PDS 2010 PROJECT PLAN

DRAFT EXECUTIVE SUMMARY

July 1, 2008[1]

NB: The audience for *this* Draft Executive Summary (DES) is the PDS Management Council (MC) — the node managers. The goal of the DES is to provide an overview of PDS 2010, which can then be discussed and refined at the MC level. As the Plan evolves, the Executive Summary will be re-targeted to a non-PDS audience, and its content will be focused more narrowly on the steps needed to implement PDS 2010 rather than the broad outlines of the system itself.

For over fifteen years, the Planetary Data System (PDS) has been NASA's official data system for archiving and distribution of data from planetary exploration missions. It has been a leader in defining data standards, working with missions and instrument teams, and developing data system technologies. The PDS has been instrumental in changing the scientific culture by working with the planetary science community to publicly release and peer review the data it captures. It has also been used as a model by other science data systems interested in establishing distributed scientific networks organized by independent discipline nodes at facilities that are doing leading-edge scientific research.

While PDS has been a leader in developing and exploiting new technologies and ideas, an increasing workload and substantial increases in the volume of delivered data are now threatening the system's ability to accomplish its primary missions of both archiving planetary science data and distributing it to working scientists. PDS identified these challenges in its Roadmap [1] published in 2006. In addition to these challenges, the ten year Roadmap outlined several goals including improving the PDS data standards, increasing user services by leveraging newer technologies and technical standards, and re-architecting PDS to ensure efficient operations of the system while supporting the increasing demands on PDS by both the data providers and end users.

In response to these challenges and goals, PDS has developed a plan for the next generation called "PDS 2010". The vision for PDS 2010, as defined by the PDS Management Council at its April 2008 meeting, includes:

- simplified, but rigorous, archiving standards that are consistent, easy to learn, and easy to use
- adaptable tools for designing archives, preparing data, and delivering the results efficiently to PDS
- on-line services allowing users to access and transform data quickly from anywhere in the system
- a highly reliable, scalable computing infrastructure that protects the integrity of

---

[1] Derived from R. Simpson 6/10/2008 modifications

data, links the nodes into an integrated data system, and provides the best service to both data providers and users

PDS 2010 is a three-year transformation comprising six critical "projects" aimed at achieving the goals identified in the Roadmap. First, and foremost, is a project for improving the current PDS data standards. After years of *ad hoc* additions and imperfectly-controlled modifications, the standards need to be overhauled so that they are more rigorous, easier to use, and better aligned with current information technology capabilities and practices. Inconsistencies must be resolved, opportunities for "creative interpretation" eliminated, and the learning curve made more manageable. Archiving must be made an easily adapted adjunct to routine data processing — not a career in itself. The data standards project will update the data model, streamline the common data dictionary, and define a (limited) set of core data formats. The data standards project will set the foundation upon which the remainder of PDS 2010 is built and is therefore the most critical step in achieving the ultimate objective.

Second will be definition of the software system architecture and implementation of the core services that will support ingestion, cataloging, and distribution of PDS products. PDS legacy software is a set of "tools" developed over time and written in a mix of computer languages. Much was designed before PDS became a distributed on-line system; maintenance is costly and inefficient. PDS 2010 will re-architect the system as an online set of distributed services that support ingestion and distribution over the Internet. Web-based technologies and standards will be leveraged and deployed at nodes to support submission, query, and retrieval of PDS data products from anywhere on the Internet. In the information technology community this is known as a "service-oriented architecture" (SOA).

The third project will be development of a set of tools for design and validation of PDS data products that are based on the PDS 2010 data standards. These tools must be in place before data providers can begin deliveries based on PDS 2010 standards. A substantial amount of reuse can be expected of the recently released Java-based tool sets for design and validation; but, in other areas, more fundamental changes will be required.

Fourth, PDS will build an automated, distributed cataloging system on top of the core software service architecture, tools, and data standards. This is essential for improving efficiency and accuracy in the ingestion, tracking, and search functions of the system. Today, management of most catalog information requires human intervention. In addition to reducing delay and error, the new system will provide end-to-end tracking, protecting integrity of the system as well as ensuring that individual data products are both locatable and uncorrupted. The new cataloging system will provide a wide variety of information on products that can now be obtained only through individual query and retrieval. It will be distributed across the PDS with management being conducted as part of the federation.

The fifth project will be development of a new web-based portal interface. Changes in the PDS data standards, catalog system, and core software services will necessitate an upgrade to the PDS home page. In 2007, PDS appointed a "user services" team that identified a need for both improved and extended support to users accessing the system [2]. While the primary purpose of

this project is to ensure that the PDS home page can work with the new software architecture, tools, and data standards, the project itself can be phased to deploy increasingly sophisticated functionality in response to the recommendations of the user services study. New resources and services, such as the distributed catalog, will be in place from earlier PDS 2010 projects allowing more comprehensive searches for and manipulations of data, for example. It also envisioned that various user services such as data format transformation and geometry functions can be provided.

The sixth and last planned project is the data movement project. With increasing volumes of data being exchanged online, PDS 2010 must adopt standards for high-speed data transfer and delivery among data providers, data users, nodes, and the deep archive. Presently, PDS uses *ad hoc* mechanisms for delivery and movement of data. This project will define standards for structuring data transfers and efficiently moving those over the network. PDS is already working with National Space Science Data Center to improve transfers to the deep archive (NSSDC) and with the San Diego Supercomputing Center (SDSC) for near real-time backup support. As PDS embarks on PDS 2010, adopting technologies to improve data transfer throughout the system will be critical to improving overall efficiency and integrity of the system and service to users.

PDS faces major challenges in transitioning from its current operations to the PDS 2010 system. A phased approach is unavoidable with the major milestone being acceptance of the first data meeting the PDS 2010 standards. This will require that PDS complete definition of the PDS 2010 data standards, build and initialize the new catalog system, deploy enough new tools to validate and distribute the new products, and update supporting documentation. In the current scheduling of projects, acceptance of new data deliveries is planned for the end of the second year. Other milestones will include releasing additional new tools, websites, and applications. The detailed schedule is spelled out in the PDS 2010 Project Plan.

The PDS Management Council will ultimately be responsible for the portfolio, selection and schedule of projects. The PDS 2010 Implementation Manager will coordinate, track and report on the status of projects to the Management Council through the PDS Program Manager. The Implementation Manager will assemble design and project teams and oversee development, integration, and initial operations of PDS 2010. An overall PDS 2010 schedule will be maintained which will identify the deliverables and milestones for each of the projects.

References

[1] PDS Roadmap 2006 – 2015, February 2006
[2] PDS4 User Services Working Group Recommendations, December 2007