ARM Standards Committee Meeting Notes

Table of Contents:

November 13, 2012

Discussion Topics

Notes

Attendees: Sherman, Brad, Raymond, Alice, Renata, Stefanie, Robin, Scott, Chitra, Brian, Sean

Does a standard provide good business sense?
- Level of effort
- Cost, development cost
- Does the standard create a more complicated web, or how hard would it be to update to a new standard with a complicated web of code?
- Could it take too much time to implement but not make it much better?

A lot of effort in the begging for one one person can save a lot of time for a lot of people in the end

It would be a good idea to understand the historical context of an old standard or older code. History matters!

Scope
- Need to keep in mind the longevity of the data
- Data needs to be able to scale to larger size
- How many exceptions do we find with the current data?

Example of a assumed standards that may not be: If a field has a missing_value attribute must have a qc_<field>. Robin said this is not true.

We need to get a list of what other programs are doing.

Post processing could handle some non-standard issues like missing_value

Need to have a valid reason for not following the recommended standards. Not just that you don't like it.

Users may have code that expects the format not to follow standards because of historical data sets. We need to keep this in mind when updating data DODs.

November 20, 2012

Discussion Topics

1. Absolute vs Recommended
2. Expanding the explanation of optional data qualifier for VAP files (ie. qcrad1long)
3. Set standard optional data qualifier options (ie. avg, mean, total, Translator name)
4. What is the definition of QC checks applied to data to move data to level b1? (does qc_time count as a QC check?) What moves a datastream from 0 level to a level?

Attendees

Ken, Robin, Chitra, Brian, Scott, Sherman, John

Notes

Stefanie Shamblin was not be able to attend the call but has given some good notes:

-In file naming, (nn) as the optional data integration period does not always refer to units of period but can also refer to instrument attributes. For example, in the case Chitra points out, rwp's values refer to the frequency at which the instrument is running (50, 915, 1290 MHz). In general, it should always refer to something that is intrinsic to the instrument.

-Currently, our storage procedures and database tables are working with a 64 character limit on filename. 23 of those characters (from data level through the end of the string) are fixed in size. We also need room for 3 characters after the file type to designate a version (.v0, .v1, etc) if it applies. That leaves 38 characters for the platform naming.

Our position is that this needs to be enforced. Increasing the limit from 64 to 80 is possible but certainly not preferred. We also agree that if 38 characters is not enough space to define an instrument platform name, then perhaps the naming should be rethought.

-We believe having 9 (10 if you include 0) sub levels of a data level is more than enough. This field is defined as varchar2 in all of our database tables. If we don't limit it somehow, the question of "where do we stop?" comes up. There are enough idiosyncrasies within data level conventions as it is.

-The Archive is currently processing both .nc and .cdf file formats for netcdf files. Either is fine going forward, but we can't retroactively conform to a standard that requires .nc as the file format for netcdf files.

- Add in something describing that data in the raw file can have issues, but ingested data must be cleaned up for time.
- Instrument named data file does not need a standard. Vendors make the choices and we can't force a change.
- Need a new global attribute describing which standard the file format
- Need to talk to Annette about issues with naming file with first time sample in file.
- Time in file name must match first time in file. Example of a difficult time mismatch is one file using end of data and one is using start of data. How close does it matter. Resolution will depend on instrument. We should encourage that the data follow the standards, allow a small number to not match but list the datastreams that do not meet the requirement. (raw files can be exceptions if not used by users)
- Recommended standard that the raw data file from instrument put time in the name for the first sample of data in the file.

December 4, 2012

Discussion Topics

1. netCDF file extension: nc vs. cdf or both?
    a. Allow upload to Archive as cdf but deliver as nc or cdf upon user's request.

    b. Which do we use in the ARM infrastructure now? Current vote is .nc extensions?

    c. What will break? ISDE and Zebra library will break, but can be fixed. cdf to ASCII conversion tool.

    d. Need to do compile the problems this will cause and then submit an ECR.

    e. ISDE will be a simple change.

2. Need to find documentation about site and facility naming
    a. In a BCR/ECR (BCR-832)

3. Changing file naming to allow easier splitting? Do we want to go down this road?
    a. This is a big deal!
    b. What is a better method? site.facility.***.nc

4. What are the penalties for not following the standards?
    a. Holding data at DMF?
    b. Delay in releasing data from Archive?
    c. Different consequences for different failures to follow standards.
    d. DQ Office does not review the data.
    e. Recommended vs Required standards

5. File time integration descriptors. Do we allow non-time descriptors (ie 915rwp vs 30met). Need to read this section for units descriptor sign off from committee.
    a. Suggest putting time at the end of the name to allow units
    b. About 50/50 with at the front vs. at the end.
    c. We need to not allow a default with no units label.
    d. The main user product will not have a time integration period.
    e. Have a separate document describing how to read an ARM datastream for the end users. Need to call out the exceptions.
    f. Code snippets (both parsing and making a name)

6. What are the requirements for elevating an a level datastream to b level? What does not require the elevation? What about QC checks applied by the instrument, not ingest?
    a. qc_time does not push an a level to a b level
        i. has in the past, inconsistently.
        ii. Need to create documentation
    b. VAPs do not have qc_time?
        i. BW did not require qc_time.
        ii. passing qc_time does not ensure time is correct.
        iii. Should time be treated as just another field?
    c. If the instrument performs QC but only replaces bad data with missing_value does not push level up.
        i. The AOS has files that do not have qc fields, but are b level. This is

because the mentor has looked at the data and provided quality control. This should be good enough to say the data has been quality controlled and allowed to go to b level?

    d. If instrument has QC field and it can not be put into ARM QC standards this will not require to push the level up. (Quite rare if possible?)

        i. Current thinking is that QC from the instrument does not force a b level datastream

        ii. Having a missing_value attribute for a b level field does not require a qc_<field>

    e. Do we allow a valid_min/valid_max/valid_delta without a qc_<field>?

        i. QC was provided by the instrument, but is not site specific. Does this push the "QC'd" data to b level. Not necessarily.

Attendees

Brian, Sherman, Justin, Alice, Renata, Chitra, Annette, Stefanie, Robin, Ken

Notes

December 11, 2012

Discussion Topics

1. In the Standards Document Ken will label Recommended Standards and everything else will be a Required Standards. Sound good?
2. Results from file name e-mail request
3. What is a summary file?
   a. Generated by simplifying the VAP QC
   b. Generated at the DMF
   c. If have a c1 then get a s1
   d. s level file can have a subset of fields and can be site specific
   e. Talk to Krista about this
4. Should qc_time be a requirement for a, b, c level data?

     a. qc_time is not the threshold for promotion to b1

     b. If it is possible to apply automated qc to a field for QC and it is not done then it's a levle.

     c.

5. What happens with c1 data when a c2 datafile is created?
6. What conflicts are people currently working with now. We need example cases
     a. HSRL
     b. MWR3C from a different vendor, ECR-968
     c. ARSCLWACR
7. Other data formats, ASCII, HDF, PNG, JPG, MPG? What standards do we make for these?
8. netCDF dimensions.
     a. time
     b. co-ordiante dimensions and variables
     c. String length dimensions

Attendees

Annette, Sherman, Robin, Brian, Ken, Stefanie, Scott, Alice, Renata,

Notes

- QC is a double sided question.
  - One is QC of the instrument
  - One is QC of the data
- QC at a b level needs to be reported at a b level
- How do we call a file that has a mentor review the data for QC
- using netCDF4? Radar data would be OK as group. Other datastreams will require some thought?

December 18, 2012

Discussion Topics

Reviewing the file naming sections of the document with the full 15 member

committee. We will vote on issues that are not immediately agreed upon or table for further discussions.

Attendees

Robin, Justin, Stefanie, Brad, Alice, Ken, Sherman, Renata, Chitra, Sean

Notes

- Need to keep a list of new changes of bigger changes
  - Position of time in filename
  - change to nc
- Historical data will not need to follow all the rules but new datastreams will. This needs be stated in document.
- Stephanie can get list of all site indicators. Will send to group.
- Renata, Chitra, Robin will discuss statistical examples for file naming
  - look at IEEE examples or other places that have addressed this issue
- SGP S01 sonde data for backup of SGP C1?
- Need to have Archive confirm 40 character limit for datastream name.

January 8, 2013

Discussion Topics

Review netCDF dimensions, global attributes and field name and attribute standards

Attendees

Sean, Brian, Stefanie, Ken, Alice, Sherman, Chitra, Renata

Notes

- Should we allow units other than seconds for arm time and time (julian day, hour, month, …)
- qc_time homework for all.

January 15, 2013

Discussion Topics

Review field name and attribute standards

Attendees

Scott, Brian, Robin, Chitra, Alice, Renata, Sherman, Justin, Tim

Notes

-

January 22, 2013

Discussion Topics

Review field attribute standards

Attendees

Brian, Alice, Chitra, Ken, Sherman, Tim, Robin, Scott

Notes

- _FillValue is the original value set. missing_value is the defined value when the instrument does not report a value and the ingest handles it correctly.
- If we only have missing_value then a _FillValue will not be filtered out by most users.

January 29, 2013

Discussion Topics

Review global attributes

Attendees

Chitra, Brian, Tim, Robin, Sherman, Ken, Renata, Stefanie, Alice

Notes

February 1, 2013

Discussion Topics

Continue review global attributes

Attendees

Robin, Scott, Ken, Tim, Sherman, Renata

Notes

-

February 5, 2013

Discussion Topics

Continue review global attributes

Attendees

Ken, Robin, Brian, Annette, Stefanie, Alice,

Notes

- Only require serial_number if the serial number comes from the raw data.

February 8, 2013

Discussion Topics

Continue review global attributes

Attendees

Robin, Ken, Alice, Stefanie, Sherman, Justin, Krista, Brian, Tim, Chitra

Notes

- Could we remove base_time, time_offset and just go with time?
- This would allow us to use time bins via. the CF convention.

February 12, 2013

Attendees

Ken, Robin, Karen Johnson (in place of Tim), Sherman, Raymond, Stefanie, Tim, Justin, Sean, Renata, Brian, Alice

Discussion Topics

Full Committee Meeting to review proposed standards of netCDF field and attribute structure.

Significant Changes:

- base_time and time_offset will no longer be required (or even recommended) method of providing time. Time will follow NCAR CF recommendations.
- qc_time will no longer be required
- Adding a new field to describe the time cell boundaries and indicate where within the cell the time values are being reported.
- Coordinate dimensions and the corresponding cell description method adopted from NCAR CF convention
- Explicit definition of Location Fields. Alt references altitude above mean sea level to ground level
- List of acceptable abbreviations for field name qualifiers
- Adding standard name attribute for primary fields
- Use of State Indicator field method
- use of _FillValue instead of missing_value attribute
- Explicit definition of sensor_height
- Addition of platform_id global attribute
- Redefine facility_id global attribute
- Addition of location_description global attribute
- Removal of global attributes:
  - averaging_interval
  - averaging_interval_comment
  - zeb_platform
  - resolution_description

Notes

- 11,000 new files sent to Archive were above the 60 character limit for file names. ECR is now proposing a 120 character limit.
  - Changing the Archive DB would be significant increase for 13,000,000 files (currently).
  - Old solution was to use tar. But files are too large to tar up. Would be one tar file per data file. Large overhead for unpacking and repacking.
  - One option is to rename and use another look up table.
  - Second option is to truncate original name to use ARMified name.
- Tim looked at original base_time time_offset discussion.
  - Requires reading two fields to create time
  - Archive uses base_time and time_offset. If those do not exist then use

time in filename.
- Write ECR about removal of base_time and time_offset and going with time only.
- Time Cell Boundary method from CF is generally "liked"
- How difficult will Time Cell Boundary arrays be to generate?
  - Add more attributes describing what the process is.

February 19, 2013

Attendees

Ken, Justin, Brian, Tim, Chitra, Robin, Alice, Renata, Stefanie, Sean

Discussion Topics

Quality Control fields

Notes

February 26, 2013

Attendees

Ken, Tim, Robin, Justin, Alice, Stefanie, Renata, Brian, Sean, Chitra, Sherman,

Discussion Topics

Quality Control fields continued

Notes

-

March 5, 2013

Attendees

Ken, Brian, Tim, Robin, Sean, Renata, Stefanie, Alice, Chitra

Discussion Topics

Continued discussion of Quality Control fields and source field
- Need to discuss pros. and cons. of requiring the reserved bits of 1 - 3 for Standard Mentor QC tests
- Using attributes to directly link corresponding fields for QC in addition to source and other linked fields.

Notes

- Check with Sherman about interpreting the PCM checks for valid_min/valid_max rolled up into one bit test.
- Will require first three bits to be reserved.
- Sean and Ken will look into issue with indicating when a test could not be performed and how to convey that information.
- Alice will add wording from Standards Policy Committee about Penalties for not following standards and exceptions committee.
-

March 12, 2013

Attendees

Ken, Robin, Sherman, Alice, Tim, Renata, Chitra, Sean, Stefanie, Brian

Discussion Topics

- Exception Process
- Test performed method
- Bit option for source field

Notes

- Send document out to developers and translators for review. Developer review by end of STM. Translators?

- Standards Committee deadline = May 1, 2013

April 2, 2013

Attendees

Ken, Justin, Annette, Robin, Tim, Chitra, Karen Johnson, Alice, Sean

Discussion Topics

- New Introduction
- New Integer QC
- Updating wording for required vs. recommended
- netCDF4 file size limit for allowance
- Keeping requirement for first three bits. missing/min/max tests
  - need to standardize the description text!
- Method of indicating bit-packed QC vs. integer QC

Notes

April 9, 2013

Attendees

Ken, Tim, Annette, Justin, Stefanie, Brian, Karen Johnson, Sherman, Renata, Sean

Discussion Topics

- Changed name of attributes from value_<#>_description/assessment to flag_<#>_descripiton/assessment
- Removal of reserved bits 1 - 3 for missing, valid_min, valid_max
- Why are test parameter value attributes listed with the data field for QC tests indicated with the QC field?
- If qc_min, qc_max are used which field do they belong to? (data or QC)
- What is the point of a "value equal to missing_value" qc test?
-

Notes

- Are there tools that use NCAR CF convention for status flags from other organizations?
  - Renata will ask around.
- Having a missing_value bit
  - Allows multiple missing_value values and only checking one test will check all at once
- Does having a missing_value declared in data require a bit/flag for qc field as well?
- For SAS VAP a case where the value can not be calculated both missing_value bit and test bit are set.
  - Should set both bits vs. allowing the bit to indicate the value is set to missing_value?
- qc_min and qc_max belong to data or qc field?
- Additional things to tackle?
  - Do we tackle Best Estimate?
  - Primary vs. secondary vs. instrument fields?
  - Statistical time?
  - How to roll out standards?

April 16, 2013

Attendees

Ken, Justin, Brian, Tim, Stefanie, Sean, Renata, Alice

Discussion Topics

- Time

Notes

- Make an example as an appendix
- Make a statement of when the "best estimate" term can be used.
- Stefanie will look into best estimate use currently.
- Prefered list of field names needs to exist. Split up the work.

April 23, 2013

Attendees

Ken, Justin, Renata, Tim, Sean, Alice, Stefanie, Chitra

Discussion Topics

- Best Estimate:
    - Asked around and there is no definition.
    - Jim is asking the Translators on the next call for a definition
    - I suggest we define what a Best Estimate is but not currently define a standard.
- cell_method ([link](#))
- New time wording
- Exception process. Should this describe the DOD building process or only the exception process. Where do we describe the DOD building process?
- Examples in Appendix.

Notes

- Can DQ Office, DMF, Archive, … query primary, secondary, … database?
    - dsdb
    - arm_int
- Any datastream that goes to the Archive must have primary, secondary, … info in the database.
- PI products also have primary, secondary, … categories?
- PMC tool can upload a ncdump header to populate the database.

April 30, 2013

Attendees

Ken, Brian, Stefanie, Tim, Robin, Justin, Sean, Alice, Chitra, Karen Johnson

Discussion Topics

- cell_method attribute
- category ranking
- examples
-

Notes

May 7, 2013

Attendees

Ken, Robin, Tim, Stefanie, Renata, Alice, Sean, Chitra

Discussion Topics

- Applying standards to aerich1.b1 datastream

Notes

-

May 14, 2013

Attendees

Ken, Robin, Justin, Chitra, Renata

Discussion Topics

- How to implement the new standards

Notes

- Move time_bounds from required to recommended?

- Krista is building MWR RET VAP for n channel?
- SAShe implementation plan for test case?
- SACR implementation plan with Karen for test case?
- VCEILPBLheight with Vic?
- ASSIST?
- KAZR b1 level?
- Standards implementation date July 1, 2013

May 28, 2013

Attendees

Ken, Justin, Karen Johnson, Stefanie, Tim, Renata,

Discussion Topics

- How to implement the new standards

Notes

- Stefanie will talk to Archive about her being the ARM Archive Exception Committee representative.
- Ask Laura about being on the Exception Committee?
- July 1, 2013 too soon?
  - Need to update PCM
  - Need to update ADI
- How fast can PCM and ADI be updated?
- Need to have a code repository area for code that is updated to the new standards.
- Bringing others into DOD building in PCM early is a good idea.
  - Who can do this
  - How to make it a known and used process
  - Create a DOD with standards implemented to compare with current DOD in the PCM tool
- Create a powerpoint document to show the development process?
  - Presented via webEx
  - Eventually can be a reference
  - Help training people with PCM tool
  - e-mail list for discussing DOD construction? Archived for reference.

- ○ Many, many examples...
- MMT?
  - ○ Used to document the work involved?
  - ○ Should be implemented early?
  - ○ ECO/EWO/... may contain information as well
- KAZRARSCL DOD is in PCM
  - ○ Should review and update to standards
  - ○ Help
- Send e-mail reminder about Standards Document review
- Create b and c level template in the PCM too
- Add standards breakout group at Developer's Meeting

June 4 , 2013

Attendees

 Ken, Scott, Karen Johnson, Tim, Robin, Sean, Chitra, Brian

Discussion Topics

- Implementing suggestions

Notes

- KAZR data calibration:
  - ○ data file from a1 -> b1 and an AUX data file from a1 -> a2?
- How to build a process for giving recommended suggestions to mentor/translator?
  - ○ need to build the lookup table for field names
- RAW file name convention may need to be relaxed to fit?
- Mike Jenson's suggestion:
  - ○ use VAP implementation plan accepted as date for VAP meeting deadline of standards

June 18, 2013

Attendees

 Ken, Justin, Renata,

Discussion Topics

- Roll out plan
- Final format of the document
- Changes made from suggestions

Notes

July 2, 2013

Attendees

Ken, Justin, Brian, Tim, David T., Karen J., Brian,

Discussion Topics

- Roll out plan
-

Notes

-

September 17, 2013

Attendees

Chitra, Renata, Krista, Justin, Ken, Laurie, Adam, Rick, Jim, Sherman, Alice, Matt, Tim, Karen, Brian,

Discussion Topics

- Issues with implementing ECO-01026

Notes

- Add a data version attribute etc.
- And we have a data flow change that will tweak a couple sentences on page 17.
- Cycle of once a year where errata is collected and a new version is published?

Need to add a web based reference for a short summary.
- Have a summary for recommended and required
- Optional with required format

Issues with understanding difference between required vs. recommended.
Understanding the required for optional parts.

Be explicit about exceptions with comments?

PCM updates are fine by deadline.
- Business decisions i.e. Type for correct field
- Attribute names
- List of field names as a pick list?

Code snippets for developers? vs. ADI library having the code implemented.
- Brian could expose the library functions
- Smaller bits of code for standards stuff or more including science codes?

What are the expectations of PCM and ADI?

ADI vs. non-ADI by implementation date?

Exception Committee?

How many DODs can we review before the deadline to understand the difficulties?

September 24, 2013

Attendees

Ken, Sherman, Brian

Discussion Topics

● Issues with implementing ECO-01026

Notes

PCM
- Remove qc_time as required field
- Add state indicator field format checking (an example currently exists in example.b2)
- Add check for datastream, Conventions, location_description, and platform_id global attributes
- Add standard_name field attribute for lat, lon, alt
- Add ancillary_variables field attribute to base_time and time_offset
- Add cell_method field attribute as an option (?)
- Add sensor_height field attribute as an option (?)
- Add ancillary_varialbles field attribute for data fields to link to qc, source and status fields
- Add flag_method field attribute to QC fields
- Add check for bit_<#>_assessment to only allow "Bad" or "Indeterminate"
- Add fields for bin bounds
- Reserved attributes for setting offsets from time or coordinate field values under time field. observation_window = -30, 30 ; bound_offsets = [-30, 30] . Will be a recommended value.

ADI
- use *.nc file extension
- Add time cell boundary field via bounds attribute
- Add coordinate cell boundary field via bounds attribute
- Add state indicator field method
- Add ancillary_varialbles field attribute for data fields to link qc, source and status fields
- Add flag_method field attribute to QC fields
- Global attribute bit definition (is there anything that needs updating?)
- Add method for using source field attribute
- Add method for using source_* fields
- Add datastream, Conventions, location_description, and platform_id global attributes

October 1, 2013

Attendees

Justin, Sherman, Ken, Laurie, Laura R., Alice, Renata, John Bell, Chitra, Sean, Matt, Krista

Discussion Topics
- Issues with implementing ECO-01026

Notes

- I believe that the paragraph after "File Name Length" on page 13 needs to have the numerical limits altered. a) The current numbers don't add up. b) The 40-character limit for "datastream" ["(sss)(inst)(qualifier)(temporal)(Fn).(dl)"] and the 33-character limit for "instrument_code" ["(inst)(qualifier)(temporal)"] will require rebuilding all of our tables that use "datastream" or "instrument_code". Were these changes (from 33 and 25 out of the previous 61-character filename) explicitly approved as part of the standard?

[To be fair, the previous limits allowed fudging the ("instrument_code+Fn") if the facility number needed two digits, and allowed only 59 characters if the version number needed two digits.] John Bell, 9/27

What does the Jan. 1, 2014 deadline mean:
- First time products (not in Archive now) must meet standards
  - not including evaluation VAP
- Older products would be grandfathered in
- External data? Not required but will do best to comply.
  - But must meet the datastream name standard.
- Updated VAP/Ingest
  - Small emergency changes do not require a full change to meet standards
- Is it an issue of ADI vs. non ADI datastreams?
  - ~30% of VAPs in ADI
  - 20/52 ingest are not ADI
- Could set up a 6 month period of testing to see how difficult it will be to implement
- Education of ARM?

- - Someone needs to create the teaching tools?
    - Someone will teach people. Eventually Developer teaches Developer
    - Need to teach Mentors/Translators?
    - Need a web page version of standards?
    - More linked examples for instructions
    - Working group tutorial for mentors/translators?
- Biggest issue is delaying the release to mentors and translators
  - Don't understand cosmetic changes
  - Details of standards are not important to people not funded by ARM
  - Start selling the standards to the user community to encourage the desire
- s1 datastream creator
  - Needs to be updated to meet standards?
  - Contact Krista about making the updates.
  - Implemented in ANCIO but looking to get rid of this method for creating the summary files.
  - How much effort to rewrite? Could be done in a week. Need to look at plan and see if possible after Oct. 8
  - Do we still need it? Is Archive creating summaries?
  - Get metrics on s1 downloads

October 8, 2013

Attendees

Ken, John B., Renata, Justin

Discussion Topics
- Issues with implementing ECO-01026

Notes

October 18, 2013

Attendees

Ken, Jim M.

Discussion Topics

●

Notes

- Need to update the required standards summary
- Need to create review committee by Jan. 1, 2014
- Create a review with people not part of development committee by Working Group
    - ~10 Powerpoint slides presented to key stakeholders
    - Need to present with enough time before deadline to allow discussion and changes

November 12, 2013

Attendees

Ken, Sean, Alice, Jim, Mike J., Robin, Matt M., Justin, Cory, Jimmy, Tom, Doug, Randy, Scott, Raymond, Stefanie, Sherman, Renata, Connor, Brian, Chitra, Laura, Shaocheng

Discussion Topics
- Review of implementation for new standards

Notes

- A product currently in the Evaluation area and released to Archive production area, does it need to meet the standards? No, only VAPs that are started after the deadline.
- The review committee should go through the evaluation VAPs to make a determination that the product should/should not be updated to meet standards when released to production.
- udunits required. Need to check on exceptions language for units that do not

currently exist. (Good to go).
- Document method of updating standards
- Document policy for reprocessing and meeting standards
- Is a new site a new datastream? (Currently no, only if a significant difference)