C-RDR Case Study: Fun with Metadata Conformance & netCDF-4.1

Art Burden – STG, Inc.
NOAA/NESDIS/NCDC
May 12, 2011

NCDC Climate Raw Data Record (C-RDR) Project

- Raw Data Records (RDRs) are level 0 data for VIIRS, CrIS, ATMS, OMPS instruments on NPP satellite
- C-RDRs = RDRs converted to :
  - Level 1a data: reconstructed, unprocessed, packaged with support data needed to calibrate and geolocate
  - netCDF-4
- simplify access to raw data
- Available through CLASS and possibly Local Data Manager (LDM)
- Likely uses: Reprocessing, improve calibration methods, post-launch checkout, raw data retrieval

C-RDR Metadata

- NCDC archive guidelines require metadata conforming to ISO 19115 (we are using ISO 19115-2 for remotely sensed data)
- C-RDRs provide unique case for metadata: uncalibrated data from satellite
- How do we best adapt current discovery and usage metadata conventions to these files?
- Ongoing work within NCDC and CF-Satellite discussion group to establish CF conventions for satellite data

C-RDR Metadata
A Hybrid Approach

- C-RDRs:
  - follow CF Conventions where applicable (global and variable attributes)
  - contain the relevant Attribute Convention for Dataset Discovery (ACDD) attributes
  - also include metadata that map to relevant NPP RDR/SDR metadata elements
- Attempt to link C-RDR metadata elements to corresponding ISO 19115-2 elements

Metadata Standard Conformance Issues

- Extension of CF conventions to satellite data not fully established
  - Swath data coordinates: band, sample, scan
  - Bounding box attribute: G-Rings (Geographic min/max values are useless for long swaths)
  - Engineering data: keep NPP names
- NPP fill values fall within data range
  - Not recommended by CF
  - No practical solution available

C-RDR Data Format

- netCDF-4: classic model vs. enhanced
- Using enhanced model (netCDF-4.1): e.g., groups, NC_STRING data type, multiple unlimited dimensions
- NPP raw data benefit from group organization (sensor data, diagnostic data, engineering, calibration, spacecraft diary)

C-RDR Format Challenges

- Many in the user community prefer classic model
  - Can use existing software
  - Groups and other new features elicit unprintable responses
  - Small hurdles can alienate potential data users
- How do we bridge gap?
  - Make enhanced netCDF-4 transparent to end user
- How easy is that?
  - It isn't. Vis/analysis software support for netCDF-4/HDF5 has been piecemeal

MATLAB Example

- Last release of MATLAB (2010b) does not support enhanced netCDF-4 with built-in routines (e.g., ncdisp, ncread). 2011a just released, also does not support enhanced netCDF-4
- High-level built-in HDF5 functions (e.g., h5info) can read some netCDF-4 data, but not non-numeric data and attributes (e.g., NC_STRING). 2011a improved, but not full support
- Low-level HDF5 functions (e.g., h5o.get_info, h5a.read) can be used to read contents of enhanced netCDF-4, but are difficult to use and documentation is minimal
- According to Mathworks, support for netCDF-4.1 is being considered for upcoming future release, but no timeline is available

readCRDR Software
MATLAB C-RDR Reader

- Wrote MATLAB software to access enhanced netCDF-4 data found in C-RDRs
- Consists of two parts:
  - Modified version of h5load.m (available on MATLAB File Exchange) reads and stores all or subset of data and metadata from C-RDR
  - File content printout similar to ncdump
- Provide software to end users that use MATLAB – data access issue? Gone.

IDL netCDF-4.1 Support

- IDL 8.0 does support some netCDF-4.1 features using the built-in hdf_browser function, for example.
- According to ITT, additional support for netCDF-4.1 is planned for the upcoming early summer release

Concluding Remarks: Chickens & Eggs

- Software developers – data providers – users…….Who makes first move?
- We pay penalty for pushing envelope
  - Added overhead to keep up with tech
  - Lag in COTS software support
  - Customer consternation
- From archive perspective, basis exists for pushing technology that prioritizes long-term stewardship over short-term comfort
  - End users often do not care about extensive metadata, just want list of data and a way to access it
  - Archive cannot offer that – learned lessons
- Five years from now, won't we be debating other access issues, and taking data groups for granted?

Contacts

- Team Lead, ATMS - Drew.Saunders@noaa.gov
- VIIRS - Jim.Biard@noaa.gov and Dan.Baldwin@colorado.edu
- Infrastructure - Linda.Copley@noaa.gov
- CrIS, OMPS - Art.Burden@noaa.gov

C-RDR Metadata Example

CF Convention Variable-level Attributes

Name	Description
name (intrinsic to variable)	The name of the variable.
shape (intrinsic to variable)	The dimensions of the variable
type (intrinsic to variable)	The data type for the variable elements
units (CF)	The units for the measurement, if any
long_name (CF)	A human-readable description of the variable
valid_min (CF)	The smallest valid value for the variable
valid_max (CF)	The largest valid value for the variable
_FillValue (CF)	A value that indicates that a variable element is missing
coordinates (CF)	Names of one or more variables that act as coordinates for values of the variable
source	One or more APID and measurement names from the CCSDS telemetry packet format descriptions that specify the telemetry elements used as the source for the variable

C-RDR Metadata Example

ACDD Convention File-Level Attributes

Name	Description
acknowledgment	A place to acknowledge various type of support for the project that produced this data.
cdm_data_type	The THREDDS data type appropriate for this dataset.
comment	Miscellaneous information about the data.
contributor_name	The names of any individuals or institutions that contributed to the creation of this data.
contributor_role	The roles of any individuals or institutions that contributed to the creation of this data.
creator_email	The data creator's email. If not present, institution is used in its place.
creator_name	The data creator's name.
creator_url	The data creator's URL.
date_created	The date and time on which the dataset was created.
date_issued	The date and time on which the dataset was formally issued.
date_modified	The date and time on which this dataset was last modified.
geospatial_lat_max	Geographic bounding box northern limit.
geospatial_lat_min	Geographic bounding box southern limit.
geospatial_lon_max	Geographic bounding box eastern limit.
geospatial_lon_min	Geographic bounding box western limit.
+ 18 additional	

C-RDR Metadata Example

RDR/SDR File-Level Attributes

Name	Description
Ascending_Descending_Indicator	Indicates the ascending/descending state for the dataset.
Auxiliary_and_Ancillary_Data_File_IDs	IDs of the auxiliary and ancillary data files needed to produce science units.
Beginning_Orbit_Number	The orbit number at the beginning of the dataset.
Beginning_Time_IET	The time of the beginning of the dataset in seconds since the TAI epoch.
Day_Night_Flag	Indicates the day/night state for the dataset.
Ending_Orbit_Number	The orbit number at the end of the dataset.
Ending_Time_IET	The time of the end of the dataset in seconds since the TAI epoch.
G-Ring_Latitudes	Latitudes of the corners of the image footprint.
G-Ring_Longitudes	Longitudes of the corners of the image footprint.
Input_RDR_Granule_IDs	The NPP granule IDs for the RDR granules used to produce the dataset.
Input_RDR_Granule_Versions	The NPP granule versions for the RDR granules used to produce the dataset.
Input_RDR_Reference_IDs	The NPP reference IDs for the RDR granules used to produce the dataset.
Input_Support_File_IDs	The IDs of the ancillary and auxiliary files used to produce the dataset.
Instrument	The short name of the instrument.
+ 29 Additional	