

February 26, 2008

PDS4 Discussion Meeting GEO/EN Notes

Attendees: Dan Crichton, Steve Hughes, Lars Arvidson, Keith Bennett, Ed Guinness, Dan Scholes, Susie Slavney,

Executive Summary

The meeting reviewed the PDS4 concept outline and project schedule that the PDS4 Planning WG has been putting together. There was general agreement that with PDS4, we should be focusing on defining and extending our services. That does not mean that PDS neglects its archive function, but rather we should be building a PDS4 for 2008+ on modern computing infrastructure that allows for constructing higher order functions to submit and interact with PDS data. In addition, PDS needs to be an integrated data system that interacts with our international partners. There was quite a bit of discussion regarding the need to build a data system from a solid set of data standards. There is concern with the existing PDS Standards Reference since in many cases it has recommendations and not requirements. Recommendations, when building a data system, are difficult to implement because it leads to different interpretations across nodes and internationally. The meeting also stressed the importance of data integrity and ensuring that PDS data and services are protected and reliable. The project schedule was also discussed and it was emphasized that PDS must build an implementation team that is empowered by the Management Council to make decisions in order to move PDS4 forward and meet the proposed time line.

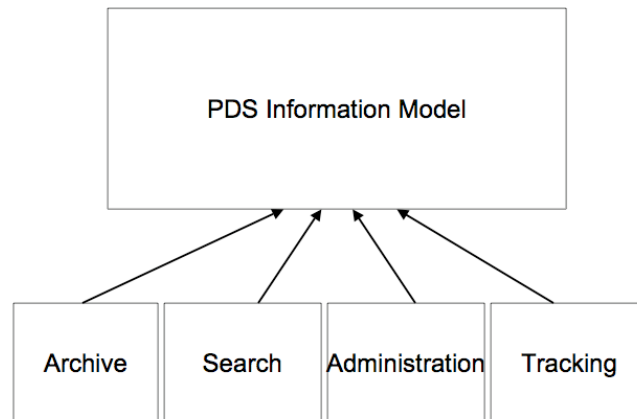
Specific Notes

The following are notes regarding the discussion of several of the items in both the outline and the PDS4 schedule.

1. Vision – Regarding the vision, there was extensive discussion that PDS4 should focus on improved services for users, however, there was also discussion that PDS can't forget its archive responsibilities. The current vision statement lacks a mention "archive" which needs to be addressed.
2. Formats - The term "formats" is used inconsistently by PDS. In particular, is PDS describing the organization of a product from the PDS model perspective or is it defining the data structure of the bits and their packaging? While there was some concern that "three formats" might be too limiting, all agree that cleaning up the organization of "PDS data formats" (i.e. those formats that are used to organize a product) is necessary. It was also mentioned that in many cases PDS has a number of very similar products that are not derived from any general hierarchical structure of core formats that would be "best practices" in terms of data modeling.

There was also discussion about PDS adopting a set of standard formats for the data objects that make up the PDS products (e.g., HDF5, CDF, etc) since there is support from other communities to work with these products. The PDS information architecture doesn't preclude this given PDS' approach to richly describe the structure of its data. However, it was also discussed that a "one-size" fits all solution won't work across PDS.

3. International – It was agreed that PDS must work with our partners and must coordinate the PDS4 development with them.
4. Standards – Regarding standards, there was a discussion that formalizing PDS standards, and resolving the issues that the Data Model WG has identified will go a long way towards addressing several of the usability issues associated with PDS standards. However, it was also discussed that the standards reference should be derived from the PDS information model explicitly identifying that the construction rules associated with using PDS standards. Phrases like "recommended" should be replaced with "required" since recommended standards make interoperability among disparate data sets, nodes and partners extremely difficult. Several would like to see the standards reference re-written based on the formalized PDS model.
5. Presentation Layer – There was a discussion regarding how data is presented to users. The issue was not related to common look and feel, but rather common views of the repository when a repository structure is made visible to the user.
6. Search at the product level – All agreed that search at the product level needs to be architected in PDS4, but there is a lack of common search keywords in products which make it difficult to search across the archive. It was discussed that the information model should logically include separate sub-models for supporting critical data system functions including search, archive, administration, etc. The example of this is below (which does not include the comprehensive set of functional requirements on the model):



There was also a discussion that the model reflected in the PDS standards should be independent of any particular notation or language. This does not mean that PDS can't create an implementation of the model using a modern approach, it just means that it can't be bound to any one technical approach.

7. Services – Common and domain –specific services were discussed. Currently, the PDS4 planning lacks a list or description of these. Example common services are find a PDS object (data set, product, etc), download a a PDS object, etc. Domain-specific services might be “stretch”, or “subset”. PDS should make a simple list, at least identifying the core services that are needed.
8. Data Integrity and Tracking – Geo is really concerned about data quality, especially as our volumes increase. We need to address the three critical components identified in the archive integrity policy: file corruption (e.g., using checksums), data availability (e.g., the three copy rule of data availability), and tracking. Geo is concerned about separating these because they should be built as an integrated capability across PDS, end-to-end.
9. Submitting small data sets to PDS – There's been quite a bit of discussion regarding the submission of small data sets (e.g., investigator-type) to PDS. It was agreed that PDS should provide a “template” for them to submit data. This is in contrast to missions which really need the concept of a design tool to provide the flexibility.
10. Submission to PDS – There was general agreement that PDS could do a better job of providing standards and tools for submitting data to PDS. Currently, there are no

standards related to the structuring and transferring data for delivery. In addition, PDS needs to provide tools to track those submissions. The number of data files associated with a submission is getting very large so we need mechanisms to ensure all files have been delivered.

11. Operations Monitoring – Operations monitoring really needs to be a “service-based” monitoring capability. As PDS shifts to offer more and more services (application interfaces, web interfaces, etc), PDS should monitor these services to ensure they meet their “quality of service” (QoS) commitments. There was also a discussion regarding checking such items as “links” which should be included.
12. Engaging DN communities - It was discussed the PDS should engage its communities as it goes through projects. This could mean helping give input on requirements and needs, however, it could also mean that savvy software-types in the community can write and share their own services and software. There was a related discussion that PDS first clean up its standards, second build basic services, and third engage the community.
13. NASAView – There was a discussion related to NASAView. This discussion had multiple threads, but effectively was resolved to a couple of key points. First, NASAView, as it is written now, does too many things. If we are going to have PDS-specific data formats and structures, then we need a NASAView type capability. Second, the implementation of NASAView is dated. It needs to be re-written. Third, NASAView was written on the Object Access Library (OAL) which was written by University of Colorado at Boulder several years ago. The purpose of OAL was to introduce “abstraction” in the PDS software architecture so there was a common application interface for reading PDS data. Standard software readers and libraries are a critical part of the tool box needed for PDS4. Where those libraries come from is dependent on the data format decisions and the availability of software to work with those formats.
14. Knowledge-base - There was quite a bit of discussion about the knowledge-base. This related to some of the other discussions regarding involvement of “users”. Keith suggested that users should be able to submit back comments on the data, and we should be able to show it. Some of the discussion related to opportunities to inject Web 2.0 type functionality into PDS (blogs, wikis, etc), but the point was to provide a mechanism for users to increase the knowledge base of the data captured within the system. There was also some concern that comments would need to be controlled to ensure that they were productive.
15. Management and Schedule for PDS4 Projects - There was a discussion related to keeping the work moving forward. There was general agreement regarding the phasing of the projects, however, there was concern efforts could stall if not managed correctly. It was suggested that PDS needs to have a decision-maker in place that keeps projects moving forward. In addition, PDS needs to have a cross-node implementation team that is empowered by the MC to implement the PDS4 projects

and that the representative on the implementation team should be empowered to speak for the node.

16. Missions and bringing them into PDS- There was a discussion of when we bring missions into PDS4. It's going to be important to identify which missions will start using PDS4 standards, however, we need to have the comprehensive tool suite in place along with good documentation about PDS4 standards. Phasing PDS4 will help, since some capabilities could be brought online while other parts of the PDS4 development are still on-going. It was also discussed that we could engage our community and discuss our plans at conferences such as LPSC.