PDS 4 Information Model
- User Stories and Declarative Requirements.

Version 0.0

Working Draft

This version:

Latest version:

Previous versions:

Editors:
Todd King (UCLA/IGPP) <tking@igpp.ucla.edu>
Steve Hughes (JPL) <Steven.Hughes@jpl.nasa.gov>
Ed Guinness (Wash U) < guinness@wunder.wustl.edu>
Lyle Huber (New Mexico State University) < lhuber@nmsu.edu>
Chris Isbell (USGS) <cisbell@usgs.gov>

Referenced Documents:

Abstract

User stories are briefly stated use cases which describe a desired attribute or character of a system. Declarative requirements are expectations of the system which are difficult to express as a use case or user story. Most declarative requirements relate to non-functional requirements of the system. User stories and declarative requirements aid in defining the overall requirements for a system and are an important part of scoping a problem and capturing the needs and expectations of users. In this document we capture user stories which will help guide the development of the PDS4 Information Model. These use cases have originated from different sources and are compiled into this document.

Table of contents

---

1. Introduction

The intent of the PDS4 Information Model is to allow a resource to be described in sufficient detail so that it may be used with clarity. The types of resources and the amount of detailed information which is needed to described a resource is identified through user stories and declarative requirements. User stories are briefly stated use cases which describe a desired attribute or character of a system. Declarative requirements are expectations of the system which are difficult to express as a use case or user story. Most declarative requirements relate to non-functional requirements of the system. In this document we capture user stories and declarative requirements which have originated from different sources. The sources include an "expectations" survey conducted during the August 2007 PDS Management Council meeting, representative datasets currently archived in the PDS, formally stated use cases from external community efforts (IPDA), locally (within PDS) defined use cases and anecdotal use cases based on general comments expressed within the community. A distinction between each type of use case in maintained in the following sections.

2. Expectations Survey

During the August 2007 PDS Management Council meeting Mitch Gordon led a session to clarify PDS4 goals. He requested that each person who was present state a significant requirement. The goal was to display the range of community expectations. The following goals or characteristics were offered during the discussion:

- simplify, simplify, simplify
- reformulate data model
- machine verifiable
- logical rigor
- restrict formats
- really distributed
- no implied volume
- expanded data model
- describe or proscribe?
- formally documented
- stricter object definitions
- improved clarity

- interoperability
- science, rather than IT, oriented
- no recommendations
- classification (science vs engr vs admin)
- data dictionary structural changes
- robust validation possible
- decouple mission-specific DD items
- easily extensible DD
- multiple paths of accessibility
- service oriented architecture
- more attractive than PDS3
- content vs. physical descriptions
- needs harvesting protocol
- a taxonomy
- needs more linkage capability
- configuration control
- effective data integrity and verification
- tighter integration with mission data systems
- better user model
- better understanding of what tools PDS should provide

3. Representative Datasets

In the Spring of 2007 Steve Hughes requested from each of the PDS nodes a list of existing PDS datasets which are representative of the type of data actively being archived in the PDS. This list was used to generate an draft of the IPDA Information Model. The following datasets were provided by the nodes:

Dataset ID	Object	Format	Description
CO-D-CDA-3/4/5-DUST-V1.0	Table	ASCII	
CO-E/J/S/SW-CAPS-2-UNCALIBRATED-V1.0	Table	Binary	Cassini CAPS data/TABLE
CO-E/SW/J/S-MAG-2-REDR-RAW-DATA-V1.0	Table	Binary	Cassini Mag data/TABLE
CO-S-INMS-2-PKT-U-V1.0	Table	ASCII	Cassini INMS data/Header and SPREADSHEET
CO-V/E/J/S/SS-RPWS-2-REFDR-WBRFULL-V1.0	Table	Binary	Cassini PWS data/Header and TIME_SERIES
CO-V/E/J/S/SS-RPWS-3-RDR-LRFULL-V1.0	Table	Binary	With bit columns
DI/EAR-C-KECK1LWS-3-9P-IMAGES-PHOT-V1.0	Image	FITS	FITS image and header
DIF-C-HRII-3/4-9P-ENCOUNTER-V1.0	Image	FITS	FITS image and header
DII-C-ITS-3/4-9P-ENCOUNTER-V1.0	Image	FITS	FITS image and header
EAR-A-5-DDR-ALBEDOS-V1.1	Table	ASCII	
EAR-C-COMPIL-5-COMET-NUC-PROPERTIES-V1.0	Table	ASCII	
GO-E-EPD-2-SAMP-PAD-V1.0	Table	ASCII	Galileo EPD data/TABLE
GO-J-MAG-3-RDR-HIGHRES-V1.0	Table	ASCII	Galileo Mag data/TABLE
MEX-M-HRSC-3-RDR-V2.0	Image	VICAR	VICAR image and header
MEX-M-HRSC-5-REFDR-MAPPROJECTED-V1.0	Image	VICAR	VICAR image and header
MGN-V-RDRS-5-DIM-V1.0	Image	Simple	
MGN-V-RDRS-5-DIM-V1.0	Histogram		
MGS-M-RSS-1-EXT-V1.0.			
NEAR-A-NIS-5-EDR-ALL-PHASES-DSREV-V1.0	Image	FITS	FITS image and header
SDU-C-NAVCAM-5-WILD2-SHAPE-MODEL-V2.1	Table	ASCII	
VG1/VG2-S-ISS-2/3/4/6-PROCESSED-V1.0	Image	VICAR	VICAR image and header
VO1/VO2-M-VIS-5-DIM-V2.0	Histogram		
VO1/VO2-M-VIS-5-DIM-V2.0	Image	Simple	

4. External (IPDA) Use Cases

TDB

5. Internal (PDS) User Stories

Multi-record format products

Missions generate data which can have more than one record format in a single file (product). In some situations the record format alternates in a known pattern. In other situations the record format is determined by the content of a known location within the record. We need to describe such products.

Externally Specified Formats

Some products which should be part of an archive have complex internal formats that are formally specified or well supported. We should be able to archive these products provided that the specifications are clearly identified and also archived.

6. Anecdotal Desires

Over the past several years various individuals have expressed comments or observations regarding the current PDS data model or system. Steve Hughes compiled the following lists (slightly edited and reformatted) of anecdotal desires. Some of these anecdotal desires expand upon the expectations listed in section "2Expectations":

General

- Make requirements more clear versus recommendations; separate requirements from recommendations (e.g. and standards from policies); Define clear set of core requirements for archives; Clearly define all elements of the model; Provide clear justifications for requirements
- Implement real Data dictionary, Data Model, and Standards versioning
- Promote data system interoperability among different data systems and agencies (ESA); promote cross-mission, cross-instrument search and data recovery (may impact selection of which keywords remain global)

Products

- Provide Clear, consistent definition of product (the "one product, one label" issue)
- Identify Core data formats and derived object classes;
    - All data objects are currently treated equally, but some shouldn't be permitted.;
    - reduce/limit number of acceptable formats; reduce/limit number of keywords in PSDD (or re-categorize many of them into local data dictionaries) ;
    - remove CONTAINER objects; no variable length records;
    - handle compound products;
    - Have consistent set of keywords in all data product labels within a data set;
    - Make object definitions more rigorous;
    - standards should reflect prescriptive (vs. descriptive) use of labels;
    - easier data product label design;
    - enforce consistency in multi-valued keywords within data sets
- Define and formally specify all relationships among data elements; support validation of relationships between keywords (eg. if BANDS = 3, then BAND_SEQUENCE must be present and must have three values)

- Make software and document first class classes; Add Software, Document, SPICE, as a type of core Data Formats.
- Make Implicit File Object explicit; Address problem of Implicit File object pointing to two explicit files; handling of RECORD_TYPE definitions for files containing multiple different object types; Use explicit FILE object so that required keywords can be defined (and validated against)
- Resolve discrepancy between PSDD and StdRef for ARRAY optional keywords
- Consider extending model to deeper level
- Define document formats (text vs PDF vs other?)

Data Elements

- Manage Data Elements better
  In the attempt to make a keyword universal, the definitions are often very broad to the point where the intent of the keyword is hard to understand; Model does not account for complicated modern instruments that may need specialized keywords that really should be used only once; The process of trying to create universal keywords often leads to endless technical discussions because of the diverse nature of the science being covered by the nodes; There are so many keywords in the data dictionary that nodes and data producers may end up using two different keywords for the same purpose, e.g., standard_product_id and product_type.

Upper Level Model

- Fully specify relationships among components of the model; Add explicit relations for links that are currently implicit or assumed by existence in a directory on a volume. (e.g. document->data_set); Associate / link all components of an archive
- Review the data model and how it currently handles anomalies and unexpected cases (one-to-many inst-to-inst_host) consider changes; Resolve the one-to-many instrument-to-instrument host problem
- Fix target model. (simplify for planets and satelites, augment with subclasses for SBN bodies and PPI regions etc)
- Clean up Instrument Model; (One magnetometer subclass, what about subclasses of magnetometer) I.e. Do we allow multiple layers of subclasses, Fix Instrument unique identifier -> Instrument_Host_Id + Instrument_Id
- Modify collection / mission relationship to collections of type = mission, campaign, experiment, support, etc.; Fix collections / missions / campaigns / etc.
- Fix the HISTORY object (has no required or optional keywords)
- Model does the WINDOW object
- Model subclasses for Mission, Data Set, etc
- Formally redefine the "logical" PDS3 volume concept as a package.
- Need clear requirements for minimum catalog information required for saved data sets

7. Conclusion

The PDS encounters a wide range of data products with many perturbations of interchange format, representations and associations. In addition to the data formats currently archived in the PDS there are many other formats which can not be archived using the current PDS standards. For example, record oriented data with non-heterogeneous record formats. Balancing a rich support of formats with simplicity of expression is challenging with the current PDS data model. For example, the addition of keywords to support mission specific purposes has added complexity to the data dictionary and obscures essential features. In the PDS4 data model defining fewer generalized classes of data with enhanced flexibility and allowing externally specified formats should create the necessary rigor to create archive quality documentation.