Developing a Data Pipeline and PDS-Compliant Archive in Response to the Discovery Announcement of Opportunity (AO)

July 18, 2014

This document is meant to assist a proposer of a Discovery mission in sizing and costing the development of a science data system and the data archive. The major effort in developing this aspect of a proposal will involve considering the end-to-end processing of the mission data that Includes

- The formulation of the data pipeline
- Identifying an approach for structural and scientific validation of the data
- Documentation of instrument calibration
- Preservation of information relating to the mission's observing log that is needed to support search and recovery from the archive as well as scientific use of the data.

Discovery missions will be required to create PDS4-compliant archives. This document provides a listing of factors to be included in estimating the cost of creating an archive, a listing of contacts within PDS for technical advice while preparing a proposal, reference to a proposer's archiving guide and a brief explanation of PDS4 data structure.

Costing the Data Preparation, Validation and Archiving

The costs that will be incurred for preparation, validation and archiving of PDS4-compliant data will depend on several factors, including the complexity of the mission and the heritage of science operations and instruments. The following check list (Table 1) summarizes major factors to consider in developing and estimating the cost of creating a data archive for the Planetary Data System (PDS). These elements should be described in a Data Management and Archive Plan as part of a Discovery Proposal.

Table 1. Check List for Developing and Costing of a Data System	
Science Operations Center (SOC)	
1. Distribution of data to and receipt of products from instrument teams	
2. An Archiving Working Group	
3. Document Development (mission data management plan, etc.)	
Individual Instrument Teams	
1. Definition of raw and processed data products	
2. Estimation of needed calibration activities	
3. Estimation the data volume and complexity	
4. Pipeline development	
5. Use of PDS4 validation software	
6. Scientific validation by team members using the data to be archived	
7. Development of archival documentation	
Interaction with the PDS	
1. Development and review of data plan and documentation	
2. Review of the design of pipeline products	
3. Establishment of a delivery schedule	
4. Peer review	
5. Lien resolution	
Staffing	
1. Estimation of adequate staffing for the SOC	
2. Estimation of required staffing for science teams to complete data development and archiving activities	

Contacting the Appropriate PDS Personnel

Each team is responsible for presenting a well-defined archive plan and accompanying budget. If needed, PDS staff members are available to provide technical advice.

Table 2. PDS Personnel Who Can Provide Technical Advice.

PDS Node *	Personnel to contact
Geosciences	Ray Arvidson, arvidson@wunder.wustl.edu, 314-935-5609 Ed Guinness, guinness@wunder.wustl.edu, 314-935-5493
Imaging	Lisa Gaddis, lgaddis@usgs.gov, 314-935-5609 Sue Lavoie, slavoie@jpl.nasa.gov, 818-354-5677
Navigation and Ancillary Information Facility (NAIF)	Chuck Acton, charles.acton@jpl.nasa.gov, 818-354-3869
Atmospheres	Reta Beebe, rbeebe@nmsu.edu, 575-646-1938 Lyle Huber, lhuber@nmsu.edu, 575-646-1862 Lynn Neakrase, lneakras@nmsu.edu, 575-646-1862
Planetary Plasma Interactions	Ray Walker, rwalker@igpp.ucla.edu, 310-825-7685 Steve Joy, sjoy@igpp.ucla.edu, 310-622-3462
Rings	Mark Showalter, mshowalter@seti.org, 650-810-0234 Mitch Gordon, mgordon@seti.org, 276-393-8822
Small Bodies	Ludmilla Kolokolova, ludmilla@astro.umd.edu, 301-405-1539 Mike A'Hearn, ma@astro.umd.edu, 301-405-6076
PDS Program Manager	Tom Morgan, thomas.h.morgan@nasa.gov, 301-286-1743

If you are selected for a Phase A study

The PDS staff will review the selected missions and determine which node will lead the archive development and what supporting nodes will be involved for each mission. During the Phase A period, you should interact with the appropriate lead node to refine your archiving plans and define a data archive working group and a delivery schedule to allow a valid assessment of this component of your mission.

Basic Steps for Planning and Preparing a Dataset are:

- Gaining a preliminary understanding of a PDS4 data bundle
- Working with PDS nodes to design products and labels
- Creation of archive products, collections and bundles
- Submission of products, collections and bundles to PDS
- Peer review of the products, collections and bundles

The PDS4 System

The PDS has developed the PDS4 system to streamline the ingestion and distribution of archived data and to take advantage of the structured data capabilities and off-the-shelf software that is available with the Extensible Markup Language (XML). A useful resource, the Proposers Archiving Guide, is available at http://pds.nasa.gov/pds4/propose/pds4-pag-20140721.pdf

Contents of a PDS4 Data Bundle

Data in PDS4 are organized into a hierarchical structure of bundles, collections, and basic products. Bundles contain logical groupings of related collections and collections contain logical groupings of related basic products (See Figure 1.) Collections may include: context information (target, spacecraft, instrument, etc.), documentation for usage of the data, science data (raw, calibrated, derived), calibration information, and linkages to XML schema and schematron (blueprints and sets of rules) used in the generation of the label files. **Development and construction of the labels is the key to constructing the mission bundles.**

An example bundle (with labels) can be found at http://atmos.nmsu.edu/PDS4BETA/phoenix/met.htm. If you have questions, seek help from your PDS contact (See Table 2.)

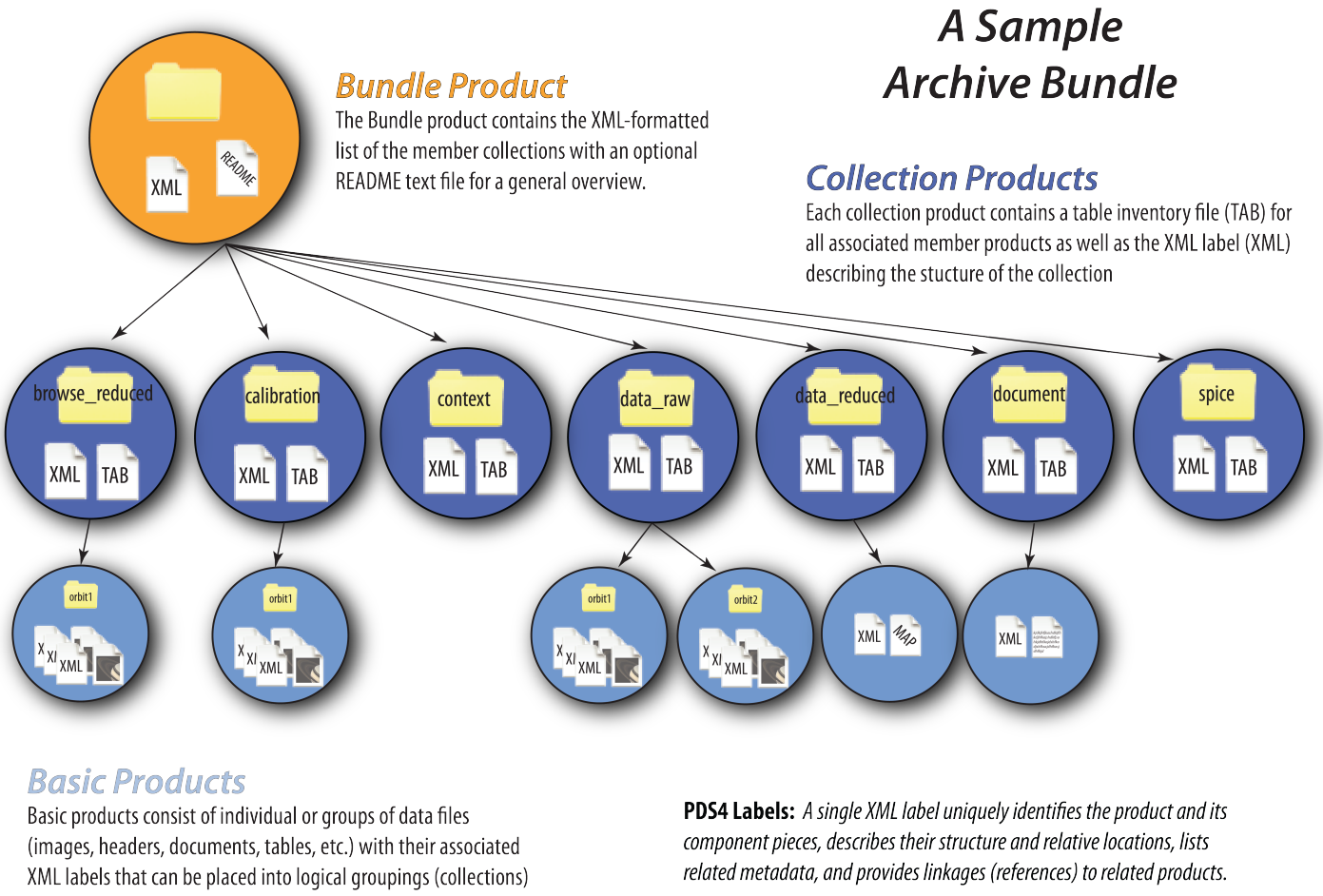

Figure 1. Structure of a PDS4 Data Bundle