

Dataset Interoperability Recommendations for Earth Science

This page contains recommendations adopted by the NASA Earth Science Data System (ESDS) Dataset Interoperability Working Group (DIWG). They are meant to increase and enhance the interoperability of Earth Science data product files by improving their compliance, discoverability, and extensibility with relevant metadata conventions. These recommendations are extracts from the recommendations documents [published](#) by the ESDIS Standards Coordination Office (ESCO). They have been copied verbatim with only formatting changes. If there are any differences between the two versions, the ESCO-published documents should be considered as accurate.

This page will be updated as new recommendations are adopted.

Glossary

ACDD	Attribute Convention for Dataset Discovery
ASCII	American Standard Code for Information Interchange
attribute	HDF5 attribute or netCDF attribute
API	Application Program Interface
CCB	Configuration Change Board
CCR	Configuration Change Request
CDM	Unidata Common Data Model
CF	Climate and Forecast Metadata Conventions
CRS	Coordinate Reference System
COARDS	Cooperative Ocean/Atmospheric Research Data Service
DAAC	Distributed Active Archive Center
DIWG	Dataset Interoperability Working Group
EOSDIS	Earth Observing System Data and Information System
ESDS	Earth Science Data Systems
ESDIS	Earth Science Data and Information System
ESDSWG	Earth Science Data System Working Groups
ESIP	Federation of Earth Science Information Partners
ESCO	ESDIS Standards Coordination Office
group	HDF5 group or netCDF group
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
HDF4	Hierarchical Data Format, version 4
HDF5	Hierarchical Data Format, version 5
HDF-EOS	Hierarchical Data Format - Earth Observing System
HDF-EOS5	Hierarchical Data Format - Earth Observing System, version 5
HPD	HDF Product Designer
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
JPL	Jet Propulsion Laboratory
MCC	Metadata Compliance Checker
NASA	National Aeronautics and Space Administration
NaN	Not-a-Number

ncks	netCDF Kitchen Sink
NCO	netCDF Operator
ncpdq	netCDF Permute Dimensions Quickly
ncwa	netCDF Weighted Averager
netCDF	Network Common Data Form
netCDF3	Network Common Data Form, version 3
netCDF4	Network Common Data Form, version 4
NUG	NetCDF User Guide
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
RFC	Request For Comments
THREDDS	Thematic Real-time Environmental Distributed Data Services
UTC	Universal Time Coordinated (also: Coordinated Universal Time)
UTF-8	Unicode Transformation Format—8-bit
variable	HDF5 dataset or netCDF variable
WGS-84	World Geodetic System 84
WKT	Well-Known Text
XML	Extensible Markup Language

Recommendations Approved by ESCO

- [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).

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [compliance](#)
- [identifier](#)
- [filenames](#)

- [Character Set for User-Defined Group, Variable, and Attribute Names](#) —

We recommend that user-defined group, variable, and attribute names follow the Climate and Forecast (CF) convention's 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.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [netcdf](#)
- [cf_conventions](#)
- [rfc-036](#)
- [metadata](#)
- [compliance](#)
- [hdf](#)

- [Consider "balanced" chunking for 3-D datasets in grid structures](#) —

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

Recommendation Details: If a dataset is exceptionally large, it is often more useful to break it up into manageable parts. This process is known as *chunking* and is used on data in datasets that are part of a grid structure. Exactly how the data chunking is done can greatly affect performance for the end user. Because the precise access pattern employed by the end user is usually unknown until the distributor analyzes sufficient requests to discern a pattern, it is difficult to determine the most effective way to chunk.

- [esdswg](#)

- [recommendations](#)
- [diwg](#)
- [gridded_data](#)
- [chunking](#)
- [rfc-028](#)
- [dimensions](#)
- [compression](#)

• [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 m, 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.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [metadata](#)
- [units](#)

• [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.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [compliance](#)
- [time](#)
- [filenames](#)
- [iso8601](#)

• [Distinguish clearly between HDF and netCDF packing conventions](#) —

We recommend that datasets with non-netCDF packing be clearly distinguished from datasets that use the netCDF packing convention.

Recommendation Details: Earth Science observers and modelers often employ a technique called "packing" (a.k.a. "scaling") to make their product files smaller. "Packed" datasets must be correctly "unpacked" before they can be used properly. Confusingly, non-netCDF (e.g., [HDF4_CAL](#)) and netCDF algorithms both store their parameters in attributes with the same or similar names – and unpacking one algorithm with the other will result in incorrect conversions. Many netCDF-based tools are equally unaware of the non-netCDF (e.g., [HDF_CAL](#)) packing cases and so interpret all readable data using the netCDF convention. Unfortunately, few users are aware that their datasets may be packed, and fewer know the details of the packing algorithm employed. This is an interoperability issue because it hampers data analysis performed on heterogeneous systems.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [hdf](#)
- [netcdf](#)
- [packing](#)
- [rfc-028](#)

• [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:

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [compliance](#)
- [filenames](#)

• [Include Basic CF Attributes](#) —

We recommend that, at minimum, the following basic Climate and Forecast (CF) Convention attributes be included in future NASA Earth Science data products where applicable.

Recommendation Details : The Climate and Forecast (CF) [Conventions](#) are widely employed guidelines for Earth Science data and metadata storage. Included in the conventions is a comprehensive list of metadata attributes that are available for use by dataset producers. Because the list of metadata attributes is so extensive, dataset producers are constantly struggling with which metadata attributes to attach to a variable.

- [esdswg](#)

- [diwg](#)
- [recommendations](#)
- [cf_conventions](#)
- [rfc-028](#)
- [metadata](#)
- [compliance](#)
- [time](#)
- [coordinates](#)
- [Include Datum Attributes for Data in Grid Structures](#) —

We recommend that Horizontal and Vertical (as necessary) Datum attributes be included for data in grid structures.

Recommendation Details: Locations on Earth are specified using coordinates, which are tuples of numbers that describe the horizontal and vertical distances from a fixed point, thereby pinpointing a particular place on the map at some level of precision. But knowing the coordinates is very different from being able to interpret them.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [gridded_data](#)
- [rfc-028](#)
- [datum](#)
- [coordinates](#)
- [Include Georeference Information with Geospatial Coordinates](#) —



This recommendation expands on the [Include Datum Attributes for Data in Grid Structures](#).

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:

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [swath_data](#)
- [datum](#)
- [gridded_data](#)
- [time](#)
- [coordinates](#)
- [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.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [swath_data](#)
- [time](#)
- [rfc-036](#)
- [coordinates](#)
- [Include Time Dimension in Grid Structured Data](#) —

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

Recommendation Details: A Time dimension is required for a single grid product file that contains many time intervals of daily, weekly, or monthly averages. In contrast, grid product files that are distributed as daily, weekly, or monthly granules that have one time range or stamp for the entire grid could be defined without a Time dimension because each file records the specific time interval being provided in both the file name and file-level attributes. We nevertheless recommend that a Time dimension be defined and used in all data fields that vary in time, regardless of whether multiple time slices are stored in the file. More specifically, we recommend that Time be defined as a record dimension, not a fixed-length dimension. This allows downstream users to more easily and efficiently aggregate data across separate files because HDF5 and netCDF4 storage geometries and APIs are designed to more easily extend record dimensions than fixed-length dimensions.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [time](#)
- [rfc-028](#)
- [gridded_data](#)
- [dimensions](#)
- [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).

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [swath_data](#)
- [time](#)
- [gridded_data](#)
- [coordinates](#)

- [Make HDF5 files netCDF4-Compatible and CF-compliant within Groups](#) —

We recommend that all HDF5 Earth Science product files be made netCDF4-compatible and CF-compliant within groups.

Recommendation Details:

- [esdswg](#)
- [diwg](#)
- [netcdf](#)
- [cf_conventions](#)
- [hdf](#)
- [recommendations](#)
- [rfc-028](#)
- [compliance](#)
- [dimensions](#)
- [coordinates](#)

- [Mapping between ACDD and ISO](#) —

We recommend use of existing mapping between ACDD and ISO developed by ESIP.

Recommendation Details: The ESIP Community supports a vast array of systems that are accessed and utilized by a diverse group of users. Historically, groups within the community have approached metadata differently in order to effectively describe their data. As a result, similar dialects have emerged to address specific user requirements. The multi-dialect approach described above hinders interoperability— as it results in different terminology being used to describe the same concepts. By clearly depicting fundamental documentation needs and concepts and mapping to them in the different dialects, confusion is minimized and interoperability is facilitated. Thus, demonstrating connections between dialects increases discoverability, accessibility, and reusability of data via consistent, compatible metadata.

- [esdswg](#)
- [diwg](#)
- [acdd](#)
- [iso19115](#)
- [esip](#)
- [recommendations](#)
- [rfc-028](#)
- [compliance](#)
- [metadata](#)

- [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.

Recommendation Details: NASA data products based on Earth Science observations are typically in HDF, trending to [HDF5](#), while Earth Science modelers generally prefer to produce data in netCDF, trending to [netCDF4](#). It is not possible to make HDF4 files look exactly like netCDF3 files. On the other hand, netCDF4 is built on HDF5 (netCDF4 is essentially a subset of HDF5), and so it is possible to construct HDF5 files that are accessible from the netCDF4 API, which is a tremendous opportunity for interoperability. While using the netCDF4 API ensures this, the recommendation also provides guidance for those using the HDF5 API to ensure netCDF4 interoperability.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-028](#)
- [dimensions](#)
- [compliance](#)

- [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](#) 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.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [hdf](#)
- [netcdf](#)

- [Order Dimensions to Facilitate Readability of Grid Structure Datasets](#) —

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

Recommendation Details: The order of the dimensions in a group structure should be carefully considered, since this can have significant impact on the ease with which an end user can read the data in the group structure. While tools such as NCO's [ncpdq](#) can re-order dataset dimensions, thereby permitting dataset designers to test the effects of their ordering choices against common access patterns, re-ordering large datasets can be time consuming and is best avoided.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [gridded_data](#)
- [dimensions](#)
- [rfc-028](#)

- [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

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [rfc-036](#)
- [hdf](#)
- [netcdf](#)
- [filenames](#)

- [Use CF Bounds Attributes](#) —

We recommend that spatio-temporal and other coordinate boundaries be specified by adding CF "bounds" attributes.

Recommendation Details: The CF conventions are widely employed guidelines for Earth Science data and metadata storage. The purpose of the CF conventions is to require conforming datasets to contain sufficient metadata that they are self-describing in the following ways: Each variable in the file has an associated description of what it represents, including physical units if appropriate; and each value can be located in space (relative to Earth-based coordinates) and time. Thus, adhering to CF guidelines will increase completeness, consistency, and interoperability of conforming datasets.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [cf_conventions](#)
- [rfc-028](#)
- [metadata](#)
- [compliance](#)
- [time](#)
- [coordinates](#)

- [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:

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [cf_conventions](#)
- [rfc-036](#)
- [metadata](#)
- [units](#)

- [Verify CF compliance](#) —

We recommend that CF compliance of NASA-distributed HDF/netCDF files be verified.

Recommendation Details: The Climate and Forecast (CF) Conventions are widely employed guidelines for Earth Science data and metadata storage, which help tool developers find and/or interpret data values, coordinates, units, and measurements in data files. Thus, it is increasingly important to adhere to CF conventions in order to benefit from analysis tools, web services, and middleware that exploit them.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [cf_conventions](#)
- [rfc-028](#)
- [compliance](#)
- [metadata](#)

- [When to Employ Packing Attributes](#) —

We recommend that packing attributes (i.e., `scale_factor` and `add_offset`) be employed only when data are packed as integers.

Recommendation Details: Packing refers to a lossy means of data compression that typically works by converting floating point data to an integer representation that requires fewer bytes for storage. The packing attributes `scale_factor` and `add_offset` are the netCDF (and CF) standard names for the parameters of the packing and unpacking algorithms. If `scale_factor` is 1.0 and `add_offset` is 0.0, the packed value and the unpacked value are identical, although their datatype (float or integer) may differ. Unfortunately, many datasets annotate floating point variables with the attributes, apparently for completeness, even though the variables have not been packed and remain as floating point values. Incorporating packing attributes on data that have not been packed is a misuse of the packing standard and it should be avoided. Data analysis software that encounters packing attributes on data that are not packed is liable to be confused and perform in unexpected ways. Packed data must be represented as integers, and only integer types should have packing attributes.

- [esdswg](#)
- [diwg](#)
- [recommendations](#)
- [packing](#)
- [netcdf](#)
- [rfc-028](#)
- [hdf](#)

Additional Recommendations Awaiting Formal ESCO Approval

- [Attach the CF `flag\_values` or `flag\_masks` Attributes Along With the CF `flag\_meanings` Attribute to Each Flag Variable](#) —

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:

- [esdswg](#)
- [diwg](#)
- [cf_conventions](#)
- [metadata](#)
- [awaiting-esco-approval](#)

- [Avoid Use of the `missing\_value` Attribute](#) —

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.

- [esdswg](#)
- [diwg](#)
- [awaiting-esco-approval](#)

- [Define the Projection Ellipsoid to Match the Reference Datum](#) —

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 WGS84. 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.

- [esdswg](#)
- [diwg](#)
- [awaiting-esco-approval](#)

- [Document Missing Granules for Instruments That Acquire Data on a Regular Basis](#) —

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).

- [esdswg](#)
- [diwg](#)
- [missing_data](#)
- [awaiting-esco-approval](#)

- [Include Only One Variable per GeoTIFF File](#) —

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).

- [esdswg](#)
- [diwg](#)
- [awaiting-esco-approval](#)

- [Indicate in CRS Metadata the Order of Elements in Horizontal Coordinate Pairs](#) —

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:

- [esdswg](#)
- [diwg](#)
- [awaiting-esco-approval](#)

- [Make a Variable's Valid Data Range Consistent Within Each Product Release](#) —

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.

- [esdswg](#)
- [diwg](#)
- [awaiting-esco-approval](#)

- [Make a Variable's Valid Data Range Useful](#) —

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.

- [esdswg](#)
- [diwg](#)
- [metadata](#)
- [cf_conventions](#)
- [awaiting-esco-approval](#)

- [Use a Number Outside of the Valid Data Range for a Variable's Fill Value](#) —

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.

- [esdswg](#)
- [diwg](#)
- [metadata](#)
- [cf_conventions](#)
- [missing_data](#)
- [awaiting-esco-approval](#)

- [Use DOIs for Referencing Documentation](#) —

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.

- [esdswg](#)
- [diwg](#)
- [metadata](#)
- [cf_conventions](#)
- [doi](#)
- [awaiting-esco-approval](#)

- [Use Double Precision When Archiving Time in Seconds Since a Specific Epoch](#) —

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.

- [esdswg](#)
- [diwg](#)
- [awaiting-esco-approval](#)

- [Use Only Officially Supported Compression Filters on NetCDF-4 and NetCDF-4-Compatible HDF5 Data](#) —

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: NetCDF4 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) netCDF4 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 netCDF4 default installation. Combining the shuffle and the DEFLATE filters can noticeably improve the data compression ratio.

- [esdswg](#)
- [diwg](#)
- [netcdf](#)
- [hdf](#)
- [compression](#)
- [awaiting-esco-approval](#)