

IPDA Prototyping Using PDS4

Final Report

Draft August 26, 2011

Introduction

At the International Planetary Data Alliance (IPDA) Steering Committee Meeting held July 2010 in Bremen, Germany, the IPDA Steering Committee felt that it was important to initiate a project to prototype agency data sets using the new PDS4 Data Standards. The goal of the project is to test the capabilities of the PDS4 Data Standards for preparing data sets for use in the international planetary science community under the auspices of the IPDA. The project "IPDA Prototyping Using PDS4" was initiated and a team consisting of S. Hughes (lead), P. Allan, M. Gangloff, D. Heather, G. Krishna, T. Roatsch, A. Sarkissian, I. Shinohara, T. Stein, M. Teresa, and J. Salgado was formed. The project had two phases.

The goal of Phase 1 of the project was to re-assessed the PDS4 Data Standards and documentation. This phase started on April 15 and ran through May 15, 2011. Phase 2 of the project focused on developing a prototype PDS4 data product using one or more data products produced by the agency. This phase started on May 25 and ran through July 15, 2011. Both phases used the PDS4 Data Standards documents that had been released for Build 1c.

Each phase was preceded by a telecon where the team was given a brief overview of the PDS4 data standards development effort and briefed on its current state. Most team members were able to participate in these telecons. All PDS4 Data Standards documents were accessible from either the PDS Engineering Node website or the PDS4 Data Design Working Group (DDWG) Wiki site.

Phase 1

The introductory telecon for phase 1 briefed the project team on the tasks performed by the PDS4 Data Design Working Group (DDWG) over the previous year. Many completed tasks addressed issues that had been identified by the IPDA in its assessment the previous year. These tasks include:

1. Rewrote or redesigned most documents
2. Wrote additional informative documents such as tutorials and user guides
3. Rewrote many attribute definitions in the data dictionary.
 - Designed the local data dictionaries schema and their management process

4. Rewrote the PDS4 data provider's handbook
5. Released the Standards Reference document

For phase 1, two assessment worksheets were provided. On the first worksheet the assessors were to include their assessment comments and suggested improvements. The assessors were to answer the following questions on the second worksheet.

- 1) Do the document provide sufficient background for the review? If not, how could they be improved?
- 2) Assess the four fundamental structures. Are they useful? Will they support your needs? Do you have products that you believe will not fit into the structures?
- 3) Assess the PDS4 core product types. Do they provide an adequate set of baseline templates for constructing new templates and new PDS4 products? What is missing?
- 4) Assess the structure and layout of the PDS4 product examples? How can it be improved?
- 5) What overall recommendations do you have for the team? Do you have suggestions for improvement?

The first phase was completed on scheduled and the original and compiled results are available at the [IPDA PDS4 Assessment and Prototyping](#) website.

Phase 2

The introductory telecon for phase 2 presented the basic guidelines for the prototyping phase. The goals of this phase were the following.

1. Primary – Attempt to produce XML labels for one or more observational data products.
2. Primary – Report on the process.
3. Secondary – Produce XML labels for a Bundle, a Collection, a Document Product, attempt pipeline set up.

The following process outline was then explained in more detail.

1. Choose a data set.
2. Design an observational data product.
3. Outline the Archive Bundle design.
4. Design and produce an observational data

The second phase was completed on scheduled with a waiver given to the PSA team. They wished to develop a more complete prototype to further test the data standards. They have completed their work and forwarded the results. The original and compiled results are available at the [IPDA PDS4 Assessment and Prototyping](#) website.

Two assessment worksheets were also provided for Phase 2. On the first worksheet the teams were to include comments and suggested improvements. The following questions on the second worksheet were to be answered.

1. How well did the process for creating PDS4 products work?
 - Is the generic product schema you chose complete and useable?
 - Is the process for creating a specific schema well documented and complete?
 - Is the process for creating a product label well documented and complete?
 - What parts of the process could be improved or what needs to be changed?
2. What tools should be developed and made available?
3. Are the PDS4 data standard documents useful?
 - What could be improved?
4. Did you find any limitations or items missing that you expected?
5. Do you have any other comments?

The Phase 2 telecon presentations and assessment worksheets are available at the [IPDA PDS4 Assessment and Prototyping](#) website.

Assessment Summary

In general the assessment is that good work is being done and that the PDS4 data standards will be useful to data providers in the Planetary Science Community.

The project team provided a large number of comments for both phase 1 and 2. These have been compiled into a single spreadsheet and were uniquely identified and appended with a status column for logging progress. The original and compiled spreadsheets are available at the [IPDA PDS4 Assessment and Prototyping](#) website. The following table provides a count for each classification.

Classification	Count
Ambiguous	13
Duplication	2
Error	12
Improvement	20
Incomplete	33
Inconsistent	31
Missing	27

Other	40
Question	27
Suggestion	74
Kudos	21
Removed	6
Total	306

A few of the key issues are summarized below.

1. As described in the PDS4 standards documentation, Local Data Dictionaries (i.e. mission and node dictionaries) will only be used for adding new classes/attributes under the Mission and Node area, and not for further restricting classes/attributes already defined in the PDS4 core dictionary. But PSA wants to define further restrictions to the attributes/classes/types in the PDS4 schemas (both, core dictionary schemas and generic product schemas), and wants those restrictions to be applicable to all products at different levels (PSA global/mission/instrument level). What would be the best approach for this? Will there be a PSA core data dictionary? How will that work?
2. We (PSA) need to understand how permissible values, restrictions and formation rules in both the Information Model and the schemas will be controlled and updated. Who will control them? How will updates be handled? Will external authorities, like PSA, be able to add new restrictions/values? How? If not, can this be implemented with namespaces/dictionaries.
3. We (PSA) need some clarifications on the usage and creation of local data dictionaries i.e. mission data dictionaries.
4. There needs to be better separation of the Concepts and DPH documents so that the DPH contains just the basic guidelines and requirements for data preparers/users and with the most common practices/examples.
5. Data Dictionary definitions need a lot of work. (Editor's note: The DDWG was performing a data dictionary attribute definition cleanup during this assessment.) Rules and recommendations have to be clearly defined in the PDS4 standards documents. Use of data dictionaries by other authorities / organizations needs clarification.
6. For IPDA, it would be nice to see references to how PDS4 should be used for data providers not delivering directly to a PDS Node. This could be a simple matter of generalizing by adding something like 'your Archiving Authority' to the sections describing management of schemas and dictionaries.

7. The XML introduction and terminology is very nice and useful to those with limited XML experience, but suggest that this is an Appendix.
8. Consider eliminating the abridged data dictionary since it was not used in the assessment.
9. The Standards document needs quite a lot of work. This will be the central resource for data providers wishing to use PDS4, so requirements need to be very clear in this document, and recommendations only used for non-critical elements. Historically, 'recommendations' are rather brutally ignored by many data providers if it doesn't suit them, and the level at which various Nodes decide to enforce the 'recommendations' has varied greatly. We would rather see the Standards Reference provide only firm rules with as little extraneous text as possible.
10. Clarification of Data dictionary management procedures including rules for editing/modifying the product types, the link between the XML schema and the Dictionary, formation rules or permissible values from the dictionary are in the schema, management of schema the Node / Archive Authority level.
11. Continue toward the use of more intuitive language especially for the sake of non-native English speakers.

The IPDA assessment project team requests that the PDS4 DDWG accepts the assessment comments and recommendations as constructive input during PDS4 Data Standards development. Editor note: These results have forwarded to the DDWG and to the individual document author's for their information.

Acknowledgements

The project team would like to acknowledge the generous help provided by the DDWG staff members, Mitch Gordon and Ron Joyner. During the prototyping phase especially, they made themselves available to provide technical advice and guidance.