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

Sent: Monday, May 02, 2016 10:46 AM

To: pds4ddwg

Subject:Notes from PDS DDWG 2016 04 28

---

title: DDWG Notes 2016-04-28

layout: default

date: 2016-04-28

April 28, 2016

Notes by Debra Kazden

Known Attendees:

R. Chen, E. Guinness, S. Hardman, L. Huber, S. Hughes, R. Joyner, D. Kazden, J. Mafi, S. McLaughlin, T. Morgan, L. Nagdimunov, L. Neakrase, J. Padams, C. Phillips, A. Raugh, R. Simpson and J. Stone

Meeting Agenda and Summary

1) CCB/SCR Statuses

-- None

2) Task Statuses - 5 Minutes each

(Not Discussed)

3) SCRs and Issues to Discuss:

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

-- URGENT - enhancement / improvement

-- Open:

(1) Needs Proposed Solution

(2) Needs Requested Changes

-- 20150730: DDWG -- Anne to think about working the solution;

-- 20150813: formed WG: J.Mafi, Ed.G, A.Raugh, RJ

-- 20160225: Anne presented 4 questions to DDWG; will update JIRA with consensus

-- 20160324: Anne posted solution to JIRA; EN to review - done

-- 20160329: email to Steve to TA; then back to DDWG for review / discussion ?

(Not Discussed)

-- 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-100: Remove Array 2D and Array 3D from File Area. (T.King)

-- Open; Under_DDWG+Review: 20150201

-- 20150519: sent email to C.Isbell asking for input (since E is not available)

-- 20150602: sent email to C.Isbell asking for input (since E is not available)

-- 20150604: C.Isbell entered IMG comment -- waiting for Steve TA

-- 20150609: TA'd by Steve with recommendation to withdraw SCR

-- 20151007: J.Padams requested to table this until after Insight; Maybe end of November/early December?

-- 20160315: J.Padams added comment to JIRA that IMG wants to retain Array 2D and Array 3D

-- S.Hughes updated TA; sent email to Todd requesting that PPI withdraw SCR

-- 20160322: Emails between Simpson & Todd to re-word SR

-- 20160420: Sent to CCB for review

(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-132: Units of Map Scale Improperly includes pixel/deg as a unit (J. Padams)

-- 20151007: Open

-- 20151007: Email to Jordan to provide explicit changes to IM

-- 20151008: I updated SCR to include specific changes required; ready for Steve to TA

-- 20151012: TA'd; email to Emily and Dick to review

-- 20151013: Email from Jordan to pull back for further discussion / work

-- 20151022: Jordan to finalize new & improved proposal before next DDWG

-- 20151104: Jordan updated SCR as comment in JIRA

-- 20151105: DDWG agreed to send to CCB

-- 20151116: Needs TA; then ready for CCB per DDWG

-- 20151117: TA'd by Steve; Emily reviewed; Dick sent email with concerns

-- 20151118: Set up telecon to discuss Dick's concerns

-- 20151119: DDWG discussion; send to CCB if no comments

-- 20151123: Needs TA; then ready for CCB per DDWG

-- 20151202: Emily and Dick reviewed -- Ready

-- 20151208: CCB e-vote; rejected

-- M.Showalter proposed name changes:

Units of Map Pixel Resolution

Units of Map Pixel Scale

-- CCB will have telecon to discuss

-- 20151222: CCB sent back to WG / DDWG to re-work

-- 20160204: MC on 2016-02-04, Jordan et al violently agreed on a workable solution

-- 20160428: updates discussed by DDWG

(Discussed-will be sent to CCB)

-- 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

(Discussed - on hold for a week)

-- CCB-143: Validate field format via regex (Lev Nagdimunov)

-- 20160210: Open & Under DDWG review

-- 20160322: EN governance; will take lead

(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

(Not 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

(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

(Not 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

(Discussed - work group formed)

-- 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'

(Not Discussed)

-- CCB-156: Inconsistent Discipline Dictionary Technique for Local Internal Reference, et al. (A.Raugh)

-- 20160418: Open

(Not Discussed)

-- CCB-157: Remove sampling parameters attribute from Uniformly Sampled. (L. Nagdimunov)

-- issue from CCB-128:

-- email from Lev requesting IM be changed before v1600 released

-- to make the following change:

-- propose everything be left alone, except sampling_parameters be removed.

-- how to document that EN made the change (outside of the scope of the original SCR)?

-- 20160418: Ready; sent to CCB

-- 20160419: Queued for Implementation

-- E-vote PASSED: 6 Yes (ATM, GEO, IMG, IPDA, PPI, RMS),1 Failed to vote (SBN)

(Not Discussed)

-- CCB-158: Restore J2C as valid format for supplemental data. (S.Slavney)

-- 20160426: Open & Under DDWG review

-- 20160428: updates discussed by DDWG

(Discussed)

-- CCB-159: Bug fixes for Version 1.7.0.0. (J.Hughes)

-- 20160426: Open & Under DDWG review

(Not Discussed)

) Topics for Discussion

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

-- Status & develop implementation plan

(Discussed)

-- 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 - April 27, 2016

There are two enclosures:

1. a list of the full topics under discussion by the DDWG 2. a presentation for CCB-142: Create Data Quality Flags to hold metadata on Quality Flags (E.Shaya)

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

-- CCB-142: Create Data Quality Flags to hold metadata on Quality Flags (E.Shaya)

-- see enclosure for a proposal / explanation (R.Simpson)

-- email from R.Simpson: PDS4 Data Design Working Group - Telecon Announcement - Mar 23, 2016 - 9:30AM PDT

Please review the following SCRs (including the TA) and be prepared (in advance of the DDWG) to discuss and possibly vote:

-- CCB-132: Units of Map Scale Improperly includes pixel/deg as a unit (J. Padams)

-- CCB-158: Restore J2C as valid format for supplemental data. (S.Slavney)

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

-- CCB-143: Validate field format via regex (Lev Nagdimunov)

-- email from L. Nagdimunov: field format, IM and SR (20160426 2:02pm)

DDWG Telecon

The short notice agenda cleverly forgot that Steve will be giving a status update on enumerated list. Will do that at the end.

People are having trouble joining the meeting - will go out of order to accommodate that.

Hopefully, everyone has reviewed all of the SCRs for today.

CCB-158 - Restore J2C as valid format for supplemental data See https://pds-jira.jpl.nasa.gov/browse/CCB-158

Hopefully, this is a no brainer.

GEO got a PDART proposal that would add a raster image and a supplemental JPEG2000 and to an existing dataset. They were going to label the raster image as primary and the JPEG2000 would be ancillary. This was fine in 1500, but doesn't work in 1600. Tried to figure out what happened. Seems there was some clean up done for enumerated lists as bug fixes. J2C was taken out of the list of acceptable formats for supplemental data.

This SCR is asking that it be put back in.

Question: Any questions?
Answer: (Silence)
Another Question: Is there any issue or any objections to sending this to the CCB for a vote? No objections? Does everyone endorse this?
Answer: (Silence)

A late comer has arrived on the call. Has objection. That CCB was turned in to make the IM consistent with PDS4 policy. Does not endorse.

~ Everyone else endorsed. It's a list.

~ The model drives what goes on the website. It's compressed and not a format that the MC agreed to. Would need to look at two years worth of minutes to find the discussion. Policy on acceptable formats.

Someone could find no evidence that management said to take J2C off that list.

Question: If there's a raster image, why not make a transformation? Could do it on the fly and then we would not have to worry about this compressed format which may not be legitimate...?

Answer: Someone is confused. JPEG and other formats are compressed.

~ PDF is not necessarily compressed.

~ Someone is trying to understand the list.

Question: These are supplemental - not archive format?

Answer: Supplemental still goes in the archive.

~ We still have to take care of them.

~ They just have to be documented.

Question: is that true - we have to accept if something is documented?

~ That would solve a lot of radio science problems.

The policy says supplemental can be in other formats as long as certain conditions are met. Has the list. It includes J2C. It's well laid out.

Objection is withdrawn.

Question: Does everyone endorse sending this SCR to the CCB?

Answer: Put RS down as abstain.

CCB-132 - Units of Map Scale is Badly Defined See https://pds-jira.jpl.nasa.gov/browse/CCB-132

Worked all avenues that needed to. There were comments regarding removing the Cartography DD - but left in the solution, even though it's not part of it.

Question: Any issues with sending this to the CCB? Any objections?
Answered with a question to Jordan: Are you happy with this?
Answer: Yes.

Cartography DD is in a shaded area, so that's okay. Someone has no objections.
~ So there are no objections.

Someone else is not sure why we are talking about this.
~ Because it was supposed to be a PDS DD. Then it became part of the Cartography DD. That was decided at the last F2F. This documents that.
~ No opinion. If the experts are happy with it, I'm happy.

##CCB-142 - Create Data Quality Flags to hold metadata on Quality Flags See https://pds-jira.jpl.nasa.gov/browse/CCB-142

Question: Is Ed Shaya here?
Answer: (Silence)

Imagen a 2D image array, 1024 x 1024 - maybe there are bad pixels in it - you might want to flag those, so people would know. It's possible they would be in all images or only in some. So, what's being proposed to go with this, along with 1024 x 1024 array would be one with flags. If the image array is floats, flag array is probably integers. Could have different problems with different flags. Everyone should see the summary that Ed Shaya sent out (See attachment to telecon announcement, April 27,

2016, 160323_explain.docx)

There is software to deal with problems somewhere. We could set up a similar system.

(Missed something - bit masks... something about populating a data quality array...)

Shaya likes integer values:
1 for Reed-Solomon problem
2 for fill value
3 if both issues

Software would then ask user which ones matter to them.

Basically, array with data values and a parallel array.

The attachment has an example. Think of data quality array as a filter - could come out with image filtered on a pixel by pixel basis.

The rest is what would need to be changes in order to implement this. It's sort of complicated and we don't have the software yet. We could just say the product is the data array and have the second array and tell users what they mean. Don't have anything like this now, but could do it.

Shaya proposes taking object statistics - putting this there and renaming. The question is if we want to pursue this. Would need a work group with image experience.

We can do this now because the product could have multiple array - one with flags. We can do this now, but there might be advantages to having this. It might be worth pursuing.
~ We might want to adopt what space telescopes are using, but it might not be specific to our images. We could put all explanations in description or table.

Someone thinks we can do it now - can put second array, quality array. Deep Impact did it. Advantage - could allow software to work with it. That doesn't mean we need it. In terms of flags, the best implementation would have something in description. Could specify whatever flag values you want and describe and ask users what they want to apply.
~ Someone likes this. Regarding flags, likes some separate structure where data providers get to say what the flags mean. No reason we couldn't put a table in the label itself as part of the data quality array. All of the ideas are very good.

We did something similar with arrays that describe the axes of data arrays. Ability to define logical relationships - created new objects to describe logical relationships between arrays. Were told to put in discipline DD. It might make sense to do something similar here. Seems these are for image data.
~ Not similar. In Deep Impact it was data, not metadata, so needed a standard method for documenting data quality flags. Not sure why would do this anywhere other then in the label. See it in several places - tabulated spectra etc. Data quality values are in the data. This is to document each position in flags.
~ Not two completely different things. Had data array and array that gave the errors - but used a local DD to describe the relationships. This information ended up in the label too.

One potential implementation - have array in data and one that describes data - has extra class and flags.

Seems like people are interested in this. Suggest we do a prototype to really look at.

Question: Who would make the prototype?
Answer and another question: Could use Ed Shaya's work and Steve could do the IM part. Needs help with the table of descriptors. Should that be separate file or in the label?
Answer: Should be in label.
~ That doesn't preclude description also being in something like a SIS document. Multiple places where people can read it.

~ **Action Item - Steve**

A data type of quality flag that is required - in flag field substructure - tag the info so parser can get the info out. Could write a program for users to actually be able to work with it.

~ Becomes packed data fields - we're just not calling it that.

~ Pretty much. It's a question of if we think it's worth the effort to support this very specialized data type.

~ Someone doesn't agree that it's like packed data.

~ Think of it as a special case of packed data.

~ A special, extremely restricted case.

Someone thinks we need a work group to move forward.

Question: Do we expect to have enough of this data to justify work on this?

Answer from IMG : Not familiar enough with our packed data sets to answer.

~ Ed Shaya must have thought this would be important to some extent.

Ed Shaya may not know that we are meeting now. Will put this on hold - need to decide if it's worth doing the work.

On hold until next week. Need to invite Ed Shaya.

Other nodes should decide if the bang is worth the buck on this.

CCB-143 - Validate field format via regex See https://pds-jira.jpl.nasa.gov/browse/CCB-143

This goes with CCB-152 (field format definition mismatch between IM and SR - see https://pds-jira.jpl.nasa.gov/browse/CCB-152)

The general idea is that field format is unclear.

Issues:

1- SR has a page of limitations and IM contradicts what's allowed. Probably an over site.

2- Lack of clarity on how to use field format.

If I have a real 10.423 - when I print I can want 10.5 or all of the significant digits. Not sure if this is for validation or to print data a different way. Not sure if there's been any agreement. SR says it can be used to validate, IM says something slightly different. Both imply data as written, not print formatted.

Someone is used to a fairly flexible interpretation. Has a FORTRAN background. Not sure what the consequences are if we tighten this us.

~ If tightened up - can be used for validation - could be very useful.

~ Can see the point on the width of the data, concerned about decimal precision.

Another revealed thing - if you have a real and want to display an integer, could get an error. If we want to require data type output is the same, seems the implication is that's already required, but it's not stated clearly.

Was hoping for a vote, but sees a disconnect between the proposed solution and the TA. The TA suggested a pattern.

~ The TA was written before this conversation. There are three issues: if a pattern is needed, synchronization of SR and IM needed, and need agreement on how this applies. Can ignore the TA for now - needs update.

There are two SCRs here. SR specifies pattern in label, but doesn't say if decimal point. Not settled by the pattern.

Question: Are we ready to vote or not so much?

Answer: There's a series of issues.

1- Syntax validation - should be a no brainer.

2- Should it be validated against the data type of field - a schematron rule?

3- Should format statements be required and used to validate data objects?

Didn't specify required in CCB. First question is if it should be used for validation or to format data.

~ SBN has been requiring format statements for years. Use them to find data errors. Never regretted it. Strongly recommends it for validation.

Question: Vote for CCB-143 as written? Good issues. Not 152. Need to add additional stuff to SCR for that. Talk about it next week...?

Answered with a question: So vote to send CCB-143 as written to CCB?

Another Question: As written - adding to schematron or pattern in model?

Answer: Patterns in schematron.

Another Question: I see pattern, not see validation against the type of field?

Answer: The question is validate or hope the data provider puts it in and it's right.

Question: Are people happy?

Answer: Yes.

Question: Any objections to sending this to the CCB?

Answer: RS is abstaining. Sounds like a temporary change that might be overwritten by another SCR.

~ We could table this until we discuss the other one.

Let's send it out.

Question: If we send it, will it be automatically approved because it's a schematron change?

Answer: RS is still abstaining.

Moving on...

CCB-154 - Promote Mission Information class to Discipline Governance Level See https://pds-jira.jpl.nasa.gov/browse/CCB-154

We discussed this a few weeks ago when some of us were on vacation. There was a lot of discussion.

This was a proposal for a mission level class being promoted to a higher level. The point was that people recognized that it was probably good to define at the common level - things like mission phase name, spacecraft clock start/stop times. Times are specific to mission - would be in the mission DD, but class would be in common. The question is discipline level versus mission level.

Someone thought we had a policy that disciplines and missions handle stuff at their level.

This doesn't make sense. Each mission has it's own definitions, different syntax. It can't be used for system search.
~ Someone agrees with all of that. There is no general consensus that this is useful anywhere other than the mission level.

Having the same names for search terms would be useful.
~ Could do that - could tell disciplines and missions some terms they need to use - could even enforce that if it makes sense to uniformly impose that - with schematron. Same way we're enforcing the query model. We could enforce it, but sceptical that it's as universal as is being claimed.

Could be namespace mission with very generic descriptions. This is a new wrinkle. Mission level, but want some governance over names used at a higher level.

We should decide first if we want to do this. Implementation should be decided second.
~ Fair enough. Some missions would have terms in all labels - others would put in bundle XML only - not

sure important it is to try to enforce this. Also worries about non-applicables. Micromanaging gets very complicated.

Question: Is the SCR mandating these be used or just standardizing the element names if used?
Answer: Depends on how you read the proposed solution. The problem statement doesn't state a problem that can be solved.

To clarify, when Santa provided this, mission class seemed to list things across many areas. Wouldn't take proposed solution as written. This has been proposed and is worth discussing.

It's nice if you're working with multiple missions to know what names to use, but need to be careful, and missions will have their own needs.

These were all optional. It was if you do use mission phase name, this is how you spell it.
~ Should have an enumerated list so you can search. I wonder if maybe we can address the problem by having certain attributes and classes created by the base LDD file.
~ That's implementation. The question is if this is useful.
~ Can see the utility, but implementation has to be appropriate of the utility is undermined - or the core DD is undermined.

This was proposed at the discipline level with higher level governance. Not a single entity is concerned with governance - another wrinkle.

The question is how you ensure things are properly defined if they are included automatically.
~ We have terminological entry - can capture alternatives that way.
~ That was for internationalization.
~ Was a PDS guidance - could be used for synonyms.

We're running out of time.

Question: Is it appropriate for a work group to be formed?

Answer: Work group - Steve, Anne, Jordan and Joe.

~ More of a discussion group - don't see an obvious answer.

~ Steve gets to be the chair.

~ Santa should be involved.

~ No. The first issue is to get it to work with PDS system. Need to design, then ask her after some ideas.

~ That works.

Enumerated list/Instrument Type Status See email from Steve Hughes - DDWG Exercise - Instrument Classification - Proof-of-Concept - April 28, 2016 9:21 AM - and attached spreadsheet

We agreed last time to a three week exercise for proof of concept to make a hierarchical type of list. Steve sent his part out today. Extracted 581 distinct instruments, 466 by instrument name only. Eleven discriminators - a new one proposed - only six being used now. Need to decide the what if situations for discriminators. Had input from GEO, IMG and radio science and radar, so about half have values assigned.

Instructions - Please don't add rows. More than one value can go in a cell. Go through the list - identify the ones you are familiar with. Look through the discriminators - choose the ones that work - add what you need.

IMG proposed acquisition method.

A lot of questions regarding specificity. Included more information, but looking for best set of discriminators right now.

Question: What output do you foresee? What will instrument type look like?

Answer: Good question. Will send something out this afternoon.

~ Asking because if the end result is useful, great, but if all this is is an exercise in sorting it may not be worth our time.

~ Steve will send the file this afternoon **(Action Item)**. Bin names don't matter yet. Astrophysics has done something similar - using concatenation from bin names.

Question: Does that help?

Answer: Can see where it might go. Not convinced yet, but willing to try.

~ It may not work. Don't know. Don't want people to spend too much time. This is a proof of concept.

Question: Eleven columns and six discriminators?

Answer: Only using six discriminators to create bins at this point.

~ When looked at the test case, was sometimes confused by the discriminators. Inferred detected versus actual detected.

~ The notes point out that you might infer something based on something you detected.

~ Confusing. In Row 3 examples, would have categorized things different. The definitions are kinda fuzzy. People may interpret different - not sure how that affects the end result.

~ This is computer science to classify - not an exact science. Will be firmed up. Looking for help from the nodes.

Only a third of the list are SBN instruments. SBN won't look up other instruments - no time.

~ This is just to see if this exercise has merit.

~ The time involved is a valuable discriminator too. Should also influence how we handle instrument type.

~ It's easier with new instruments - older ones require more time and work.

Don't touch anything you don't know. Ignore them. Don't do a lot of research. This is for proof of concept.

Someone is worried this will be built on mission instruments and ignore ground based.

IMG did the exercise. Took an hour or two. Discussed it with USGS. Did a brief walk through - focused on discriminators - walked through it - did it ff the top of their head.

~ If someone's data were all images they would expect it to take less time.

~ GEO has many instrument types - will take a shot. See where it leads us. Will put in about an hour on the test case.

~ More than an hour is too much - it's a test case.

Question: Is there a way you want us to note updates?

Answer: Yes. IMG and GEO both updated one - this is being done in a database. People get notified.

~ So highlight the text in some color.

~

Question: Any issues for next week?

Answer: SETI issues next week.

Another Question: Any SCRs? Will add 142 and 152 to the list.

Question: Anything else?

Answer: (Silence)