. prescribe a granule file naming schema, or
2. address the issue where one granule might be generated more than once and each of these files must be differentiated within the same dataset release (version).

The decisions about the above issues are left to data producers to make based on their particular processing, distribution, and user requirements. We assume here that each dataset release will eventually end up with only one copy for all of its granules and that these granules' filenames should then be unique across different dataset releases.

Of the three recommended fields in granule file names, the dataset identifier and the date-time of the first data observation in the file are constant for the same granule. What secures that granule's file name uniqueness is the dataset's release identifier. It is important the identifier's format can express sufficiently wide range of cases that could result in a new dataset release.

3.10 Adopt Semantically Rich Dataset Release Identifiers

We recommend that dataset release (version) identifiers can at least represent the following information:

- *Changes in the file content that only improve its compliance with the already published documentation ("patch" release).*
- *Changes in the file content that reflect a different scientific methodology ("major" release).*
- *Any changes in the file content not falling in the above categories. For example, changes that improve interoperability of the content with certain convention or software tool ("minor" release).*

Furthermore, for any dataset identifier scheme the ordering of its fields and their formatting should allow for lexical ASCII sort order of granule file names from the oldest to the most recent dataset release.

Recommendation Details:

The dataset release (version) identifier is usually the final field that can assure uniqueness of a granule file name if all other information for that granule remains the same. This recommendation neither prescribes nor endorses any existing scheme for dataset release identifiers. Further, it does not prohibit providing any additional (more specific) release information. What constitutes a major, minor, or patch dataset release is left up to the data producer to decide based on the recommendation's broad descriptions.

3.11 Date-Time Information in Granule Filenames

We recommend that date-time information in granule file names adheres to the following guidelines:

- *Adopt the ISO 8601 [11] standard for date-time information representation.*
- *If describing a date-time interval, the start date-time should appear before the end date-time.*
- *Date-time fields representing the temporal extent of a granule's data should appear before any other date-time field in the file name.*
- *All date-time fields should have the same format.*

The recommended date-time formats are:

- Full resolution

YYYYMMDDThhmmss[.]f+Z

YYYYMMDDThhmmssZ

- Reduced resolution

YYYYMMDDThhmmZ

YYYYMMDDThhZ

YYYYMMDD

YYYY-MM

YYYY

Where:

- Y is a digit used in the time element “year”
- M is a digit used in the time element “month”
- D is a digit used in the time element “day”
- T is the separator between the date and time parts
- h is a digit used in the time element “hour”
- m is a digit used in the time element “minute”
- s is a digit used in the time element “second”
- [, .] the decimal mark – either the comma or the full stop.
- f is a digit used in the decimal fraction of a second
- + means "one or more digits"
- Z is UTC time zone designator

Recommendation Details:

It is commonplace to find date-time information in Earth Science granule file names; however, its formatting varies. ISO 8601 [11] is an international standard for representing dates and times of the Gregorian calendar in many interoperable formats and with differing level of detail. The recommended formats imply:

1. Date-time information should always be in the UTC time zone.
2. There are no limits on the highest date-time resolution.
3. The delimiter between the seconds and the fraction of a second can only be the comma or the full stop. The choice should take into account which of these two characters is considered safe for the adopted granule file naming scheme.

4 SUPPORT FOR PROPOSED CF CONVENTIONS EXTENSIONS

The Dataset Interoperability Working Group officially endorsed two proposed extensions to the CF Conventions: on the use of HDF5 groups⁴ (hierarchies) [12] and the encoding of satellite swath data [13]. Both of these extensions are very important for NASA-produced Earth Science datasets. Their use is recommended by the Working Group although they have not yet been officially accepted into the CF Conventions.

⁴ The extension of the CF Conventions to include groups has been accepted (see <http://cfconventions.org/cf-conventions/cf-conventions.html#groups>).

5 AUTHORS' ADDRESSES

2016/17 and 2017/18 Dataset Interoperability Working Group

Contact:

Aleksandar Jelenak, Technical Chair
The HDF Group
410 E. University Avenue, Suite #200
Champaign, Illinois 61820

2015/16 Dataset Interoperability Working Group

Contact:

Peter J.T. Leonard, Technical Chair
ADNET Systems, Inc.
7515 Mission Drive, Suite A100
Lanham, MD 20706

Appendix A - Glossary

ASCII:	American Standard Code for Information Interchange
API:	Application Programming Interface
CCB:	Configuration Change Board
CCR:	Configuration Change Request
CF:	Climate and Forecast Metadata Conventions
COARDS:	Cooperative Ocean/Atmosphere Research Data Service
CRS:	Coordinate Reference System
DAAC:	Distributed Active Archive Center
DIWG:	Dataset Interoperability Working Group
EOSDIS:	Earth Observing System Data and Information System
ESDIS:	Earth Science Data and Information System
ESDSWG:	Earth Science Data System Working Groups
ESO:	ESDIS Standards Office
GSFC:	Goddard Space Flight Center
HDF5:	Hierarchical Data Format Version 5
HDF-EOS5:	Hierarchical Data Format-Earth Observing Systems Version 5
IEEE:	Institute of Electrical and Electronics Engineers
ISO:	International Organization for Standardization
NaN:	Not-a-Number
netCDF-4:	Network Common Data Form Version 4
NUG:	NetCDF User Guide
OGC:	Open Geospatial Consortium
UTC:	Universal Time Coordinated (also: Coordinated Universal Time)
UTF-8:	Unicode Transformation Format—8-bit
WGS-84:	World Geodetic System 84
WKT:	Well-Known Text