

From: Debra Kazden <dkazden@igpp.ucla.edu>

Sent: Monday, June 06, 2016 9:15 AM

To: pds4ddwg

Subject: Notes from PDS DDWG - 2016-06-02

title: DDWG Notes 2016-06-02

layout: default

date: 2016-06-02

June 2, 2016

Notes by Debra Kazden

(No notes from Debra Kazden for May 26, 2016)

Known Attendees:

M. Gordon, E. Guinness, S. Hardman, L. Huber, S. Hughes, C. Isbell, R. Joyner, D. Kazden, J. Mafi, S.

McLaughlin, L. Nagdimunov, J. Padams, C. Phillips, A. Raugh, R. Simpson and J. Stone

Meeting Agenda and Summary

1) CCB/SCR Statuses:

-- Voted by CCB:

-- CCB-65: Need additional Target Identification/type values (A. Raugh)

-- E-vote PASSED: 5 Yes (ATM, GEO, IMG, RMS, SBN), 2 Failed to vote (IPDA, PPI)

(Not Discussed)

2) Task Statuses - 5 Minutes each

(Not Discussed)

3) SCRs and Issues to Discuss:

-- CCB-77: Augment Product Update with File Area Update - S. Hughes

- Open: under DDWG discussion
- has been TA'd
- 20141002: There is now a tiger to work Update in general that will start in a few months
- 20150519: Waiting for M.Gordon ?
- 20150922: DDWG discussion topic; SCR needs to be updated by Mitch
- 20160324: Mitch prefers to supersede this SCR and add new SCR
- ** (Not Discussed) **
- CCB-125: The bit mask attribute seems to be misplaced and possibly missing where needed (A.Raugh)
 - 20150915: Open; needs DDWG discussion
 - 20151008: Jordan to provide example label that uses bit mask
 - 20160323: WG: J.Padams, R.Simpson, A.Raugh, R.Joyner
- ** (Not Discussed) **
- CCB-131: Missing constraint on Special Constants attributes (A.Raugh)
 - 20150922: Open
 - 20160223: under DDWG discussion
 - 20160322: EN governance; will take lead
- ** (Not Discussed) **
- CCB-133: Special Constants class precludes the ability to specify multiple invalid/missing constants (J. Padams)
 - 20151012: Open
 - 20151021: Under DDWG review
 - 20151022: WG -- Jordan, Steve and RJ; sent email to WG with proposed changes
 - 20151105: Jordan -- special constants needs to be specified per "band" not per "axes"
- ** (Not Discussed) **
- CCB-138: Mismatch between context object types and values of type in Observing System Component class (A.Raugh)
 - 20151202: Open; under DDWG review
 - 20151203: WG: Anne, Steve, Dick, Jordan, and RJ
 - 20160310: until someone volunteers to lead the effort -- on hold

****(Not Discussed)****

- CCB-142: Create Data Quality Flags to hold metadata on Quality Flags (E.Shaya)
 - 20151229: Open;
 - 20160126: Under DDWG review
 - 20160322: Ed didn't like Simpson's CCB-142 implementation
 - Ed wants a lot of specifics embedded into XML
 - Simpson trying to figure out how to make it 'simpler'
 - 20160323: Simpson generated presentation for DDWG review / comment
 - tabled until next session (20160410)
 - 20160428: updates discussed by DDWG
 - requires IMG & others expertise to carry forward
 - 20160505: discussed by DDWG; E.Shaya led discussion; 2 competing implementations
 - address 2 issues:
 - bang for buck in terms of worth doing
 - implementation recommendation
 - WG: Steve, Anne, Jordan & Chris, Lev

****(Not Discussed)****

- CCB-144: Some examples in Examples collection are incorrect or out-of-date (Lev Nagdimunov)
 - 20160210: Open
 - 20160322: EN governance; will take lead
 - 20160323: may be augmented by CCB-155
 - 20160526: discussed by DDWG; "fixed" as part of updates to v1600

****(Discussed)****

- CCB-149: Should PDS4 allow packed data? (E.Shaya)
 - 20160309: Open & Under DDWG review
 - 20160310: Sent email to E.Shaya asking that he upload his version of the IM for packet data to JIRA
 - DDWG will review and provide comments
 - PPI has volunteered to attempt to convert a PDS3 product using the Packed Data class
 - 20160322: dependency on CCB-153; and vice-versa

-- 20160428: sent email to J.Mafi for status on providing examples

****(Not Discussed)****

-- CCB-151: Bundle Member Entry and Internal Reference do not require either LID or LIDVID. (A.Raugh)

-- 20160309: Open & Under DDWG review

-- 20160322: EN governance; will take lead

****(Not Discussed)****

-- CCB-152: field format definition mismatch between IM and SR. (L.Nagdimunov)

-- 20160309: Open & Under DDWG review

-- 20160322: EN governance; will take lead

-- 20160428: dependency on CCB-143

-- 20160511: 2nd email to Steve to ask for TA; to make ready for vote

-- 20160512: Lev & Dick to revise SCR to include adding POSIX to IM

-- then back to DDWG for review

-- 20160526: Lev to present to DDWG

-- Steve to update TA to include modifying regex to add 'E' ; back to DDWG

-- 20160601: TA'd; back to DDWG for vote

****(Discussed)****

-- CCB-153: SR Needs Additional Description of Packed Data Fields. (E.Shaya)

-- 20160321: Open

-- 20160322: dependency on CCB-149; and vice-versa

****(Not Discussed)****

-- CCB-154: Promote a Mission Information class to Discipline Governance Level. (S.Hughes)

-- 20160321: Open & Under DDWG review

-- 20160428: updates discussed by DDWG

-- formed WG: Steve, Anne, Jordan, Joe

-- 20160512: emailed Steve asking if the WG had met and is SCR ready to go back to DDWG for discussion?

****(Not Discussed)****

-- CCB-155: Need "Example Set" to include program test data. (A.Raugh)

- 20160323: came from discussion of CCB-144
- 20160323: Open; request to provide additional examples; to include 'test data'
- ** (Discussed) **
- CCB-156: Inconsistent Discipline Dictionary Technique for Local Internal Reference, et al. (A.Raugh)
 - 20160418: Open
 - ** (Not Discussed) **
- CCB-159: Bug fixes for Version 1.7.0.0. (J.Hughes)
 - 20160426: Open & Under DDWG review
 - ** (Not Discussed) **
- CCB-160: URN prefix for the Japan Aerospace eXploration Agency (JAXA). (E.Law)
 - 20160517: Open & Under DDWG review
 - 20160531: Queued_for_Implementation:
 - E-vote Consent item PASSED: 7 YES (ATM, GEO, IMG, IPDA, PPI, RMS, SBN)
 - ** (Brief Update - Passed by the CCB) **
- CCB-161: Allow DOY format in start date time and stop date time (L.Huber)
 - 20160517: Open & Under DDWG review
 - 20160601: TA'd; back to DDWG for vote
 - ** (Discussed - voted to endorse the SCR as written) **
- 4) Topics for Discussion
- Proposal: CCB-1xx: Remove Enumerated List from Instrument.type (L.Huber)
 - Status & develop implementation plan
 - ** (Discussed - voted to have an enumerated list) **
- SETI Issues (R.Simpson et al)
 - Status
 - Issues in XLS have been vetted by SETI notes
 - Issues to be "consolidated" & prioritized
 - ** (Not Discussed) **
- IPDA PDS4 Project: 2014-2015 Final Report (S.Martinez, S.Hughes)
 - Status & develop implementation plan

****(Not Discussed)****

Notice sent before the telecon in email from R. Joyner - June 1, 2016 See enclosures:

-- a list of the full topics under discussion by the DDWG

The results of the CCB 20160529 e-vote are as follows:

-- CCB-160: URN prefix for the Japan Aerospace eXploration Agency (JAXA). (E.Law)

-- E-vote Consent item PASSED: 7 YES (ATM, GEO, IMG, IPDA, PPI, RMS, SBN)

This week's agenda will focus on the following topics.

(1) Please review and be prepared to discuss and possibly vote:

-- CCB-152: field_format definition mismatch between IM and SR. (L.Nagdimunov)

-- 20160309: Open & Under DDWG review

-- 20160601: TA updated to include 'E'; ready for DDWG vote

-- CCB-161: Allow DOY format in start date time and stop date time (L.Huber)

-- 20160517: Open & Under DDWG review

-- 20160601: TA'd; ready for DDWG vote

-- CCB-144: Some examples in Examples collection are incorrect or out-of-date (Lev Nagdimunov)

-- 20160210: Open

-- 20160322: EN governance; will take lead

-- 20160323: may be augmented by CCB-155

-- 20160526: discussed by DDWG; 'fixed' as part of updates to v1600

-- 1 week to review / comment; back to DDWG

-- CCB-155: Need "Example Set" to include program test data. (A.Raugh)

-- 20160323: came from discussion of CCB-144

-- 20160323: Open; request to provide additional examples; to include 'test data'

-- PPI will provide ARRAY example

-- request for color RGB image; check if IMG can provide

(2) Proposal: CCB-1xx: Remove Enumerated List from Instrument.type (S.Hughes / L.Huber)

-- Status & develop implementation plan

(3) Exposing elements in the PDS4 Common XSD (S.Hughes)

-- Discussion & implementation plan

(4) SETI Issues (M.Gordon, et al)

-- Reconciling Local Data Dictionary Versions with New System Builds

DDWG Telecon

CCB/SCR Statuses

CCB-160 passed. Went to CCB. Vote was pretty unanimous. This was the JAXA thing.

SCRs and Issues to Discuss

CCB-152 - field format definition mismatch between IM and SR See <https://pds-jira.jpl.nasa.gov/browse/CCB-152>

Question: Has everyone reviewed CCB-152? We decided last week that the TA would be reviewed. Any issues with the SCR as written?

Answer: Yes. Two issues - they are not new.

1) For binary data the format statement gives precision, but some tables might have different precision for different rows.

2) Upper case E and lower case e interchangeable? Would think if upper or lower case is in format statement the same would be in the data.

~ For input, upper or lower case is not significant. Is for output. In terms of varied precision, it's rare, but when it happens one node has their DPs prepare a statement about it.

~ Someone isn't sure they understand the problem with upper case E.

~ If POSIX differentiates between upper case and lower case E in output - in input interchangeable.

~ Agrees. Would make the validator spit out a warning.

~ Someone was concerned about validation. Wouldn't validate output.

Question: What does the POSIX standard say? We should go with that.

Answer: Begins with if you begin with a lower case e it converts it to upper case E.

~ Someone is pretty sure that's right, but still unsure what problem we are trying to solve.

~ How stringent we want to do the format statements.

~ Someone would do a validation warning. Thinks we only need to nail down 95 percent.

~ Okay with a warning, but we could also not say anything. Lower case e in format statement would force lower case e in data and vice versa.

~ Not sure how we would say that.

~ If lower case e in format, we don't want an 'F' in the data.

~ We've covered that.

Question: There are two issues on the table. Are we ready to vote or do we want to get back to addressing the issues?

Answered with Another Question: Does anyone care?

Answer: Someone was late. Not sure what the first issue was.

~ First issue was concern over precision indicated by format statement might not always be the same in tables.

~ Someone also has that concern.

~ We could drop the wording. Not sure.

It is important to have indication of the precision in binary files. Should be well defined. On the rare occasions when it's not the same that should be documented. The POSIX standard costs 300\$ to look at. Input is not significant for reading data values.

Question: Should we change the wording about precision?

Answer: If we have to. DPs should be describing precision in fields.

~ Not all columns have the same precision.

~ True, so we need to provide a description.

~ Someone is not sure this can be solved with field format. The SR should continue to say what it does, but not super opposed to a change.

~Someone would like to see the SR say that the format statements reflects the highest precision and that there's a description for lower precision.

Question: Do others agree?

Answer: Agree.

****Action Items**** - Dick will write it up. Lev will update the SCR. We will bring this back next week. Dick will look at upper case versus lower case E again.

CCB-161 - Remove DOY format for start date time and stop date time in Standards Reference See
<https://pds-jira.jpl.nasa.gov/browse/CCB-161>

Question: We decided last week that this would be TAed and we would vote this week. Has everyone reviewed it? Is everyone ready to vote?

Answer: Crickets.

Another Question: Vote?

Answer: Someone is still unsure that the wording is clean, but is ready to vote.

~ The SCR doesn't validate the changes in 101.

****The Vote - to Endorse the SCR as written:**

ATMOS - Endorse

EN - Endorse

IMG - Endorse

GEO - Abstain

PPI - Endorse

NAIF - Not Here

SBN - Endorse

RS - Endorse

Rings - Endorse**

Cool. It goes to the CCB now.

CCB-144 - Some examples in Examples collection are incorrect or out-of-date See <https://pds-jira.jpl.nasa.gov/browse/CCB-144>

(See email: PDS4 Example Products -- updated to IM v1600 - from Ron Joyner, May 25, 2016)

We elected last week to review the examples and send comments to Ron. Only got comments from RS.
~ GEO will send comments today or tomorrow.

Someone got the general feeling that once the explicit issues are fixed we should move on. New SCR if need more. Found a few issues. One was an issue with a tool. Would like the issues fixed. Will send Ron comments today.

One concern is that there is no indication in the examples of what is required and what is optional. Also, users might not know what labels are supposed to look like. Unclear that the examples reflect our best practices. Not ready to show these to others. Many mistakes.

~ Someone agrees. Things like array 1D are deprecated, so there's no point having an example of that.

Question: How will we know that all of the examples are reviewed?

Answer: Two people were working on them, but one was re-directed. We do what we can do.

~ One of the two didn't look at everything - tried to cover the ones that were the most relevant to the most users. Documents need more work too. Tried to do it on the side, but it's all fallen to EN. Very underfunded in terms of time and effort.

Someone says there's a level of usefulness. Thinks it's more important that the structure is correct.

Waiting for comments. If we need to do documents we will need a volunteer.

~ Suggestion: Maybe this can be added to the document report for the MC. The document team is down to one person. The MC needs to know that.

~ Someone agrees. Presentation to MC will show only Ron's name on the document team slide. Will be attempting to raise the issue.

~ Don't need to hint - this is the small tip of the iceberg. We need more people and time. It's not realistic to have only one person. Not enough manpower. Not sufficient.

~ It's also important to have other viewpoints.

Let's move on.

Someone will support this at the MC when it's reported as a real problem.

~ Yes, the issue is being raised of who is responsible for each document. We need to look at the governance level and the resources.

CCB-155 - Need "Example Set" to include program test data See <https://pds-jira.jpl.nasa.gov/browse/CCB-155>

Last week we left it that PPI would provide an array example. Also a PPI packed data example.

~ PPI hasn't had time yet. Steve Joy promised the packed data example.

~ There's also something expected from IMG.

The SCR is in limbo. Templates were sent (see email: PDS4 Product templates - From Ron Joyner, June 1, 2016)

Was an MC directive. Want to see templates for each product type. EN generated them.

~ The templates have lots of problems.

~ Someone agrees.

Question: What problems?

Answer: For example, in document additions, it says there are 50 additions, but only one is listed.

~ Templates have random values. They are templates. Don't suggest that users use them - but they can see in Oxygen what's required or optional.

~ Not very helpful. Array one had the wrong number of axes. Definitely not ready for prime time.

~ Things may pass Oxygen and still not be correct. Don't want templates that are nonsense - should focus on the examples.

In XML, every field that isn't fixed should be left blank - not filled in - shouldn't validate - empty places for values - that removes the problems - except for problem that if you don't put in number of repetitions won't know how many to add.

Typical user can't use the templates. Can't tell what's required versus what's optional. There's no indication of why things are there.

Not sure how to solve this. If we prefer examples without values they can be made. Did these with a simple python script. Can do it with no values.

Question: Think this is premised on the idea that the MC requested templates. Does anyone actually remember that?

Answer: MC note-taker doesn't recall this recently.

~ Dan asked for it twice.

~ Not sure the MC is behind this.

~ Maybe we should ask the MC who is responsible for getting this done. Then they can make a plan - then generate and review based on the plan.

~ Someone agrees. We don't want to have maintain examples - they should be very stable. Has users who wanted templates before they understood how to use XML aware editors. Templates are still useful for data structures, but people seem happier generating from schema on the fly. Users might need

templates, but will have to understand what goes and what doesn't.

~ Templates can't be too complicated. MC needs to guide more.

Let's move on.

CCB-1xx: Remove Enumerated List from Instrument.type

At the last meeting we said we would try to make a decision on an implementation that would push this forward. Wants to vote on if we can move forward on modeling approach.

~ That's not the right question. The question should be that we have three options...

1) Steve's classification

2) Go with a slightly different version - the hierarchical approach

3) Unenumerated

Question: Did everyone get that? That's 1 - the Steve approach - to model all values in a classification scheme, 2 - hierarchical, and 3 - unenumerated..?

Answer: Options 1 and 2 are very close. Hard to vote. Needs to be a discussion on how close they are.

~ There is preferential voting.

~ We could vote on enumerated versus unenumerated and then vote on options 1 or 2.

~ How you govern the values is the question. Really thinks there are only two options.

~ Someone likes the idea of enumerated versus unenumerated vote and then discussing hope to proceed with the implementation.

Initial Vote - Enumerated or Not Enumerated ATMOS - Unenumerated EN - Enumerated IMG - Enumerated GEO - Not Enumerated PPI - Enumerated NAIF - Not Present SBN - Not Enumerated RS - Enumerated Rings - Enumerated

Three votes for not enumerated, five for enumerated. Enumerated is the winner. So now we need to decide on the Steve approach or the hierarchical.

The classification identifies the top level part of the hierarchy list. Thinks the classifier approach is hierarchical. The question is how to specify the next levels.

Question: Do we need a second or third level?

Answer: That might be where text fields come into play. Classifiers led to the top level.

If we want to be clear, we are looking at Steve's systematic approach versus a hierarchical common sense free form list.

~ Someone isn't sure they would qualify it that way. At this point it looks like we have a dozen or so higher level items and then we need free form qualifiers or keywords to further identify them.

Someone strongly supports Steve's approach. It's not what we did in PDS3. Steve's approach has three aspects: 1 - we embrace the decision tree, 2 - Instrument type is multi-valued in context products so we don't need so many values, and 3 - The list of 20 or so in the second iteration was pretty good.

Supporting Steve's approach - thinks it is pretty complete - wants to keep it simple. If you have an instrument that can be typed - don't think we need too many levels.

ATMOS and GEO have very large numbers of potential instruments. If put something like atmosphere's instrument with qualifiers could live with that, but concerned that people aren't understanding ATMOS's issue with the keyword. All they get is atmospheric science instrument - then something and and then you have to hope users can find it.

~ Could add more values to the list if it's not complete. Can tell Steve what other bins to add.

~ Will probably never be complete. There needs to be a mechanism for adding to it that isn't onerous.

Also, these terms are pretty generic. Likely in search to get too much. Bins will be dumped together.

Struggles with how practically this can be used.

~ Context products, but would affect search.

No actual use cases considered here. If really used for search need a better understanding of use cases.

We have 20 or so from classifier - issue is extending down to make more specific. That's the list of 80 that was distributed. Has different imagers and stuff and hooks to GEO or ATMOS things like temperature instruments. Believes that what's being said is that we want the top level and a text field. We need to make a decision.

There's already a description field in the context products. Agree that we need to look at use cases. Spending a lot of time. One solution is just use the top level.

~ On use cases - idea from model perspective was to describe the types. To type the instruments from modeling perspective. Sean is using it for search - for grouping. Will show that at MC.

If model is rich enough there will be lots of use cases. Right now we are harvesting and information is being grouped based on instrument type.

Question: Are we ready to vote - Steve versus hierarchy? Is everyone clear?

Answer: No.

~ Think we've said that a majority of us are not ready to give up enumerated list, but how to move forward to find best enumerated values - if going to continue - should focus on one approach.

We could vote for the decision tree - yes or no.

Let's vote to endorse the decision tree.

~ The issue is if people will put the time and effort into it.

~ Steve has given the information he has so far - the list of 20 is reasonable solid until there are new instruments. The question is if the list is sufficient.

~ The issue is how much more time this merits. If need 42 more man hours maybe we should go with unenumerated. Maybe we need to see Sean's tool first, but not all of us will be at MC.

To clarify the middle of the road - nodes that wanted specific values got them. Also have the ability with hook to provide own list - like to add specific instrument names. Seems the best of both worlds. Not a lot of additional work to do, except for people creating their own names.

~ So, proposing the 20 plus a new parameter.

~ Yes. Could have text field controlled by nodes - that they would validate against. It's one possible implementation. Not sure what else we might do.

Question: So, what are we calling this?

Answer: Middle ground. Nodes who want formal list get it and others can use text field. Gives all nodes what they wanted.

Question: Does everyone understand?

Answer: There's an issue of discipline level governance for the second level of the of the first 20.

~ Nodes would need to agree on governance. Issues to work out.

~ Yes, there would be issues to solve.

~ Yes, but the work resulted in a pretty solid list.

We are not resolving this today. No meeting next week because of MC. Need to sit with Sean to see how hard it is to find things in search.

~ SBN doesn't care about Sean's tool. Wanted unenumerated because instrument type is not a parameter for search for SBN users. Need broad categories of end type. Thinks this requires a lot more analysis.

~ But a lot more time on analysis might not be worth it.

~ Someone isn't sure about that.

~ The hybrid solution - accepts that some nodes can easily type their instruments and has hook for those who can't easily type their instruments - they can come up with their own lists. Thought it addresses both issues.

~ The hybrid wasn't fully explained before today. Optional sub instrument type or whatever - giving additional description. Would allow nodes to put in what they need.

~ Exactly. One of the issues raised was that all values had to have CCB review.

~ CCB process for adding to enumerated list is beyond stupid.

Question: To clarify, hybrid approach - decision tree or something different?

Answer: We came up with a short list for instrument type - very specific. Hybrid is in part two - value like atmospheric properties that is not specific at all indicates the need to go to another type. A sub-type where you can add your own values. Hierarchical for those who need that.

Another Question: So is this the decision tree?

Answer: The decision tree is only for the first level.

~ In reality, we would run the entire decision tree to see if only top level. We need to capture all of this in writing. Has to leave in five minutes. Sees no reason to vote. Will write this up on plane next week if people are okay with that.

Question: Anyone opposed?

Answer: (Silence)

****Action Item - Mitch****- to write up the enumerated list discussion.

No meeting June 9 because of the MC.

Question: Everyone around for June 16?

Answer: Joe might be unavailable.

Have a good couple of weeks.