# PDS 2010 PROJECT PLAN

## *DRAFT EXECUTIVE SUMMARY*

May 31, 2008

**Disclaimer: The audience of this executive summary for the PDS 2010 Project Plan is the PDS node managers. The purpose of this document, at this point, is to help calibrate the project plan across the PDS by providing a short overview of the plan. Subsequent versions will be created in concert with the project plan as it progresses. The audience may be expanded in the future to include non-PDS personnel as deemed necessary by the PDS Management Council.**

For over fifteen years, the Planetary Data System (PDS) has been NASA's official data system for archive 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 to support the archiving and distribution of planetary science data, an increasing workload and substantial increases in the volume of data being delivered threaten the ability for PDS to maintain a system capable of supporting its primary mission of both science data archive and distribution. 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 user.

In response to these challenges and goals, PDS has developed a plan for the next generation PDS called "PDS 2010". The vision for PDS 2010, as defined by the PDS Management Council, is as follows

*"PDS will provide the community with planetary science archiving standards that are consistent and simple to use. PDS 2010 will provide online services for using its data archives, allowing users to access and transform data quickly from across the federation of PDS nodes. Data providers will be given adaptable tools to design, prepare, and deliver data efficiently to PDS. PDS 2010 data and services will be managed and delivered from a highly reliable and scalable computing infrastructure that is designed to protect the integrity of the data and virtually link PDS nodes into an integrated data system."[1]*

---

[1]     Accepted by PDS Management Council, April 4, 2008

The PDS 2010 plan is a three-year plan identifying six critical projects aimed at beginning to address the above vision along with the goals and challenges identified in the PDS roadmap. First, and foremost, is a project for improving the current PDS data standards. Due to years of continuous additions and modifications, the standards need to be overhauled to simplify the learning curve, resolve inconsistencies, and make it easier for data providers to produce PDS-compliant data products. The data standards project will define an updated data model, a new data dictionary, a standard structure for defining data products, and a set of core data formats. Its purpose is to greatly reduce the time and investment in adopting, using and managing the data standards for both PDS and its suppliers. The data standards project will set the foundation on which PDS 2010 will be built and therefore is a critical dependency that must be addressed in the first phase of the project.

Second will be definition of the software system architecture and implementation of the core services that will support the ingestion, cataloging and distribution of PDS products. The current PDS legacy software uses a mix of computer languages and tools developed over time. Maintenance of the legacy software is costly and inefficient to extend as PDS evolves. Much of it was not architected for the distributed nature of the PDS as an online data system nor the newer computing environments. PDS 2010 will re-architect the system as an online set of distributed services that will support ingestion and distribution over the Internet. Web-based technologies and standards will be leveraged and deployed at nodes to providing computing services to support submission, query, and retrieval of PDS data products from distributed locations on the Internet. These services will form what is known in the information technology community as a "service-oriented architecture" (SOA). The purpose is to allow remote applications to access data and computing services at the nodes.

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 data deliveries that are based on PDS 2010 data standards. It is anticipated that a substantial amount of reuse can be leveraged from the current Java-based tool set for design and validation that PDS has recently released.


Fourth, PDS will build the cataloging system on top of the core software service architecture and the new PDS 2010 data standards. Building on top of the service architecture will allow for a distributed cataloging system. This is essential for improving efficiency and accuracy from the ingestion, tracking, and the search functions of the system. Presently, updates and management of catalog information is done with "human-in-the-loop". The intention of building a distributed catalog system is to reduce human involvement in updating and validating catalog information. In addition, it will enable PDS to build an integrated end-to-end tracking system that is critical to protect the integrity of the system by ensuring that data is tracked from receipt all the way to delivery to the deep archive. The cataloging system will be capable for providing access to a variety of information that is catalogued within the PDS. One of the key challenges will be migration of the current catalog system to the new catalog system using PDS 2010 data standards.

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 held a study team that identified the need for both improved and extended support for users accessing PDS [2]. While the primary purpose of this project is to upgrade the PDS home page to ensure it can work with the new software architecture and data standards, the project itself can be phased to deploy increasing levels of functionality in response to the recommendations that the user services study team identified. New software services, including the distributed catalog infrastructure, will be in place from earlier PDS 2010 projects allowing for more comprehensive user services such as search, access and transformation across the PDS.

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 between data providers, nodes, and the deep archive. Presently, PDS uses ad hoc mechanisms for delivery and movement of data. This project will evaluate, select and deploy standards for structuring data deliveries and moving those deliveries over the network. This will allow for more efficient electronic data deliveries from data providers as well as to the deep archive. It is envisioned that PDS and the NSSDC can work together to adopt technologies for efficient data transfer from PDS nodes to the NSSDC. PDS has already been testing technologies for high-speed data transfer as part of evaluating the online storage services and software from San Diego Super Computing Center (SDSC). As PDS embarks on PDS 2010, deploying technologies to improve the transfer of data online is going to be critical for improving the overall efficiency and capability of the system.

PDS 2010 also faces major challenges in transitioning from its current operations to the PDS 2010 system. It is very clear that PDS will need to transition in phases. The major milestone will be preparing to accept data deliveries formatted for PDS 2010. This will require that PDS complete definition of the PDS 2010 data standards, build and migrate to a new catalog system, deploy new tools for design and validation of PDS-compliant data products, and update supporting documentation. In the current scheduling of projects, the transition to accept new data deliveries is planned for the end of the second year. Other major transitions will include transitioning tools, websites, and applications. These transitions are defined in the detailed 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. The implementation manager will assemble design and project teams and oversee development, integration and operations of PDS 2010. The project schedule will be maintained as a living document so as to ensure that PDS meet its operational commitments while rebuilding PDS for the future.

**References**

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