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

Sent: Wednesday, April 27, 2016 9:22 AM

To: pds4ddwg

Subject:Notes from DDWG 2016 04 21

---

title: DDWG Notes 2016-04-21

layout: default

date: 2016-04-21

April 21, 2016

Notes by Debra Kazden

Known Attendees:

R. Chen, M. Gordon, E. Guinness, S. Hardman, L. Huber, S. Hughes, C. Isbell, R. Joyner, D. Kazden, T. King,

J. Mafi, 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

-- 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

   (Not Discussed)

2) Task Statuses - 5 Minutes each

(No Updates)

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

(Discussed - Voted not to endorse the SCR)

-- 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

(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

(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)

(Not Discussed)

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

    -- 20160210: Open & Under DDWG review

    -- 20160322: EN governance; will take lead

    (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

    (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

(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'

(Not Discussed)

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

-- 20160418: Open

(Not Discussed)

4) 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 19, 2016

There are two enclosures:

1. a list of the full Agenda

2. Enumerated Lists – Pros & Cons (L.Huber)

This week's agenda will focus on two topics:

1. CCB-100: Remove Array 2D and Array 3D from File Area

CCB-100 was submitted more than 15 months ago. It would remove Array 2D and Array 3D as possible choices in File Area, effectively making them 'abstract'.

Data providers could then choose either Array for 'generic' arrays or subclasses such as Array {2,3}D Image, Array {2,3}D Map, etc. when the array has specialized content. CCB-100 has been discussed by DDWG previously but the results were inconclusive.

Please read the CCB-100 entry in JIRA (including the TA), be prepared for a 5 minute discussion, and then be ready to vote.

2. Enumerated Lists

-- Please review the enclosure and the email from S.Hughes to the DDWG on 4/15 that provides comments

-- Please be prepared in advance of the DDWG to discuss

-- The Plan is to have group consensus on an implementation so that an SCR can be written

Should time permit, a discussion on the SETI issue(s) will be next...

DDWG Telecon

Today we will mainly focus on the ever popular enumerated list.

CCB-100 - Remove Array 2D and Array 3D from File Area See https://pds-jira.jpl.nasa.gov/browse/CCB-100

We will spend five minutes on this.

Todd sent email with the SCR history. Would like to suggest some of that be discounted. Wants people to look at the SCR as written. Changes are out of scope to the SCR as written. Liked the email.

Didn't think, given the evolution of the IM and SR, that array 2D or 3D add any value. The real value is

when there's a higher order - more specialized - that's what SCR 100 was trying to address.

~ The original thinking was at least if you had a 2D array - there might be generic things you could try on it. Not sure we gain anything with array 2D, except maybe programming short cuts.

The other thing, in 9F ... (Interrupted) ~ Stop. That's outside CCB-100.

~ Fair enough.

There has been a lot of email discussion between two parties - see two design paths - descriptive and prescriptive. For this SCR, we have to consider this. If we want to be less prescriptive, could move 9F into DPH. The point is, really a common sense way of looking - most specific - when to allow array - then question when to use array or something more specific. Would have to tell provider to use descriptive or prescriptive. 9F is common sense.

Descriptive versus prescriptive, not necessarily more prescriptive. With images we have subclasses for resolution etc.

Conflating this. There's no disagreement that array 2D is more prescriptive than array. It's two approaches. Least prescriptive versus most. We shouldn't be jerry rigging array - this is confusing the issue.

~ Someone disagrees. Array 2D and 3D are unnecessary. That's what the SCR is about.

Time to take a vote. Vote to endorse SCR to remove array 2D or 3D or not endorse.

~ This is to remove then from file area, not the model. Abstract versus concrete classes.

A vote to endorse is to remove 2D and 3D array from file area - it will still be in the model.

** The Vote:

ATMOS - not endorse

EN   - not endorse

IMG   - not endorse

GEO   - abstain - not sure what problem we're solving

PPI   - endorse

NAIF - not present

SBN   - not endorse

RS   - not endorse

RINGS - not endorse**

So, having concluded that, on to enumerated list...

Enumerated Lists

See EnumeratedListsB.PPTX attached to meeting agenda - sent April 19,2016

Hope everyone has already looked at the PowerPoint. Will summarize key points - hopes we have a way forward to solve this keyword, which is currently broken. We need a solution.

One reason for enumerated list is for search tools, also validation, also restriction on choices.

Characteristics of a good enumerated list - it's complete, all values have the same level of specificity, no overlap, orthogonal and additions are rare.

PowerPoint includes examples on slides 4 and 5.

In slide 8, "How do we build the optimal enumerated list?", regarding additions to the list - new values - we need a better template for adding. Node should set limit for specificity and be able to justify that it is the right value.

What's happening is that values are going to CCB - and they ask provider over node recommendation.

Sees this as a very bad practice.

We need a way forward. Three options - from slide 10:

"Three suggestions:

1 - Create a mulch-tiered instrument.type with a small enumerated list at the top level which has matching

  specificity levels and a schematron controlled lower level to provide flexibility and ease incompleteness worries.

2 - Set aside the current enumerated list and start over requiring the nodes to provide the components that would

  serve their community.

3 - Make the list unenumerated. Trust the nodes to instruct providers what proper values should be."

The third option would be backwards compatible and go along with Google type search.

There are currently 54 values. Some can be gotten rid of, but several are missing. If we do the math there will be 60 to 80 values if we keep this level of specificity.

Question: Of the three, which do you prefer?

Answer: Option three - unenumerated - because we have no idea what's coming in the future. We know of two values that are not yet included - Seismometer on InSight and MOXIE on Mars 2020. They are not in the list yet. Not sure what we would call MOXIE. An unenumerated list will solve the problem of having to go to the CCB for a standard value.

In PDS3 things were unenumerated and everyone threw their values on the list. Option 3 - worried it might end up like PDS3.

Question: So, to use the tree type example, what's the mechanism? What would be the control to keep one group from putting Doug Fir versus another putting Douglas Fir?

Answer: Nodes have to approve.

Another Question: But what's the control without enumeration?

Another issue is that this requires reasonable communication between the nodes. They archive different instruments.

~ True. Unlikely.

~ Missing agreement on what level of distinction we want. For magnetometer, for example, there are fluxgate and vector helium etc., but we only need magnetometer on the list.

~ True. Question is what is right level of specificity.

There are very few simple instrument types. You could have an imager or a spectrometer or an imaging spectrometer.

~ Could have multiple values.

~ Someone is torn.

Someone's boss suggested the tree approach - could start with a small set and move to a controlled list.

Years ago, read a book on Murphy's Law. Gordon's Law is that "If a job is not worth doing at all, it is not worth (doing) well". Concerned. A single enumerated list or a tree type is going to take up a lot of time of the DDWG. It's not trivial to clean this up. Want to make sure there's enough value to do this. Not enumerated is no work. Keeps coming back to the question - why do this?
Non-enumerated is used for search, classification hierarchy fundamental to the model - not sure that fits. So, have to decide what is the real reason to have enumerated values. Would like Sean to answer. Wants to know if we had a good enumerated list if Sean would use it.

Someone needs to interrupt - needs to correct something - Sean has been using instrument type to map to a list he has.

Sean says he cares deeply.

Currently, high level search does use instrument type - uses the PDS3 types now... (Interrupted) ~ This is for a context type for context object - not for labels.

~ Yeah. So, search at top level, then can filter out to more specific - that's how we do it now. Can search by instrument type - more advanced. Would include PDS3 or PDS4. Not sure anyone is using it. Would get context object and associations. Love enumerated lists - they are the greatest thing in the world, but feels the pain of keeping the list maintained. Free string text is useful, but not as useful. Will still get close, get top 90 percent in search. Intrigued by the hierarchy, but doesn't want something as complicated as Primary Results Summary.

Question: If we do enumerated list - imaging spectrometer?
Answer: Don't care how we tag it.

Question: Are we using context products? Metadata consistency team's query model uses metadata from context products and has been implemented.

Someone wants to bring up a new point - if enumerated list, the process for getting new values in could take six months to a year depending on where we are in the build cycle. It could be very important for missions or individual providers. If they are developing an archive without a value they need they have to wait for the value to be added. SBN had an issue with BOPPS instrument.

~ Yes, SBN had to postdate the dataset.

~ Didn't realize this was an issue. At some point the CCB said all enumerated lists are common. Evolved a process where all enumerated lists are treated the same. We should work the process to make this easier.

The real problem is that the IM only comes out twice a year. Maybe we need incremental releases for trivial changes. Not major, but little updates as necessary to populate enumerated values.

~ Makes sense. We don't want PDS3 problems in PDS4. Could come up with something to allow more frequent changes.

First we need a good baseline list. We need a list that doesn't need so many additions.

~ That doesn't exist. Due to the nature of space exploration the list will change regularly.

If we assign instruments to nodes there's no ambiguity. We can do this. There are pretty basic instrument types. Geo-science instrument, atmospheres instrument, etc.

Question: Each node would have a list?

Answer: No. The first level of instrument type could be type of node. Geo-science instrument, atmospheres instrument.

~ The exercise was to find the top level of type list that wouldn't change. This is a common classification. We focus on implementation and what's intuitive. We could have a small list - stable over time. Until we try an analytic approach we won't know if it will work. We need to identify the discriminators. Figuring it out is the process - not a different task.

~ This reminds someone of the taxonomy hierarchy of journals. Keywords - controlled vocabulary - most general to more specific. Structure collapsed to string of controlled terms.

Discriminators - we need to figure out which bin things fit in. Top level is defined by discriminators and decision tree. Very analytic and formal.

Question: How many values would you expect to arrive at?

Answer: Starting with about 20 discriminators, some are binary - formal way to make decisions.

Approximately 500 unique instruments in PDS3 - been able to classify about half.

Another Question: So the answer is a couple hundred? If we try to make an enumerated list based on an analytic tool - how many?

Answer: Depends on how many we want. Depends on the level of specificity. Not worried about implementation yet.

To try to summarize: came in leaning towards unenumerated cause the weight of effort versus return,

but that might not be true. Should explore a way forward. Need to expand the work group. See what's being put together - also need to consider incremental builds for the IM. This is two separate issues.

Question: Is this a reasonable summary?

Answer: Somewhat, but still concerned. Still 25 to 200 values. Maintaining that is probably not where we want to go. Probably we all kind of know where the number is at the common level of specificity.

Someone is open to this. Need to do some of the work to figure out if it's worth doing. Can make a decision later - need more information on what we get from starting over.

Question: Do we think this is easier than what we currently have?

Answer: Yes. Have been doing it intuitively, no formal way - decision tree changes that. Not sure how many values are at the top level. Could specialize or not. Nodes would decide. Can control some of it. If we go by the nodes there are just 8.

Another Question: Decision tree leads to enumerated, not values - it's discriminators. Are we trying to decide about the decision tree - what will sort it all out?

Answer: Creates a set of bins to put instruments in.

There's a fundamental flaw here - if you assume all instruments fall in one bin.

Question: Is the naked eye instrument in PDS3 electronic or not? Decide then determine what will be arriving in the future.

Answer: Concerned. We have instruments now that don't fit in a single bin.

~ The short list could be very simple - active versus passive.

Question: How close to being able to show us the discrimination tree are you?

Answer: Has one now - could send in a few days.

~ A proof of concept iteration.

~ Yes, could see if convergence.

Someone wants to get back to things fitting in more than one bin. Also, unsure what active versus passive or insitu versus remote mean. Plus, there could be more we might want to use.

Definitely would allow multi-valued. Prism on LRO is an example of one that would be multi-valued.

Single instruments would go in two or more bins - multiple values in actual label. Each thing goes in a bin.

We're hypothetically discussing this. It's hard to understand. It's worth our time to look at the work that has been done before continuing.

Someone has seen it - has difficulty imagining a user would search for a passive instrument. Thinks the primary consideration should be how the users would describe the type. If a user is searching for data - how they would describe the instrument. That's how they would design the taxonomy.
~ Good point, but how do you know what bin - might not be a part of search. Bin might be say remote, passive, or imager and be a visual imager. Bins could be useful. Nodes could add specification at a lower level.
~ Bins aren't working for everyone. This is not a pigeon hole process. Thinks of facets, descriptives, independent properties, so looking for a descriptive keyword field with some constraints - the various ways to describe these things then would collect the strings of descriptors. Bin metaphor doesn't work.

Question: SBN does classifications all the time, how do you do that?
Answer from SBN: Good example. We have several classification schemes that contribute terms to what they are classifying.
~ Visualizing - so a root - at bottom is bins, categories based on discriminators. Maybe more descriptive the the values.
~ Feels like tagging, not binning.
~ Someone thinks that hit the nail on the head - would like not to have what we have - something semi-

enumerated, small list that can be added to as appropriate.

We need some way to determine the categories.

~ At some point, whatever the values we put, needs to be useful to us and users.

~ Names don't matter yet - need to decide how many categories.

We should look at something concrete before we continue. Steve H and Dick have started something, we should look at that.

This is time critical. Suggests an action item to Steve - in one week to provide the proof of concept to the DDWG and then we can go over it in two weeks.

~ Steve can get it out, but needs expertise to help him out. Could use the node dictator or could send to all nodes. Probably needs 3 or 4 weeks. It can't be done in one week.

~ That's too long.

Question: Are you going to work full time on this?

Answered with another question: Is it worth the time to solve this or go another way?

~ Warning - the work is based entirely on the name of the instruments, so just a proof of concept. Will need to start over. Keep that in mind. Also, some things just don't fit in any category.

The proof of concept - just go through the list. Don't worry about anything you don't know about. The nodes could get a week and then Steve could take a week to compile.

So, something could go to the nodes in a week - they could have a week to respond and then it could take a week to compile.

~ Depends on how many instruments belong to a node.

~ The first list had 1500 instruments, it was taken down to 480.

~ Someone is not sure which is better.

~ Probably the short list because this is still a proof of concept.

So, in three weeks we will have something to look at.

Action Item Steve will send proof of concept to the nodes, they will respond in a week, and in three weeks the DDWG will look at it.

Question: Extra telecon?

Answer: We should have weekly telecons to address everything on the agenda.

~ Someone isn't available next week. Plus, next week isn't enough time for this.

~ There's a list of SCRs we can talk about.

~ Yes, and the IPDA list of changes.

~

Question: So, one more question - have had some differences of opinion on who is in charge of this meeting...?

Answer: Ron sets the agenda - not sure who's in charge. There's a process for SCRs - not an in charge person.

Question from management: Thought Steve was chairman. Is that incorrect?

Answer: Dan said the DDWG is an EN advisory group. If Steve is the chair he can hand off responsibility.

Another Question: Are we sure we want Ron to have that power?

Answer: The authority of managing the agenda was delegated to Ron.

~ When Ron was asked who sets the agenda - he said it wasn't him.

Action Item Steve, Ron and Dan need to sit down and decide.

Someone doesn't want authority.

This agenda came from ATMOS.

This was civilized meeting.

If people want something on the agenda - email Ron. This worked out well.

We will meet next week.

Question: Anything else?

Answer: (Silence)