

TSSD-22423
October 28, 2013

To: William P. Knopf, NASA Planetary Data System Program Executive

From: Brian P. Duncan, JHU/APL, PDS4 ORR Chair

Subject: Minutes of the PDS4 Operational Readiness Review, held
September 17-19, 2013

Attachments: (1) PDS4 ORR Review Panel Members
(2) Requests for Action (RFAs)

An Operational Readiness Review (ORR) for PDS4 was held September 17-19, 2013 at the NASA Goddard Space Flight Center. The purpose of the ORR was to assess the readiness of a limited release of PDS4 for data providers for the LADEE and MAVEN missions.

The PDS4 team provided two days of very good, detailed presentations. PDS history was presented to complement the new development for PDS4, which was quite helpful given the mixed level of PDS experience on the review panel.

The Review Panel commended the PDS4 team for several engineering decisions that helped drive the implementation towards operational readiness for LADEE and MAVEN:

1. The excellent engineering decision to design and manage the PDS Information Model separately from the technical model (i.e. implementation architecture). This insulates the Information Model, which will persist long-term, from the implementation (and underlying technology), which will evolve over time.
2. The design decision to use an XML-based approach, even if it means some degree of education for the user and development community is necessary. The Information Model is well-designed and demonstrates the benefit of applying a strong system engineering process. It is refreshing to see model-based engineering applied to the Information Model development. The early benefits have been demonstrated by leveraging COTS tools such as Oxygen, and by the ease with which modifications to the model can be applied and artifacts generated. Long term, the XML-based approach will facilitate compatibility with international users of the PDS data.
3. The selection of an iterative/spiral software development methodology. This keeps the community engaged and provides ongoing feedback into the implementation process. Testing has included the nodes and missions, which helps the users become familiar with the system, as well as benefiting the project by finding issues as early as possible.
4. The detailed and thorough documentation of the software architecture documentation, providing a variety of perspectives into the system design.

PDS4 will be long-lived and the team will need consistently high quality regardless of changes in development personnel for the life of the system. To achieve this, a level of formality is needed for the PDS development cycle (e.g. documented and repeatable) due to the size of the team, the distributed nature of the discipline nodes (and their relationship with the engineering node), and the disparate needs of the community. Review panel members who supported previous PDS reviews noted the increased rigor applied to the software development life cycle by the PDS4 team.

However, the panel identified three areas that warrant additional work, most notably in the areas of requirements management and the development of test artifacts. These include:

1. Using PDS4 build 4a to more rigorously document system testing, and clearly trace test results to the requirement level (PDS4ORR-RFA1)
2. Clarify the relationship between ICDs and SISs and any interdependencies (PDS4ORR-RFA2)
3. Ensuring that delivered bundles include all information required to understand and interpret the data (PDS4ORR-RFA3)

In addition, there are opportunities to clean-up and clarify aspects of the PDS4 documentation. These are called out in detail in the recommendations in the “Requests For Action” section.

Finally, the live “demo” of the system was informative. It was a significant milestone to have PDS4 components in place now, proving the approach is successful for supporting data formats for PDS3 and PDS4 during the upcoming transition period. It is terrific to see a more consistent look-and-feel across web pages, the user experience, and making the searches less dependent on knowing (in advance) which node to use.

The PDS4 ORR was a short review but a focused and productive one. We would like to commend the PDS team on a truly excellent piece of system and software engineering, and recognize that you have figured out how to successfully navigate and manage a potentially very difficult distributed and diverse community.

The outcome of the PDS4 ORR is a small list of actions for the team to address to fully demonstrate the operational readiness of PDS4 limited release (PDS4ORR-RFA1, PDS4ORR-RFA2, and PDS4ORR-RFA3). Upon resolution of these items, we are confident that the PDS4 will be capable of successfully supporting both LADEE and MAVEN.

Attachment 1

Review Panel Members

Brian P. Duncan, JHU/APL, Chair
PJ Clark, JHU/APL
Dave Heather, ESA/ESAC
Dave Linick, JPL (via telecon)
Tom McGlynn, GSFC
Lauren McIver, Virginia Bioinformatics Institute (via telecon)
Marilyn Newhouse, MSFC/CSC
Peter Shames, JPL

Attachment 2

This report uses the term “lien” to indicate a significant finding levied against the PDS4 that indicates it may not be ready for LADEE or MAVEN unless something is changed or further shown. An “action” indicates a finding or concern that warrants a response from the PDS4 team to this panel, but not in such a timeframe as to prevent forward progress for the project. A “recommendation” is a finding that suggests a change or action be taken, but no response is needed back to the review panel. A “comment” is simply an observation.

Requests for Action

PDS4ORR-RFA1

Originator: Brian Duncan, JHU/APL; Marilyn Newhouse, MSFC/CSC

Category: Lien

Statement of Concern: It is difficult to understand the testing performed (and results generated) to validate and verify the core tools to support LADEE and MAVEN.

Explanation: The scope specified in the ToR for this ORR includes the “Test plan and results for this Limited Release”. The entrance criteria specified in the ToR indicates that the test plan and results will be made available to the review participants in advance. The assumption of the review panel going into the ORR was that the testing of the system for LADEE and MAVEN was complete and documented. Through discussion during the ORR it was understood that testing was on-going, development of the information model and software was on-going: release 3b is more or less closed out, but release 4a is preparing for the test cycle and is needed/desired for LADEE/MAVEN. To support common data format (CDF) files and data product-level searches, development and testing must continue for the next revision of the information model and the 4a release of the software. For the review panel to gain confidence that the testing approach is sufficiently mature that a stable 3b system would not be jeopardized by changes made in 4a, the panel dug into the testing performed on 3b, the requirements traceability, and the artifacts generated as a result of the testing. From this research, it was not easy to tell what testing had been done, the details of the system under test, and the results of that testing.

Recommendation: The review panel requires that some modifications to the capture and presentation of test material be adopted during the 4a test cycle. If the following are applied to the 4a test cycle and documented, then this lien will be addressed for closure.

- a. A new requirements trace to the test case has been supplied, but we were still not able to tell pass/fail for individual requirements, only for overall test cases. In the test results report, please indicate the pass/fail of each requirement covered by that test case. If a test case fails, for example, it does not necessarily mean that all of the requirements failed.

- b. In the test results report, it is important to indicate the version under test. If the test result report shows the effect of testing multiple internal software builds over time, then this is most easily handled by indicating the software build version number in the test case result section (where you already have the date).
- c. In the anomalies section (i.e. JIRA) of the test results report, we would like to see the data more clearly organized and with some further information:
 - a. Separate the list of issues so that it is clear which were submitted during this round of testing. We should know, for example, which issues were submitted new during testing of release 4a (versus ones that were submitted back in 3b, etc.)
 - b. Separate (or at least show in the table) the issue type so that it is clear what is a software defect, what is a software improvement, etc.
 - c. Please show in the table the severity of each issue; a list of 5 minor defects may not be as significant as one showstopper defect
 - d. Please show in the table the title of the issue, so that the reader can get a sense of the types of issues being found during test
- d. An executive summary at the end of the test result report showing a requirements verification matrix, and what has passed or failed for each requirement provides the health of the release at a glance. The Build 3b system test plan very nicely shows the requirements and test case IDs that test each (in section 4). This table with a test **result** for each would give us a succinct view of the state of the release at a glance.

PDS4ORR-RFA2

Originator: Peter Shames, JPL

Category: Action

Statement of Concern: The document relationships specified in the LADEE LDEX ICD and the SIS are not clear. The ICD references other SIS documents that may or may not exist.

Recommendation: Clarify the ICD document in the areas where it refers to an SIS. Is this as simple as only referring to a single SIS that now exists? Perhaps provide a template for an ICD and the necessary SIS documents to support it. This should clarify what is expected of other data providers.

PDS4ORR-RFA3

Originator: Dave Heather, ESA/ESAC

Category: Action

Statement of Concern: The current / proposed set up for MAVEN and LADEE focuses on web-based tools, software and some documentation external to the bundles being archived in order to provide the best user service. This is fine for currently active missions but requires significant maintenance and implies much information and many tools may be lost if the www system is no longer managed. This goes against the long term usability of the data.

Recommendation: Ensure that all information (documents, algorithms, calibrations etc) from the web pages that may be useful to end users is included within the bundle itself, so that data can be used and understood fully with the standalone bundle should the full online support be discontinued. This could become a standard check made as part of the PDS4 Peer Review process.

PDS4ORR-RFA4

Originator: Peter Shames, JPL

Category: Recommendation

Statement of Concern: The overall handling of requirements in the project could have benefited from some of the same level of rigor that has been applied, with great effect, to the information model.

Recommendation: Use of any formalism (DOORS, SysML, or even a more traditional and complete spreadsheet) would have been a better way to provide traceability down to design and to the test suite. Given the present information there is no strong evidence that all requirements have been satisfied.

PDS4ORR-RFA5

Originator: Lauren McIver, Virginia Bioinformatics Institute

Category: Recommendation

Statement of Concern: It would be good to clarify what a “browse” collection is in the Standards Reference (pg 9) and the Concepts document (pg 8).

Recommendation: I would suggest adding examples to these documents as was done with the other collection types in the Standards Reference (pg 10). Another option would be to include references in the Standards Reference and the Concepts document to the definition found in the Data Providers handbook (pg 24).

PDS4ORR-RFA6

Originator: Dave Linick, JPL; Peter Shames, JPL

Category: Recommendation (to PDS program management)

Statement of Concern: While the PDS is a “federation” of nodes, and while recognizing that the needs of the nodes vary one from another, the PDS4 architecture offers opportunities to more fully exploit commonality among the nodes and to better integrate the PDS into a true system.

Recommendation: The PDS project would really benefit from having a higher level of top down architecting & engineering to guide and focus the Engineering Node and Discipline Nodes for common look and feel of web sites, common/adaptable tools, and tool sharing.

As a recommendation for Program Management: Develop a PDS vision and a Strategic Plan for achieving that vision. The vision should be bold, recognize the full possibilities of the PDS, and exploit the PDS4 architecture to its fullest. It should include a roadmap that defines, at a top-level, how the Strategic Plan should be achieved over time. The Strategic Plan should be owned by the primary stakeholders, including the Program, the

nodes, and the science community that uses the PDS. This would serve as an alignment vehicle and would provide a basis for proposing a logical and defensible evolution of the PDS. Much progress has been made recently and this is an opportunity to build on that progress.

PDS4ORR-RFA7

Originator: Lauren McIver, Virginia Bioinformatics Institute

Category: Recommendation

Statement of Concern: The term “node” has multiple definitions.

Recommendation: In the concepts document, section “1.2 The archiving process” and “7.3 Validation” it seems a “consulting node” and “discipline node” are PDS staff. In the same document, section “6.4 Using data dictionaries” it refers to “node data dictionaries” that are “needed by users of that node for searching or manipulating the data” in which a node seems to refer to a group of resources. Since “consulting/discipline node” seems to be a commonly used term throughout PDS, it is probably not possible to change “node” to “staff” for instances in which it refers to an individual. Instead it might be useful to define the term in the Concepts document when it is first used. I would also suggest adding the definition of “discipline node” to the PDS4 Glossary at the end of the Concepts document as currently a “consulting node” is defined as a “discipline node” but there is no definition for a “discipline node” or “node”. The definition could state it is a group of resources that include both PDS staff and data sets. This would help in clarifying these terms for individuals who are not familiar with the PDS terminology.

PDS4ORR-RFA8

Originator: PJ Clark (JHU/APL); Dave Heather (ESA/ESAC)

Category: Recommendation

Statement of Concern: The PDS documentation should have sections clearly describing restrictions placed on external file standards such as FITS and CDF that will allow them to be made fully PDS4-compliant.

Recommendation: A specific section should be included in the Standards Reference and/or the Data Provider’s Handbook to list the restrictions that are placed on all acceptable external file standards. This should cover both CDF and FITS as needed for MAVEN. Ensure that the information contained in the backup slide entitled “Requirements for Archivable CDF” becomes a registry document at any discipline node that archives CDF files. Ensure that all labels that pertain to CDF encoded products contain a reference to this document in the labels’ Reference_List.

PDS4ORR-RFA9

Originator: Dave Heather, ESA/ESAC

Category: Recommendation

Statement of Concern: If a restricted form of CDF is going to be an accepted format for MAVEN data, and accepted by PDS in general, the tools, software and documentation developed by PPI specifically to accommodate this should be centralized or endorsed by

the PDS as a whole so other future nodes using the format will not have to develop their own tools for the same purpose.

Recommendation: Centralize the tools and documentation being developed by PPI for the usage of CDF within PDS4. At least the PDS, at a high level, should endorse them and ensure that the same set of tools is used for future CDF deliveries to the PDS.

PDS4ORR-RFA10

Originator: Lauren McIver, Virginia Bioinformatics Institute

Category: Recommendation

Statement of Concern: Running some of the PDS tools on a windows platform I came across a section in the generate, validate, and transform tools code that would benefit from a warning message.

Recommendation: Line 26 of the windows batch file for these tools sets the JAVA_HOME environment variable to a fake path (set JAVA_HOME=\path\to\java\home). If JAVA_HOME is not defined in the environment and the user does not alter line 26 of the code, the code fails with an error message (a default message printed when the system tries to execute at a non-existing path) that is not the best description of the issue for the failure. This message is “The system cannot find the path specified.” I would suggest putting in a warning message at line 26 instead of setting the variable to a fake path. This warning can direct the user to the readme which describes how the user should set the path. This would make the code more user-friendly.

PDS4ORR-RFA11

Originator: Dave Heather, ESA/ESAC

Category: Recommendation

Statement of Concern: It is unclear if the data to be delivered by LADEE and MAVEN are going to flow through the full validation cycle or if it will evolve significantly from what was shown during this review, and therefore it is difficult to judge whether the system is ready now to allow full ingestion and product level searches.

Recommendation: PDS could require data providers to run the full validation tool as it evolves to check bundles and collections. Wherever possible the validation tools and procedures should be centralized.

PDS4ORR-RFA12

Originator: Marilyn Newhouse, MSFC/CSC

Category: Recommendation

Statement of Concern: The goals for the PDS4 development, while laudable, are not presented as metrics that can be used to prove the completion and success of the PDS4 development.

Recommendation: The PDS4 goals (for example, see slide 12 of the presentation) should be quantified, and the progress against these goals measured. This provides several benefits: 1) a clear definition of when the PDS4 has realized its goals and is “complete,” 2) measurement of progress and early identification of concepts or approaches that may be straying from the desired path, and 3) data useable by NASA HQ to justify and defend PDS4 progress and funding. The metrics should look at the PDS holistically, that is, from a perspective that includes development, maintenance, and mission effort and cost to ensure that emphasizing improvements in one area do not simply shift effort to another area. Some historic data may need to be estimated at this point, but defining the metrics by which the PDS4 improvements will be measured will identify what data should be collected in the future. Some possible metrics, provided mainly to begin a brainstorming process, are:

<p>Improve efficiency and support to deliver high quality science products to PDS</p>	<ul style="list-style-type: none"> • Effort required (PDS & mission personnel) to define the products and interface for mission data delivery (should decrease) • Effort required (PDS & mission personnel) to deliver and validate the first and subsequent data deliveries (should decrease) • Number of redeliveries required to provide validate mission data (should decrease) • Percent budget allocated to maintenance (should decrease) versus development of new capabilities (should increase)
<p>Improve user support and usability of the data in the archive</p>	<ul style="list-style-type: none"> • Total number of data requests (should increase) • Total number publications using archived data (should increase) • Amount of data retrieved vs. session duration (should increase) • Number of requests for assistance in finding and retrieving data (should decrease)

PDS4ORR-RFA13

Originator: Tom McGlynn, GSFC

Category: Comment

Statement of Comment: The PDS4 Software Tools and Services for LADEE/MAVEN describe a capability in the Search Interface of: Provides the second line of metadata harvesting within the system in order to facilitate discovery of products. If this is a

future expansion capability to enhance the search metadata, we would suggest that it be controlled in a similar manner to other software design/implementation changes.

PDS4ORR-RFA14

Originator: Dave Linick, JPL

Category: Comment

Statement of Comment: APPS, the Automated Pipeline Processing Service from AMMOS, is likely to be available in time to support MAVEN data ingestion into PDS. There was no mention of this possibility in the ORR. This is not inappropriate, since APPS is not a pre-condition for operational readiness for MAVEN, but APPS should be considered in the eventual plan for MAVEN.