PDS4 Vision and Projects

2/29/2008 – Augmented PDS4 vision to include archiving based on comments from GEO and Imaging.
2/22/2008 – Begin discussions on the management plan; reorganize projects
2/19/2008 – PPI Input and Vision Statement; reorganization of PDS4 projects
2/7/2008 – Updates from Joy, Martin and PDS4 Planning WG
1/23/2008 – Core PDS4 Concepts for the White Paper
1/16/2008 – Input for the PDS4 Concept Paper

PDS4 Planning WG: Beebe, Crichton, Joy, LaVoie, Martin, Stein

Characteristics of PDS4 need to be identified in order to construct the PDS4 Concept Paper. The following are concepts pulled from the PDS Roadmap, PDS4 WGs, and MC discussions.

1. PDS4 Vision

Vision Statement: With PDS4, PDS will become an online provider of services, allowing users to quickly find, access, and transform data from across the PDS enterprise. Its archiving standards will be international, unambiguous and simple to adopt and use. Its data providers will be given adaptable tools that enable it to design, prepare and deliver data efficiently to PDS for archiving and distribution. Its 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.

Characteristics of PDS4

- (1) Unambiguous Standards: Standards are explicitly and unambiguously defined so that they can be consistently used and implemented by Discipline Nodes, data providers and international partners. A small number of data formats[1] are included in the recommended PDS standards reference.

- (1) International: PDS is implemented to support collaboration with international partners and missions. It's data standards and software are developed to enable data sharing based on the IPDA standards. PDS and IPDA will share a common core.

- (1) A Distributed Online System: PDS should be built from a solid foundation of distributed services and functions. These services are

[1] PPI is recommending three which have associations to image, table and cube

geographically distributed, built on formally defined interfaces and protocols enabling a highly scalable, highly reliable, virtual system. PDS functions are offered as online, server-side functions (i.e. service-oriented model) that support both data delivery and computational processing (e.g., data discovery, data access, format conversion, coordinate system conversion, subset, etc). Search and access should be to the product level and/or sub-product, catalog or record level. Users should be able to get data in a usable format that may be different from their archive formats.

- (1) Archive/Data Integrity: PDS ensures that all data is preserved and moved error-free and accessed across the enterprise regardless of the medium in which it is stored. Three copies of the data exist in a reliable storage environment.

- (1) Integration with Data Producers: PDS provides tools and standards for data producers, both missions and DAP investigators, working with them as early as possible to ensure they can efficiently design products and archive data in the PDS.

- (2) Efficiency: PDS functions for ingestion and inventory management are automated to ensure the interface between operating units within PDS and missions is efficient and timely.

- (2) High-Speed Data Exchange: PDS adopts technologies and standards that allow for the online exchange of very large data sets and products. PDS should also provide standard mechanisms for electronic data delivery and packaging.

- (2) Tracking and Reporting: PDS tracks data throughout the system from delivery to PDS through to the deep archive to ensure that all data holdings are accounted for. This includes deliveries from a data provider to PDS, data captured at the PDS nodes, and data delivered to the deep archive. Reports on are available on demand which provide status and metrics (e.g., delivery reports, archive reports, usage reports, etc)

- (2) Coordination & Management: PDS will have well defined procedures and policies for the operation of the system. This includes standards, interfaces, protocols, operations, and infrastructure. The governance of the system is clear.

- (2) Flexible Search Facilities: Search tools allow users to search across the PDS enterprise (breadth and depth) using modern techniques (e.g., Google-like, Map-based, Facet-based, etc) enabling discipline-specific search capabilities. Results, from across the PDS enterprise, are collate-able and capable of being presented in multiple views including discipline-

specific views. Search services will be built on top of the distributed online system providing a uniform search architecture.

- (3) Highly Scalable, Reliable Computing Infrastructure: PDS deploys a computing infrastructure, based on technology standards, that is highly reliable and scalable. PDS, as a system, monitors the availability to ensure continuous operations.

- (3) Knowledge-base: In contrast to the archive metadata, the catalog can be updated to add curatorial information and associations as part of the catalog system.

2. Assessment of where we are with the PDS3 Implementation

- Ambiguous standards: PDS standards are captured as narrative text rather than explicit specifications which lead to different interpretations and implementations that do not work well together. We need primer documents for using our standards, but also explicit specifications for those building systems.

- Ad hoc storage infrastructure: PDS data is maintained in a variety of storage environments lacking consistent backup and recovery solutions.

- Growth in data sizes: Data delivered to PDS is continuing to grow rapidly which is putting pressure on PDS as a system affecting the overall storage of the data as well as the transfer of data to PDS, to the end user, and to the NSSDC. Large data volumes is affecting the time to validate, transfer and release data. It is also affecting decisions regarding optimal sizes of archive volumes as well as hardware.

- Built pre-internet age: PDS was "architected" as an offline model and has evolved to support online data distribution. Little consistency exists in terms of consistent technology standards or a reliable, scalable computing infrastructure.

- Varying degrees of search services: Search services vary widely across nodes in terms of their usability, coverage, and functionality. Search is generally limited to a particular area. Users are more more "tech savvy" and expect modern tools and interfaces.

- Not well integrated: PDS nodes are not well integrated. It is difficult to currently query catalogs across PDS nodes.

- International missions and interoperability challenges: Missions are much more international; PDS standards are being adopted internationally, but interpreted differently.

- Changes to our data standards are inefficient: Changes to standards can be slow and costly which makes it difficult to respond in the era of increasing diversity of missions, instruments and data providers.

- Mission support challenges: Missions are more frequent, they have a larger number of "first time" data providers who do not have a history with PDS. They also have shorter design/launch windows which mean data providers have to complete pre-launch work in less time. Archive support is often a lower priority.

3. How do we go from PDS3 to PDS4?

- Understand PDS3: PDS must understand PDS3 as a system before it attempts to finalize a PDS4 system specification. PDS must ensure that there is a clear transition plan for moving to PDS4 that applies lessons learned, builds on necessary functionality, and resolves existing problems.

- Well defined system specification: PDS4, from the outset, requires a well-defined system specification for the architecture, data standards and technology standards. The architecture needs to identify the interfaces both within PDS and to/from PDS. The standards represent a planetary science core that is shared internationally. The implementation projects build on these standards.

- System-level Requirements: PDS needs system level requirements for PDS4 that identify the gaps in the current implementation along with new proposed requirements for meeting the characteristics of PDS4.

- PDS Work Plan: The PDS4 project needs to include a work plan that identifies the roles of all the nodes and plans for their involvement. This includes system engineering, implementation, and operations.

- Multi-Phase Project Approach: Funding and operational limitations require a multi-phase project approach. The phases need to be identified in context with the operational transition from PDS3 to PDS4. Projects should be based on priority and the transition from PDS3 to PDS4 should occur over time. Requirements and critical specifications should be defined prior to implementation of any of the phased projects. Each project should contain a project plan, schedule and list of deliverables.

- Build New System Components: Rather than try and "port" existing system components, PDS4 should implement new system components

using modern system engineering practices and implementation languages. The goal is to not be constrained by the existing system implementation.

- Schedule Drivers - PDS needs schedule drivers that will set project priorities and drive them to completion. Overall, PDS has a senior review in 2009 that requires that the project be underway. In addition, PDS must identify schedule drivers that are tied to mission support. Additional drivers include the international community and IPDA.

4. PDS4 Projects

PDS4 Projects will be conducted in phases. They will include both full scale projects for deployment to PDS as well as feasibility projects which will evaluate a specific technology and/or solution within a subset of the PDS. The phases are planned to last from FY09 – FY11. Subsequent phases beyond FY11 would include building out specific user services once the infrastructure and support is in place.

- *PDS4 Phase I* – Phase I of PDS4 will focus on projects that will be used to create the foundation of PDS4, particularly standards and securing the data holdings of PDS. In addition, feasibility projects will be introduced earlier in the PDS4 development to provide evaluate design decisions early and help to provide initial capability quickly for PDS4. Feasibility projects will have a full deployment to PDS in the future, if successful.

  Phase I projects will be planned for FY09.

  - PDS4 Data Standards Project: The data standards project will focus on releasing v4 of the PDS Standards Reference. This should include a core set of data formats and simplified, unambiguous version of the standards reference per the recommendations of the PDS4 Data Model WG (*Data Model PDS3 Issues List*[2]*, User Model White Paper*)

  - PDS4 Technical Standards Project: The technical standards project will focus on defining v1 of the PDS Technical Standards. These will include hardware and software standards. (*Architecture White Paper*)

  - Data Integrity Project: The data integrity project will implement the PDS MC policy and the archive integrity requirements including data integrity, tracking and availability. It will address the "three" copy rule of data and ensure that data properly

[2] PDS3 Issues, Ed Guinness and Steve Hughes, February 5, 2008

archived and can be accessed as needed. (*Architecture White Paper*)

- Distributed Infrastructure and Services Project (Year 1): The Distributed Infrastructure and Services Project will evaluate technical solutions for moving PDS to a fully distributed, "service-oriented" architecture. This will be used to support integrated search, distribution, data delivery, and data processing (transformation, subsetting, etc). The project will focus on evaluation and testing within a subset of PDS and then schedule full deployment as a project in a later phase of PDS4. (*Source: Architecture White Paper)*

- Data Dictionary Enhancement: The data dictionary project will clean up the data dictionary, based on the PDS4 adopted standards. It will perform a streamlining to ensure that there is a core set of keywords. In addition, it will coordinate the data dictionary with the IPDA to distinguish between international, PDS, and node/mission-specific keywords. (*Source: Architecture White Paper*)

- Data Movement Feasibility Project: The Data Movement Feasibility Project will evaluate data movement technologies for PDS by selecting and implementing a data movement capability for a subset of PDS where high-speed data throughput is critical. (*Source*: *Architecture White Paper)*

- *PDS4 Phase II* – Phase II of PDS4 will focus on migration of the existing data and tool infrastructure to the PDS4 design and standards. This will include alignment of the technical infrastructure, cleanup of the data dictionary, migration of the repository to a PDS4 structure, construction of the PDS4 catalog system and migration from PDS3, and tools to support the submission of data to PDS.

  Phase II projects will be planned for FY10.

  - Data Migration & Clean Up Project: The Data Migration and Clean Up project will migrate data repositories from PDS3 to PDS4 as needed. The purpose is to be forward-compatible with PDS4. The PDS MC will need to consider the policy associated with migrating data. (*Source: Architecture White Paper*)

- PDS Catalog System: The PDS Catalog System project will implement a distributed catalog system for PDS, based on the PDS4 information model, and built on the PDS4 distributed services. (*Source: Architecture White Paper)*

- Tool Migration Project to PDS4: Changes to the PDS4 standards will most likely result in changes to the existing tools and system. This project will focus on the migration of the PDS tools, particularly VTOOL, LTDTOOL and NASAView, to support PDS4. (*Source: Architecture White Paper*)

- Distributed Infrastructure and Services Project (Year 2): The Distributed Infrastructure and Services Project will evaluate technical solutions for moving PDS to a fully distributed, "service-oriented" architecture. This will be used to support integrated search, distribution, data delivery, and data processing (transformation, subsetting, etc). The project will focus on evaluation and testing within a subset of PDS and then schedule full deployment as a project in a later phase of PDS4. (*Source: Architecture White Paper)*

- Portals, Search and Distribution Feasibility Project: The Portals, Search and Distribution Feasibility Project will evaluate various approaches for better integrating PDS and allowing product level search and access to data. This will also include evaluation of product level vs record level access to data. The results of this feasibility project will be used to guide full deployment to PDS in a later phase. (*Source: User Services White Paper*).

- *Phase III* – Will focus on implementing several of the feasibility projects for PDS4 which move PDS to a virtual system including distributed, service-oriented infrastructure, an integrated search architecture for "one-stop shopping", and high-speed data movement across the PDS enterprise.

  Phase III projects will be planned for FY11.

  - Technical Infrastructure/Operations Project: The Technical Infrastructure and Monitoring Project will focus on upgrading the technical infrastructure of PDS, based on the technical standards recommendations. It will also employ a monitoring capability for monitoring the quality of service for PDS operations. (*Source: Architecture White Paper*)

- Portal, Search and Distribution: The Portal, Search and Distribution project will focus on "re-architecting" PDS such that there is an integrated technical approach with respect to the user experience in navigating portals, searching for data, and requesting PDS data holdings. This needs to include the ability to support cross-node searching and access to data. (*Source: User Model White Paper*)

- Data Movement and Delivery Project: The Data Movement and Delivery Project will focus on adoption and deployment of technologies and standards to support high speed data movement across the PDS enterprise as well as to/from data producers and customers. It will focus on addressing movement of large-scale data products and will build on the feasibility project. (*Source*: *Architecture White Paper*)

- Ingestion and Data Producer Tool Suite Project: The Ingestion and Data Producer Tool Suite will focus on delivery of tools to data producers from missions to DAPs providing a comprehensive tool suite to support electronic submission of data to the PDS. (*Source: User Services White Paper*)

- *Phase IV* – Will focus on extended services once the service architecture is fully in place. These services are TBD. This will also focus on increased cataloging mechanisms and tools to support capture of user input into a knowledge-base.

  Phase IV projects will be planned for FY12.

5. Management Plan

- Implementation Team: Implementation team consists of members involved in development and integration. It crosses nodes of the PDS. Nodes are critical to the integration aspects of the implementation

- Operations: Develop an integrated view of our system operations and provide periodic presentations to PDS MC.

- Reporting: Reporting to the MC will be cross-cutting covering development, integration and operations which will span all nodes