PDS 2010 Project Document

NASA Planetary Data System Operations Concept

Change Log

Revision	Date	Description	Author
0.1	2010-01-14	Initial draft release.	E. Law
0.2	2010-01-26	Added the Information Model and Archive Lifecycle operations scenarios.	R. Joyner
0.3	2010-02-01	Added the System Service operations scenarios.	S. Hardman
0.4	2010-02-15	Synched up the use cases and moved the Archive Lifecycle use cases to the end of the document.	S. Hardman
0.5	2010-03-09	Removed the Operations Environment section.	S. Hardman

Contents

1 Introduction

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

Key design decisions and recommendations, which arose in support of the above challenges / goals, include:

- Replace PDS3 ad hoc information model with a PDS4 information model that is now managed in modern tools

- Replace PDS3 ad hoc product definitions with PDS4 products that are defined in the model
- Require product formats to be derivations from a core set – support transformation from a core set
- Replace "homegrown" PDS data dictionary structure with an international standard – ISO11179 RIM
- Adopt a modern data language / grammar (XML) where possible for all tool implementations
- Adopt system of registries in support of improved tracking and access
- Support remote access to data and services to bring the PDS federation closer together both for ingestion and distribution

1.1 Purpose

The purpose of this document is to describe the operations concepts for PDS 2010 and how the PDS data architecture (e.g., data model, data structures, data dictionary, etc) and the PDS software system (online data services, distributed catalog, etc) will work across the PDS federation throughout the PDS Archive Lifecycle.

1.2 Scope

This document describes the PDS 2010 system operational concept. The focus of this document is on the PDS 2010 system operations throughout the PDS Archive Lifecycle, the functionality users need from the PDS 2010, and use scenarios. It does not address implementation and design issues.

1.3 Document Revision

Revisions of this document will be held in the PDS Engineering Node website through the use of its document history functionality. Previous versions of this document can be accessed through the use of that tool.

1.4 Applicable Documents

1.4.1 Controlling Documents

[1] Planetary Data System Strategic Roadmap 2006 - 2016, February 2006.

[2] Planetary Data System Level 1, 2 and 3 Requirements, August 2006.

[3] PDS 2010 Project Plan, February 2010.

1.4.2 Referenced Documents

[4] PDS 2010 Project Executive Summary, July 2008.

[5] Planetary Data System (PDS) 2010 System Architecture Specification, February 2010.

2 System Overview

PDS is an active archive providing documented data to the planetary science community. It was established in response to requests for increased availability of planetary data and to concerns that data from past NASA missions were becoming unusable. Current PDS functions include ingestion and distribution of data at 'nodes' specializing in certain scientific disciplines and/or technical skills, a central catalog and web accessible system that provide high level information on node holdings. PDS 2010 system will modernize PDS leveraging a modern, online, distributed services design to provide new and improved capabilities to meet the following objectives:

- Improve efficiency of archiving data with PDS
- Improve stability of the PDS archive and usability of data in the archive long term
- Improve access and usability of PDS overall

2.1 System Requirements and Constraints

The top-level requirements for the PDS project are defined by the PDS Level 1, 2 and 3 requirements [2] that were signed off by the Management Council in August 2006. The PDS 2010 architecture and services are derived from these as well as from the PDS 2010 drivers and directives from the PDS Management Council in order to meet the objectives of the project.

2.2 System Descriptions

The PDS 2010 system is designed using a Service-Oriented Architecture. The system infrastructure provides core services that allow the system to be extensible and enable new functionality to be added over time. The following is the set of core services for PDS 2010:

- Harvest/Ingest Service: to support product ingestion.
- Inventory Service: for cataloging ingested products.
- Dictionary Service: for managing the PDS data dictionary.
- Search Service: to facilitate remote access to catalog and product resources.
- Subscription Service: for notifying data consumers of data availability.
- Report Service: for centralized metrics collection and reporting across all

nodes.
- Monitor Service: to support system monitoring and tracking
- Security Service: for managing username/passwords so common tools that require authentication can reuse.

In addition, a set of PDS 2010 tools will be provided to support:

- Design of PDS product labels/schemas
- Generation of PDS products
- Validation of PDS4 product labels and digital products
- Transformation of product formats/types (into PDS)
- Transformation of product formats/types (distribution from PDS)

3 Operations Concept Stakeholders / Users

The PDS 2010 system operations concept affects a number of stakeholders. They are identified as follows:

Data Provider
This group represents the mission, instrument team and NASA-funded researcher who are involved with delivering data to the PDS.

Data Consumer
This group represents the Planetary Scientist, which includes those experienced with solar system exploration missions and those who are mission-naïve. They include graduate students.

Data Engineer
This group represents a portion of the PDS Technical group that curates the data before and after it enters the PDS system.

Operator
This group represents a portion of the PDS Technical group that is responsible for configuring and monitoring the system.

Standards Coordinator
This group represents a portion of the PDS Technical group that is responsible for maintaining the Standards Reference, for conducting the Standards teleconferences, and for maintaining the Data Providers Handbook.

4 PDS 2010 Operations Concept

The PDS 2010 Operations Concept closely follows the PDS Archive Lifecycle. This document relates PDS policies, processes, system services and tools usage to the functions that are performed throughout the PDS Archive Lifecycle phases.

4.1 Archive Life Cycle

PDS, as a project, participates in all phases of development of a mission, from the proposal stage (pre-phase A) through operations (phase E). These phases are mapped to an end-to-end PDS Archive Lifecycle that includes planning, definition and design, production and distribution and maintenance of the archive data. The PDS Archive Lifecycle identifies the functions that are conducted during each of these phases. A high-level diagram is provided below:
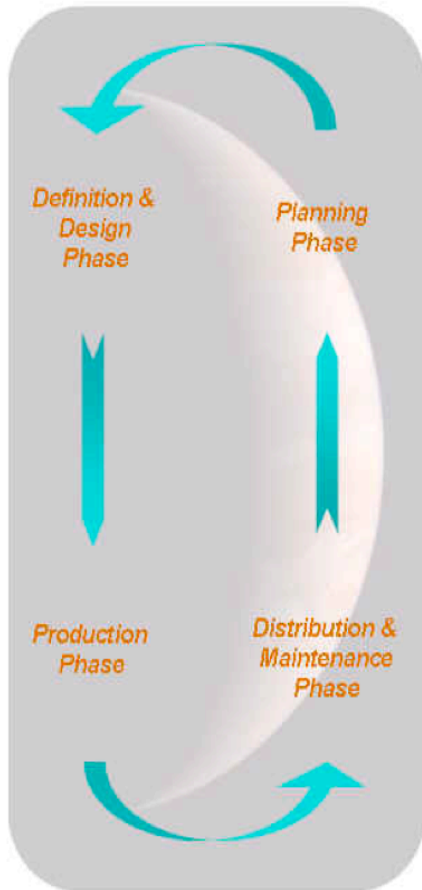
Planning Phase:
- Data archiving requirements written into mission Announcement of Opportunity
- Pre-proposal briefing on PDS data archiving requirements given to potential proposers
- Proposal data archiving section reviewed by PDS
- PDS orientation to flight project staff
- Data archiving working groups formed

Definition & Design Phase:
- Project Data Management and Archive Plans define data to be archived
- Data Product and Volume Organization Software Interface Specifications 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 & 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 media refreshes, addition of new / updated data products

Definition & Design Phase Planning Phase

Production Phase Distribution & Maintenance Phase

4.2 Operations Policy, Process and Procedure

It is envisioned that most the existing PDS policies, processes and procedures will continue to be applicable and exercised in the evolution towards the PDS 2010 era. Others will be either improved or created to enhance operations and usage of PDS 2010 system services and tools.

Planning Phase

Very little change from how this phase is currently approached is anticipated. Resources describing processes and procedures available to Data Providers and Data Engineers such as the existing Proposers Archiving Guide (PAG) and the Archive Preparation Guide (APG) will be revised in concert with the evolution towards PDS 2010.

Definition and Design Phase

Data Providers and Data Engineers will need to understand the PDS4 data model and how to create and ingest product label schemas. Processes, procedures and user guide will be captured in the Data Providers Handbook and the PDS4 Standards Reference.

Production Phase

Data Providers and Data Engineers will need to understand the PDS4 data model and how to create and ingest products. Processes, procedures and user guide will be captured in the Data Providers Handbook and the PDS4 Standards Reference.

Distribution and Maintenance Phase

By following the data release process, a Data Engineer will work with the Operator on data release allowing Data Consumers to access PDS data. Data release notification will be sent to Data Consumers via the Subscription Service. Data Engineer will deliver products to the deep archive facility NSSDC by following the NSSDC delivery process. Metrics reports will also be periodically generated during this phase.

4.3 System Services and Tools

Planning Phase

For the most part, there will be no interaction with the PDS 2010 system services or tools during this phase of the lifecycle. The possible exception will be a project/mission portal where the archive working groups would have access to a repository of mission information.

Definition and Design Phase

During this phase, the PDS 2010 design tool will be utilized for creating PDS4 product label schemas and example labels. The PDS 2010 validation tool, which will utilize the product label "specific" schema to specify the syntactic and semantic constraints, will be used to validate the product labels. A representation of the data dictionary, managed by the PDS Dictionary Service, will be utilized during schema creation. The service will import and export dictionary content using the ISO/IEC 11179 standard. The resulting schemas and initial catalog information (e.g., mission) will be ingested into the system using the PDS 2010 Registry Service. This phase will also be the entry point to update the "catalog" information. A PDS-internal portal will be provided to support this interaction. The portal will allow Data Providers and Data Engineers to create, update and modify catalog information. This portal

interface will allow users to manage catalog content within their namespace. The Registry/Inventory Service will be used to manage the catalog information across the Nodes.

Production Phase

During this phase, the key activities include archive product peer review, generation and ingestion. The PDS 2010 product validation tool will be utilized to validate the products. The Harvest/Ingest Services, PDS-internal portal and Registry/Inventory Service will be used to ingest those products into PDS and ensure the metadata descriptions are properly registered and cataloged into the PDS Registry. The Registry/Inventory Service will allow PDS to track and locate what have been ingested and archived into PDS 2010 system, and ensure integrity of the archival products and their metadata.

Distribution and Maintenance Phase

During this phase, Data Providers continue producing and delivering archival products/collections to the PDS. The product validation tool and the Ingest Subsystem will continue to be used to support archive of those products, and make them available to users via the PDS 2010 Distribution Subsystem. The Distribution Subsystem will support coordination of data release, ensure back up and deep archive copies are delivered to their designated target.

In addition, a PDS 2010 Report Service will be available to capture and report metrics generated by PDS 2010 services as well as metrics generated by commercial components such as web logs and FTP logs.

5 Operations Scenarios

An operations scenario represents a capability of the system and specifies how and why the user (actor) interacts with the system. The operations scenario describes the interaction at a high enough level so as to not reveal or imply the internal structure of the system.

The following definitions are useful in understanding terms used within this document:

PDS4 Schema
An XML Schema Document (XSD) – a document written in the XML schema language. Like all XML schema languages, XSD can be used to express a set of rules to which an XML document must conform in order to be considered "valid" according to that schema.

Generic Schema
An XML Schema Document (XSD) where, by definition, the metadata includes optional classes and/or attributes.

Specific Schema
An XML Schema Document (XSD) where, by definition, the metadata includes only required classes and attributes. The Specific schema represents the overall structure and format of the product and defines, in the strictest sense, the greatest latitude permissible in the validation of the archived product to ensure PDS compliance.

Product Schema
The Product Schema, which can be either generic or specific in character, is representative of one of the basic fundamental component classes (e.g., Image_Grayscale, Table_Binary, Header, Document, etc). This schema is typically the parent to any referenced discipline schemas. Generic-product schemas are the precursors for specific-product schemas.

Discipline Schema
The Discipline Schema, which can be either generic or specific in character, is representative of a class that is applicable to (re)use for a particular discipline (e.g., node, mission, instrument, etc). These schemas are representative of a more singular-use / particular-use schema (e.g., camera_model, cartography, rings_information, etc). Generic-discipline schemas are the precursors to specific discipline schemas.

5.1 Actors

An actor is an object (e.g., organization, person, application, etc.) outside the scope of the system but has some interaction with the service(s) provided by the system. See Section 3 for a list of the actors / stakeholders / users.

5.2 Information Model Operations Scenarios

This section captures operations scenarios relating to how the Information Model either directly or indirectly will be used within the system. Figure 5.2-1 depicts the lifecycle of a product label, as derived from the Information Model.
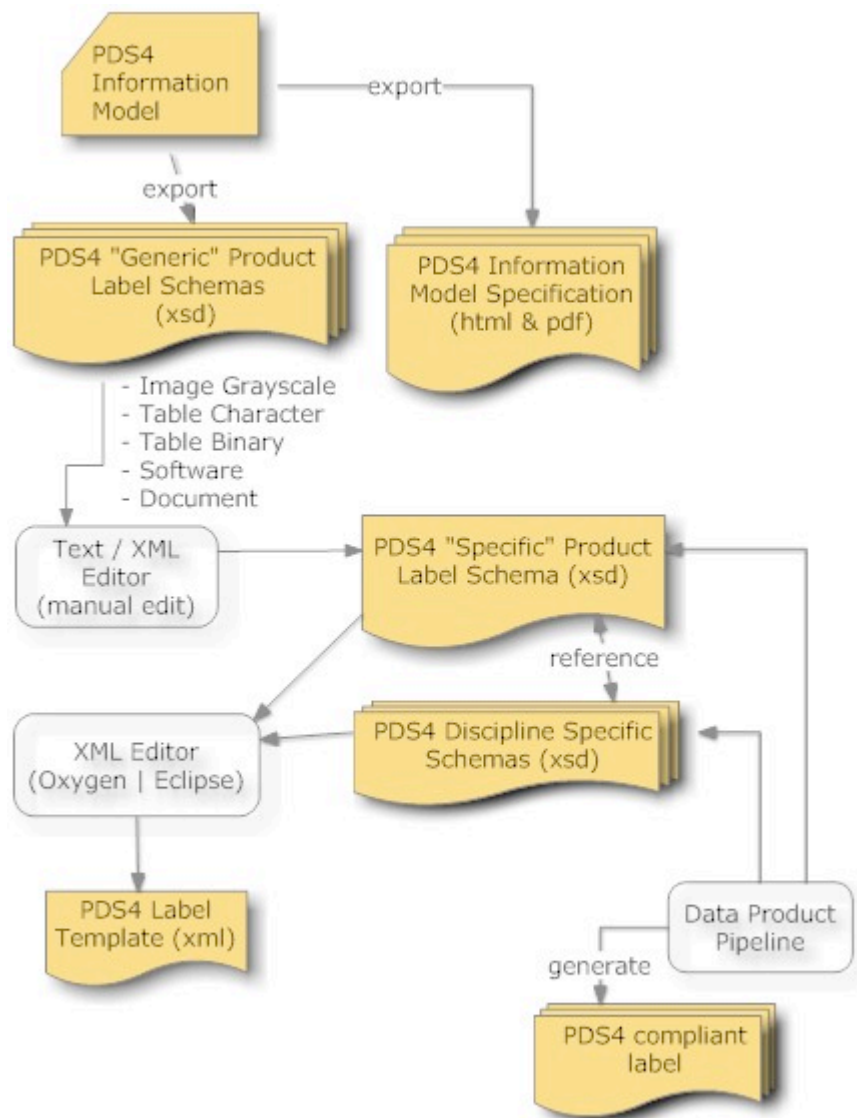
Figure 5.2-1.   Lifecycle of a Product Label Schema

From Figure 5.2-1, Lifecycle of a Product Label Schema, use cases can be derived and are detailed in the following sections.

5.2.1 Manage Information Model

The PDS receives a request that affects the content of the PDS Information Model.

1. Data Provider wants to create a new class, make an update to an existing class, or delete an existing class.
2. Data Engineer accesses the PDS Information Model (via the protégé tool) interface and performs the requested operation. The Data Engineer will need to be authenticated and have sufficient rights / privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.
3. Data Engineer notifies the Data Provider that the updates have been made.

Alternative: Create Operation
At step 2, the operation is to create a new class.
   a. Data Engineer enters identifying metadata into the Information Model.
   b. Data Engineer adds the identifying metadata as one or more classes; including the parent / child relationships of the classes.
   c. Return to primary scenario at step 3.

Alternative: Update Operation
At step 2, the operation is to update an existing class.
   a. Data Engineer enters identifying metadata into the Information Model.
   b. Data Engineer optionally adds or removes the identifying metadata to/from one or more classes.
   c. Return to primary scenario at step 3.

Alternative: Delete Operation
At step 2, the operation is to delete an existing class.
   a. Data Engineer removes identifying metadata from the Information Model; including the parent / child relationships to other classes.
   b. Return to primary scenario at step 3.

5.2.2 Publish Information Model

The PDS has made updates to the Information Model in response to requests that affected the content of the PDS Information Model.

1. Data Engineer determines that sufficient changes have been made to warrant a publication of a newer version of the Information Model.
2. Data Engineer publishes the Information Model, in the form of revised schemas, to the schema repository.
3. Data Engineer exports Information Model to an online searchable document.
4. Data Engineer notifies the PDS Data Provider community that a newer version is available.

5.2.3 Manage Data Dictionary Model

The PDS receives a request that affects the content of the PDS Data Dictionary Model.

1. Data Provider wants to create a new namespace, a new attribute, make an update to an existing attribute, or delete an existing attribute.
2. Data Engineer accesses the PDS Data Dictionary Model (via the protégé tool) interface and performs the requested operation. The Data Engineer will need to be authenticated and have sufficient rights / privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.
3. Data Engineer notifies the Data Provider that the updates have been made.

Alternative: Import Local Data Dictionary Operation
At step 2, the operation is to import discipline-specific attributes and standard-values.
   a. Operator imports locally-defined metadata into the Data Dictionary Model.
   b. Return to primary scenario at step 3.

Alternative: Create Operation
At step 2, the operation is to create a new namespace, a new common attribute, or standard-value.
   a. Operator enters identifying metadata into the Data Dictionary Model.
   b. Operator adds the identifying metadata against one or more attributes
   c. Return to primary scenario at step 3.

Alternative: Update Operation
At step 2, the operation is to update an existing namespace, common attribute, or standard-value.
   a. Operator enters identifying metadata into the Data Dictionary Model.
   b. Operator optionally adds, updates, or removes the identifying metadata against one or more attributes.
   c. Return to primary scenario at step 3.

Alternative: Delete Operation

At step 2, the operation is to delete an existing namespace, common attribute, or standard-value.

    a. Operator removes identifying metadata from the Data Dictionary Model.

    b. Return to primary scenario at step 3.

5.2.4 Publish Data Dictionary Model

The PDS has made updates to the Data Dictionary Model in response to requests that affected the content of the PDS Data Dictionary Model.

1. Data Engineer determines that sufficient changes have been made to warrant a publication of a newer version of the Data Dictionary Model.
2. Data Engineer publishes the Data Dictionary Model, in the form of a revised schema, to the schema repository.
3. Data Engineer exports Data Dictionary Model to an online searchable document.
4. Data Engineer notifies the PDS Data Provider community that a newer version is available.

5.2.5 Update Standards Reference and Data Providers Handbook

In tandem with the PDS publishing a new version of the Information Model, the Standards Reference and Data Providers Handbook will be updated.

1. Standards Coordinator updates the Standards Reference and Data Providers Handbook in tandem with changes made to the Information Model.
2. Standards Coordinator publishes the Standards Reference and Data Providers Handbook documents.
3. Standards Coordinator notifies the PDS community that a newer version of each document is available.

Note that the Standards Reference and Data Provider Handbook may be published independently of changes having been made to the Information Model. Software Change Requests (SCRs) will continue to be the driving force for updates to the Standards Reference and the Data Providers Handbook.

5.2.6 Retrieve Product Schema from Schema Repository

The Data Provider wants to use a generic-product or (re)use a specific product schema as a starting point for creating a specific product schema that more closely describes the data to be archived.

1. Data Provider, with the possible assistance from a PDS/DN, ascertains which product schema most closely matches the format of the data to be archived.
2. Follow the steps from use case 5.3.2 Retrieve Schema(s) from PDS.
3. Data Provider uses an XML editor (e.g., Oxygen or Eclipse) or an application (e.g., java, python, etc) to ensure the schema is fully formed (i.e., ensure there are no errors present in the XML).

Note in order for schemas to be searchable/accessible, they must have been previously registered and stored in the PDS4 online schema repository.

5.2.7 Retrieve Discipline Schema from Schema Repository

The Data Provider wants to use a generic discipline or (re)use a specific discipline schema as a starting point for creating a schema that more closely describes the metadata to be archived.

1. Data Provider, with the possible assistance from a PDS/DN, ascertains which discipline-schema(s) are applicable to the metadata to be archived.
2. Follow the steps from use case 5.3.2 Retrieve Schema(s) from PDS.
3. Data Provider uses an XML editor (e.g., Oxygen or Eclipse) or an application (e.g., java, python, etc) to ensure the schema is fully formed (i.e., ensure there are no errors present in the XML).

Note in order for schemas to be searchable/accessible, they must have been previously registered and stored in the PDS4 online schema repository.

5.2.8 Create Specific Product Schema from Generic Schema

The Data Provider wants to use a generic-product schema as a starting point for creating a specific product schema with a goal of having the specific schema describe the overall structure and format of the archived product. The specific schema defines, in the strictest sense, the greatest latitude permissible in the validation of the product labels to ensure PDS compliance.

1. Data Provider saves a copy of the generic product schema as the initial cut at generating the specific product schema.
2. Data Provider opens the copy of the generic product schema in an XML editor (e.g., Oxygen or Eclipse) or in a text editor (e.g., UltraEdit or BBEdit) to examine the XML and begin tailoring the schema to more closely describe the structure and format of the product to be archived.
3. Tailoring includes, but is not limited to:
    • Restricting the value set for appropriate keywords.
    • Restricting ranges for appropriate keywords.

- Removing unnecessary instances.
4. Data Provider optionally incorporates Node-specific or Mission-specific content.

Alternative: Create Specific Product Schema from product label(s)
At step 1, the operation is to gather a representative set of product labels.
   a. Data Provider puts together one or more product labels that are representative of all combinations/permutations of the range of values for each product attribute (i.e., the goal is to identify/define all possible values for each keyword in the label).
   b. Data Provider submits the set of product labels to the PDS for formal "schematizing" (i.e., for inclusion into the PDS4 Information Model).
   c. Data Engineer reviews the set of product labels.
   d. Data Engineer accesses the PDS Information Model (via the protégé tool) interface and creates the specific schema from the product labels. The Data Engineer will need to be authenticated and have sufficient rights/privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.
   e. Data Engineer generates specific schema from the model
   f. Data Engineer and Data Provider iterate on the schema definition until both are satisfied that the specific-schema, defined in the PDS4 Information Model, defines the overall structure and format of the archived product.
   g. Data Provider and the PDS/DN complete the design of a new specific-schema.

Alternative: Create Specific Product Schema from scratch
At step 1, create an XML-compliant schema from scratch.
   a. Data Provider drafts an XML-compliant schema using an XML editor (e.g., Oxygen or Eclipse) or in a text editor (e.g., UltraEdit or BBEdit).
   b. Data Provider uses an XML editor (e.g., Oxygen or Eclipse) or an application (e.g., java, python, etc) to ensure the schema is fully formed (i.e., ensure there are no errors present in the XML).
   c. Data Provider submits the draft product schema to the PDS for formal "schematizing" (i.e., for inclusion into the PDS4 Information Model).
   d. Data Engineer reviews the draft schema.
   e. Data Engineer accesses the PDS Information Model (via the protégé tool) interface and creates the specific-schema. The Data Engineer will need to be authenticated and have sufficient rights/privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.
   f. Data Engineer generates specific schema from the model
   g. Data Engineer and Data Provider iterate on the schema definition until both are satisfied that the specific schema, defined in the PDS4

> Information Model, defines the overall structure and format of the archived product.
> h. Data Provider and the PDS/DN complete the design of a new specific schema.

5.2.9 Create New Generic Product Schema

The Data Provider, with the assistance of a PDS/DN, has ascertained that there isn't a generic-product schema that can describe the overall structure and format of the product to be archived.

1. Data Provider, with the assistance of a PDS/DN, ascertains that there isn't a generic-product schema that can describe the overall structure and format of the product to be archived (e.g., movie)
2. Data Provider and the PDS/DN design a new generic-schema.
3. PDS/DN submits the new generic-schema to the PDS for technical approval.
4. If approved, the PDS will:
    a. Incorporate the new generic-schema into the PDS4 Information Model.
    b. Generate a generic-schema from the model.
    c. Register the generic-schema with the PDS4 schema repository.
    d. Inform the PDS community.
5. If not approved, the PDS will either:
    a. Make recommendation as to which PDS4 product schema is appropriate to use, or
    b. Make recommendation as to which PDS compliant data format the Data Provider should be using to archive with the PDS

5.2.10 Create Specific Discipline Schema from Generic Schema

The Data Provider wants to use a generic discipline schema as a starting point for creating a specific discipline schema with a goal of having the specific schema provide the necessary descriptive information that the Data Provider would like to include as part of the product metadata. The specific schema defines, in the strictest sense, the greatest latitude permissible in the validation of the product labels to ensure PDS compliance.

1. Data Provider saves a copy of the generic discipline schema as the initial cut at generating the specific discipline schema.
2. Data Provider opens the copy of the generic schema in an XML editor (e.g., Oxygen or Eclipse) or in a text editor (e.g., UltraEdit or BBEdit) to examine the XML and begin tailoring the schema to provide the necessary descriptive information that the Data Provider would like to include as part of the product metadata Tailoring includes, but is not limited to:

- Restricting the value set for appropriate keywords.
- Restricting ranges for appropriate keywords.
- Removing unnecessary instances.

3. Data Provider optionally incorporates Node-specific or Mission-specific content.

Alternative: Create Specific Discipline Schema from another Specific Schema
At step 1, repeat the above steps using a specific discipline schema as the starting point instead of a generic schema.

5.2.11 Create New Discipline Specific Schema

The Data Provider, with the assistance of a PDS/DN, has ascertained that there isn't a discipline-specific schema that can provide the necessary descriptive information that the Data Provider would like to include as part of the product metadata (e.g., a new cartography model as yet not archived by the PDS).

1. Data Provider, with the assistance of the PDS/DN that has cognizance over the discipline, ascertains that there isn't a discipline-specific schema that can provide the necessary descriptive information that the Data Provider would like to include as part of the product metadata.
2. Data Provider, with the assistance of the PDS/DN searches the PDS4 schema repository for a close-matching-schema that can be used as a starting point for creating the new discipline-specific schema
3. Data Provider drafts an XML-compliant schema using an XML editor (e.g., Oxygen or Eclipse) or in a text editor (e.g., UltraEdit or BBEdit).
4. Data Provider uses an XML editor (e.g., Oxygen or Eclipse) or an application (e.g., java, python, etc) to ensure the schema is fully formed (i.e., ensure there are no errors present in the XML).
5. Data Provider submits the draft discipline-schema to the PDS for formal "schematizing" (i.e., for inclusion into the PDS4 Information Model).
6. Data Engineer reviews the draft schema.
7. Data Engineer accesses the PDS Information Model (via the protégé tool) interface and creates the specific schema. The Data Engineer will need to be authenticated and have sufficient rights/privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.
8. Data Engineer generates specific schema from the model
9. Data Engineer and Data Provider iterate on the schema definition until both are satisfied that the specific schema, defined in the PDS4 Information Model, provides the necessary descriptive information that the Data Provider would like to include as part of the product metadata
10. Data Provider and the PDS/DN complete the design of a new discipline-specific schema.

11. The PDS will:
    1. Register the discipline-specific schema with the PDS4 schema repository
    2. Inform the Data Provider that the change has been incorporated into the PDS4 Information Model and that the schema is available through the PDS4 schema repository.

Alternative: Create Discipline-Specific Schema from scratch

At step 2, create an XML-compliant discipline-schema from scratch (e.g., a new cartography model as yet not archived by the PDS).

a. Data Provider drafts an XML-compliant schema using an XML editor (e.g., Oxygen or Eclipse) or in a text editor (e.g., UltraEdit or BBEdit).
b. Data Provider uses an XML editor (e.g., Oxygen or Eclipse) or an application (e.g., java, python, etc) to ensure the schema is fully formed (i.e., ensure there are no errors present in the XML).
c. Data Provider submits the draft discipline-schema to the PDS for formal "schematizing" (i.e., for inclusion into the PDS4 Information Model).
d. Data Engineer reviews the draft schema.
e. Data Engineer accesses the PDS Information Model (via the protégé tool) interface and creates the specific schema. The Data Engineer will need to be authenticated and have sufficient rights/privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.
f. Data Engineer generates specific schema from the model
g. Data Engineer and Data Provider iterate on the schema definition until both are satisfied that the specific schema, defined in the PDS4 Information Model, provides the necessary descriptive information that the Data Provider would like to include as part of the product metadata
h. Data Provider and the PDS/DN complete the design of a new discipline-specific schema.
i. The PDS will:
    1. Register the discipline-specific schema with the PDS4 schema repository
    2. Inform the Data Provider that the change has been incorporated into the PDS4 Information Model and that the schema is available through the PDS4 schema repository.

5.2.12 Tie Specific Product Schema to Discipline-Specific Schema(s)

The Data Provider wants to tie/link the set of specific schemas together.

1. Data Provider opens the set of specific schemas in an XML editor (e.g., Oxygen or Eclipse) or in a text editor (e.g., UltraEdit or BBEdit) to examine the XML.

2. Data Provider reviews the PDS instructions on "how to link schemas together".
3. Data Provider follows the instructions and "links" the set of specific schemas.
4. Data Provider uses an XML editor (e.g., Oxygen or Eclipse) or an application (e.g., java, python, etc) to ensure the schemas are fully formed (i.e., ensure there are no errors present in the XML of any of the schemas and the "linking" of the schemas is correct).

5.2.13 Create Label Template from Specific Schema(s)

The Data Provider wants to create a label-template from the set of specific schema(s).

1. Data Provider opens the set of specific schemas in an XML editor (e.g., Oxygen or Eclipse). Both Oxygen and Eclipse provide a capability to "export/create" an XML label from a schema (XSD).
2. Data Provider uses the "export" feature of the XML editor to create an XML label-template.
3. Data Provider uses the save feature of the XML editor to save a copy of the label-template to local storage.

5.2.14 Validate Label Template using Specific Schema(s)

The Data Provider wants to ensure the label-template is fully formed and can be validated against the set of specific schema(s).

1. Data Provider opens the label-template in an XML editor (e.g., Oxygen or Eclipse). Both Oxygen and Eclipse provide a capability to validate a XML document against a schema (XSD).
2. Data Provider uses the "validate" feature of the XML editor to ensure the XML label-template is fully formed (i.e., ensure there are no errors present in the XML of the label-template).
3. Data Provider uses the save feature of the XML editor to save a copy of the label-template to local storage.

5.2.15 Design/Implement Product Pipeline

The Data Provider wants to design/implement software that is capable of auto-generating collections of products.

1. Data Provider writes software that will auto-generate XML product labels for each product in the data set.

2. The Data Provider iterates on a design/implementation until the Data Provider is confident that the auto-generated product labels are PDS compliant and the pipeline production software is sufficiently robust at producing collections of product labels.

Note that the label generation software can either validate each auto-generated product label as the label is being produced. Or, the product labels can be validated after the fact. In either case, every product label must be validated against the set of schemas to ensure PDS compliance.

5.2.16 Create Sample Products using Product Pipeline

The Data Provider wants to use the label generation software pipeline to auto-generate a sample collection.

1. Data Provider uses the software to auto-generate "sample" product labels that accurately describe the products.
2. The Data Provider bundles the "sample" products into one or more PDS compliant sample collections.
3. The Data Provider uses collection validation software to ensure the sample collection is PDS compliant
4. The Data Provider prepares the sample collection(s) for submission to a peer-review committee.

Note that the sample collections will contain the product schemas. The directory structure for schema placement has yet to be worked out. But, the schemas could reside within a "schema" sub-directory of each directory that contains products (e.g., for instance, the "index" directory would have a "schema" sub-directory where the schema files would reside for each associated product label).

5.2.17 Peer-Review Sample Products and Product Pipeline

The Data Provider submits sample collections to a peer-review committee for review and approval.

1. Data Provider assembles a peer-review committee for the purpose of critically reviewing the data.
2. The peer-review committee reviews the sample collections and ensures PDS compliance; which includes, but is not limited to validating the product labels against the schemas.
3. The peer-review committee identifies any liens. Any liens identified during the review are documented and classified as major or minor, with a recommendation as to the disposition of the lien

4. Data Provider resolves liens.

Note that the end result of the peer-review committee is to approve the pipeline production of archive quality products/collections to the point where a complete validation of archive collections would be unprofitable.

5.2.18 Create Archival Quality Product Labels using Product Pipeline

The Data Provider wants to use the label generation software pipeline to auto-generate archival quality collections of products.

1. Data Provider uses the peer-review-approved software pipeline to auto-generate collections of product labels that accurately describe the products.
2. The Data Provider bundles the products into PDS compliant archival quality collections.

5.3 System Service Operations Scenarios

This section captures operations scenarios relating to the operation of the system and the interaction with the services.

5.3.1 Register Schema(s) with PDS

The Data Provider wants to ensure that their schema is searchable/accessible by other data providers. The Data Provider will need to be authenticated and have sufficient rights/privileges for the target namespace of the schema. Once a product has been registered with the PDS, the product is searchable/accessible across the PDS federated repository.

1. Data Provider accesses the Registry (Document) service via the Operator Portal to register one or more schemas.
2. Follow the steps from use case 5.3.7 Authenticate/Authorize User Access.
3. Data Provider submits metadata about the schema and the schema file to the Registry (Document) service.
4. Registry (Document) service registers the schema.

5.3.2 Retrieve Schema(s) from PDS

The Data Provider wants to retrieve a previously registered generic or specific schema.

1. Data Provider uses the Registry (Document) service to query for one or more registered schemas.
2. Registry (Document) service returns a result set corresponding to the query criteria with a web-accessible link for each schema result returned.
3. Data Provider uses the provided link to download the desired schema(s) to local storage.

5.3.3 Register Catalog Description(s) with PDS

The Data Provider / Data Engineer wants to ensure their catalog description is searchable/accessible by other data providers. Once a product has been registered with the PDS, the product is searchable/accessible across the PDS federated repository.

1. Data Provider / Data Engineer accesses the Registry (Inventory) service via the Operator Portal to register one or more catalog descriptions, which includes but is not limited to:
    a. Mission
    b. Instrument
    c. Data Set
2. Follow the steps from use case 5.3.7 Authenticate/Authorize User Access.
3. Data Provider submits the catalog description to the Registry (Inventory) service.
4. Registry (Inventory) service registers the catalog description.

5.3.4 Deliver Product(s) to PDS

The Data Provider delivers a package of archival quality products to the cognizant PDS Discipline/Data Node.

1. Data Provider delivers the package to the cognizant PDS Discipline/Data Node via the Transport Service. Examples of transport mechanisms include FTP, Data Brick, etc.
2. Data Engineer from the cognizant PDS Discipline/Data Node acknowledges receipt of the package.

5.3.5 Verify Product Delivery

The Data Engineer verifies the package of archival quality products delivered from the Data Provider prior to registration.

1. Data Engineer uses the validation tool to verify the contents of the delivered package.

Alternative: Validation Fails
At step 1, the package fails validation.
   a. Data Engineer rejects the delivered package.
   b. Data Engineer notifies the Data Provider of the validation failure.

5.3.6 Register Product(s) with PDS

The Data Engineer registers the package of archival quality products to facilitate tracking, search and access of the products. Once a product has been registered with the PDS, the product is searchable/accessible across the PDS federated repository.

1. Data Engineer initiates the Harvest service to register the archival quality products from the delivered package; which includes but is not limited to:
   a. Data
   b. Documents
   c. Software
2. Follow the alternative steps from use case 5.3.7 Authenticate/Authorize User Access.
3. Harvest service registers each product with the Registry (Inventory) service.
4. Data Engineer accesses the Registry (Inventory) service via the Operator Portal to review the registrations.
5. Follow the steps from use case 5.3.7 Authenticate/Authorize User Access.
6. Data Engineer approves the registrations accordingly.

Alternative: Registration Fails or Not Approved
At step 3, one or more products fail registration.
   a. Data Engineer deletes the registrations associated with the package.
   b. Data Engineer corrects the problem.
   c. Return to primary scenario at step 1.

5.3.7 Authenticate/Authorize User Access

A user of a system service/application must be authenticated where appropriate. For the most part, this is only required when adding or modifying content in one of the Registry service instances.

1. User (e.g., Data Provider, Data Engineer, or Operator) requests access to a restricted service/application and submits authentication credentials (user name and password).

2. Service/Application submits authentication credentials to Security service where credentials are verified and authorized access is determined.
3. Service/Application receives authentication/authorization from Security service and grants access to the user.

Alternative: Service-to-Service Authentication/Authorization
At step 1, a service is requesting access to another service.
   a. Requesting service submits authentication credentials via a certificate.
   b. Responding service grants access to the requesting service.

5.3.8 Respond to Query for Products

The Data Consumer is interested in discovering and retrieving PDS products. This request can be made from the main PDS portal or from a Node-specific portal.

1. Data Consumer provides query criteria using the Data Consumer Portal.
2. Search service accepts the query criteria and returns a corresponding set of results to the portal.
3. Data Consumer selects the desired products from the result set and requests delivery using the Data Consumer Portal including any format transformation.
4. Transport service accepts the delivery request and returns the requested products utilizing inline transformation software where applicable.

5.3.9 Monitor System Operations

The Operator monitors the status of the system.

1. Monitor service receives constant status updates from the system services.
2. Operator accesses the Operator Portal to view the system status.

Alternative: Service Failure
At step 1, a service failure is detected in the system.
   a. Operator is notified via configured mechanism of the failure.
   b. Return to primary scenario at step 2.

5.3.10 Report Metrics

The Operator / Data Engineer is interested in generating a metrics report.

1. Report service receives constant metrics submissions from the system services.

2. Operator / Data Engineer accesses the Operator Portal to generate a report.

5.3.11 Deliver Products to the Deep Archive

The Operator at the Discipline/Data Node delivers a package of products to the deep archive.

1. Operator initiates a request via the Operator Portal to deliver a package to the Deep Archive.
2. Transport service accepts the delivery request, creates the package and delivers it to the Deep Archive.
3. Operator updates the affected products in the Registry (Inventory) service to reflect the delivery to the Deep Archive.

5.4 Archive Lifecycle Operations Scenarios

This section captures operations scenarios relating to the archive lifecycle processes and the interactions between those processes and the Information Model. Figure 5.3-1 and Figure 5.3-2 depict two different approaches to the PDS archive lifecycle:

- Traditional archiving approach
- Accumulating archiving approach

The traditional archiving approach/process is where the entire set of (raw) data is on hand and ready to be processed in a single-pass-thru. Figure 5.3-1 depicts the traditional archiving process.
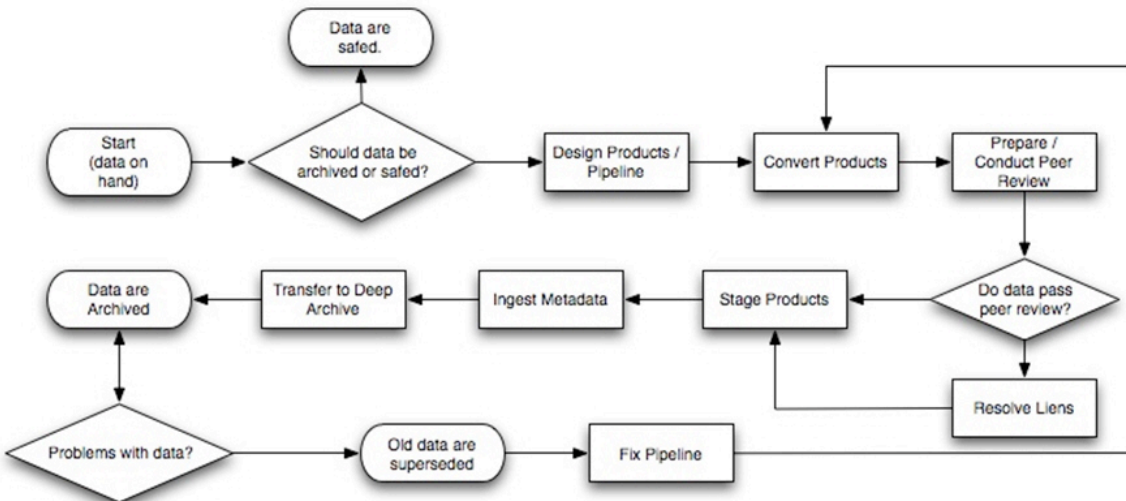
Figure 5.3-1. Traditional Archiving Process

The accumulating archiving approach/process is where the (raw) data becomes available in chunks, the data is processed in chunks, and processing occurs over multiple processing periods. Figure 5.3-2 depicts the accumulating archiving process.
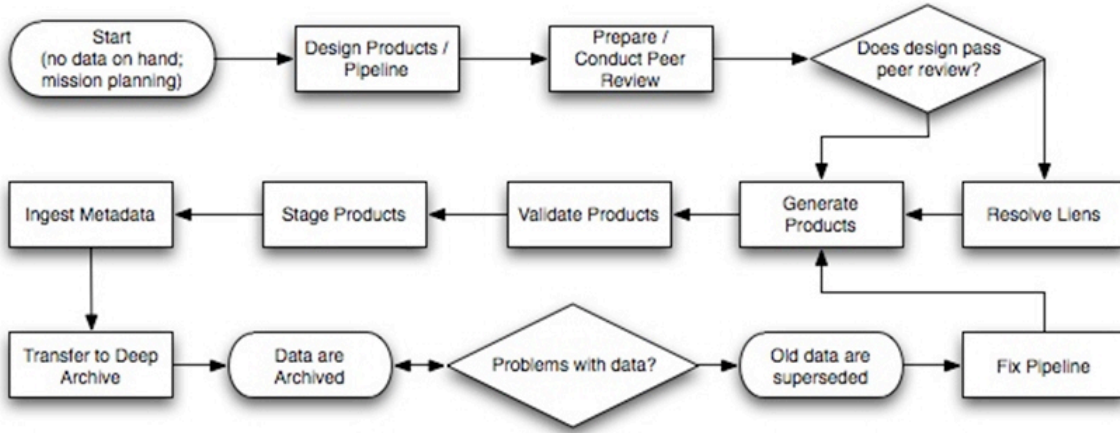

Figure 5.3-2. Accumulating Archiving Process

From Figure 5.3-1 and 5.3-2, use cases can be derived and are detailed in the following sections.

5.4.1 Conversion of PDS3 Dataset to PDS4 Dataset

This use case describes the traditional approach to archiving a dataset where the entire set of data is on hand and ready to be processed. In this case, the dataset was previously archived under PDS3 and the dataset was identified for migration to PDS4.

In sequence, follow the steps outlined in Use Cases:

Use Case	Description
5.2.6	Retrieve Product Schema from Schema Repository
5.2.7	Retrieve Discipline Schema from Schema Repository
5.2.8	Create Specific Product Schema from Generic Schema
5.2.10	Create Specific Discipline Schema from Generic Schema
5.2.12	Tie Specific Product Schema to Discipline-Specific Schema(s)
5.2.13	Create Label Template from Specific Schema(s)

5.2.14	Validate Label Template using Specific Schema(s)
5.2.15	Design/Implement Product Pipeline
5.2.16	Create Sample Products using Product Pipeline
5.2.17	Peer-Review Sample Products and Product Pipeline
5.3.1	Register Schema(s) with PDS
5.3.3	Register Catalog Description(s) with PDS
5.2.18	Create Archival Quality Product Labels using Product Pipeline
5.3.4	Deliver Product(s) to PDS
5.3.5	Verify Product Delivery
5.3.6	Register Product(s) with PDS
5.3.11	Deliver Products to the Deep Archive

Alternative: Create Specific Product Schema from product label(s)
At use case 5.2.6, the operation is to gather a representative set of product labels.

a. Data Provider puts together one or more product labels that are representative of all combinations/permutations of the range of values for each product attribute (i.e., the goal is to identify/define all possible values for each keyword in the label).

b. Data Provider submits the set of product labels to the PDS for formal "schematizing" (i.e., for inclusion into the PDS4 Information Model).

c. Data Engineer reviews the set of product labels.

d. Data Engineer accesses the PDS Information Model (via the protégé tool) interface and creates the specific schema from the product labels. The Data Engineer will need to be authenticated and have sufficient rights/privileges. The Data Engineer will ensure that the Data Provider has the authority to request the change.

e. Data Engineer generates specific schema from the model

f. Data Engineer and Data Provider iterate on the schema definition until both are satisfied that the specific schema, defined in the PDS4 Information Model, defines the overall structure and format of the archived product.

g. Data Provider and the PDS/DN complete the design of a new specific schema.

h. Continue with Use Case 5.2.10

5.4.2 Create PDS4 Dataset from Single Data Release

This use case describes the traditional approach to archiving a dataset. In this case, the Data Provider wants to create a PDS4 compliant dataset from (raw) data that is on hand and is ready to be processed (i.e., the Data Provider wants to use the label

generation software pipeline to auto-generate archival quality collections of products).

Note that the following steps are identical to those outlined in *"Conversion of PDS3 Dataset to PDS4 Dataset"*.

In sequence, follow the steps outlined in Use Cases:

Use Case	Description
5.2.6	Retrieve Product Schema from Schema Repository
5.2.7	Retrieve Discipline Schemas from Schema Repository
5.2.8	Create Specific Product Schema from Generic Schema
5.2.10	Create Specific Discipline Schema from Generic Schema
5.2.12	Tie Specific-Product Schema to Discipline-Specific Schema(s)
5.2.13	Create Label Template from Specific Schema(s)
5.2.14	Validate Label Template using Specific Schema(s)
5.2.15	Design/Implement Product Pipeline
5.2.16	Create Sample Products using Product Pipeline
5.2.17	Peer-Review Sample Products and Product Pipeline
5.3.1	Register Schema(s) with PDS
5.3.3	Register Catalog Description(s) with PDS
5.2.18	Create Archival Quality Product Labels using Product Pipeline
5.3.4	Deliver Product(s) to PDS
5.3.5	Verify Product Delivery
5.3.6	Register Product(s) with PDS
5.3.11	Deliver Products to the Deep Archive

5.4.3 Create PDS4 Dataset from Multiple Data Releases

This use case describes the accumulating approach to archiving a dataset. In this case, the Data Provider wants to create a PDS4 compliant dataset from (raw) data that will be received periodically and will both accumulate and be processed over time (i.e., the Data Provider wants to use the label generation software pipeline to auto-generate archival quality collections of products as the products are received).

In sequence, follow the steps outlined in Use Cases:

TBD

6 Appendix A: Acronyms

DDWG	Data Design Working Group
DN	PDS Discipline Node
EN	PDS Engineering Node
FTP	File Transfer Protocol
NASA	National Aeronautics and Space Administration
NSSDC	National Space Science Data Center
PDS	Planetary Data System
PDS3	Version 3.8 of PDS Data Standards
PDS4	Version 4.0 of PDS Data Standards
PDS 2010	PDS 2010 Project
PDS MC	PDS Management Council
RIM	Registry Information Model
SDSC	San Diego Supercomputer Center
SDWG	System Design Working Group
XML	Extensible Markup Language