

Dataset Interoperability Recommendations by Category

Recommendations

Compliance and Metadata Recommendations

- [Adopt Semantically Rich Dataset Release Identifiers](#)
- [Attach the CF flag_values or flag_masks Attributes Along With the CF flag_meanings Attribute to Each Flag Variable](#)
- [Character Set for User-Defined Group, Variable, and Attribute Names](#)
- [Consistent Units Attribute Value for Variables Across One Data Collection](#)
- [Date-Time Information in Granule Filenames](#)
- [Ensure Granule's Filename Uniqueness Across Different Dataset Releases](#)
- [Include Basic CF Attributes](#)
- [Make a Variable's Valid Data Range Useful](#)
- [Make HDF5 files netCDF4-Compatible and CF-compliant within Groups](#)
- [Mapping between ACDD and ISO](#)
- [Maximize HDF5/netCDF4 interoperability via API accessibility](#)
- [Use a Number Outside of the Valid Data Range for a Variable's Fill Value](#)
- [Use CF Bounds Attributes](#)
- [Use DOIs for Referencing Documentation](#)
- [Use the Units Attribute Only for Variables with Physical Units](#)

Grid Data Recommendations

- [Order Dimensions to Facilitate Readability of Grid Structure Datasets](#)
- [Include Datum Attributes for Data in Grid Structures](#)
- [Include Time Dimension in Grid Structured Data](#)
- [Include Georeference Information with Geospatial Coordinates](#)
- [Keep Coordinate Values in Coordinate Variables](#)
- [Consider "balanced" chunking for 3-D datasets in grid structures](#)

Swath Data Recommendations

- [Include Georeference Information with Geospatial Coordinates](#)
- [Include Time Coordinate in Swath Structured Data](#)
- [Keep Coordinate Values in Coordinate Variables](#)

NetCDF Recommendations

- [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.

- [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.

- [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

- [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_C](#) [AL](#)) 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.

- [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.

- [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:

- [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.