CF Unleashed

OGC CF-netCDF Status and Plan – George Persive (sp?) and Ben Domenico

* 3 documents have been adopted by OGC standards
	+ NetCDF Core Encoding Stand
	+ NetCDF Enhanced Ata Model Extension
	+ CF-netCDF3 Data Model
* Possible future
	+ CF-netCDF encoding for WCS (need encoding… may be netCDF encoding)
	+ NcML and ncML-GML encoding specs
	+ Uncertainty – based on uncert in L – mark-up language – set of encoding using statistics
* Coordination among OGC - OPeNDAP, HDF, ESA
* CF-netCDF OGC standards – have core and extensions – more modular approach to standards and thus software of standards
	+ Not sure about formalizing of application profiles
* George – role is as chief engineer – how OGS meets need of members
* 92 people in OGS working group for netCDF
* Slide is from Jan 2013
* Ted – important – netCDF and CF use the term community standard –used to divide standards and communities – community > OGS > ISO
	+ OGS is an active standard organization
	+ ISO is more model based
* Q – are you working with any other develops
	+ Uncert web – Aston (???) (<http://www.uncertweb.org/>)
	+ Able to run monte-carlo processes
* Q – is there a relationship between the ISO14064 (climate change) and the CF
	+ (does not know that ISO)
	+ Don’t know - ? policy sections

CF Unleashed on Satellite Data – Aleksandar Jelenak (NOAA/NESDID-UCAR)

* Forecasting people have grids and have beautiful data – jealous of this for satellite
* Want to improve netCDF-CF files
* Satellite data – level 1 or 2 data in sensor geometry projects (not gridded - lower level)
* Use cases (provided links in PowerPoint)
* Case #1 – multiband imagery
	+ Multiband 2 D observation – lots of cases (probably the most common type of satellite)
	+ CDL Example
	+ Dimensions – 1) along\_track, 2) across\_track, 3)band
	+ Comment – Peter Conillian – instant – along\_scan and across\_scan
		- Not problem either way
	+ Band can be a wavelength or other information
	+ Variables – “coordinate variables”
		- X&Y because they are mutually orthogonal axis
		- Ex. Float lat (along\_track, across\_track) –
		- Lat and long different for each pixel
		- Time is dependent on scan or pixel
		- Swath\_data and swath\_band\_data (sensor observation)
	+ This would save 80% of cases – would work with current CF conversion
* Q – don’t understand CF – but if you like – then propose and then adopt
	+ Not that simple because proposing “feature type”
	+ Discrete sampling typologies – took some time to come to agreement
	+ Want support from this community (and others)
	+ Historically CF was not focused on satellite data (only modeling)
* Q (Ed) – GRIS has implement an adapted CF (same idea)
* Case #2 – Hyperspectral Imagery
	+ Has few thousand band – not able to use single field of view
	+ Each sensor of the group are “field of regard” – similar to field of view
	+ Graph from EUMETSAT – now becoming more mainstream (6 or 7 years)
	+ For each yellow ellipse have different lat/long – gets more complicated
	+ 3 approaches
		- Use Case#1 for each field of regard – so 4
		- Incorporate fields of regard into across\_track (problem then can have missing values)
		- Intro new term then #2
* Case #3 – Hyperspectral Sounder EDR
	+ NOAA unique product – sent to National Weather Service
	+ Example of problem where don’t have best practice – need to avoid
	+ Data from hyperspectral data that has been processed into geophysical parameters (2D) (ex. surface pressure) or 3D (atmospheric profile)
	+ Lots of specific info – but have not followed CF convention
	+ Have directive from GOSARD for NetCDF4 and CF compliant
* Need use cases to develop a pattern
* Q – do you think CF is sufficiently rich to define complex data (15 products with 2000 parameters)
	+ He thinks it is good
	+ Problems seen with CF – have multirate data (1 Hz or 50 Hz)… CF not handle well
	+ For each point need lat/long value
	+ Have data group… rate groups – when try to identify specific time, x, or y with CF – CF does not like
	+ CF does not include groups – how to fix it (send an email)
	+ Ed – need fine grain coordinate system
	+ CF is focused on modeling – so no groups
* Q – Swath – 1) had band as a dimension – does that require order to the band (by frequency)
	+ Yes have to be sorted numerically (coordinate variable has to be increasing or decreasing)
	+ If you alphanumeric version – then doesn’t matter
* Q – why isn’t this a discrete geometry
	+ Because include buoys, sounding balloons – didn’t show up when thinking about satellites
	+ Use x & y – then a discrete geometry
* Ted – it is possible to deal with OGC than CF community

CF extensions for satellite data – Ed Armstrong (NASA JPL)

* Extensions for documenting level 1
	+ Wavelength and frequency are not elegantly represented
	+ Often put it in variable name or comment section or create your own attribute (gets messy quickly) – not machine/tool readable
* <http://wiki.esipfed.org/index.php/Standard_Names_For_Satellite_Observations>
* spectral response of channel 5 of NOAA-17 AVHRR/3 – want to describe the mid-point (what it formalized
	+ normalize spectral response of a frequency of a spectral response function
	+ this becomes an instrument parameter itself
* GRIS project has SST dataset in netCDF since 2005 – implemented CF best for level 2 dataset
	+ Similar to case#1 (Alek)
* For level 1 variable – have band – essential wavelength “sensor\_wavelength”
* Also able to include level 2 – without band information - combined in 1 file
* Recognition automation – from the tools
	+ Identify variable dimension is part of band/channel list
	+ Find variable described
	+ Read wavelengths
	+ Apply as “dimensions”
	+ Should be relatively simple
* netCDF4 – can band wavelength be a pointer to netCDF group structure – way to package relevant variables (did not investigate if this could be done)
* Alek thinks this is a great way to incorporate groups
	+ Ted – grouping metadata – HDF5 allows groups in metadata
	+ CF community always a tooling argument (grad student not here to re-write tools)
	+ May not need major changes to incorporate groups…
	+ NCML – for external netCDF file – write in ncml to THREDS to look like CF 🡪 can take forward looking file – take ncml to move CF to netCDF file with groups (need to propose inelegant solutions that they don’t like so they move forward)
	+ HDF group is active partner in moving CF to group
	+ Ed – maybe lobby HDF to create groups

CF Unleashed: Introduction to CF/Radial – Joe VanAndel - NCAR

* CF – Climate and Forecast – intended for model-generated and observational datasets
	+ Nothing for radial – all Cartesian
* Want to support radar/lidar community for data providers and tool creators – provide libraries and tools, conversions, and display data
* CF/Radial is a set of extension of radial radar/lidar – submitted request to CF
	+ If you submit and it stalls – not sure what happens next (nothing wrong but no blessing either)
	+ Useful for atmospheric science – supports assimilation into forecast models
* Types of instruments – wide variety – scanning, staring, vertical, and fixed
	+ Ex. S-Pol Radar (stationary) with Ka-Band (1 degree beams)
		- Scanning radar scans in azimuth and radiation
	+ Doppler on wheels (mobile) – can go anywhere there is a kind of road – used for hurricanes, hydrology in mountains of Italy – doesn’t scan while moving, but scientist can’t resist (drive to site and set-up)
	+ HIAPER Cloud Radar (research air craft operated by NCAR) (airborne scanning) - when have airborne platform, have more conversions to worry about because have more plans (not level, not straight line, not in same place)
	+ High Spectral Resolution LIDAR – can point in different directions (also airborne)
	+ NCAR Profilers (449 Mhz and 915 Mhz) – these are fixed – each have multiple beams
* NetCDF means you have operation system independence
* Advantage byte order independent (past had to byte flip to get data)
* Staggered 2D storage of gates and range
	+ Q – this is a ragged array – does NetCDF support
		- In 2 ways – in NetCDF4 it is explicitly support
		- But want NetCDF3 – for a given variable for an entire sweep – all gates stored in array and encode start-index and # of rays
* NetCDF4 uses HDF5 – provides transparent compression (client doesn’t need to deal with compression – library deals with this)
	+ In the past, compression algorithm – then had to uncompressed
	+ Can be up to 20% of original – in the past NetCDF took up too much space
* Sample data – reflectivity field, hotter color = higher precipitation, ½ degree scan
* Range height radar data – bottom is range, vertical is height – cross section of a storm
* Lidar Data – different than radar – point in one direction – either they move or atmosphere moves over them – here lidar is fixed and different air masses flow over (range)
* Have data fields (moments) for each instrument – reflectivity, velocity, polarization)
* Each ray has metadata
* If moving then need more metadata
* Defined multiple coordinate conventions (mobile vs. airborne)
* Current tools
	+ Radx C++ library
	+ Several of these read/write are older and binary
* Future work – incorporate NODC and ACDD,
	+ creating some new libraries (python, matlab, IDL, community archive)
		- These are more approachable for students
* Q (Alek) – submitted proposal to CF (18 mos) – no response … has it been accepted?
* Q (Ed) – what about future satellite mission – SWAT, Mable? – see applications to those instruments - (Jeff) – model doesn’t work – they have multiple beam, push-broom
* Poster of “CF Unleashed” including unstructured conventions to CF

The National Oceanographic Data Center’s Application of CF Conventions for In-Situ Data – Mathew Biddle (NODC)

* Attribute Conventions for Dataset Discovery (ACDD)
* Use all CF attributes
* Highlighted have examples on THREDS and CDL for insitu observations – these are CF definitions
* Q – Difference between trajectory and trajectory profile is ?
	+ Interested in SST fronts - ? not included – line at surface of ocean (contour)
	+ #3 has no temporal order
	+ Does a trajectory need to have time – Peter has constant time – monotonically constant (other variable is distance)
	+ For the convention – decision in time not space based on use cases
		- What about generalize it to a monotonically increasing variable
		- These are CF conventions (except swath)
* Combining CF and ACDD – provides robust document – not standard – assistance and guidance on how to populate NetCDF file with documentation
	+ Provide a decision tree between different templates
* NODC added attributes (global and variable level) (some can be both)
	+ NODC\_name – attribute under geophysical variable, in R controlled vocabulary table (such as instruments)
	+ Platform and instrument (at both levels) – more info about various platforms for instruments collected from (ex. Calibration date, make, model)
	+ Uuid – unique id for netCDF file – changes with updates
	+ Sea\_name
	+ Nodc\_template\_version – which template used to create file
* NODC file populated by NODC terms – NODC manages most of these (except sea\_names)
* Relationship between attributes and variables
	+ Use “cf\_role” to bring in CF under “station\_name”
	+ Added ancillary\_variables for QC flags
* Q (Ted) – talked earlier for group – this is example of group – instrument2 is a “int” – NetCDF is a container (generic) – groups sometimes called variables (code knows it is not really a variable) – this is an “un-natural act” with variables
* Q(Alek) – why netCDF3
	+ Because it is CF compliant
	+ (Ken) – recommend netCDF4 – but if too many “un-natural acts” then go for it (use more logical structure)
* Rubric to compare datasets (pre and post NODC) applied NODC template (only evaluates completeness not quality)
* Benefits of NODC templates
	+ QC in file, standardize data, re-use beyond original intent
* Ongoing – provide tools for convert data into templates and providing a validator
* Q (Jonathon Blythe)– what tools are you developing
	+ Pearl and matlab (not had much time to develop broad based tools) – difficult because have data in different formats and data providers
* Q (Ed) – what % of new providers are using template
	+ This is just a recommendation – they can submit any way they want
* Q (Ed) – pushing it to industry – marine, instrument manufacture
* Community is trying to move standards forward (Ted)