

From: Debra Kazden <dkazden@igpp.ucla.edu>  
Sent: Wednesday, January 20, 2016 11:14 AM  
To: pds4ddwg  
Subject:Fwd: Notes from PDS DDWG 2016 01 14

---

title: DDWG Notes 2016-01-14

layout: default

date: 2016-01-14

---

#January 14, 2016

Notes by Debra Kazden

Known Attendees:

M. Gordon, E. Guinness, S. Hardman, L. Huber, S. Hughes, C. Isbell, D. Kazden, T. King, J. Mafi, S. McLaughlin, A. Raugh, R. Simpson and J. Stone

##Meeting Agenda and Summary

1) CCB/SCR Statuses/Issues

-- CCB-66 - Refine value domain definitions for ASCII Numeric Base2 and ASCII Numeric Base8 data types

(A.Raugh)

-- 20160107: E-Vote PASSED: 6 Yes (ATM, CIS, GEO, IPDA, PPI, SBN)

-- 20160107: Queued for Implementation

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

2) Task Statuses - 5 Minutes each

\*\*(Focused on Product-Update, quick Geometry update)\*\*

3) SCR 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

\*\* (No progress yet) \*\*

-- CCB-72: Need Pixel Scale Units (A.Raugh)

-- URGENT - bug fix

-- Open: needs "Requested Changes" before can be TA'd by Steve

-- under DDWG discussion

-- 20140930 from A.Raugh: This one didn't come up at DDWG.

Someone who knows what appropriate pixel scale units are should do this - that's not me.

-- 20150730: Created WG - Mitch, Chris, Jordan & RJ

-- 20150917: Jordan provided proposed solution / definition --- back to DDWG for discussion

-- 20150922: DDWG discussion & agreement on proposed solution

-- 20150922: Steve to TA

-- 20151021: Back to DDWG for discussion & agreement

-- 20151022: Back to DDWG -- need Anne to weigh in

-- 20151118: Under discussion by WG (Jordan, Chris, Ron)

-- 20151207: agreement in WG (Jordan, Chris, Ron, & Dick) -- either withdraw or supersede

\*\* (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

**\*\*(Product-Update was 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: Setup 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

**\*\*(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-139: Resolve Standards Inconsistency in Bundle 'Readme' File Format (R.Simpson)

-- 20151207: Open; under DDWG review

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

-- CCB-140: SR Section 6E.3 levies unnecessary requirements for attribute values (J.Padams)

-- 20151229: Open;

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

-- CCB-141: ASCII Numeric Base16 is inconsistent with xs:hexBinary definition (J.Padams)

-- 20151229: Open;

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

-- CCB-142: Bit Mask should be moved into a new object called Data Quality Flags (E.Shaya)

-- 20151229: Open;

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

#### 4) Topics for Discussion

-- SETI Issues (R.Simpson etal)

-- Status

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

-- The Stream Text parsing standard id (Steve, Simpson)

-- Status

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

-- Rules for Collections of type Data (Steve)

-- schematron rules

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

##DDWG Telecon

Ron is on vacation. Steve is running the meeting.

Build 6B is coming at the end of March. On February 26, five weeks prior to the build, need to release the schema for testing. So, by February 26 need to identify the final set of SCRs for that build. February 26 is the deadline.

Question: So things need to be approved by the CCB by the 26th of February?

Answer: Exactly.

Question: What is the official release date of 1.6?

Answer: March 31. Will freeze 1.6 and put it in the release directory.

Question: So we can use it for real deliveries and it will include all that has been approved up to now?

Answer: Not necessarily. Some approved things haven't made it in yet. On February 26, 1.6 will go to the development directory for testing.

**\*\*Action Item - Steve\*\*** will identify SCRs that have been approved and figure out what will be included.

Maybe SRC 77.

##CCB-129 - Implement PDS4 Query Models See <https://pds-jira.jpl.nasa.gov/browse/CCB-129>

This all started with CCB-61 (Metadata for Data Collections - see <https://pds-jira.jpl.nasa.gov/browse/CCB-61> ). Collections didn't have the required metadata for search - attributes needed for search weren't included. The metadata consistency team was set up to address the issue. They had lots of discussion and came up with a query model that uses property maps.

See email from Steve, January 13, 2016, "CCB-129 - Implement PDS4 Query Models"

It was suggested that the DDWG would need to see the schematron rules for this - that was included in yesterday's email - a prototype of the schematron rule based on query model.

Would like to include this in the common model. Believes that schematron rule is necessary to assure that the necessary metadata is present. If the DDWG agrees, would like to send this to the CCB for approval for 1.6. There is still work to be done at EN. Will implement a prototype for testing the set of attributes. Maybe there could be an email vote over the next week.

Question: How will this be applied? What does this mean for validation or the label preamble?

Answer: If it's in the common model then it becomes part of the validation.

Another Question: So it's transparent to the user?

Answer: True.

~ Would be part of the core schematron file.

Question: Does this change any other schematron rules?

Answer: Have to implement it to check.

~ That won't be enough to check it all - very logically complex test.

We currently do apply rules in schematron validation.

~ We don't test every possible path.

~ Someone has confidence that the schematron works correctly.

~ Not worried about the schematron, worried about the rules. They sometimes don't fire cause of things like typos.

~ We need to test.

Someone is worried that when we check a label, like a MAVEN label, against the new criteria that the other rules might not fire.

~ That's not the issue. The question is do we have the right query model and attributes. Most nodes were involved in this. Want everyone to agree. There are 26 attributes. These rules do not include everything. We need final confirmation for this now. As far as testing, yes, it's a very hard problem too test all schematron rules. We need to begin testing and come up with a plan.

Someone thinks it will be impossible to avoid rule contention. A divide and conquer approach might be the best way - if these validated separately then the rules couldn't be hidden - wouldn't have to worry about collisions.

~ Good Idea.

The code that is used combines rules that would have stepped over each other - rules with similar context - have tried to address that.

We don't want to not see a problem where there is one.

Question: Are we ready for prime time or do we need more discussion?

Answer: Concerned that schematron rules will make lots of currently optional things required.

~ Yes, that will be necessary. That's what CCB-61 asked for.

~ So, the concern is that we need to evaluate the impact of this change - need a list of what was optional and would be required. There might not be appropriate values in enumerated lists for all the newly required. This may lead to several necessary fixes.

~ Yes, exactly. This forces the issue. We discussed this years ago. Not sure everyone wanted primary results summary.

Wavelength has to go away. It can't be required everywhere. There are places where that makes no sense.

This is for mission science data collections. The tiger team figured out this list of attributes.

~ "I was on the tiger team and I never would have required wavelength."

We need all of the nodes to scrub this list. Maybe we can say that some things should throw a warning.

The wavelength comment is excellent. Maybe we could make an alternative that something is there - wavelength, temperature - whatever is appropriate.

Maybe we need something in the SR to let people know that some of these things really aren't optional. Like product-update.

~ This will be very hard to do.

~ Someone would like to look at this. We're not providing as much information as we should.

~ Then document editors will need more help with documents.

The issue of alternatives can be looked into, but we also need to look at if the constraints on a collection are appropriate. The question is if sub-setting is correct. If wavelength isn't correct, need to know how to subset for when it is required.

As to documentation of the schematron rules - they have text - can put it anywhere we need it. Can make it more accessible to users. Some rules are already documented in the IM.

~ Someone would like to see them and to look at this to see if we got the rule correct. Feeling confused, in target type the lines in the block are identical - wants to understand why. (See email from Steve, January 13, 2016, "CCB-129 - Implement PDS4 Query Models") ~ Someone else noticed that too.

Wonders if it was a cut/paste thing or one of those things where the intricacies are beyond us.

That's an error. Query model - what the team agreed on - is attached to the email. Everyone should look at that attached excel spreadsheet to determine if those are the attributes we want to validate against. We can then make the schematron rule consistent with it.

Question: Is system:archive status the only one that isn't included?

Answer: Primary results summary science facets not there.



Want each node to review the query model as represented in the spreadsheet. **\*\*(Action Item)\*\*** ~ SBN can't do it in the next week.

Question: Can the rest of the nodes review it? Don't want to implement something and then change fifty percent of it.

Another Question: This list is really intended for discover-ability for search, right?

Answer: Yes.

Another Question: So, if not all of them are present in the label, will it be left blank in the search engine?

Answer: Blank fields wouldn't be used.

The issue is what is required in a label versus what is necessary for search.

~ The issue came from Mike Martin. People weren't including metadata that they could have.

Rules can fire an error or a warning. The question is when a collection is valid. If you get an error in the validation then you can't proceed - the collection fails - so maybe we should use a warning.

~ We don't want errors that want attributes that make no sense, like wavelength.

~ Right. That's why we want people to see what we want and do require - what we want to require for reasonable searches.

Question: So, if wavelength shouldn't be required, but many products should have it, but we don't require it - how do we make sure it's there?

Answer: Can do sub-setting for collections that you want to have a certain attribute in.

~ Could have separate schema/schematron for system for discover-ability for each node.

~ Maybe we need a tool to check that schematron rules are being followed for each type of data - by node, mission, etc. Might be faster.

~ Separate schematron allows configuration tied to each system.

Question: Doesn't that violate multi level governance? Isn't there a minimal set of metadata for all

collections? That's what we're asking - what's the minimal set that's required. The only thing that's required now is identifiers. We can subset by type, like data collection, so we can say the minimal attributes required for a data collection are... (Interrupted) Another Question: Can subset so that you include what's applicable, but how do you identify what to use to begin with?

Answer: Several types of data are problems.

~ But we could have groupings.

~ We're going back three years.

~ We could just require primary results summary and that it has to be valid.

But there was a problem with early deliveries where things were optional so they weren't included. We wanted to make those things required.

~ We could just trust the nodes to do the right thing.

Someone wants to see a prototype - even if we decide not to use this. We need to test how much this solves.

Suggestion that we all need to test this - then come back to discuss it again. **\*\*(Action Item?)\*\***

There's a lot of flexibility. It's fine to let nodes do their own thing, but we're at the level of throwing the baby out with the bathwater. Have to have some required attributes for search.

Primary results summary. We can solve this by saying primary results has to be present and correct.

~ Someone likes this, but still wants feedback. Wants to know the required set.

Question: This is for mission science data sets?

Answer: To give an out for things that are not missions.

~ None of SBN's ground based data fits this.

We can use primary results summary for this. CCB-129 was to provide concept for search. Once we have

the concept down and have the tools we can design a query model for other types of data.

~ If this works it solves problems for a lot of the archive.

This has been a great - long - discussion.

Question: To clarify on us reviewing the spreadsheet and schematron... Looking at attribute wavelength, if we don't think it should be included we should recommend that it comes out?

Answer: Yes.

Another Question: There are products with no relevant start or stop time... (Interrupted)

Answer: Data would have a start or stop time.

~ There are ways of including things.

~ That's not how people would search.

People want to use natural search parameters to find their data.

**\*\*Action Item to everyone\*\*** - do your best to identify what you think should be required for search. We will discuss this again.

##Versions - Mitch

No SCR has been written yet, but the SR doesn't contain anything that tells when a version number has to be changed.

Mitch sent email on this topic. (See email from Jan. 12, 2016, 1:55 p.m., "Re: FW: PDS4 Data Design Working Group - Telecon Announcement - Jan 14, 2016 - 9:30AM PDT -- updated agenda")

Question: Does anyone have any comments?

Answer: "Draft" needs to be a different word. Also, nodes wanted to do versioning differently. A black hole.

~ All that's being said is that if we make any change, have to change the version number. Don't want products that aren't identical to have the same version.

EN tracks the version every time something things come in.

~ That's an internal version number for when things come in.

~ VID in label - EN has no control over that.

~ But they track things as they come in.

~ We shouldn't be relying on EN for the version.

~ That's tracking what's in the registry - that's not PDS documentation.

We have had two products with identical LIDVIDS with different checksums.

~ Then that's an error and EN should throw it back.

~ NSSDC too.

Nothing in the SR says when a version should be changed.

~ This has been discussed at NSSDC a lot. This needs to be addressed.

~ The SR needs to be clear on this.

Question: Regarding things with the same version number, do you have an actual example? Do you know that this has occurred?

Answer: Ron (who is on vacation) has said it happened.

~ It's a valid concern, but doesn't believe it was released. The registry should bounce it.

Someone was surprised that the SR doesn't have a prohibition on this.

**\*\*Action Item - Mitch\*\*** will write an SCR to address this. Doesn't expect any impact on the IM or system, other than changing the word "draft" to "preliminary."

Anything that gets to the registry has to have a unique LIDVID.

People will read the SCR and send comments.

Mitch will put it in jira on Monday. If people have comments, let Mitch know.

Question: When is the next CCB meeting?

Answer: (Silence)

In the SR, we say throughout that all products have a unique LIDVID, so unsure why this isn't known.

~ Want to make it explicit.

~ Not sure a paragraph in the SR will make much difference.

##Product Update - Mitch

Inclined to request that CCB-77 be superseded by a new one. CCB-77 has been around so long - it's not all relevant.

See <https://pds-jira.jpl.nasa.gov/browse/CCB-77> - Augment Product Update with File Area Update

Question: Objections?

Answer: (Silence)

~ "Silence is golden."

One thing we said, that we would figure out how to do product update and then decide if we would do it. We need to think that through. Will ask for feedback in a week.

There was a question about how likely the system was able to handle this and use it.

~ Would get into search, but the possibility of recreating labels on the fly will not happen. EN will not rewrite the labels on the fly to add the metadata to the labels.

Someone doesn't think users will be happy with this.

At Rings, users get metadata for entire volumes of images of rings - it's simple to extract metadata from a simple table.

~ That's fine for that, but not for tools or readers. If there's a list of ten thousand changes for a product, not sure readers would handle that.

Someone envisions search at the nodes - nodes might not have the tools to deal with this.

This wouldn't affect Rings. They would just put it in search database pipeline. Thought it would be easy for other nodes too, rather than requiring providers to re-submit labels.

Product update advantage was to make it easier to update labels.

Not looking for an answer today. Mitch would like feedback over the course of the week so that if he does need to submit something to the CCB, that he has time to. **\*\*(Action Item)\*\*** ~ Definitely wants to put something in for LIDVID, an SCR.

Question: Product update for MAVEN?

Answer: PPI would rather see as a policy that they redeliver their labels. Understand how this would work for Rings, but not seeing any use for it at PPI.

If this is really to update geometry, maybe we need a separate product.

~ It's not just for geometry. It would also be used for labels delivered with the wrong information. Could use it to include supplemental information that was not originally in the label.

Looking to have an email discussion on this.

Question: Did LADEE have a geometry update?

Answer: ATMOS is expecting it, but the mission is over. Millions of labels. Will be more efficient for them to provide a product update.

~ It might be useful as input to a tool to update the labels. Not comfortable just adding it to the collection. Product update could be an internal class.

Question: How would we keep that information going forward?

Answer: Apply it to the labels. Different for primary versus secondary products.

~ Two uses - update labels or add supplemental content.

Rings gets very limited index tables for New Horizons LORRI data then generates updated expanded indexes. Users like them.

Question: What level of support is PDS going to provide for this? How will it merge the information for the users?

Answer: PDS could update the labels, but users won't object to more information.

ATMOS is unsure of the time line for the new LADEE updates.

~ All columns in the update table have to be defined, even if it's not a label update.

##Geometry - Mitch

Geometry DD final revision should be out next week. It includes changes to phase angle and a few other straight forward changes.

Should be made available in the release directory.

Good progress.

---

Question: Other priority issues?

Another Question: When is the next DDWG? We hardly touched the agenda today.

Answer: Two weeks.

Another Question: Was there something we never covered?

Answer: Was a lot we keep pushing back - will never finish.

Suggestion - How about if each node sends a list of priorities? We could have a telecon next week.

We should get everyone to set priorities for what we need for the next build. Going every two weeks we have five DDWDs left - we could add an extra.

We have a lot of work to get things to the CCB in time for the release.

Will suggest an extra meeting to Ron.

Many things on the agenda have been there for months. An extra meeting won't help more than people actually working on these things.

SBN can't be on another telecon next week.

~ But there are things we can discuss without SBN.

**\*\*Action Item - Steve\*\*** will email Ron about adding another telecon next week.

Question: Anything else?

Answer: (Silence)