ROSES Data Management Plans for PDS Archiving

Eric E. Palmer
Carol Neese
PDS Small Bodies Node

Ed Guinness
PDS Geosciences Node

Mitch Gordon
PDS Ring-Moon Systems Node

Nancy Chanover
Lynn Neakrase
PDS Atmospheres Node

Dick Simpson
PDS Radio Science

Why You Need a DMP

- ROSES 2016 C1, sec. 3.5
  - Requirement to archive data produced by the proposed project and include a Data Management Plan (DMP) in the proposal
- Archiving -
  - Stable and long-term supported data repository
    - Non-PDS (such as journals)
    - PDS, which is today’s focus
- Archiving takes effort, and the archiving cost must be realistically estimated and included in the budget
- NASA has established and funded PDS to provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof. The archive is publicly accessible, providing the fruits of NASA’s missions to the public at large. The mission of the PDS is to facilitate achievement of NASA's planetary science goals by efficiently collecting, archiving, and making accessible digital data produced by or relevant to NASA’s planetary missions, research programs, and data analysis programs.

What should be in the DMP for a ROSES proposal? (from ROSES C1, sec. 3.5)

- A description of data types, volume, formats, and (where relevant) standards
- A description of the schedule for data archiving and sharing
- A description of the intended repositories for archived data, including mechanisms for public access and distribution
- A discussion of how the plan enables long-term preservation of data
- A discussion of roles and responsibilities of team members in accomplishing the DMP
- (If funds are required for data management activities, these should be covered in the normal budget and budget justification sections of the proposal.)

Costing the archiving effort

- Most of the contents of the DMP can be created by the proposer without help from PDS
- EXCEPT
  - Costing - How much effort is required to archive data in PDS?
  - Schedule - How long does it take?
- Budget for archiving is not part of the DMP itself, but it is necessary for the proposal and the proposer needs help from PDS to estimate the effort.

Steps for Archiving Data in PDS

- Produce data products in acceptable PDS format (currently PDS4)
- Produce PDS labels, which under PDS4 are XML files (templates may be available)
- Assemble supporting documentation
- Organize data, labels, documentation, etc. into an archive package
- Validate labels using PDS provided tools
- Participate in peer review
- Make improvements as necessary, based on peer review recommendations
- Deliver final package to PDS.

Costing Guidelines — Introduction

- PDS4 is new
  - Limited experience so far on which to base costing estimates.
  - Estimates will improve as we gain experience with the new standard.
- Universal guidelines
  - Difficult to give because the scope and complexity of data can be so variable.
  - Here we offer some guidelines for simple data types
  - **If you have any questions or concerns, discuss your proposed project with your PDS node.**

Different Methods for Archive Preparation

- By hand
  - Prepare labels using an off-the-shelf XML editor such as Eclipse with PDS4 schema to design labels
  - Write code or script to generate full label set, and assemble bundle.
  - Validate using PDS4 Validate tool.
- LACE (in development)
  - Archiving tool which facilitates design and generation of PDS4 product labels.
  - Works for any PDS4-compliant data format, but it is not available yet for general use. It has been tested by some early PDS4 data archivers such as OSIRIS-REx.
- On-Line Archiving Facility (OLAF)
  - Enter supporting information into a web form and upload the data files, and OLAF creates the labels and assembles the archive for you. OLAF also automatically does validation against PDS standards.
  - No need for you to learn PDS4 standards.
  - The catch? OLAF supports only a limited set of formats. If you want to use an unsupported format, you can’t use OLAF.

OLAF - Overview

- Created by SBN over ten years ago in PDS3 standard, it now can produce labels and archive bundles in the PDS4 standard
- Now mostly used by SBN, but can be used by any PDS node.
- Limited to small data sets (< 10 Gb)
- Mostly for R&A projects. It has been used for some mission data (LADEE, Cassini, Hayabusa) but not suitable for most mission data
- OLAF automatically validates its output against PDS standards and validates data labels against the data files
  - No additional validation is required.

OLAF - Currently Supported Data Types

- Images
    - FITS images with no extensions
- ASCII Tables
    - Spectral tables
    - Time Series (e.g. lightcurves)
    - Complex tables (e.g. shape models)
    - General (user-defined) tables in fixed-width ASCII
- Documents
    - PDF/A (Archival PDF)
    - Plain text

Additional data types planned for OLAF

- Some will be available by the time your project is funded
- Images and cubes
    - FITS images with extensions
    - ISIS images and cubes
    - General PDS4-compliant 2D and 3D arrays
- Tables
    - CSV
    - FITS tables
    - General PDS4-compliant binary tables
    - XLSX spreadsheets (OLAF converts to PDS4-compliant table format)

OLAF Costing Guideline

- Time and effort to submit a data set through OLAF is variable
    - See handout
        - Super simple (single table) — 3 days + 1 day for review and liens resolution
        - Very complex — 6 weeks + 2 weeks for review and liens resolution
    - Depends on size and complexity.
    - The costing guideline suggests how to account for this variability.
- OLAF reduces time and effort
    - Past experience shows as much as a factor of ten over doing it by hand. (Estimate based on PDS3 data sets. No by-hand generated R&A datasets)

Validate Tool

- Engineering node created software for validating PDS4 product labels and product data.

- Features:
  - Validates labels against associated XML Schema and Schematron files generated from the PDS4 Model.
  - Supports Bundle and Collection referential integrity checking.

- Status:
  - Version 1.8.0 supports PDS4 Model 1.5.0.0 and earlier. [1]
  - New releases every 6 months.

[1] https://pds.nasa.gov/pds4/software/validate/

Summary - Steps toward your PDS DMP

- Read the ROSES guidelines for the DMP (ROSES C.3.5)

- Use the PDS web resources for proposers (given on the next slide) for the information you need to prepare the DMP.

- Guess which PDS node is right for your proposed data and contact them. (If you guessed wrong, they will put you in touch with the correct node.)

- Your PDS node can provide you with a letter of support for the proposal, and can provide guidance for your DMP if needed, particularly in estimating costs.

PDS resources for R&A Proposers

- Individual's Proposers Archiving Guide - in revision
- PDS Node Contacts - https://pds.jpl.nasa.gov/contact/contact.shtml
- Help for Proposal Writers (PDS Geosciences Node) - http://geo.pds.nasa.gov/dataserv/proposerhelp.html
- Help for Proposal Writers (PDS Imaging Node) - http://pds-imaging.jpl.nasa.gov/help/proposals.html
- Help for Archiving Derived Data Products (PDS Atmospheres Node) — http://pds-atmospheres.nmsu.edu/Derived_Data_LPSC2016_Brochure.pdf
- Planning and Developing PDS4-compatible Derived Data (PDS Atmospheres Node) — http://pds-atmospheres.nmsu.edu/pds4_comp.html
- How to Prepare Data for PDS (PDS Small Bodies Node) - http://sbn.pds.nasa.gov/howto/prepare.shtml
- SBN PDS4 Wiki (Help for creating PDS4 archives) - http://sbndev.astro.umd.edu/wiki/SBN_PDS4_Wiki
- PDS4 Concepts Document - https://pds.jpl.nasa.gov/pds4/doc/concepts/Concepts_150909.pdf

Questions?

Note: The PDS Geosciences Node has a booth in the exhibits area. PDS reps will be there to answer questions from Monday afternoon through the end of the Thursday poster session.

Examples

- Here are two examples. The costs are higher than the typical R&A proposal.
- Both are from missions
    - Complex datatypes
- Both are early adopters
    - Lack the expected tools
    - PDS 4 standards changed

Costing Example - OSIRIS-REx

- OSIRIS-Rex used LACE to design labels for all the data products of their four pipeline instruments.
- OSIRIS-REx is not yet generating individual data products and labels, so this effort is not included.
  - It took 40-60 hours to learn PDS4 standards and LDDTool
  - Creating label template for the first product type took 40 hours, subsequent product types took 20 hours
  - Assembling the bundle: the first one took a few weeks, subsequent a few days

Costing Example - BOPPS

- BOPPS designed data labels by hand using Eclipse, then wrote their own custom pipeline to generate and populate individual data labels. Times are person-weeks.
  - 5 weeks to understand PDS4, create LDD, generate template XML files and iterate w/ SBN to get them in a state ready for pipeline production
  - 2 months to create custom pipeline to generate XML product labels
  - 3 weeks to generate PDS4 archive products and inventory files
  - 3 weeks to incorporate liens from PDS and regenerate labels as needed