CDF Review & Responses

J. Mafi & T. King

PDS Management Council Face-to-Face
Los Angeles, CA
February 5, 2016

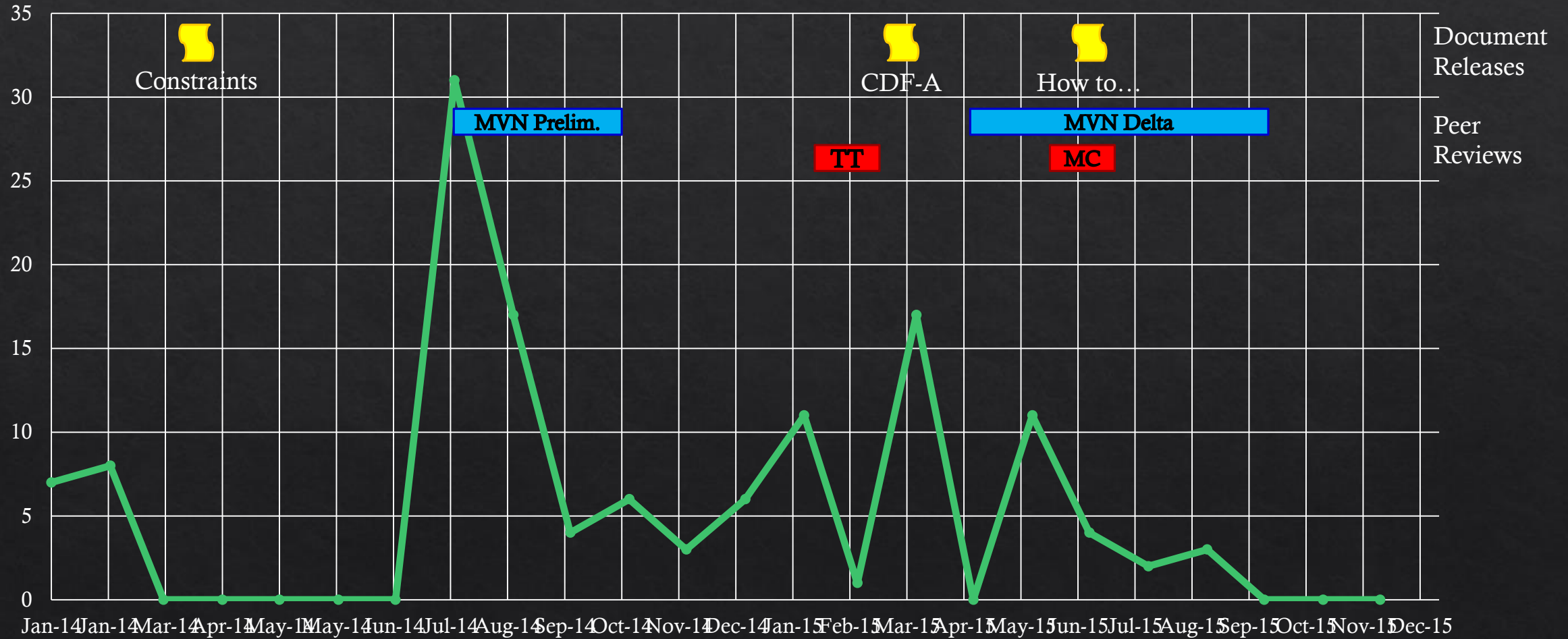# Overview

◈ PDS4 CDF archiving timeline

◈ Comment Overview & Statistics

◈ Select Issues

PDS4 CDF Archiving Timeline

- 22 Jan 2010   MAVEN initial contact (MAVEN PSG, Boulder)
- 07 May 2013   PDS4 V1.0.0.0
- 10 Jun 2013   MC Telecon: CDF to be described using PDS4 Array objects
- 26 Jul 2013   PDS4 CDF White Paper
- 18 Nov 2013   MAVEN launch
- 13 Mar 2014   MAVEN PDS4 archive CDF constraints document
- Jul-Sep 2014   MAVEN preliminary peer reviews
- Jan-Feb 2015   PDS CDF Tiger Team Review
- 26 Feb 2015   CDF-A Specification document
- Apr-Sep 2015   MAVEN delta peer reviews
- May-Jun 2015   PDSMC MAVEN CDF review (2015-04-21/02)
- Jun-Jul 2015   MAVEN Release #1
- 8 Jun 2015   How to Guide for Reading PDS4 Labeled Array Data

Comment Timeline

Comment Count

CDF Review Commenters

- MAVEN Archive Peer Reviews
  - MAVEN instrument/project personnel
  - Science data users
  - NASA Space Physics Data Facility (SPDF)
- PDS Internal Review
  - DDWG
  - PDS MC
  - CDF Tiger Team

CDF Review Comment Overview

Type	Context
Existential	CDF Tiger Team
Archive Documentation	MAVEN Peer Reviews PDS CDF Documentation Reviews PDSMC CDF Review
CDF Metadata	MAVEN Peer Reviews
CDF Data	MAVEN Peer Reviews
PDS4 Labels	MAVEN Peer Reviews PDSMC CDF Review
PDS4 Tool Support	CDF Tiger Team MAVEN Peer Reviews PDSMC CDF Review

CDF Review Comment Status Statistics

Type	TOTAL	Addressed	Open	Closed	Superseded
Existential	5	0	0	5	0
Archive Documentation	36	25	7	3	1
CDF Metadata	34	29	1	4	0
CDF Data	5	5	0	0	0
PDS4 Labels	29	23	1	5	0
PDS4 Tool Support	3	0	3	0	0
Unclassified	21	10	3	6	2
TOTAL	133	82	8	16	3

MAVEN Review Comment – Row #41

Context	Subject	Source	Date	Reviewer	Status
MAVEN Preliminary Reviews	PDS4 Labels	LPW Prelim. Review	8/12/2014	Martin	Addressed
Comment	There is a lot more product metadata in the CDF labels than is provided in the PDS4 labels. If we are really using PDS4 to create archive products that can be used hundreds of years from now then all the metadata buried in the CDF labels needs to be exposed in the PDS4 labels. Otherwise, this is not a PDS4 archive, it is a CDF archive. Possibly the detailed metadata could be extracted from the CDF's and put into some kind of PDS4 supplementary table.				
Response	Expanded content of PDS4 labels to include all relevant CDF metadata.				

MAVEN Review Comment – Row #42

Context	Subject	Source	Date	Reviewer	Status
MAVEN Preliminary Reviews	PDS4 Labels	LPW Prelim. Review	8/12/2014	Martin	Addressed
Comment	Regarding all the array products. From the labels there is really no way to know how the array components relate to each other. Each array could be completely independent. It seems like there should be some kind of explicit association that indicates that the several 1d array elements correspond (time[0] goes with density[0]).				
Response	Developed Discipline_Area.Particle_Observation class to provide associations between array objects.				

MAVEN Review Comment – Row #64

Context	Subject	Source	Date	Reviewer	Status
CDF Constraints Document	PDS4 Labels	email	9/16/2014	Simpson	Open
Comment	Only CDF 3.4 HEADER objects are allowed (not 3.5); if you don't think that's important, I'm willing to let it go.				
Response					

MAVEN Review Comment – Row #84

Context	Subject	Source	Date	Reviewer	Status
MAVEN Preliminary Reviews	Archive Documentation	email	2/5/2015	Wilson	Addressed
Comment	CDF software requires a CDF leapseconds file (not the same as an SPICE LSK kernel), the format of which is specified by SPDF. To insure accurate results data users must use the same CDF leapseconds file that was used by the data provider. Availability of current leapseconds files is dependent upon SPDF.				
Response	MAVEN used SPICE LSK's for their leapsecond calculations. LSK used is identified in both the PDS4 label and CDF metadata.				

MAVEN Review Comment – Row #88

Context	Subject	Source	Date	Reviewer	Status
MAVEN Preliminary Reviews	Archive Documentation	email	2/5/2015	Wilson	Open
Comment	Have the order of multidimensional data have the same dimensional order in both Matlab and IDL. Really this is a row-major/column-major issue – I think… I can't test to confirm this one. When I peer-reviewed the Maven CDF data using Matlab my dimensions were in the reverse order to those listed in the SIS (seemed they used IDL) - was that a typo or a row/column issue? We believed a row/column issue in my chosen reader (Matlab) of CDF files at the time.				
Response	This issue needs to be included in the archive documentation.				

CDF Tiger Team Finding – Row #89

Context	Subject	Source	Date	Reviewer	Status
CDF Tiger Team	CDF Tiger Team Report	2/9/2015	CDF Tiger Team	Addressed
Comment	MAVEN archives should include CDF files with PDS4 labels as PPI has designed and that conform to the constraints that PPI has defined (PPI white paper How To Create PDS4 Compatible File in the CDF Format, rev. 2014-10-15).				
Response	This approach has been followed for all of the MAVEN CDF data sets.				

CDF Tiger Team Finding – Row #90

Context	Subject	Source	Date	Reviewer	Status
CDF Tiger Team	PDS4 Labels	CDF Tiger Team Report	2/9/2015	CDF Tiger Team	Addressed
Comment	The PDS4 labels should be debugged and improved (e.g., to better define relationships between arrays).				
Response	PDS4 label debugging was accomplished by means of the MAVEN Delta Peer Reviews, and PDSMC MAVEN CDF Review. Improvements included the creation of the Particle_Observation, and Parameter objects				

CDF Tiger Team Finding – Row #91

Context	Subject	Source	Date	Reviewer	Status
CDF Tiger Team	CDF Tiger Team Report	2/9/2015	CDF Tiger Team	Closed
Comment	There was not a consensus on whether MAVEN CDF files should be converted to another format that would be the primary archive product, making the CDF files a supplemental product.				
Response	The approach taken with the MAVEN archive has been to used the CDF data as the primary archive, describing them as binary array objects.				

CDF Tiger Team Finding – Row #92

Context	Subject	Source	Date	Reviewer	Status
CDF Tiger Team	Existential	CDF Tiger Team Report	2/9/2015	CDF Tiger Team	Closed
Comment	Even if the CDF file can be described by a PDS4 label, it would still violate the basic "simplicity principle" of PDS4.				
Response					

MAVEN Review Comment – Row #104

Context	Subject	Source	Date	Reviewer	Status
MAVEN Delta Reviews	Archive Documentation	SWIA Delta Review	4/29/2015	Martin	Addressed
Comment	It would be nice to have a layperson description of the significance of the different data collections. They all look about the same to me in the autoplot displays.				
Response	Sections were added to the SWIA and SWEA SIS documents to address this recommendation.				

PDSMC CDF Review – Row #120

Context	Subject	Source	Date	Reviewer	Status
PDSMC CDF Review	PDS4 Labels	email	6/9/2015	Gordon	Addressed
Comment	1D-1 field arrays should not be identified in the labels (assuming my understanding is correct, and these are all single valued for the entire file). For someone using the XML label and software other than CDF-A, that information is given as an attribute in the labels. Rather than list those arrays in the XML label, consider them as part of the intervening embedded headers.				
Response	"Single-element array" values have been added to the PDS4 labels as metadata and are not defined as array objects within the data files.				

PDSMC CDF Review – Row #121

Context	Subject	Source	Date	Reviewer	Status
PDSMC CDF Review	PDS4 Labels	email	6/29/2015	Gordon	Addressed
Comment	<name>dindex</name> <description>Deflection Index for CDF compatibility</description> There must be a better way to describe this array. You should be describing it in the XML label for non CDF users. The array has values [1,2,3,4]; it is used as an array axis for multiple arrays, and I still do not see an association between it's values and something more substantial, like deflection angle values.				
Response	The PDS4 labels for these products have been modified to use the appropriate multi-dimensional array for describing the data array axes.				

Context	Subject	Source	Date	Reviewer	Status
PDSMC CDF Review	PDS4 Labels	email	6/29/2015	Gordon	Addressed
Comment	<Array> <name>theta_coarse</name> This array has two axes, of lengths 4 and 48, both with axis names of "index". Same arguments as above; for a non-CDF user trying to understand the data, this appears unnecessarily obscure.				
Response	The value "index" was used in cases where "the parameter is itself an independent variable". However, since the axis variable references have been moved out of the File_Area_Observational and into the Discipline_Area, the approach is now to simply use a descriptive value for the axis_name.				

PDSMC CDF Review – Row #123

Context	Subject	Source	Date	Reviewer	Status
PDSMC CDF Review	Archive Documentation	email	6/29/2015	Gordon	Addressed
Comment	Generally if you have a 3D array, you need values for three axes, for a 4D array, values for four axes would seem to be sufficient. So why the "...and, possibly, a face plane"? ...the overview [says] the primary data is in"... 3 Dimensional arrays with the axis of the look directions (Phi and Theta) and energy level..." with time as a fourth dimension. However it appears that Theta is not a axis of the 4D arrays, it as a face of the 4D arrays which really have axes of Phi, dindex, Energy, and Time.				
Response	The SWIA data arrays are 4-D arrays, with axes: energy, 2 angles (theta, phi), and time. However, the theta angle is itself dependent upon energy, making it a 2-D array which aligns with the energy-theta plane of the data array.				