Individual
Proposer's Archive Guide (PAG)

Last Revision: April 21, 2016

Contents

1 Introduction

NASA's Planetary Science Division supports a wide diversity of projects ranging from investigations by individual researchers to large flagship missions. Across the full spectrum of projects NASA requires the data and other resources that are produced to be preserved in the Planetary Data System (PDS) or an equivalent archive. This guide describes the process and requirements when archiving data with the PDS. While the PDS archiving requirements are the same for all types of projects, the scale, roles and responsibilities are different for an individual researcher's (i.e., small scale) project. Being familiar with PDS concepts and the general design of a PDS compliant archive is key to writing a successful Data Management Plan (DMP) in a research grant proposal. This guide describes the items that you, as an individual proposer, need to consider while preparing a proposal to a NASA Research and Analysis (R & A) program. This guide also provides a brief summary of information found in PDS standards and will help you to plan and estimate the effort of preparing your archive for submission to the PDS.

2 PDS Background

The PDS archives and distributes scientific data from planetary missions, astronomical observations, and laboratory measurements. Its purpose is to ensure the long-term access and usability of data and to support and stimulate advanced research. All PDS data are peer-reviewed, well-documented and publicly available. That is, the data may be exported outside of the United States under the "Technology and Software Publicly Available" (TSPA) classification.

The PDS is a federation of teams with expertise in different science disciplines. Each team is called a “node”. Within the PDS, there are nodes focusing on the scientific disciplines of Atmospheres, Geosciences, Cartography and Imaging Sciences, Planetary Plasma Interactions, Ring-Moon Systems, and Small Bodies. Additionally, there are two support nodes of the PDS: the Engineering Node and the Navigation and Ancillary Information Facility. For a current list of contacts for each node see

https://pds.nasa.gov/contact/contact.shtml

3 A Typical PDS Archive

When data are placed (or archived) in the PDS they must have sufficient documentation to enable others to find, read and use the data. Part of that documentation includes a description of which instrument was used to make the observation, what platform was the host of the instrument and which objects were observed. Collectively these provide the “context” of the observation. Each context is described with PDS-defined metadata that conforms to a prescribed information schema. Additional documentation may include papers or other documents that describe the instrument, calibration steps or processing that was performed. For the data, the structure and format must be described in sufficient detail to enable others to read the data. This too is described with conforming PDS-defined metadata. It is not possible to avoid the metadata requirements by expecting external software (public or proprietary) to read the data; data must be readable via the required metadata. The only exception to this rule is SPICE kernels which are part of the Navigation and Ancillary Information Facility (NAIF) system. The context, documentation and data (observational and calibration) are logically grouped and those groupings are placed in the PDS and maintained as part of the archive. To ensure a quality archive each grouping is reviewed prior to its public release.

4 Creating a New PDS Archive

Proposals responding to NASA’s Research Opportunities in Space and Earth Sciences (ROSES) announcement that will generate data that are to be archived with PDS must be archived in compliance with the PDS4 standards. In a ROSES proposal the proposer must demonstrate an understanding of the work involved in preparing data for the PDS. The main focus of the following sections is to provide sufficient details about the PDS4 standards to enable you to write the appropriate data management plan (DMP) or data management sections in a proposal. When you have determined what items you will be archiving in the PDS, you should communicate with the appropriate PDS Discipline Node (see Section 2 for contact information). Choose the Discipline Node which is most closely associated with the type of data you plan to archive. When contacting the node, be prepared to discuss in general terms the focus of your research and the types and quantity of data you expect to generate. Node staff can then advise you if they are the appropriate contact point and the place to deliver your data. If they are the right node, they can provide you with a letter of support indicating that you discussed the type of archive and what is involved in archiving data with the PDS. Even if a letter of support is not required by the program element, it is good practice to include one.

5 PDS4 Details and Definitions of Terms

PDS4 is an integrated system designed to improve discoverability and access to data across multiple storage locations. Each item stored in a PDS archive is called a product and consists of metadata stored in a label file and (in most cases) additional files that contain the data or document being archived. Every product is assigned a unique identifier so it can be referenced and managed.

PDS4 uses the Extensible Markup Language (XML) to create product labels. Products are organized into Collections, and Collections are organized into Bundles. While designing a PDS4 compliant archive it is important to keep in the mind the following general principles.

Everything is a Product.

A product is anything with a label and each product is assigned a unique identifier. Products can be created for data, documents, context files, XML schema, and even delivery websites.

Collections are groups of similar products.

A collection consists of products that share one or more attributes. These attributes can be the type (or class) of product, a time range, acquisition method (i.e., instrument), processing level or any other logical attribute. Every archived product must belong to a collection, so every project or individual researcher that submits data to PDS will have at least one collection.

Some examples of collections are:

- Document Collection – Contains documentation necessary for understanding and using the data
- Data Collection – Contains observed, derived or generated data, logically grouped by shared attributes
- Calibration Collection – Contains calibration data separated from the data collection(s) and/or more detailed information about complex calibration efforts

Bundles consist of related collections.

A bundle contains collections that are related by a high level concept such as project, acquisition method, processing level, or utility. The collections in a bundle are often complementary to each other. Every data submission to PDS will have at least one bundle.

For example, a bundle may include:

- all collections (data, document, calibration) for a particular instrument, or
- data collections from a single lab or lab project, or
- all collections from a single field campaign, or
- all collections from a research or analysis project

The relationship among products, collections and bundles is shown in Figure 1 and is described in more detail at http://pds-atmospheres.nmsu.edu/bundle_diagram.html.

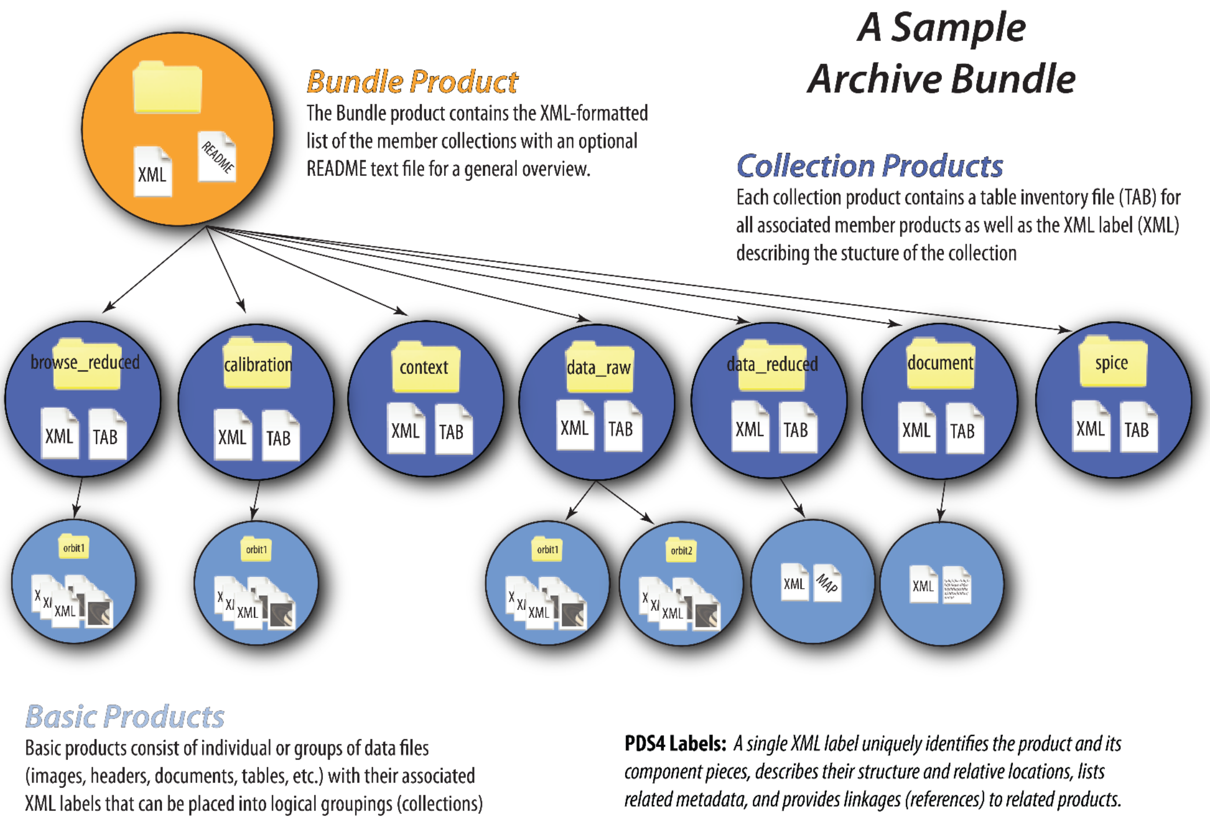
A Sample Archive Bundle
Bundle Product
The Bundle product contains the XML-formatted list of the member collections with an optional README text file for a general overview.
XML
README
Collection Products
Each collection product contains a table inventory file (TAB) for all associated member products as well as the XML label (XML) describing the stucture of the collection
browse_reduced
calibration
context
data_raw
data_reduced
document
spice
XML
TAB
orbit1
orbit2
MAP
Basic Products
Basic products consist of individual or groups of data files (images, headers, documents, tables, etc.) with their associated XML labels that can be placed into logical groupings (collections)
PDS4 Labels: A single XML label uniquely identifies the product and its component pieces, describes their structure and relative locations, lists related metadata, and provides linkages (references) to related products.

Figure 1 Relationship of product, collection and bundle.

The PDS recognizes that it takes some effort to learn how to produce XML labels and conform to PDS4 standards, and we will help you with this. Please feel free to communicate with us as often as needed to ensure that you approach archiving in an efficient manner. We can assist you in identifying the components for your planned archive, assist in developing templates for labels, and help you organize the collection and bundle structure.

6 Archive Planning

Most NASA ROSES programs require the data and other resources that are produced to be preserved in PDS or an equivalent archive. How the data are to be preserved is described in a Data Management Plan (DMP). The DMP is a very important document for successful proposal and may be something new to you. NASA maintains a list of DMP Frequently Asked Questions which may be helpful. This list is maintained at

http://science.nasa.gov/researchers/sara/faqs/dmp-faq-roses/

In addition, PDS offers this guidance. A DMP should contain the following elements in adequate detail for review:

• 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.

The topics above are suitable for most ROSES proposals. Be sure to check for any requirements specific to the program to which you are proposing because some programs may have additional or more specific requirements for the contents of the DMP.

For those proposers who plan to submit data to the PDS, the PDS has the following specific recommended topics for your DMP or, if appropriate, elsewhere in your proposal.

- Mention the PDS Node that you plan to work with and describe how you will interface with the Node. Include a letter of support from the node stating that you have discussed your proposed data products and that the Node will work with you to archive them.
- Describe a schedule for design, product generation, validation and delivery of your products, and be sure to include time required for participating in all steps of the archive process (see Section 7).
- Characterize the full scope and complexity of the archive, including all types of products to be delivered. Include total size estimates for all archive components, including images, documents, etc.
- Summarize the products you intend to archive, including the data products, documentation, and ancillary information (e.g., information on how the data were obtained, processed, calibrated---include anything that a user would need to know to use your data as a scientific product).
- Outline the design of your archive, including the bundle(s) and collection(s) as well as any derivatives of the delivered products (i.e., browse images) or supplemental products (i.e., SPICE kernels).

7 PDS Archiving Process

The PDS archiving process consists of multiple steps, beginning with identifying the data products, designing the collections and bundles, and continuing with the creation of context products (if appropriate products do not already exist), labeling of products, delivery of the products to PDS, conducting a peer review, resolution of any issues discovered during peer review and a final delta review.

A peer review is itself a process and is necessary in ensuring a quality archive. PDS requires that all data submissions from data providers pass a peer review before being archived. This review occurs near the end of the archiving process and is coordinated by the PDS. During a peer review, reviewers will evaluate the usability of the data by utilizing the metadata (in labels) to read and perform a scientific assessment of the data. In addition, the completeness and accuracy of other information in the archive submission will be assessed. Any issues discovered during a review are expressed formally as a “lien” against the products. Major liens are those severe enough to render a product to be unsuitable for scientific research. Minor liens are those which may affect ease of use, completeness of documentation, quality of information (i.e., typos in documentation or metadata) or other similar issues. All liens must be addressed before data products can be archived. A data provider (that’s you) is required to see the process through to having the products archived, including allotting enough time for iterative work on the preparation of products, supporting a peer review and completing any lien resolution. After all liens are resolved there will be a delta review to confirm that the liens were addressed. You need to allow sufficient time after the peer review to resolve the liens and to support a delta review in order to complete the archiving process.

8 Estimating Effort and Cost of Creating a PDS Archive

The task of creating an archive is shared between the proposer (PI) and the PDS. In general the division of effort is as follows:

The PI performs the following tasks (using funds from the proposed effort):

- Produces data products in acceptable PDS format (currently PDS4).
- Produces PDS labels, which under PDS4 are XML files.
- Writes supporting documentation.
- Organizes data, labels, documentation, etc. into collections and bundles.
- Validates labels using PDS provided tools.
- Participates in peer review (often done via web, email and phone).
- Makes updates to the products, as necessary, based on peer review recommendations.
- Participate in delta review (if needed)
- Delivers final package to PDS.

The PDS Node performs the following tasks (using PDS funds):

- Provides advice on PDS standards and requirements.
- Assists in designing PDS labels, if needed.
- Helps create context products.
- Provides available PDS tools.
- Sets up and conducts a peer review.
- Accepts the final package and integrates the data into its archives, including making it available on the Node web site.

We have found that this division of tasks helps to have an efficient archiving process; however, there may be situations where this division of tasks is not practical or fully achievable within the scope of the project. PDS is here to help, so contact the appropriate node and discuss how they might assist you.

Our experience indicates that preparation of a simple archive, consisting of one bundle with a few collections, may take one person up to a month of effort spread over the full duration of the project. You cannot wait until the end of the project before beginning the archiving process and expect the archiving process to go smoothly. Someone familiar with PDS standards and archiving procedures would likely take the least amount of effort to create an archive. Naturally, more complex archives take more time and effort. For example, creating multiple similar collections will take some amount of time for the first collection and a fraction of that time for each new collections (incrementally more time), whereas creating collections that are each unique will take about the same amount of time to create each collection (proportionally more time). The time spent on creating a PDS archive is spread across the design, product generation, validation, delivery of your products, peer review, lien resolution and delta review. In general, for archives created by an individual, more time is spent on product generation and validation than the other phases. Lien resolution can be time consuming if the product generation goes astray, but PDS4 is designed to minimize this likelihood and the PDS Nodes will give you guidance on how to avoid any surprises. PDS also provides tools to aid in the design, generation and validation of products (see https://pds.nasa.gov/pds4/software/index.shtml).

PDS is your partner in preparing an archive. While preparing an archive you should:

- Be prepared to follow the archiving process through to the end
- Expect to communicate often and to iterate with a PDS Node on archive design and product formats, labels, etc.
- Expect to validate your data submission using PDS-supplied software
- Be prepared to participate in a peer review of your data submission, with the support of PDS Node personnel
- Be prepared to resolve any liens.
- Be prepared to participate in a delta review of your data submission

9 Additional Information

We understand that not all data providers have extensive backgrounds in XML or PDS standards and processes. While this guide reflects the information found in a variety of PDS documents and specifications, you should refer to these additional resources for more details and suggestions:

- PDS4 Concepts (https://pds.nasa.gov/pds4/doc/concepts)
- PDS4 Standards Reference (https://pds.nasa.gov/pds4/doc/sr/current)
- PDS4 Information Model Specification (https://pds.nasa.gov/pds4/doc/im/current)
- Small Bodies Node PDS4 Wiki (http://sbndev.astro.umd.edu/wiki/SBN_PDS4_Wiki)

For writing a Data Management Plan you may want to look at these resources:

- NASA’s Frequently Asked Questions (FAQ) page on Data Management Plans (https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=499685/solicitationId=%7B96D0CCC2-2EF8-D528-B203-4269C960B788%7D/viewSolicitationDocument=1/PSDDMPFAQ030116.pdf)
- Geo Node’s information on writing DMP (http://pds-geosciences.wustl.edu/dataserv/proposerhelp.html)
- Imaging Node's Draft Archive Plan for a NASA Research Proposal (http://pds-imaging.jpl.nasa.gov/help/Draft_ArchivePlan_research_proposal_6-25-15_cei_skl_lg.pdf)
- Small Bodies Node Data Management Plan Tips (http://sbndev.astro.umd.edu/wiki/ROSES_Data_Managment_Plan_Tips)

- Atmospheres Node’s How-to Guide for Archiving Derived Data (http://pds-atmospheres.nmsu.edu/Derived_Data_LPSC2016_Brochure.pdf)

Tools and other resources

- PDS4 Software (https://pds.nasa.gov/pds4/software)
- PDS/PPI Software (http://ppi.pds.nasa.gov/software)
- PDS/SBN Software (http://sbn.pds.nasa.gov/tools/software.shtml)

And remember, PDS staff are available to help you find the information you need to ensure PDS4-compliance and successful archiving with PDS. If you have any questions, contact a PDS Node representative.

10 Referencing an Existing PDS Archive

In mid-2013 the PDS released version 4 of its archiving standards (known as “PDS4”). Most of the existing archives comply with version 3 of the standards (PDS3). New projects must create archives that comply with PDS4 even when using existing PDS archived data that may comply with other versions of PDS standards. When referencing PDS data in a proposal it is best to refer to it by its unique identifier. In PDS3 a grouping of data is called a “data set” and each data set has a unique Data Set ID. In PDS4, a grouping of data is called a “collection” and it has a unique Logical Identifier (LID) and version identifier (VID). These two identifiers can be combined to form a LIDVID, which references a specific version of a collection. In addition, with PDS4 every archived item (products, collections and bundles) has a unique identifier, so it is possible to refer to PDS4 items at a much finer granularity than with previous versions of PDS.

When creating new PDS4 products that rely in some way on existing PDS3 products it may be necessary to migrate some PDS3 products to PDS4 so that the products can be properly referenced by new products. For example, some of the documentation from a PDS3 data set might require migration into an equivalent PDS4 document collection. PDS can help you determine if this is necessary, and a PDS Node will do most of the work related to the creation of equivalent PDS4 collections.

11 Revision History

April 21, 2016; PDS; Initial version.

Appendix A: Glossary

archive: A place in which public records or historical documents are preserved; also the material preserved — often used in plural. Sometimes capitalized when referring to all of PDS holdings — the PDS Archive.

bundle: A list of collections. For example, a bundle could list a collection of raw data obtained by an instrument during its mission lifetime, a collection of the calibration products associated with the instrument, and a collection of all documentation relevant to the first two collections.

collection: A list of products, all of which are closely related in some way.

data object: A physical, conceptual, or digital object.

data provider: A person or organization that assembles archival data for delivery to PDS.

identifier: A unique character string by which a product, object, or other entity may be identified and located. Identifiers can be global, in which case they are unique across all of PDS (and its federation partners). A local identifier must be unique within a label.

information model: A representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse.

label: The aggregation of one or more description objects such that the aggregation describes a single PDS product. In the PDS4 implementation, labels are constructed using XML.

label template: A text file which serves as a pattern for constructing labels.

lead node: One of several consulting nodes designated as the PDS coordinator and primary contact with a project.

lien: A formally expressed issue with one or more products in a collection. Minor liens are those intended to improve the data set, but which are not considered critical to the understanding and use of the data. Major liens are those severe enough to render the data set (or increment, in the case of dynamic data sets) to be unsuitable for scientific research.

logical identifier (LID): An identifier which identifies the set of all versions of an object.

metadata: Data about data — for example, a ‘description object’ contains information (metadata) about an ‘object.’

product: One or more tagged objects (digital, non-digital, or both) grouped together and having a single PDS-unique identifier. In the PDS4 implementation, the descriptions are combined into a single XML label. Products are the smallest addressable granular unit in the PDS holdings.