Individual Proposer's Archive Guide (PAG)

Last Revision: April 20, 2019

Contents

1 Introduction 1

2 PDS Background 1

3 A Typical PDS Archive 2

4 Proposals for Using an Existing PDS Archive 2

5 Proposals for Creating a New PDS Archive 2

6 PDS4 Details and Definitions of Terms 3

7 Archive Planning 4

8 Estimating Effort and Cost to Create a PDS Archive 6

9 Additional Information 7

10 Revision History 8

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 PDS or an equivalent archive. 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 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. It will help you to estimate the effort (and cost) of preparing your archive for submission to the PDS. It reflects the more detailed information found in the PDS standards.

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 usability of data and to stimulate advanced research. All PDS data are peer-reviewed, well-documented and, when released, publicly available; and may be exported outside of United States under the "Technology and Software Publicly Available" (TSPA) classification.

The PDS is a federation of teams with expertise in different science discipline. 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 PDS they must have sufficient documentation to enable others to 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. 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 also is described with PDS-defined metadata. Requiring software (proprietary or public) to read the data is **not** allowed in a PDS archive. There is only one exception to this rule and that 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 PDS and maintained as part of the archive. To ensure a quality archive each grouping is reviewed prior to its public release.

4 Proposals for Using an Existing PDS Archive

Research opportunities typically require that data used in proposed research are available in PDS or equivalent public archive at least 30 days prior to proposal submission date. In the PDS archive, data that are considered archived and publically released have a "Certified" status. The listings of data available from a PDS Node will typically show whether a collection of data is certified. If it's not clear, contact the appropriate Node and they can help you determine the status of the data you are interested in using.

In mid-2013 the PDS released version 4 of archiving standards (commonly referred to as "PDS4"). Most of the existing archives comply with version 3 of the standards (commonly referred to as "PDS3"). New projects must create archives that comply with PDS4, but when using existing PDS archive it may comply with either PDS3 or PDS4. In PDS, each collection of data is assigned a unique identifier. In PDS3 a collection is called a "data set" and each data set has unique Data Set ID. In PDS4, a collection 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. Additionally, in PDS4 every archived item has a unique identifier, including bundles and products.

5 Proposals for Creating a New PDS Archive

For proposals responding to NASA's ROSES that will generate data that will 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

this guide is to provide sufficient details of the PDS4 standards to enable you to write the appropriate 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 https://pds.nasa.gov/contact/contact.shtml for contact information). Choose the Discipline Node which is most closely associated with the type of data you plan to archive. Be prepared to discuss in general terms the focus of your research and the types and quantity of data you expect to generate. The Node can then advise you if they are the appropriate contact point and the place to deliver your data. The Node 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 as it can strengthen your proposal.

6 PDS4 Details and Definitions of Terms

PDS4 is the current PDS archive implementation, and it is an integrated system designed to improve 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) the data or document being archived. Every product is assigned a unique identifier.

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 keep in the mind the following general principles.

Everything is a Product, and Products belong to Collections.

> Products are not limited to data; documents, context files, XML schema, and even delivery websites are considered products and will have labels and unique identifiers.

Collections are groups of similar products.

Some examples are:

- Document Collection – Contains all documentation necessary for understanding and using the data
- Data Collection(s) – Contains all of the data, logically grouped, and could consist of multiple directories
- Calibration Collection – Could contain calibration data separate from the data collection(s) and/or more detailed information about complex calibration efforts

Bundles consist of related collections.

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

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

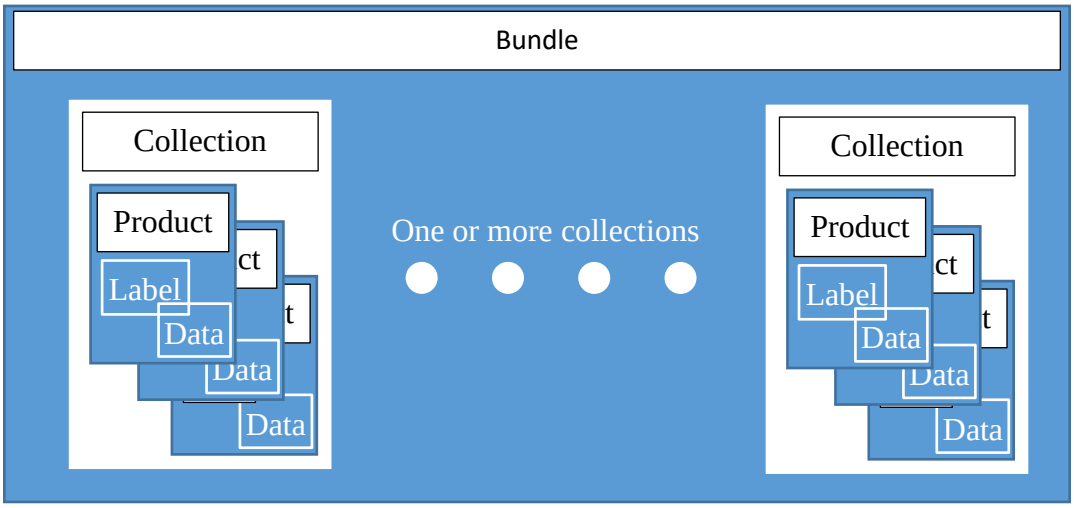

Figure 1: Relationship of product, collection and bundle in a PDS4 archive.

You can find examples of PDS4-compliant products, collections and bundles on the PDS Home page under PDS4, Information for Providers (see https://pds.nasa.gov/datastandards/documents/). If you don't find an example that you think is relevant to your proposed archive, ask any PDS Node for help. For a current list of contacts see https://pds.nasa.gov/contact/contact.shtml.

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 this in an efficient manner. We can assist in developing templates for labels, help you organize the collection and bundle structure and identify the components for your particular data set.

7 Archive Planning

Proposals to NASA ROSES programs are required to include a Data Management Plan (DMP). Some program elements may have specific requirements for the contents of the DMP, but in general the DMP must contain the following elements, as appropriate to the project, 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 above requirements apply to all ROSES proposals whether or not the proposer intends to archive data in the PDS.

For those proposals that do include the creation of a PDS archive, the PDS has the following specific recommendations for proposers, including information to include in your DMP.

The PDS archiving process consists of multiple steps, beginning with designing the collections and bundles and continuing with the creation of context products (if appropriate products do not exist), labeling of products, delivery of the products to PDS, a peer review, and resolution of any liens resulting from the review. A data provider is required to see the process through to the end, including allotting enough time for iterative work on the label creation process and enough time for the peer review and lien resolution at the end of the grant period.

In the DMP or elsewhere in your proposal, you should discuss the following things related to creating a PDS archive.

- Mention the Node in PDS you plan to work with and describe how you will interface with the Node. Include a letter from the PDS Node stating that you have discussed your proposed data products and that the Node will work with you to archive them. Even though such a letter may not be a NASA requirement, it can strengthen your proposal.
- Describe a schedule for development, testing, validation and delivery of your products, and be sure to include time required for participating in all steps. (see “Cost” section)
- Characterize the full scope and complexity of the archive, including all versions of products (i.e., whether you intend to include interim products, multiple mission phases, etc.) to be delivered. Include total size estimates for all archive components, including images, documents, labels, 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 and directory layout of your archive, including the bundle(s) and collection(s) as well as any needed ancillary products (i.e. browse images, SPICE kernels).

And if processing the data are part of the archiving process,

- Describe the processing that is necessary to update, revise, reformat or otherwise restore the products you will archive.
- Explain whether the processing is complicated or simple, and discuss how long it will take you to develop procedures and/or scripts to do the processing. (If procedures and scripts already exist, be sure to mention that.)
- Describe the computing resources necessary to do the proposed work, including data storage space, processing capacity and speed, etc. and explain whether you have them or have a plan to get them.

In the final phase of delivering an archive, PDS will coordinate a peer review of the data. A peer review is very helpful in creating a quality archive. Reviewers will often take a closer look at the prepared data than a data providers has the time or resources to do. Because of this, peer reviews almost always result in liens. You need to ensure you allow sufficient time post peer review to resolve the liens in order to

pass the peer review. PDS archiving efforts require that all data submissions from data providers pass a peer review before being designated as Certified Data. Data is not consider archived until it is certified. Additionally, achieving certified status is important because Certified Data have the added benefit of being recognized for use in future proposal efforts across planetary science.

8 Estimating Effort and Cost to Create a PDS Archive

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

What The PI does (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 an archive package.
- Validates labels using PDS provided tools.
- Participates in peer review (often done via web, email and phone).
- Makes updates, as necessary, based on peer review recommendations.
- Delivers final package to PDS.

What the PDS Node does (using PDS funds):

- Provides advice on PDS standards and requirements.
- Helps 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.

Our experience indicates that preparation of a simple archive, consisting of one bundle with multiple collections may take one person up to a 1 month of effort spread over the full duration of the project. You can't wait to end the project to begin the archiving process and expect to have certified data. Someone familiar with PDS standards and archiving procedures would likely take the least amount of effort to create an archive. More complex archives take proportionally more time and effort. The time spent on creating a PDS archive is spread across the development, testing, validation, delivery of your products, peer review and lien resolution. In general, for archives created by an individual, more time is spent on development than the other phases. In the development phase the design and details of the data and the labels is set along with how the data will be organized into collections and bundles. For the testing and validation phase, PDS provides tools to make this task easier (see https://pds.nasa.gov/pds4/software/index.shtml).

While preparing an archive you should:

- Expect to communicate often and to iterate with a PDS Node on archive design and product formats, labels, etc.

- Expect to validate your archive (PDS-supplied software exists to do this)
- Be prepared to provide sample data for and to participate in a peer review of your archive, with the support of PDS Node personnel
- Be prepared to follow the archiving process through to the end in partnership with a PDS Node, and to commit to resolving any liens noted during peer review

Note that if you intend to submit new (possibly higher-order) products or reprocessed data already in the PDS, you will need to do some additional tasks to ensure your new products will be PDS4 compliant. For example, some of the linking documents and mission documentation from a PDS3 dataset might require migration into a PDS4 document collection under the original mission. A PDS Node will do most of the work related to the creation of PDS4 linkages that your project may require, but the PDS Node does need to know that this work is needed so that the proposed archive submission can be completed on schedule.

9 Additional Information

We understand that not all data providers have strong backgrounds in XML or PDS organization. While this guide reflects the information found in a variety of PDS documents and specifications you may want to look at these additional resources for more details and suggestions:

- Small Bodies Node PDS4 Wiki (http://sbndev.astro.umd.edu/wiki/SBN_PDS4_Wiki)
- PDS4 Concepts Documents (https://pds.nasa.gov/datastandards/documents/)
- PDS4 Standards Reference (https://pds.nasa.gov/datastandards/documents/)
- PDS4 Information Model Specification (https://pds.nasa.gov/datastandards/documents/)

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

- Small Bodies Node Data Management Plan Tips (http://sbndev.astro.umd.edu/wiki/ROSES_Data_Managment_Plan_Tips)
- 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)
- Atmospheres Node's How-to Guide for Archiving Derived Data (http://pds-atmospheres.nmsu.edu/Derived_Data_LPSC2016_Brochure.pdf)
- 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)

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.

Tools and other resources

- PDS4 Software (https://pds.nasa.gov/tools/about/)

- PDS/PPI Software (https://pds-ppi.igpp.ucla.edu/software/index.jsp)
- PDS/SBN Software (https://pds-smallbodies.astro.umd.edu/tools/software.shtml)

10 Revision History

March 8, 2016, T. King. Initial version.

March 11, 2016, S. Slavney. Revised section titles and order of appearance. Added sections and some text, and notes where more text is needed.

March 16, 2016, T. King. Added text to next sections, addressed comments from S. Slavney and C. Neese.

March 18, 2016, N. Chanover. Edited some text and added some links.

March 21, 2016, T. King. Consolidate comments and produce new version of document.

March 22, 2016, T. King. Updates based on comments from M. Gordon, D. Simpson, B. Semenov and L. Nagdimunov.

April 20, 2019, E. Law. Updated links.