



# NODC NetCDF Templates

Upendra Dadi



# Outline

- CF 1.6 and Discrete Sampling Geometries.
- NODC NetCDF Templates – What are they?
- Choosing the right template.
- NODC guidance on CF & ACDD attributes.
- NODC defined attributes.
- Compliance Verification and metadata quality.
- Process followed in developing the templates.
- Questions?



# Climate Forecast(CF) Conventions

CF conventions:

- Metadata convention independent of data format.
- Metadata is part of the file itself unlike FGDC or ISO metadata encodings.
- More of a “use” metadata than discovery metadata.



# Climate Forecast(CF) Conventions...contd.

Prior to CF 1.6:

- CF is largely a convention for gridded data.
- Didn't allow multi-feature types for in-situ data.
- Dealt primarily with attributes and not as much on how data is packed in the variables.



# CF 1.6 and Discrete Sampling Geometries

- CF 1.6 introduces the global attribute `featureType` which could have these possible values: **profile**, **timeSeries**, **trajectory**, **timeSeriesProfile** and **trajectoryProfile**.
- All geophysical variables in the file are assumed to be present at each coordinate (x,y,z,t).
- Uses multi-dimensional coordinate variables of CF. The coordinates for a point with indices (i,j,k) would be of the form  $\{(cvar1(i,j,k), cvar2(i,j,k), cvar3(i,j,k))\}^*$ .

\*see <http://www.unidata.ucar.edu/software/netcdf-java/CDM/index.html#CoordinateSystems>



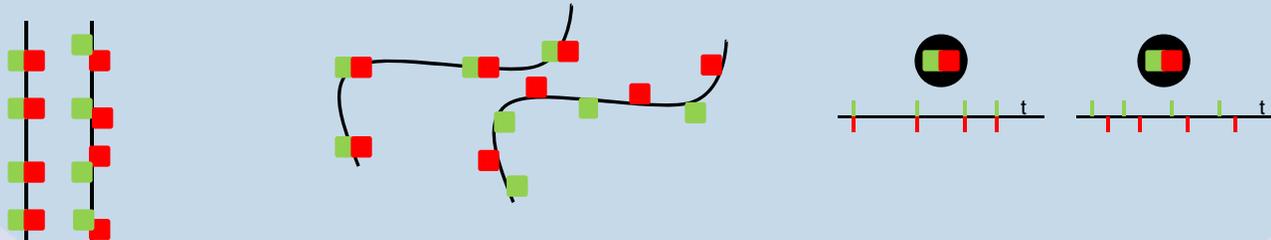
# NODC NetCDF Templates – What are they?

- More specific guidance for the data providers to submit data in standardized format.
- CF discrete sampling geometries is more complex than other parts of CF. Templates make it easier for the user.
- Guidance on text for several attributes – license, publisher, acknowledgement, email, etc.



# NODC NetCDF Templates – What are they?...contd.

- Templates incorporate new attributes, not present in CF – instrument, platform, call\_sign, etc.
- Several examples using real data.
- Guidance on NetCDF issues.
- Often data does not fit nicely with one of the featureTypes, what to do then?





# Selecting the right CF FeatureType Representation



The dataset consists of a group of profiles, time-series or trajectories which need to be combined into a single dataset using CF-convention for discrete sampling geometries.

Observational Data

Do you want to use single feature type representation or multi-feature type representation?

Single

Multi

- Data values from:
- Feature Instance 1
  - Feature Instance 2
  - Feature Instance 3
  - Feature Instance 4

Yes

A 9.2.5  
A 9.3.5  
A 9.4.4

Indexed Ragged Array Representation

No

A 9.2.4  
A 9.3.4  
A 9.4.3  
A 9.5.3  
A 9.6.3

Contiguous Ragged Array Representation

A feature instance is a single profile, time-series or trajectory within a group forming the dataset.

Does each feature instance have identical number of coordinates along the element axes of the features?

Is the number of coordinates along the element axes of the individual feature instances highly variable or is there a constraint on the availability of storage space?

The element axis of the features is the axis along the depth(or height) in case of profiles, time axis in case of time-series and the axis which follows the path of the trajectory in case of trajectories. In case of trajectoryProfile and timeSeriesProfile, there are two element axes.

No

Does each feature instance have identical coordinates along the element axes of the features?

No

A 9.2.1  
A 9.3.1  
A 9.5.4  
A 9.2.2  
A 9.3.2  
A 9.4.1  
A 9.5.2  
A 9.6.1  
A 9.6.2

Multidimensional Array Representation

Incomplete

Complete

Orthogonal

B is a special case of A



Datasets with multiple axes (e.g. trajectoryProfile) could be using orthogonal representation along one axis but more general multidimensional array representation along another.

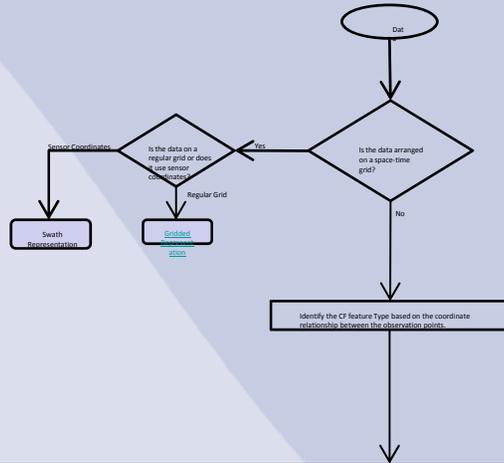
Orthogonal Multidimensional array representation is a special case of complete representation. In the first case, the number of coordinates along the element axes are same, but in the later case, the number of coordinate values is same, but not necessarily the values.

A 9.2.3  
A 9.3.3  
A 9.4.2

Single Array Representation

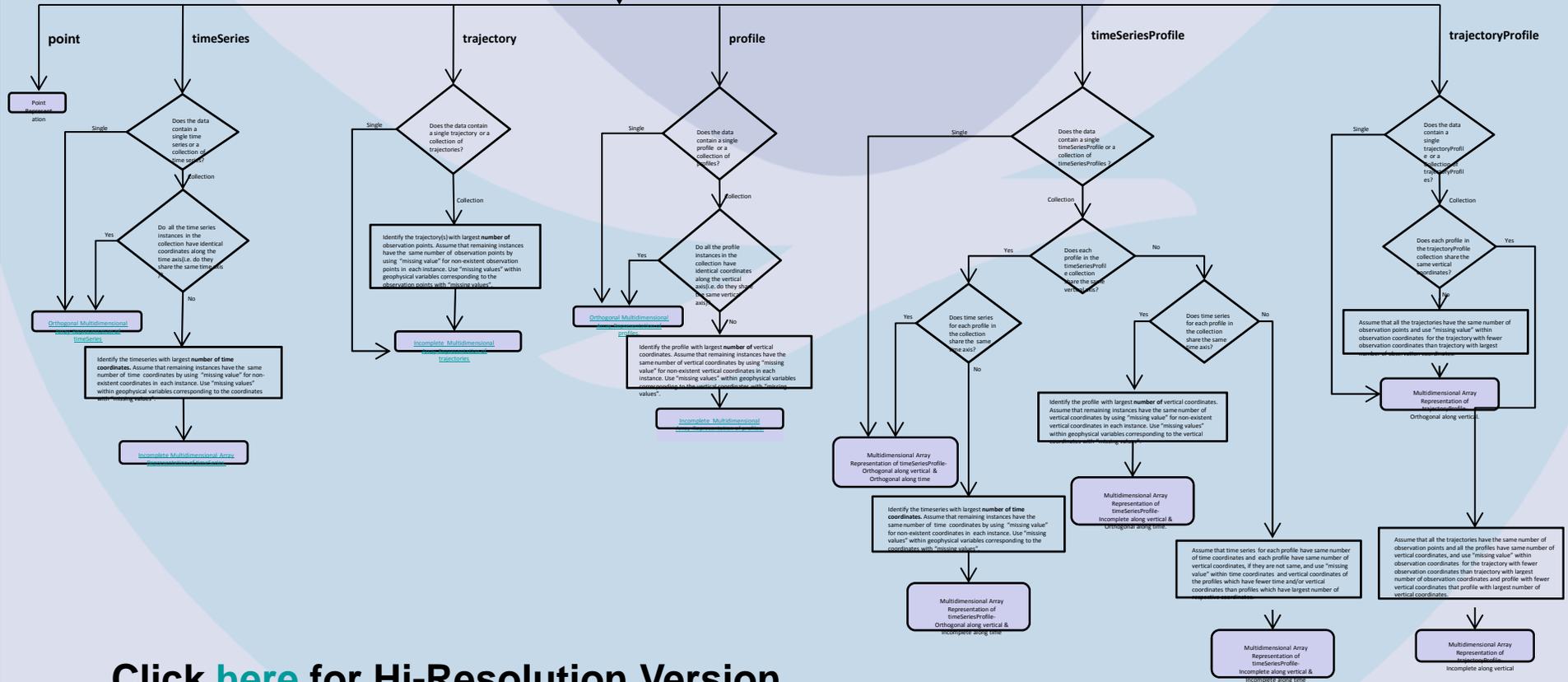


# Select CF feature type – NODC recommendation



**CF definitions of feature Types:**

- Point** - a single data point (having no implied coordinate relationship to other points)
- timeSeries** - a Series of data points at the same spatial location with monotonically increasing times
- trajectory** - a series of data points along a path through space with monotonically increasing times
- profile** - an ordered set of data points along a vertical line at a fixed horizontal position and fixed time
- timeSeriesProfile** - a series of profile features at the same horizontal position with monotonically increasing times
- trajectoryProfile** - a series of profile features located at points ordered along a trajectory.



Click [here](#) for Hi-Resolution Version



# Guidance table on CF & ACDD attributes



naming_authority	ACDD	global	string	The combination of the "naming authority" and the "id" should be a globally unique identifier for the dataset.	Recommend using the reverse URL of the institution. (ex., gov.noaa.nodc)
processing_level	ACDD	global	string	A textual description of the processing (or quality control) level of the data.	This attribute should also identify the authority which defined the processing level along with the processing level itself. The processing levels defined by NOAA for satellite data can be found here: <a href="https://www.ngdc.noaa.gov/wiki/index.php?title=NOAA_Processing_Levels">https://www.ngdc.noaa.gov/wiki/index.php?title=NOAA_Processing_Levels</a> . Similar list for NASA data can be found here: <a href="http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/">http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/</a> .
project	ACDD	global	string	The scientific project that the data was collected under.	Use the NODC projects ( <a href="http://www.nodc.noaa.gov/cgi-bin/OAS/prd/project">http://www.nodc.noaa.gov/cgi-bin/OAS/prd/project</a> ) name when applicable.

## NODC controlled vocabulary used wherever possible

creator_email	ACDD	global	string	Email address of the person or institution that collected the data.	The email of the person or institution may be found in the NODC tables for persons( <a href="http://www.nodc.noaa.gov/cgi-bin/OAS/prd/person">http://www.nodc.noaa.gov/cgi-bin/OAS/prd/person</a> ) and institutions( <a href="http://www.nodc.noaa.gov/cgi-bin/OAS/prd/institution">http://www.nodc.noaa.gov/cgi-bin/OAS/prd/institution</a> ). Use the short name of the institution if available.
creator_name				Name of the person who collected the data.	Use the name from the NODC persons( <a href="http://www.nodc.noaa.gov/cgi-bin/OAS/prd/person">http://www.nodc.noaa.gov/cgi-bin/OAS/prd/person</a> ) table when applicable.
creator_url				The URL of the institution that collected the data.	The url of the institution can be found in the NODC institutions( <a href="http://www.nodc.noaa.gov/cgi-bin/OAS/prd/institution">http://www.nodc.noaa.gov/cgi-bin/OAS/prd/institution</a> ) table.



# Guidance table on CF & ACDD attributes

Attribute	Convention	Global / Variable	Format	Description	NODC Recommended Use and Comments
<a href="#">acknowledgment</a>	ACDD	global	string	This optional free text field identifies the funding agency and other relevant information.	Please identify the funding agency and proposal number.
<a href="#">cdm_data_type</a>	ACDD	global	string	This attribute is used by THREDDS to identify the feature type, what THREDDS calls a "dataType". The current choices are: Grid, Image, Station, Swath, and Trajectory.	These data types do not map equally to the CF feature types. If the CF feature type = Trajectory Time Series, use "Trajectory"; if Point, Profile, or Time Series Profile, use "Station".
<a href="#">comment</a>	ACDD	global	string	Miscellaneous information about the data, that cannot be described in any of the other available attributes.	Note the difference between comment and long_name for a variable: While long_name is intended for a phrase or a one line description of what the data represents, comment could be much longer and is intended to provide in greater detail any miscellaneous information important to interpret the data.
<a href="#">contributor_name</a>	ACDD	global	string	The name of any individuals or institutions that contributed to the creation of this data. Please use commas to separate the names. When applicable, use the conventions described above when identifying persons and/or institutions.	Note that there is also an attribute called creator_name. This can have multiple comma separated values.



# NODC defined attributes

uuid	NODC	global	char	<p>Machine readable unique identifier for each netCDF file. A new uuid should be created whenever the netCDF file is changed. Review the history attribute listed below for further information about tracking older versions of the data file. From wikipedia: A UUID is a 16-byte (128-bit) number. The number of theoretically possible UUIDs is therefore about <math>3 \times 10^{38}</math>. In its canonical form, a UUID consists of 32 hexadecimal digits, displayed in 5 groups separated by hyphens, in the form 8-4-4-4-12 for a total of 36 characters (32 digits and 4 hyphens). (ex., 550e8400-e29b-41d4-a716-446655440000)</p> <p><a href="http://en.wikipedia.org/wiki/Universally_unique_identifier">http://en.wikipedia.org/wiki/Universally_unique_identifier</a></p>	Numerous tools exist to rapidly and easily create UUIDs. See the wiki page for a list.
platform	NODC	global / variable	char	<p>This attribute can be used with a geophysical variable to identify the platform that was used in the collection of the data. The value of the attribute should be set to another variable which contains the details of the platform. There can be multiple platforms involved depending on if all the instances of the featureType in the collection share the same platform or not. If multiple platforms are involved, a variable should be defined for each platform and referenced from the geophysical variable in a comma separated string. NODC recommends the use of the following attributes for those platform variables: call_sign, ices_code,</p>	<p>Information on ICES codes is available at <a href="http://www.ices.dk/datacentre/requests/Login.aspx">http://www.ices.dk/datacentre/requests/Login.aspx</a></p> <p>Information on WMO codes is available at <a href="http://www.wmo.int/pages/prog/amp/mmop/wmo-number-rules.html">http://www.wmo.int/pages/prog/amp/mmop/wmo-number-rules.html</a></p> <p>We recommend you avoid using ambiguous phrases like "NA" or "N/A" which might mean different things (e.g., Not Applicable, or Not Available).</p>
instrument	NODC	global / variable	char	<p>This attribute can be used with a geophysical variable to identify the instrument that collected the data. The value of the attribute should be set to another variable which contains the details of the instrument. There can be multiple instruments involved depending on if all the instances of the featureType in the collection come from the same instrument or not. If multiple instruments are involved, a variable should be defined for each instrument and referenced from the geophysical variable in a comma separated string. NODC recommends the use of the following attributes for those instrument variables: make_model, serial_number, calibration_date, factory_calibrated, user_calibrated, calibration_report, accuracy, valid_range, and precision.</p>	Please follow the guidance in the Description
sea_name	NODC	global	char	The names of the sea in which the data were collected.	See the list of NODC sea names here: NODC Sea Names ( <a href="http://www.nodc.noaa.gov/General/NODC-Archive/seanamelist.txt">http://www.nodc.noaa.gov/General/NODC-Archive/seanamelist.txt</a> )
nodc_name	NODC	variable	char	The NODC controlled vocabulary name for the variable if different from the standard name attribute or the standard name attribute does not exist.	See the list of available NODC data type names ( <a href="http://www.nodc.noaa.gov/cgi-bin/OAS/prd/datatype">http://www.nodc.noaa.gov/cgi-bin/OAS/prd/datatype</a> ).
nodc_template_version	NODC	global	char	This attribute tracks the version of the feature type created by NODC. (for example: nodc_template_version = "netCDF_single_trajectory_v1.0")	Please follow the guidance in the Description



# NODC defined attributes...contd.

An example usage of “instrument” attribute:

```
float temp(time);
    temp:instrument = "tmsr" ;
    ...other required and optional attributes
float sal(time);
    sal:instrument = "tsg" ;
    ...other required and optional attributes
int tmsr;
    tmsr:long_name = "hull mounted thermistor" ;
    tmsr:make_model = "SBE 48" ;
    tmsr:serial_number = "4893-C23-4B39D" ;
    tmsr:calibration_date = "20110325" ;
    ...other required and optional attributes
int tsg;
    tsg:long_name = "thermosalinograph" ;
    tsg:make_model = "SBE 45" ;
    tsg:serial_number = "G5K2-D84-7N3G5" ;
    tsg:calibration_date = "20110318" ;
    ...other required and optional attributes
```



# Compliance verification and metadata quality

The files created using the templates were checked using several tools:

- Online CF compliance checkers – 1.6 is not yet available.
- NetCDF-Java ToolUI was used to see if the files could be read properly by NetCDF java.
- NetCDF ACDD report – the rubric.
- Manual checking.



# Process followed in developing the templates

## 1) Internal discussion

- Weekly meetings
- Used Wiki as the platform for collaboration
- Used real data as examples.

## 2) Consultation with experts

- Got feedback from experts in CF including those developing CF based applications.
- The feedback was carefully studied and changes were made based on the feedback.



## Process followed in developing the templates...contd.

### 3) Release of draft for public comment.

- The template were released as version 0.9.
- The release was announced on several mailinglists and Ocean Sciences Meeting.
- A wiki page to record users comments was setup.
- The feedback has been very positive.

### 4) Incorporate comments and publish v1.0.

- The templates and guidance tables was improved based on the comments.
- The netCDF examples were improved.



# More Information

Primary web-site:

<http://www.nodc.noaa.gov/data/formats/netcdf/>

GeoIDE wiki:

[https://geo-ide.noaa.gov/wiki/index.php?title=NODC\\_NetCDF\\_Templates](https://geo-ide.noaa.gov/wiki/index.php?title=NODC_NetCDF_Templates)

User Comments:

<https://docs.google.com/a/noaa.gov/spreadsheet/viewform?formkey=dDdWdEdBb2xOdVV3OXJjcmFFYjRBMXc6MQ#gid=0>



# Questions?