

Operations Concept

PDS 2010 System Review
March 22-24, 2010

Sean Hardman

Topics

- Overview
- Planning Phase
- Definition and Design Phase
- Production Phase
- Distribution and Maintenance Phase
- Wrap Up

Overview

- The Operations Concept relates the existing PDS policies, processes and procedures to the services and tools that constitute the PDS 2010 system.
- The policies, processes and procedures that currently exist for PDS will be carried forward and optimized for the PDS 2010 system.
- The approach taken here is to follow the PDS Archive Lifecycle.

Lifecycle of a PDS Archive

A PDS data archive has several phases, which are often cyclical as data is augmented or updated by the data suppliers.

Planning Phase

- Data archiving requirements written into mission Announcement of Opportunity
- Pre-proposal briefing on PDS data archiving requirements given to potential proposers
- Proposal data archiving section reviewed by PDS
- PDS orientation to flight project staff
- Data archiving working groups formed

Definition and Design Phase

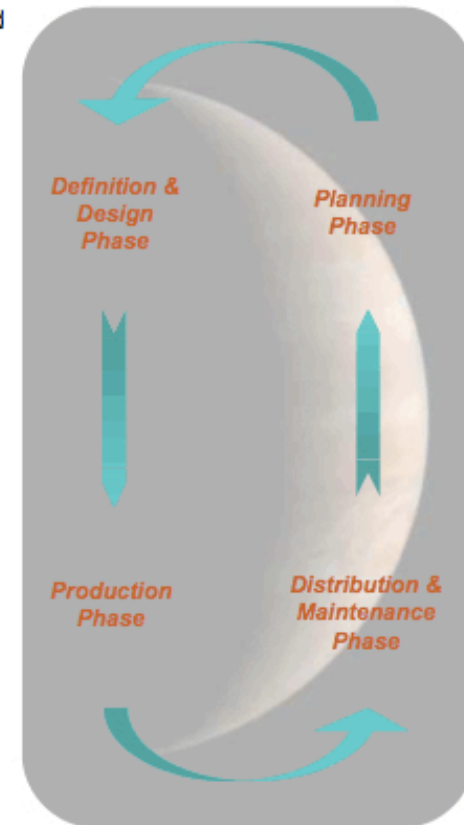
- Project Data Management and Archive Plans define data to be archived
- Data Product and Volume Organization Software Interface Specifications (SISs) detail the data and volume structure
- Preliminary metadata labels loaded into PDS catalog

Production Phase

- Raw and processed data products, labels (metadata), and documentation produced
- Preliminary and quick-look data made accessible via Project and PDS web pages
- Data archive products validated and peer-reviewed; liens corrected

Distribution and Maintenance Phase

- Final data products made available on-line
- PDS add the data to the archive
- Physical copies sent to NSSDC
- PDS provides data, documentation, and science expertise to users
- Data archive maintained via periodic refreshes, addition of new/updated data products



Topics

- Overview
- Planning Phase
- Definition and Design Phase
- Production Phase
- Distribution and Maintenance Phase
- Wrap Up

Overview

- There is minimal interaction with the system or tools during this phase of the lifecycle.
- Discipline Node staff continue to work with proposers and Data Providers and make existing resources available:
 - Proposers Archiving Guide (PAG)
 - Archive Preparation Guide (APG)
- Very little change from how this phase is approached now.

Topics

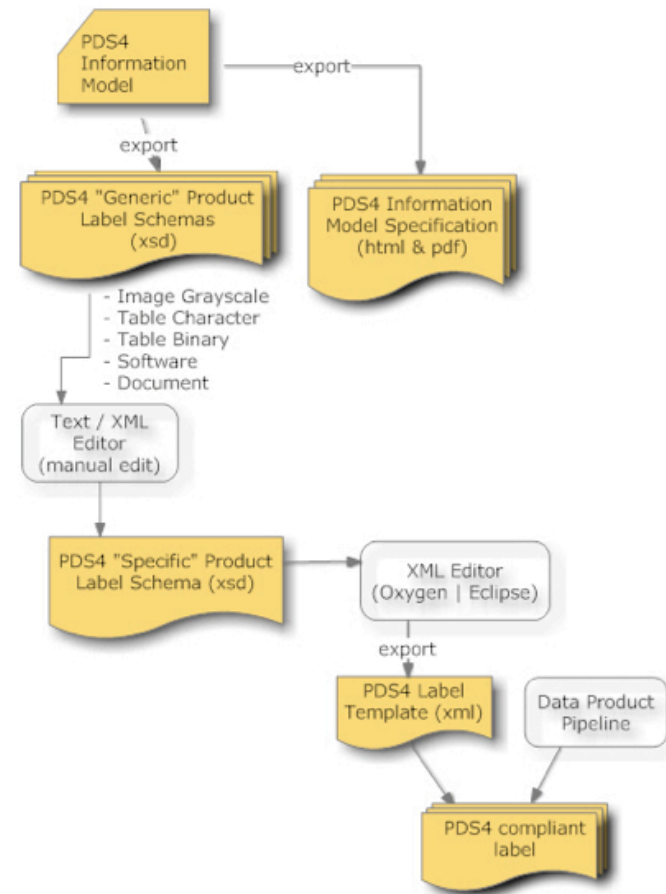
- Overview
- Planning Phase
- Definition and Design Phase
- Production Phase
- Distribution and Maintenance Phase
- Wrap Up

Overview

- This phase sees the first interaction with the system and tools.
 - Data Providers utilize tools for design of product label schemas and creation of example product labels.
 - A representation of the data dictionary, available through a service, is utilized during schema creation.
 - The resulting schemas and initial catalog information (e.g., investigation, observation system) are ingested into the system.
- New and updated resources are available to Data Providers and Data Engineers that include:
 - Data Providers Handbook
 - Standards Reference

Product Label Schema Lifecycle

- Generic schemas are derived from the Information Model.
- These schemas can be used as templates for defining new products via specific schemas.

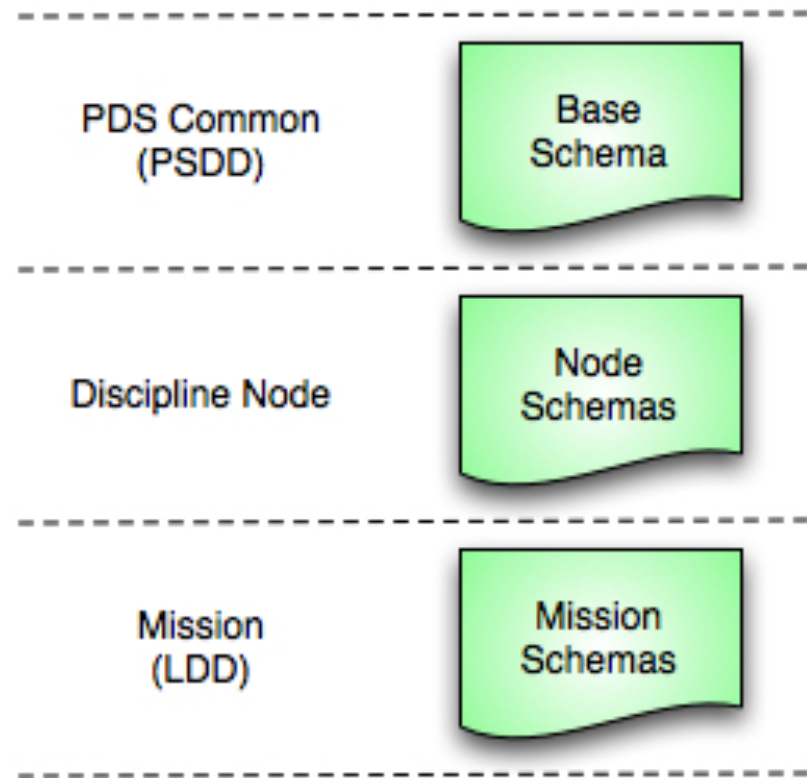


Steps to Creating a Product Schema

1. Select an appropriate generic product schema (e.g., Image Grayscale) from the repository.
2. Use the Design Tool (XML Editor) to tailor accordingly.
 - Tailoring examples include restricting keyword value set, removing unnecessary keywords, etc.
3. This may require a few iterations.
4. Create a sample label from this schema.
 - The label will be valid when generated from the schema. It should also be valid according to the generic schema.
 - Review the label for structure and content.

Namespaces

- Used to minimize global changes to the data dictionary and to coordinate at multiple levels.
- Enables governance allowing management at each level.



Data Dictionary

- The data dictionary captures object, group and element definitions for use in product label schemas.
- These definitions are associated with defined namespaces and are managed in the Information Model utilizing the ISO 11179 standard.
- A service contains data dictionary content exported from the model to support search and reference.

System Interaction

- User interaction with the system is limited to the following:
 - Initial entry and update of “catalog” information.
 - Registration of specific product label schemas.
- The System provides a PDS-internal portal to support these interactions.
- The portal allows Data Engineers to create, update and modify information in the registry.
- The registry manages the catalog and product information across the Nodes.
 - Part of a federated registry implementation.

Improvements

- Generic schemas derived from the model enable consistent metadata capture which enhances product-level search and discovery.
- Specific schemas provide for consistent quality and validation of product labels.
- A standards-based implementation for the data dictionary enables more explicit definitions and realizes namespaces.
- Common registry implementation facilitates federated registries.

Topics

- Overview
- Planning Phase
- Definition and Design Phase
- Production Phase
- Distribution and Maintenance Phase
- Wrap Up

Overview

- This phase sees the following interaction with the system and tools:
 - Data Providers and Data Engineers utilize tools for generation, transformation and validation of products.
 - Data Engineers utilize services for receiving and ingesting product deliveries.
 - The system captures and manages catalog and product metadata.

Product Preparation

- A suite of tools are available to aide in product preparation including:
 - Generation of product labels within the data pipeline.
 - Transformation of product data into conformant PDS4 formats.
 - Validation of product labels and data.
- The preference is that the Data Provider performs these functions pre-delivery.
- Some arrangements may call for PDS Data Engineers to perform transformation post-delivery.

Product Delivery

- Product delivery for the first build will not change much from current practice.
- Products are transferred to the cognizant Node via current mechanisms.
 - This includes FTP and data brick.
- Other data movement options are planned in future builds.

Archive Preparation

- One of the goals for PDS 2010 is to minimize this activity.
- For now, the assumption is that a final validation would be performed prior to ingestion.

Ingestion

- The focus is on harvesting metadata from the online archive.
- A configurable tool supports varying package structures and metadata mapping where necessary.
 - Enables harvesting from PDS3 archives.
- The registry captures product metadata.
 - Facilitates tracking and notification.
- Metadata in the registry can be augmented over time to enhance search and understanding.

Improvements

- Common tool-based software available for generation of product labels and transformation of data.
- Common service-based software for capturing and managing product metadata that can be deployed at the Nodes.
- Facilitation of metadata augmentation (post-registration) in order to enhance the search experience.

Topics

- Overview
- Planning Phase
- Definition and Design Phase
- Production Phase
- Distribution and Maintenance Phase
- Wrap Up

Overview

- This phase sees the following interaction with the system:
 - Data Consumers search catalog and product information via API-based and portal interfaces.
 - Distribution of data from search results and delivery of data to the deep archive.

Search

- The system generates specialized search indices from the federated registries for improved query performance and specific search scenarios.
- A common interface accepts queries for catalog and product-level content.
- Each Node may develop their own specific interface tailored to an index of their data.
- Cross-Node searches are enabled where it makes sense to do so.

Data Distribution

- Existing distribution mechanisms remain in place (e.g., FTP, HTTP, Data Brick, etc.).
- New mechanisms for ordering and distribution will incorporate the transformation functions of format conversion, coordinate transformation, subsetting and packaging.
- Specific packaging for distribution to the Deep Archive is also supported.

Improvements

- Metadata captured in the registries is indexed to facilitate fast query response and specialized query scenarios.
- Common service-based software enabling local and cross-node searching deployed at the Nodes.
- Enhanced delivery of data incorporating transformation and updated data movement technologies.

Topics

- Overview
- Planning Phase
- Definition and Design Phase
- Production Phase
- Distribution and Maintenance Phase
- Wrap Up

Wrap Up

- Although PDS 2010 is a significant upgrade with regard to data formats, standards adoption and software deployment, the overall design strives to minimize the impact on the Data Providers and the Nodes.
 - More details provided when discussing the System Architecture and Service Design.
- Data Consumers on the other hand will be impacted, but in a good way. They will see better integration amongst the Nodes and improved access to data.

Questions / Comments