



Transition, Migration, and Impact Discussion

PDS 2010 Tech Session

June 11, 2009

Discussion Topics

- Definitions
- Impact Process
- Example System and Data Trade-offs
- Migration Options
- Transition Options
- Build Plan
- Quantitative Analysis

Definitions

- Migration
 - Migration is the process of migrating existing data in the archive from PDS3 to PDS4
- Transition
 - Process of moving from the current system to a PDS 2010 system which includes ingestion and distribution of PDS4 data
- Impact
 - Assessment/implication of design, implementation and project decisions to PDS (DN/EN), data providers and users

Determining Impact

- Identify major design decisions and options
- Identify impact to PDS (DN/EN), data providers and users
- Identify impact to the PDS system
- Identify impact of transition and migration plans
 - Migration of PDS3->PDS4
 - Transition to PDS 2010

Impact Assessment

- Use set of criteria for evaluating decisions and impact
 - Resources (Staffing/Level of Effort)
 - Schedule
 - Training
 - Usability
 - Efficiency
- Evaluate major decisions and project plans
- Develop quantitative measurements
 - Not only cost, but also quantitative measures like # and type of data

Example Major Design Decisions

- PDS4 Data Design and Standards Mgmt
 - PDS data/information model
 - PDS grammar (XML vs ODL/PVL)
 - Data Dictionary (to ISO 11179)
- PDS System Design
 - PDS Distributed Infrastructure (Registry, Data Dictionary, etc)
 - Format Transformation Tools and Services (to/from PDS)
 - Federated Registries (including Product-level)

**What other critical decisions
should we analyze?**

Example Design Decisions

	Resources	Training	Usability	Efficiency
PDS4 Model	Short term impact on developing/migrating model and dictionary; long term costs to maintain standards decrease with more stability in the model	Data engineers, providers will need to be trained in order to develop PDS4 compliant data products	Usability for data providers and users improved if format transformation is in place	Improved archiving efficiency; better support for design, submission and use of PDS data
XML	Short term impact to re-write software for XML; longer term costs decrease by using open source libraries	Data engineers, providers and software developers will need to be trained in XML	XML data can be transformed for viewing purposes, etc	Will streamline software development of tools and services (e.g., data dictionary service)
ODL+/PVL	Short term impact to upgrade software for ODL+/PVL; longer term, costs will be incurred to maintain grammar language and software	Some re-training will be required	Will require development of specific ODL/PVL tools	Long term, ODL/PVL is a less efficient approach than XML due to its limited adoption, support and obsolescence risk

Example Design Decisions

	Resources	Training	Usability	Efficiency
Distributed Infrastructure	Major short-term impact will be on system development; Improved efficiency, tracking, access to data realized	Primary training issue is in developing software which can access online services over the network	Better support for design, validation and submission to PDS; more unified PDS-wide access to data; International access and support will be enabled.	Improved efficiency by reducing the number of versions of ad hoc tools and information
Format Transformation Tools and Services	Primary impact will be on developing tools to convert to/from PDS.	Primary training issue is in developing software that can perform transformation to/from PDS4	Improved usability by allowing users to access data in user preferred formats.	Improved efficiency for data providers and users in enabling more flexibility in the provider pipelines; long term efficiency improvement in the archive with standard formats

Example Design Decisions

	Resources	Training	Usability	Efficiency
Data Dictionary (ISO 11179)	Large impact on migrating PDS3 dictionary to a new structure and setting up a service (much of this is not driven by ISO 11179)	Training issue will be on managing the data dictionary	Improved validation; greater consistency; local governance	Streamlined and simplified process for managing data dictionary
Federated Registries (Including Product-level)	Primary impact will be on system development	Primary training issue is in developing software that uses registries	Improved tracking and management of PDS information; improved discovery of registered objects	Better overall management and integration of registered objects across PDS

Example Design Impacts

	Data Providers	Data Users	Discipline Nodes	Engineering Node
PDS4 Model	High short-term impact to data providers in software development; Improved efficiency in terms of substantial reduction in model changes and acceptance of standard formats	Low impact to data users in terms of training. Substantial benefit in terms of receiving data products in user-friendly format	High impact to data engineers and software development short term; Improved efficiency in terms of substantial reduction in model changes and standards support	High impact to software development; Medium impact to DEs; Long term reduction in standards management due to more stable model and data dictionary
XML	High short-term impact for existing software; long-term reduction in level of effort and recruiting individuals with knowledge of XML	Low technical impact, particularly if PDS can display labels in multiple structures (PVL could be one of them)	High short-term impact in node tools; long-term reduction in level of effort to maintain tools	High short-term impact in software development; long-term reduction in level of effort maintenance if using public domain libraries
ODL+/PVL	Medium impact on upgrading existing software; long-term impact on maintaining skill set and tools for ODL/PVL	Low impact except for data users who have tools that parse PDS labels	Medium impact on upgrading existing software; long-term impact on maintaining skill set and tools for ODL/PVL	Medium impact on software development; long term sustaining costs in maintaining grammar libraries

Example Design Impacts

	Data Providers	Data Users	Discipline Nodes	Engineering Node
Distributed Infrastructure	Minimal impact on data providers except to integrate with new ingestion tools/ services	Minimal impact on data users	Minimal short-term impact; Integration can occur over time to improve efficiency, tracking and search/ access	Primary short-term impact on Engineering Node to develop core infrastructure
Format Transformation Tools and Services	Minimal impact on data providers except to integrate with transformation tools/ libraries	Minimal impact on data user; expectation is that they will download data through online web services which improves usability.	Minimal impact on nodes except to integrate with libraries and tools and to support development of new transformation algorithms	Primary impact on Engineering Node to setup services and package libraries; the number of transformations can scale over time, as directed by PDS management

Example Design Impact

	Data Providers	Data Users	Discipline Nodes	Engineering Node
Data Dictionary (ISO 11179)	Minimal impact on data providers	Minimal impact of data users	Short term impact to learn how to interact with the data dictionary in a new structure and service	Impact to build and implement dictionary service and model
Federated Registries (Including Product-level)	Minimal impact	Minimal impact	Most services will not impose a major impact. Product-level registries will have an impact as it is hosted at the DN, but should be integrated easily with existing repository.	Impact to build and deploy registries, however, will reduce long term database management operations.

Major Migration and Transition Decisions

- Migration Decisions
 - What to migrate in the archive from PDS3->PDS4
- System Transition Decisions
 - Separate or combined support for PDS3 and PDS4
 - Phasing of PDS 2010 to support ingestion, distribution, user services

Migration Options

- No Migration
 - Users continue to have access to PDS3 and PDS4 data
 - Software/tools are in place to support both PDS3 and PDS4
- On-demand Migration
 - Users continue to have access to PDS3 and PDS4 data
 - Software/tools are in place to support both PDS3 and PDS4
 - Some data sets migrated and redelivered to NSSDC
- Full Migration
 - Users have access to data in PDS4 data format only
 - Users Software/Tools only support PDS4
 - All data migrated and redelivered to NSSDC
 - Accept PDS3 and PDS4 data from data providers

Data Migration Decisions

	Resources	Training	Usability	Efficiency
No Migration	No impact on resources to convert data, however, PDS software will need to support PDS3 and PDS4	Users will need to be capable of working with PDS3 and PDS4 data	Limited support for working with PDS3 data in the future	Most cost-effective solution
On Demand Migration	Impact on conversion of a subset of critical data sets; PDS software will need to support PDS3 and PDS4	Users will need to be capable of working with PDS3 and PDS4 data	Limited support for working with PDS3 data in the future; critical data sets will be converted to PDS4 to improve usability	More costly than “No Migration”, however, improves usability following a pragmatic approach
Full Migration	Substantial impact in converting data, redelivering to NSSDC, and developing supporting software	Users will ultimately need to only be familiar with PDS4	Usability would be improved since PDS data will be brought up to date	Substantial costs in migrating all data

Data Migration Impacts

	Data Providers	Data Users	Discipline Nodes	Engineering Node
No Migration	No impact	Users will need to be familiar with PDS3 and PDS4 data formats	Nodes will need to continue to provide support for PDS3 data	Engineering Node must continue to provide software support for PDS3 and PDS4
On Demand Migration	No impact	Users will need to be familiar with PDS3 and PDS4 data formats. However, critical data sets can be migrated to enhance usability.	Nodes will need to continue to provide <u>limited</u> support for PDS3 data; minor impact in migrating critical data sets	Engineering Node must continue to provide software support for PDS3 and PDS4
Full Migration	No impact	Data users will need to eventually only learn PDS4	Substantial impact in migrating and redelivering all data to PDS4	PDS3 tools and services can be retired once migration is complete

Comments?

Transition Options

- Big bang
 - Entire PDS 2010 system is stood up and then PDS transitions
- Incremental
 - Minimal PDS 2010 system is put in place and then PDS transitions
 - Increasing levels of capability are put in place over time to support enhanced user services and greater efficiency
 - Acceptance of PDS4 data can occur in stages (e.g., certain missions, etc)
- Both can address continued support for PDS3

System Transition Decisions

	Resources	Training	Usability	Efficiency
Big Bang	Resources will be fixed, however, the schedule can scale which can delay the release of PDS 2010	Not a major transition issues	If all services are in place first, then this will increase usability. However, deployment of PDS 2010 could be delayed longer affecting the ability to put PDS4 model in place	Less efficient of the two approaches since it delays rolling out PDS 2010
Incremental	Resources and the timeline can be scaled with PDS budget (whether overguide is provided or not)	Not a major transition issues	Critical capabilities are put in place earlier, but the trade off is number of functional capabilities that must be supported prior to ingestion/distribution	More efficiency of the two approaches since it allows for delivery of increasing capability and the MC can then determine functional capabilities required in order to accept and distribute PDS4 data

System Transition Impacts

	Data Providers	Data Users	Discipline Nodes	Engineering Node
Big Bang	Capabilities and services are delayed, however, greater functionality would be in place when they begin to deliver PDS4 data	Data users would get data in PDS4 format later, but could potentially have more tool support	Larger impact on the nodes if all nodes need to put services and capabilities in place first	Larger impact on the Engineering Node to put services and tools in place and ensure PDS-wide readiness to accept data for all new missions in PDS4 at once
Incremental	Capabilities can be put in place earlier and begin flow of PDS4 data and support	Data users would get data in PDS4 format earlier, but would possibly have less tool support	Less impact on the nodes if data is accepted in stages	Less impact on the Engineering Node and improved project performance since results can be realized earlier and PDS can deploy services and train over time

Transition Proposal

- Transition to PDS as a series of builds with increasing functionality
- Continue support for PDS3 ingestion and distribution
- Migrate PDS3 catalog to PDS4 registry services
- Impact: Principal impact will be on “migrating” central catalog to the new infrastructure and on building PDS3/PDS4 compatible tools. Should minimize impact to users, data providers and discipline nodes so the impact occurs more gradually.

Initial Build/Release Phasing

	Description	Impact	Timeline
Build 2010-1 (Draft Capabilities)	Initial infrastructure services released; draft release of PDS Standards Reference	Principal impact is on Engineering Node for the infrastructure release and the DDWG to release draft standards reference	
Build 2010-2 (Minimal Infrastructure to support PDS4 data ingestion and distribution)	PDS4 Data/Information Model release and standards reference Data Dictionary Service Upgraded validation tool Standard grammar Catalog registry Search/web portal changes	Principal impact will be getting tools and service and standard reference in place; PDS MC will need to determine phasing plan for acceptance of data	
Build 2010-3 (Increased User Capabilities)	to/from PDS transformation libraries and service Reconfigured web infrastructure to integrate with product registry Improved search across PDS	Improved services will be deployed to nodes including tracking and access to data products	

Quantitative Analysis (estimates for the MC)

- Capture some critical numbers to help in the decision-making process
- Capture FTEs rather than cost with estimates at across DNs and EN for major tradeoff items
 - Recommend we be careful about the granularity
- Provide rolled up detail to MC (in 2 charts for example) at Aug MC

Quantitative Analysis (estimates for the MC)

- Cost metrics
 - FTE support for critical decisions such as
 - Level of effort of upgrading software in PDS (DN and EN) to support PDS4 ?
 - Level of effort of developing toPDS and fromPDS services for transformation?
 - Level of effort of moving to XML vs PVL/ODL?
 - Question: to what granularity do we want to do this?
- Data metrics
 - What percentage of PDS3 formats will continue to be accepted in PDS4?
 - ISIS 2 Qube, ISIS 3 Qube, VICAR, FITS, ...
- What else?

Backup

