

Information Model Version 2.0.0.0

PDS4 Tech Session

Steve Hughes
February 13, 2018

Background

- The Planetary Data System (PDS) went operational in 1990.
- Science data from over forty years of solar system exploration has been ingested into the archive.
 - The majority of these data products were created using PDS3 data standards.
- PDS4 Version 1.0. was released in early 2013.
- Early in the design of PDS4 it was suggested that a “few simple formats” could describe about 85% of digital objects in the archive.
- There has been very little analysis of the remaining 15%.

Fundamental Data Structures¹ (FDS) (A Few Simple Formats)

- There are four fundamental data structures that may be used for archiving data in the PDS.
- All products delivered to the PDS must be constructed from one or more of these structures.
- These four fundamental structures are described using four base classes:
 - Array (used for homogeneous N-dimensional arrays of scalars)
 - Table_Base (used for repeating records of heterogeneous scalars),
 - Parsable_Byte_Stream (a stream of bytes that can be parsed using standardized rules),
 - Encoded_Byte_Stream (an encoded stream of bytes).
- All other digital object classes in the PDS are derived from one of these four base classes.

[1] Planetary Data System Standards Reference, JPL D-7669, Part 2, Version 1.4.0, September 22, 2015.

FDS Class Hierarchy

- *Tagged_Digital_Object*
 - *Byte_Stream*
 - Array
 - Encoded_Byte_Stream
 - Parsable_Byte_Stream
 - Table_Delimited
 - *Table_Base*
 - Table_Binary
 - Table_Character

The few simple structures



Array

Table_Delimited

Table_Binary

Table_Character

Information Model Version 2.0.0.0

- Information Model Version 2.0.0.0 should represent a relatively stable version of the PDS4 information model.
 - This suggests that IM V 2.0.0.0 should be sufficient for the remaining 15%.
 - Exception: the data that will not ever be migrated
- Suggests that all unique structures in the remaining 15% have been identified and classified.

Proposal

- All unique structures in the remaining 15% need to be identified and classified.
 - Also suggest that at least two thirds of these have been successfully converted to PDS4, at least as a prototype.
 - No information loss should occur.
 - New information is allowed.
 - Schedule should fall out of PDS3 to PDS4 migration effort
- Extend the requirement in the SR.
 - From - All other digital object classes in the PDS are derived from one of these four base classes.
 - To - “All *other* digital object classes in the PDS are **composites** of one of these four base classes”.

Proposed Classification Scheme

1. Structures that are normalized with respect to PDS4 FDSs.
 - E.g. Proposed ISIS3 Archive Format
2. Structures that are not normalized with respect to PDS4 FDSs but that can be normalized in a straight forward manner using “sound practical sense”
 - E.g. Three color image with interleaved bands normalized to three distinct 2D arrays.
3. Structures that can only be normalized through exceptional efforts.

[2] Normalization – a) A standard to which a database schema should conform in order to minimize redundant data and to prevent update anomalies. b) Normalization allows the comparison of the corresponding normalized items.

Modeling Methodology (Sound Practical Sense)

- Normalize the structure – standard data modeling technique
- Define a class for each normalized element of the structure using the Object Oriented paradigm
 - A class has named Attributes
 - A class has named Associations (three types)
 - Composite, Aggregation, simple association
- *This methodology is enforced by Protégé, the modeling tool used to design the IM common dictionary*
- *Also used to define the Ingest_LDD template, the input file to LDDTool*
- *Also used to design Composite Structure*

Modeling Methodology (Sound Practical Sense)

- Structure Properties
 - Is not interleaved
 - Is not self describing (e.g., variable length records with record length in the record)
 - ...
- Enumerated reference types
 - has_component
 - has_primary_component
 - has_axis_values
 - has_backplane
 - has_column_headers
 - has_display_settings
 - has_spectral_characteristics

Draft Plan

- Identify the structures in the archive that we believe cannot be described using a single, or a set of loosely coupled, FDSs
 - Classify the structures.
 - Model the structures.
 - Prototype a significant percentage of the structures.
- This effort should be coordinated with the PDS3 to PDS4 migration effort

Backup

Quick Modeling Review

- Class - The set of attributes (including a name) which defines a family. A class is generic - a template from which individual members of the family may be constructed.
- Attribute - A property or characteristic that provides a unit of information. For example, 'color' and 'length' are possible attributes.
- Association - An attribute that establishes a unidirectional relationship between two classes.
- Types of associations³
 - Composite – A child can not exist independently without it parent.
 - Aggregation - A child can exist independently without it parent.
 - Simple association – The weakest relationship, i.e., not one of the above.
- Alternate modeling terminology
 - Properties³
 - *Object properties* link individuals to individuals.
 - *Datatype properties* link individuals to data values.

[3] Object-Oriented Terminology.