PDS4 Core
January 2013

I.   Introduction

The PDS has been incrementally releasing system builds as it matures the PDS4 system. In November 2011, the PDS Management Council held an Operational Readiness Review and determined that, with resolution of a set of liens, the PDS4 standards would be released to a <u>scoped set of data providers</u> planning to develop PDS4 data products for new missions selected for flight. The release would also be used internally by PDS and by international partners to plan migration efforts and continue to validate the implementation. The Management Council convened a teleconference in late January 2012 to review the results of the liens and proceeded to work with the first PDS4 missions at the release of build 2b in February 2012.

A significant number of improvements have been made as a result of actively working with the PDS4 data standards releases. A stable "core" is now emerging. At the November MC meeting, it was determined that PDS should begin planning for "public release" of the PDS4 data standard. This document defines the core aspects of the data model and the associated software that will be used to support data suppliers in a version 1.0 public release of PDS4. Releases of public versions of PDS4 will be aligned with the system build schedule beginning with build 3b and following every six months. The 3b build is planned for March 2013 which will follow an acceptance test process and review by the PDS prior to release.

II.   Core Information Model

The classes from the PDS4 Information Model that are being proposed for the Version 1.0 candidate release are summarized in the following table. A more detailed breakout can be found in Appendix A. The Data Design Working Group (DDWG) made the initial determination based on those classes needed to support the LADEE and MAVEN missions, primarily the 2D image, the table classes, and the Product Observational component classes.

The concept of backwards compatibility has been taken into consideration for the determination of what classes are to be included in the initial release and those that can be delayed. For example, the Array_3D classes are subclasses of the Array class but can be delayed to V1.1 since they are backwards compatible by definition. In other words, if the Array class is accepted for Version 1.0, then only the additional features provided by the Array_3D subclasses need to be reviewed since their core components were previously accepted.

Data Structures
The four fundamental data structure classes define the structures of all PDS4 digital objects.

Class	Subclasses	Description
Array	Array_2D, Array_2D_Image	The Array class defines a homogeneous N-dimensional array of scalars.
Table ~~Base~~	Fixed-width Character Table, Fixed-width Binary Table	The Table ~~Base~~ class defines a heterogeneous repeating record of scalars.
Parsable Byte Stream	Delimited Table, Header, SPICE Kernel, Stream Text	The Parsable Byte Stream class defines byte streams that have standard parsing rules.
Encoded Byte Stream	Encoded_Image, PDF/A	The Encoded Byte Stream class defines byte streams that must be decoded by software before use. These byte streams must only use standard encodings.

Product

Product is a uniquely identified object that is managed by the registry.

Class	Subclasses	Description
Product Observational	N/A	A Product_Observational is a set of one or more information objects produced by an observing system.
Product Document	N/A	The Product Document class describes a document.
Product XML_Schema	N/A	The Product_XML_Schema is used to describe XML schemas.

Aggregates

The aggregate products are used to group a set of base products.

Class	Subclasses	Description
Product Collection	N/A	A Product_Collection has a table of references to one or more basic products.
Product Bundle	N/A	A Product_Bundle has a table of references to one or more collections.

III.   Core Software

The software being proposed for the Version 1.0 candidate release consists of services and tools for ~~driving~~ the PDS data system in addition to providing support for

preparation, review and use of PDS4 data. The core system software consists of the following functions:

Data Registration
Software for capturing metadata describing data products registered in the PDS data system. Satisfied by the following components:

Component	Description	Status
Registry Service	Provides functionality for capturing metadata from PDS products to enable search and discovery of these products within the PDS data system.	Software available from EN
Harvest Tool	Provides the means for registering products with the Registry Service. This tool reads a PDS4 product label, extracts specified metadata and registers the product with the registry. Also reads a PDS3 product label, extracts a minimal set of metadata and registers a proxy product with the registry for tracking purposes.	Software available from EN
Catalog Tool	Provides the means for registering PDS3 catalog files with the Registry Service to support ongoing PDS3 data deliveries.	Software available from EN

Data Search
Software for discovering data products registered in the PDS data system. Satisfied by the following components:

Component	Description	Status
Search Service	Provides functionality for accepting queries from data consumers for registered "catalog" products and includes functionality for retrieving search results.	Software available from EN

Support for product-level search in development by EN with initial capabilities available in Build 3b |
| Search User Interface | Provides the user interface for interacting with the Search Service supporting facet-based searching. | Software available from EN |

In addition to the core system software, the following functions provide support for preparation, review and use of PDS4 data:

Label Generation
Software for generating PDS4 labels from existing PDS3 labels in bulk. Although initially intended for migration, it should be adaptable to generating labels for new PDS4 data products. Future plans include merging the following two components into a single implementation:

Component	Description	Status
Generate Tool	Command-line based tool that generates a PDS4 label from a PDS3 label or a PDS3-specific DOM object.	Software available from EN

Additional support for Imaging data sets in development by IMG/EN |
| pds.docgen | Very similar to the Generate Tool, in that it is based on Apache Velocity. | Software available from PPI |

Local Data Dictionary Generation

Software for generating a Local Data Dictionary in the form of an XML Schema file that can be referenced by product labels. Satisfied by the following component:

Component	Description	Status
LDDTool	Currently generates a Local Data Dictionary XML Schema from a populated template. This tool has been made available to the DDWG for beta testing.	Software available from EN for Beta Testing

Initial release planned for Build 3b |

Label and Data Validation

Software for validating PDS4 product labels and product data. The associated specific schema for the product label specifies syntactic and semantic constraints. The product label itself specifies the constraints for the data. Satisfied by the following component:

Component	Description	Status
Validate Tool	Currently validates PDS4 product labels. Plans include adding support for validating the data along with collection and bundle structures.	Software available from EN

Support for initial content validation in development by EN and planned for Build 3b |

Label and Data Transformation

Software for reading and writing PDS4 data products that also includes transforming data and label files into various formats. See the Format Transformation sheet of this document for the list of requested transformations. Satisfied by the following components:

Component	Description	Status
PDS4 Tools	Java library for reading and writing PDS4 data, including support for Array 2D Image, Binary Table, Character Table and Delimited Table.	Software available from EN Support for data model updates in development by Ames and planned for Build 3b
Transform Tool	Command-line based tool for transforming PDS3 and PDS4 product labels and data into common formats, including PDS4 to PDS3. This tool calls the transformation functions from the PDS4 Tools and Transcoder libraries. Additional transformation support prioritized by PDS Node staff.	Software available from EN Support for additional transformations in development by EN and planned for Build 3b

Bundle Browser
Software for browsing labels and data within a bundle. Includes incorporating capabilities for transforming labels and data to viewable formats from the Label and Data Transformation software.

Component	Description	Status
TBD	Planned for web based and desktop deployment.	In design by EN with initial release planned for Build 3b

IV.   Core Documentation

The documentation being proposed for the Version 1.0 candidate release consists of the Information Model Specification, Data Dictionary, Concepts Document, Glossary, Standards Reference, and Data Providers Handbook.

Appendix A – Core Classes

The following table includes the classes for candidate release V1.0 and the future releases. The legend provides a brief description of each of the color nodes.

Context classes such as instrument and investigation are included in the Version 1.0 candidate release ~~since~~ they are being used for the LADEE and MAVEN missions. However the DDWG did not believe that a formal review was required ~~since~~ they are managed internally to the PDS and are already being used. Any change to a context class will have minimal impact on an associated data product because data products reference context products using unique identifiers.

Most of the remaining classes are held for candidate release V1.1. These classes are currently under development and many are mature, however further testing is desired.

	Description	Available For Testing	Projected Release
Legend	Candidate V1.0 - For Public Use - Minimum Required	Build 3b - 4/1/2013	7/15/2013
	Candidate V1.1 - For Public Use	Build 3b - 4/1/2013	Build 4a
	Candidate V1.0 - For Internal Use	Build 3b - 4/1/2013	7/15/2013
	Candidate V1.n		
	Requires advocate		

	Class	Subclass	Subclass	Key Component	Description	Node/Note
			PDS4 Core Classes - Candidate Releases			
	Class	Subclass	Subclass	Key Component	Description	Node/Note
Data Structures						
	Array				The Array class defines a homogeneous N-dimensional array of scalars. The Array class is the parent class for all n-	EN/This class must be included since it is a parent of a candidate class.
		Array_2D			The Array 2D class is the parent class for all two dimensional array base classes.	EN/This class must be included since it is a parent of a candidate class.
			Array_2D_Image		The Array 2D Image class is an extension of array_base and defines a two dimensional image.	
				Display_2D_Image	The Display_2D_Image class provides attributes to enable the display of a 2 dimensional image.	
				Special_Constants	The Special Constants class provides a set of values used to indicate special cases that occur in the data.	
				Object_Statistics	The Object Statistics class provides a set of values that provide metrics about the object.	
			Array_2D_Spectrum		The Array 2D Spectrum class is an extension of array_base and defines a two dimensional	EN/This class is currently a placeholder. The definition and use of this class need to be discussed further.
			Array_2D_Map		The Array 2D Map class is an extension of array_base and defines a two dimensional map.	EN/This class is being included since significant cartography work has been done. The class is being tested and documented. The class is needed for migration.
		Array_3D			The Array 3D class is the parent class for all three dimensional array base classes.	EN/This class must be included since it is a parent of a candidate class.
			Array_3D_Image (multi-banded images)		The Array 3D Image class is an extension of array_base and defines a three dimensional image.	EN/This class is being included since significant work has been done. The class is being tested and documented. The class is needed for migration.

			Array_3D_Spectrum (e.g. Hyper Spectral Qube)		The Array 3D Spectrum class is an extension of array_base and defines a three dimensional spectrum.	EN/This class is being included since significant work has been done. The class is being tested and documented. The class is needed for migration.
				Calibration	The Calibration class is a placeholder for forthcoming Calibration classes.	
				Cartography	The Cartography class is a placeholder for forthcoming Cartography classes.	
				Geometry	The Geometry class is a placeholder for forthcoming Geometry classes.	
			Array_3D_Movie		The Array 3D Movie class is an extension of array_base and defines a movie as a set of two dimensional images in a time series.	
	Table_Base				The Table Base class defines a heterogeneous repeating record of scalars. The Table Base class is the parent class for all heterogeneous repeating record of scalars.	EN/This class must be included since it is a parent of a candidate class.
		Fixed-width Character Table (aka Table_Character)			The Table Character class is an extension of table base and defines a simple character table.	
				Field_Character	The Field_Character class defines a field of the record.	
				Group_Field_Character	The Group_Field_Character class allows a group of table fields.	
				Uniformly_Sampled	The Uniformly_Sampled class provdes parameters for a uniformly sampled table.	
		Fixed-width Binary Table (aka Table_Binary)			The Table Binary class is an extension of table base and defines a simple binary table.	
				Field_Binary	The Field_Binary class defines a field of the record.	

				Group_Field_Binary	The Group_Field_Binary class allows a group of table fields.	
	Parsable_Byte_Stream				The Parsable Byte Stream class defines byte streams that have standard parsing rules. The Parsable Byte Stream class is the parent class for all parsable byte streams.	EN/This class must be included since it is a parent of a candidate class.
		Stream_Text			The Stream text class defines a text object.	PPI & GEO/ need in green for inventories and data
		PDS/DSV Table (aka Table_Delimited)			The Table_Delimited class defines a simple table (spreadsheet) with delimited fields and records.	DDWG 12/13 - Required for the Collection's Inventory Table
			Field_Delimited		The Field_Delimited class defines a field of the record.	
			Group_Field_Delimited		The FIeld_Group_Delimited class allows a group of delimited fields.	
		Header			The Header class describes a data object header.	
		SPICE_Kernel			The SPICE Kernel class describes a SPICE object.	DDWG 12/13 - LADEE will be generating their own SPICE Kernels. LADEE - possibly treat as an 'external standard' - green
	Encoded_Byte_Stream				The Encoded Byte Stream class defines byte streams that must be decoded by software before use. These byte streams must only use standard encodings. The Encoded Byte Stream class is the parent class for all encoded byte streams.	EN/This class must be included since it is a parent of a candidate class. PDF-A required.
		Encoded_Image		PDF/A	The Encoded Image class is used for ancillary images in standard	EN/This class must be included since it is a parent of a candidate class.
		Header_Encoded			The Header Encoded class describes a header that has been encoded using an encoding scheme that is compliant to an external standard.	

Labels						
Labels	Product				A Product is a uniquely identified object that is managed by a registry/repository.	
	Product_Observational					EN/These classes are required for all observational products.
		Identification_Area			The identification area consists of attributes that identify and name an object.	
				Citation_Information	The Citation_Information class provides specific fields often used in citing the product in journal articles, abstract services, and other reference contexts.	
		Observation_Area and Context Area			The observation area consists of attributes that provide information about the circumstances under which the data were collected.	
				Investigation_Area	The Investigation_Area class provides information about an investigation (mission, observing campaign or other coordinated, large-scale data collection effort).	
				Observing_System	The Observing System class describes the entire suite used to collect the data.	
				Target_Identification	The Target_Identification class provides detailed target identification information.	
				Time_Coordinates	The Time_Coordinates class provides a list of time coordinates.	
				Primary_Result_Summary	The Primary_Result_Summary class provides a high-level description of the types of products included in the collection or bundle	

		File_Area (Observational, Observational_Supplemental, XML_Schema, Text)			The File Area Observational class describes, for an observational product, a file and one or more tagged_data_objects contained within the file.
			File		The File class consists of attributes that describe a file in a data store.
		Reference_List			The Reference_List class provides lists general references and cross-references for the product. References cited elsewhere in the label need not be repeated here.
			Internal_Reference		The Internal_Reference class is used to cross-reference other products in the PDS registry system.
			External_Reference		The External_Reference class is used to reference a source outside the PDS registry system.
		Discipline_Area			The Discipline area allows the insertion of discipline specific metadata.
		Mission_Area			The mission area allows the insertion of mission specific metadata.

Document and Support Products	Product_Document				The Document class describes a document.	
				Document_Format_Set	The Document Format Set class is a set consisting of a document format and associated files.	
				Document_File {external standard formats}	The Document File class describes a file which is a part of a document.	
	Product_XML_Schema				The Product_XML_Schema class describes a resource used for the PDS4 implementation into XML.	
	Product SPICE Kernel				The Product SPICE Kernel class defines a SPICE kernel product.	
Browse Products	Product_Browse				The Product Browse class defines a product consisting of one encoded byte stream digital object.	
	Product_Thumbnail				The Product Thumbnail class defines a product consisting of one encoded byte stream digital object.	
Aggregates	Product_Collection				A Product_Collection has a table of references to one or more basic products. The references are stored in a table called the Inventory.	
				File_Area_Inventory		
	Product Bundle				A Product_Bundle is an aggregate product and has a table of references to one or more collections.	
Context	Agency, Facility, Instrument, Instrument_Host, Investigation, Node, Other, PDS_Affiliate, PDS_Guest, Resource, Target, Telescope				The Context classes describes something that provides context and provenance for an observational product.	The Context products are well tested and considered fairly stable. Any changes will have limited impact since Context products are referenced.

Package	Product_AIP				The Product AIP class defines a product for the Archival Information Package.	Something is needed for initial use, but it doesn't need to be stable
	Product_DIP				The Product DIP class defines a product for the Dissemination Information Package.	Something is needed for initial use, but it doesn't need to be stable
	Product_SIP				The Product SIP class defines a product for the Submission Information Package.	
	Product_DIP_Deep_Archive				The Product DIP_Deep_Archive class defines a product for the Dissemination Information Package for the deep archive.	