title: DDWG Notes 2016-12-01
layout: default
date: 2016-12-01

December 01, 2016
Notes by Debra Kazden

Known Attendees:
R. Alanis, M. Gordon, E. Guinness, S. Hardman, L. Huber, S. Hughes, C. Isbell, R. Joyner, D. Kazden, J. Mafi, L. Nagdimunov, L. Neckrase, A. Raugh, R. Simpson and J. Stone

(The full meeting agenda attachment that usually accompanies the meeting announcement was not included this week.)

Notice sent before the telecon in email from R. Joyner - November 29, 2016

CCB/SCR Statuses:
-- CCB-166: Deprecate bit mask from IM for new sample bits attribute (J.Padams)
    -- 20161122: Queued for Implmentation:
    -- Item PASSED: 6 Yes (ATM, GEO, IMG, PPI, RMS, SBN); 1 Failed to vote (IPDA)

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

(1) Please review and be prepared to discuss:
-- CCB-131: Missing constraint on Special Constants attributes (A.Raugh)
    -- 20150922: Open; under DDWG discussion
    -- 20160310: until someone volunteers to lead the effort -- on hold
    -- 20160929: back to DDWG to form WG; candidate for sunset clause
    -- 20161010: from Anne: Jess Stone (jstone@psi.edu) will be the SBN rep for this
      -- need participants to form WG
    -- 20161013: WG formed: J.Stone (Chair); Lev & me (participants)
      sent email to Dick & Jordan to ask if they will participate on WG
    -- 20161026: sent email to Dick to see if discussions between Dick and J.Stone are solid enough to update
       SCR
    -- 20161026: Dick to update SCR; will split off NaN & INF; to be addressed in separate SCR -- see CCB-170
    -- 20161109: Dick has made extensive updates to SCR; needs Steve to TA
    -- 20161109: At DDWG, Dick dropped out of WG; Lev/Anne to define competing implementation(s); send to
       DDWG
    (Brief Discussion)
-- CCB-170: Deprecate IEEE 754 NaN and Inf in Favor of PDS4 Special Constants (R.Simpson)
    -- 20161110: Open & Under DDWG review
    (Discussed - decided to resolve CCB-131 first)
-- CCB-167: Remove xmlns:pds recommendation from DPH (L.Nagdimunov)
    -- 20160906: Discussed with Sean & Steve; Sean s/w team will ascertain impacts on removing 'pds' as
       namespace from XML labels

-- on leave until end of Sept.
    -- 20161110: Sean ascertained impact on the EN tools
    -- 20161110: Open & Under DDWG review
    (Not Disscussed)

(2) Email exchanges and other stuff

-- PDS4 equivalent for the PDS3 keyword SOURCE PRODUCT ID (S.Slavney / J.Padams)
    -- 20161129: Open; under DDWG discussion; form WG
    (Briefly Discussed - will discuss again)
-- ASCII REAL (L.Nagdimunov)
    -- 20161012: Requested Lev provide cogent statement of the issue to be resolved; send to DDWG to start
      email exchange & discussion
      -- lots of email discussion; inconclusive as to whether issue is converging or diverging ?
    -- 20161026: Can a problem statement be written into an SCR where there is some measure of certainty that
      this is the issue to be resolved ?
       -- at least one email questioning whether this issue is worth the time / effort to discuss / resolve
    -- 20161109: Lev requested another week to work issue; before additional DDWG discussion
    (Discussed- will form a work group next time)
(3) Product_Update (M.Gordon)
       -- 20161109: Mitch sent out Product Update email; send to DDWG to start email exchange & discussion
      (Discussed - will continue to discuss)

DDWG Telecon

WebEx finally seems to be working!

We will jump around some today.

Officially, Anne is on vacation- even though she is here. We are expecting Sean today.

CCB/SCR Statuses

CCB-166 - Went to CCB - Passed - Good

CCB-131 - Missing Constraints on Special Constants Attributes See https://pds-jira.jpl.nasa.gov/browse/CCB-131

Where we left off last time was that RS dropped out and SBN took on the task of defining implementations and were going to send them to the DDWG.
~ SBN were to write up all of the possibilities, pick a favorite and send email. Didn't send email yet. Not totally certain which is their favorite yet. Should put this on hold for now - it's still in progress.
~ Work is progressing.
~ It may be more than two weeks. Maybe a month. Will try for sooner.

(Sean has joined the call)

Product-Update

There was a request for Sean's input on this.

The way the first round went was that there were two examples. Speaker is fine with separate product that provides updated new information in a single file. Other way was to have label indicate there were updates with a keyword attribute. There is concern that that has limited utility. Would like to know if harvester can replace values in search.
~ If there's a product update out there, it would be used to update fields or add data for new fields - maybe not at harvest - would be in search index. Label wouldn't be rewritten - maybe that would be a future requirement.

Question to Sean: Would you need product update to tell you there are better values?
Answer: Unclear if knowing matters to the software. It will just march through.
Another Question: So when the software marches through and decides what gets overwritten it's worried about slots?
Answer: Yes. Slots, fields.
Another Question: Is there a standard list of slot names?
Answer: Yes. Sean has it.
Another Question: Is it known to the nodes so they can all use the same names?
Answer: Not yet.
~ Nodes need it to prevent ambiguity. X-paths would be completely explicit.
~ That's what's in the example. Tool will replace the x-path. Concept of property maps is a good place to capture the mapping.
~ That's not what someone would have guessed.
~ The idea came up later in the process - it's still being worked on.

Question: So it won't be a munged x-path? Have some very complicated data - concerned about the configuration file. Looking for a way to cut it down. It's kind of a secondary issue. Whatever the solution is can be discussed later.
Answer: Sean is aware of the issue and will give it some thought.

Regarding the examples, no real distinction between new metadata or updating labels, but new metadata should be mapped back.

Question: Say I'm providing metadata that's not included in a label - no real x-path - how do I tell EN? Mechanically, at Rings, they make multiple entries for a single product - each row a target. Is that going to complicate things?
Answer: Certainly. Need to see a good example of that. Not sure off top of head how we're handling single, distinct issues.

Question to Sean: Do you think product update is a good thing to work on for the model?
Answer: In simple form, yes. Some things in the write up already exist. ATMOS already needed to support new metadata. Most interested in supporting our current use cases.

Another pass could be taken at the document, to take out the more complicated part.
~ Not sure how useful that is.
~ Maybe move to extras so we don't forget we discussed it.
~ Good idea. Rings will also try to better describe the file they use for the opus database.

Question: Does anyone want to weigh in on having a product update?
Answer: Someone votes for new labels. Thinks this will be a real mess.
~ Some products give a huge hook to the index table.
Another Question: Why not just regenerate the index table?
Answer: Basically we are, but with more information. A separate product - but updated with a new pointer so no real update.
~ We're getting caught up in English here. It could be that we are adding better metadata. We do it in PDS3 and could do it in PDS4 with product update.

Not everyone is on the same page here.Not everyone understands the different concerns.

There will be a discussion outside this telecon by Mitch, Sean, Dick and Lev.

Will discuss again next time.

Question: Anything else on this?
Answer: (Silence)

ASCII Real

We left this with the idea of creating an unbounded ASCII Real type.
~ Discussed mostly by email. There's still a problem for software for reals - to see if they are bound or not. Need to check every file - very slow process. Best solution someone could find was to split ASCII REAL into bound and unbound. Ninety nine percent would be bound. No ambiguity. Software will know how to handle them. An SCR was submitted on this. CCB-171. It outlines the proposed changes. (See CCB-171 - Split hardware-compatible ASCII numeric types from the unbounded - https://pds-jira.jpl.nasa.gov/browse/CCB-171)

Question: Is all that in the SCR?
Answer: Yes. If you want more included in the SCR let Lev know or let him know if there's a better way.
~ Someone read the SCR as a complaint that ASCII REAL is unbounded. Left a comment in jira. Didn't see a proposal to redefine the data types.
~ It's there. Starts with complaint then makes proposal for changes- lists everything for this to work.
~ Someone one will read it again.

Someone is not sure if they want to weigh in, but agrees that the original understanding was ASCII REAL unbounded, but had some problems. Question is that there's an understanding and then there's the programmatic validation. So, based on how much data will have this issue - wonders if it's worth having this discussion. Such a small amount of data that wouldn't pass programmatic validation. Not sure it's really worth the time and effort of this group.
~ It's not just validation. It's also the reading of the data. Looking at other standards that have this issues - this is important to handle. Difficult and important. We need to address this.
~ Microsoft employs hundreds of people - if they can't solve this in Excel we shouldn't waste time on it.

~ If we can't handle it we shouldn't allow it.
~ We are an archive. Excel can be thrown away.
~ Not everyone understands.
~ Schema validation isn't good enough.
~ Thought we relied on XML schema for that reason.
~ It's not good at range validation.
~ XML schema is supposed to be doing it.
~ It might be worth testing the PDS canonical validator to see if it catches it.
~ We need a validator that works. Maybe we can add constraints.
~ That's why the SCR was written. It's not super different.

Question: Anyone else?
Answer: Someone doesn't want to walk into the swamp.
Another Question: What is Sean's opinion?
Answer from Sean: No opinion yet. Assumes at some point we may need to add something to validation. Waiting to see what we decide. Will write the necessary code.

Question: Anyone else?
Answer: (Silence)

Action Item - Everyone look at CCB-171. Next time we will form a work group.

Someone is looking now and and has a simple request - would like a few sentences added about when you would use an unbounded versus bounded. When you would use them and how you would choose. This is in the detailed weeds, but unclear what to use.

Question: For the SR?
Answer: Yes. Wants to know when to use them.
Another Question: Is it clear to everyone else?
Another Question: If you use an ASCII REAL unbounded - what does the system do with it?
~ That's a different issue. Just wants some clarity.
~ It wasn't included. Author thought it was self evident. Not exactly sure where the description would go. As to what we would do with it - you can store as numbers, as big decimals to work with these. They are numbers. It's a way to handle them.
~ Someone would like that that in the SCR. Doesn't want DPs wondering what to use.
~ Author will think about where to put it. Will ask people.
~ Not concerned about the SR - want it clear to the DDWG first. Yes, we want it clear to everyone, but in the SCR first, everyone else later.
~ It could go in the SR.

Author will try to add something. Hopes it's clear to the DDWG why you would want to separate these. The point is that software will handle them different. Need to treat them specially. Will try to edit the SCR.

Big decimals were mentioned. Someone has used them in java, someone else is a FORTRAN guy, another someone likes Python. It's hard to understand how to use big decimals in different languages.
~ Many libraries in FORTRAN and others that will know how to use them. The issue is that if we can't handle them we shouldn't allow them.

Question: So, the proposal is that we want everyone to read CCB-171 and we will form a work group in two weeks?
Answer: AGU -PPI meeting is on Thursday.
Another Question: That takes us to December 22, but I would assume no one is going to be around. December 29?
Answer: Nope. People are closed.
Another Question: January 5?
Answer: Yes.
~ Cool. Yay. We can take this as a blessing and an early Christmas present.

PDS4 equivalent for the PDS3 keyword SOURCE PRODUCT ID

This was added at the request of IMG and GEO. They need an equivalent for source key.

Someone looked at it and thought reference list was supposed to handle that. Assumes they want this for HiRISE.
~ No, that's not it. This is a derived product where other products are archived in PDS3. Been emailing about it. GEO decided to use external reference. Don't want to see this topic go away, but can wait until January. This will come up for others who have derived products.
~ Someone agrees.
~ We might need a different solution to more easily refer to PDS3 products.
~ We will discuss this in January.

CCB-170 - Deprecate IEEE 754 NaN and Inf in Favor of PDS4 Special Constants See https://pds-jira.jpl.nasa.gov/browse/CCB-170

This followed from CCB-131 (Missing Constraints on Special Constants Attributes - see https://pds-jira.jpl.nasa.gov/browse/CCB-131). Appears they are in some data sets. The proposal is to use special constants. No proposal for new attributes, but could consider some - we could create some. Someone would define his own objects. That frees us from IEEE754 stuff. Really for programmers. Shouldn't be passing that on to science users. Not a number might be good enough.
~ That just tells you it's not a number.
~ Someone thinks the IEEE754 values are useless to our community.
~ That's a mischaracterization of the point. We can tell it's not a number after the first bit or two. The point is that it's a class of numbers for processing environments we would treat as garbage. Could be useful for binary. IEEE754 is a well defined external standard that we chose.

Someone says for binary they have gotten ASCII - tells them no and they need to make a special constant.
~ Yes. Not allowed for ASCII table.
~ NaN specification not allowed for ASCII real.

Question: Is this a two parter?
~ Confusion.
~ The SCR is just trying to make it unacceptable in PDS across the board.

For binary integers it's not a problem. Could be used for ASCII.

~ Someone has DPs who use IDL and they have to tell them they can't use NaN. Have to make special constants.
~ Someone agrees that the that we are just trying to make it unacceptable in PDS.
~ Someone else is confused because we aren't allowed to do it now. It's syntactically invalid. For binary you have to convert to something you can check.
~ Maybe we need to add a sentence to the SCR.
~ Someone agrees, but a separate issue is if NaN should be allowed for binary formats.
~ We could go either way on that. Programmers may want it. That is okay because it's part of the accepted standard, but if it's possible it will get more subtle over the years and may need something more explicit. Fine, but this is only for binary data.
~ Someone basically agrees, but NaN is part of a proposed solution for special constants.

Someone would like to wait to hear from programmers.

Question: Does anyone else want to weigh in?
Answer: (Silence)

Suggestion: Let's wait until we resolve CCB-131.
~ Yeah, and since the explicit constant values follow a pattern in PDS3- if we split them down the middle- we should go with that. It has worked for 25 years.
~ Could do that in PDS3.
~ Right. Think the default solution is that they are not allowed.

So we will wait on CCB-131.

_____

We will wait to discuss CCB-167. That will probably be a long discussion.

Next meeting January 5, 2017